A Structural Model of Electoral Accountability*

S. Borağan Aruoba     Allan Drazen     Razvan Vlaicu

First Draft: January 31, 2015
This Draft: December 26, 2016

Abstract

This paper proposes a structural approach to measuring the effects of electoral accountability. We estimate a political agency model with imperfect information in order to identify and quantify discipline and selection effects, using data on U.S. governors. We find that the possibility of reelection provides a significant incentive for incumbents to exert effort, that is, a disciplining effect. We also find a positive but weaker selection effect. According to our model, the widely-used two-term regime improves voter welfare by 4.2% compared to a one-term regime.

JEL Classification: D72, D73, H70, C57
Keywords: discipline, selection, political agency, elections, structural estimation, maximum likelihood

---

* Arouoba (aruoba@econ.umd.edu): University of Maryland. Drazen (Corresponding author, drazen@econ.umd.edu): University of Maryland, NBER, and CEPR. Vlaicu (vlaicu@iadb.org): Inter-American Development Bank. The authors thank Jim Alt, Tim Besley and Shanna Rose for their assistance with data and general feedback, Ethan Kaplan, Nuno Limao, Emel Filiz Özbay and seminar participants at University of Maryland, Paris School of Economics, LSE, Bocconi University, École Polytechnique, Northwestern University, Wallis Institute Conference, LACEA Annual Meeting, SEA Annual Meeting and IDB Research Department for useful comments, and Seth Wechsler, Pablo Cuba-Borda and Camilo Morales-Jimenez for research assistance at various stages of the project.
1 Introduction

In a democracy elections are meant to make policymakers accountable for their performance. When elected officials are judged by the outcomes they produce, elections can improve policymaker performance in two key ways. They give incumbents who want to be reelected incentives to exert effort to improve outcomes, thus disciplining poor performance (Barro [1973], Ferejohn [1986]). Elections also serve a selection function by screening out low performers (Banks and Sundaram [1993], Fearon [1999], Smart and Sturm [2013], Duggan and Martinelli [2015]).

One may then ask how effective elections are in performing these functions, that is, how to measure the disciplining and selection effects of the electoral mechanism. An obvious way would be to compare the performance of incumbents who can be reelected at the end of their terms to the performance of those who are barred from seeking reelection, for example by term limits. However, a simple performance comparison – which is the essence of the reduced-form approach to studying electoral accountability in general and the effects of term limits in particular – cannot fully separate and thus quantify different effects of elections. Doing so is crucial to studying electoral accountability empirically and, more specifically to assessing both the positive and the normative effects of imposing or changing term limits.

This paper proposes a structural approach to measuring the discipline and selection effects of elections. We set out a political agency model with adverse selection and moral hazard. We design our model to mimic those U.S. states where governors have a two-term limit in office, currently the most prevalent regime. In the model, governors are of two types: “good”, who have intrinsic incentives to exert high effort; and “bad”, who would exert high effort only in the presence of external incentives to do so, such as the possibility of another term in office. Neither the effort level chosen by governors nor their type are observable to voters. Instead, they observe incumbent performance, an outcome that depends stochastically on effort. Voters use observed performance to decide whether or not to reelect the incumbent governor.

Estimation of the structural parameters of the model allows us to quantify discipline and selection effects and to assess their importance without relying on strong identification assumptions. This allows us to study not only the positive effects of policies such as limits on the number of terms an elected official may hold office, but also their normative effects.

---

1 There is a large empirical literature on the effect of elections on outcomes, termed political economic cycles. Brender and Drazen (2005, 2008) summarize key findings for political budget cycles. Welfare implications of opportunistic policymaker behavior are studied by Maskin and Tirole (2004), among others.
Term limits have become a popular policy prescription to those who think that individual politicians have become too “entrenched” in office and thus unresponsive to voter concerns. However, as indicated above, eliminating the possibility of reelection may have negative effects on policymaker disciplining and selection (as well as implying a loss of the benefits of the experience gained by veteran lawmakers). Hence, assessing how elections (and more specifically term limits) affect policymaker performance and ultimately voter welfare requires separating possible effects.\(^2\)

To address these questions, we estimate a baseline of no electoral accountability, that is, where there is no possibility of reelection. On the basis of this, we can measure how much electoral accountability improves outcomes, as well as whether improvements come mainly through discipline or through selection. This proves to be relevant since the small effect of electoral accountability found in the reduced-form analysis outlined in Section 2 hides fairly large and distinct discipline and selection effects. Disentangling these effects is crucial in addressing the issue of electoral accountability in the political agency model, a workhorse model in political economy. Our analysis is one of the few tests of the empirical relevance of this model.

The structural model also allows us to run experiments to assess the welfare effects of alternative regimes. Using parameters estimated from governors limited to two terms, we consider alternative regimes (such as a one-term limit or one where the voters observe an imperfect signal about the effort of governors) where both the governors and voters in the economy optimally respond to the changed incentives implied by the different electoral regimes. The assumption of invariance of structural parameters to the electoral regime is essential in avoiding the Lucas (1976) critique.

Our main findings are as follows. We find that 52\% of governors are good and exert high effort independent of which term they are in. The possibility of reelection provides a significant incentive for some bad governors to exert high effort in their first term in order to increase their chances of reelection. Compared to the case with a one-term limit, allowing a second term leads 27\% of bad governors to exert high effort in their first term of office, implying a 13 percentage point increase in the fraction of all governors who exert high effort in their first term. Discipline would be stronger were it not for a stochastic relation between effort and performance (high effort does not always lead to high performance), as well as an exogenous random component to election outcomes, that is, success or failure in reelection uncorrelated with performance. The two-term-limit regime leads to an increase in voter

\(^2\)These effects may also characterize indirectly-elected policymakers, as in Vlaicu and Whalley (2016).
lifetime welfare of 4.2% relative to the case of a one-term limit. About 2/3 of this gain in welfare comes from the disciplining of bad governors. The remainder comes from the selection effect, that is, more good governors surviving to the second term because better first-term performance stochastically signals high effort and hence a higher probability that the governor is of the good type. The selection effect is reduced by a mimicking effect in that high first-term effort by bad governors makes it harder for voters to identify them as such. In the absence of mimicking, discipline and pure selection effects would be roughly the same size, but mimicking reduces the latter by about 30%.

We then consider a version of the model where effort is at least partially observable. This leads to increased discipline, but as in the case with fully unobservable effort, this effect is limited due to the stochastic nature of election outcomes: since bad governors know that they can still get reelected due to a favorable election shock, they do not have a strong incentive to exert high effort. Even if effort were fully observable, only 42% of bad governors would be disciplined, leading to a modest 2.1% increase in welfare relative to the case of unobservable effort. If, on the other hand, the increase in transparency is accompanied by election outcomes that are less stochastic, because, perhaps, elections are now won and lost more on substance that is measurable (by the observers) rather than unobserved characteristics, this would increase discipline considerably and lead to much larger welfare gains. In the extreme case where we shut down election shocks and make effort fully observable, welfare goes up by 10.4% since all bad governors choose to exert high effort.

The plan of the paper is as follows. In the next section we discuss the reduced form approach by reviewing the literature on empirical estimation of the effects of electoral accountability and its limitations. In section 3 we present our political agency model with a two-term limit. Section 4 describes the model’s solution, the estimation methods, and the data. We then present and discuss our estimates and their implications in Section 5. The final section presents conclusions. An online appendix contains technical details.
2 Literature on Estimating Effects of Electoral Accountability

2.1 Reduced-Form Estimation

There have been a number of papers using reduced-form estimation to measure the effects of term limits on politician performance. For example, Besley and Case (1995, 2003), Besley (2006), and Alt, Bueno de Mesquita and Rose (2011) consider fiscal policy outcomes under U.S. governors (the last paper also looks at economic growth), List and Sturm (2006) look at environmental policy in U.S. states, and Ferraz and Finan (2011) consider fiscal corruption of Brazilian mayors. The methodology is generally to compare, within a jurisdiction, the performance of reelection-eligible governors and lame-duck governors, that is, governors who are in their last legal term in office. These papers find clear and statistically significant differences in outcomes, but this comparison cannot in itself reveal the relative strengths of discipline and selection effects in generating these outcomes. As will be clear shortly, comparing reelection-eligible and lame-duck politicians can only reveal a net effect.

Some of the above research makes further assumptions to try to disentangle the effects. For example, Besley (2006) argues that U.S. lame-duck governors are more in tune with voter preferences, as measured by interest group ideological rankings, suggesting that performance differences reflect a selection effect. List and Sturm (2006) argue that discipline effects will dominate selection effects if the fraction of voters who vote primarily on environmental issues is sufficiently small (see footnote 8 of their paper). Ferraz and Finan (2011) argue that by comparing performance of second-term mayors with that of first-term mayors who were subsequently reelected, one can control for unobserved heterogeneity. Based on this, they argue that changes in levels of corruption largely reflect discipline rather than selection. Finally, Alt, Bueno de Mesquita and Rose (2011) argue that discipline can be measured by relative performance of incumbent governors in the same term, comparing the performance of those who are eligible to run again with those who are not (since all have survived the same number of elections), while selection over characteristics is reflected in the relative performance of term-limited incumbents in different terms (since each has been elected a

---

3 A different approach is natural experiments, as in Dal Bó and Rossi (2011). They use two episodes in the Argentine Congress when term lengths were assigned randomly to study the relation between term lengths and politician effort. Consistent with our findings for U.S. governors they find that longer terms induce higher legislator effort due to a longer horizon over which to capture the returns to high effort.

4 Ashworth (2012) makes a similar point in his excellent survey of research on electoral accountability.
different number of times but cannot be reelected again).\footnote{Gagliarducci and Nannicini (2013) estimate how increasing politicians’ wages affects the composition of the candidate pool and the reelection incentives of those elected. Using a regression discontinuity design and Italian mayoral elections data they find that higher wages increase performance and do so disproportionately through attracting more competent types.}

As suggestive as these arguments are, they often rely on specific untestable assumptions to tease out effects. Moreover, they do not fully allow separation of the discipline and selection effects. For example, Alt, Bueno de Mesquita and Rose (2011) cannot reject the hypothesis that the discipline and selection effects are equal in magnitude. Structural estimation allows us to do that and, moreover, does not rely on comparison of outcomes across regimes, e.g. number of terms to which reelection is possible, where arguably other things have changed.

\section*{2.2 An Analysis of the Reduced-Form Approach}

A standard analysis of the effect of elections, for example Ferraz and Finan (2011), compares the performance of politicians who are eligible for reelection with those that are not (lame ducks), controlling for various observable characteristics of politicians or the electorate. Differences in performance are then associated with different effects via specific identification assumptions. As an illustration, consider average (expected) performance of a governor who has a two-term limit. Average performance in the first term can be written as baseline + discipline while average performance in the second and last term is baseline + pure selection – mimicking. Here “baseline” captures the level of performance that would be observed in the absence of electoral accountability, that is, independent of the effect of elections. Using terminology in line with our model, “discipline” reflects the increase in performance of bad governors induced by the desire to be reelected; “pure selection” shows the increase in average performance of second term governors due to a higher fraction of “good” governors being reelected; and “mimicking”, the decrease in average performance of second term governors resulting from bad governors having mimicked good governors in the first term – thus increasing their probability of reelection – and then putting in low effort in their second terms. Selection as commonly used in the literature refers to what we consider pure selection minus mimicking.

If one simply computed the performance differential between reelection-eligible governors and lame duck governors, or, equivalently, regressed gubernatorial’ performance on a dummy showing whether the governor is eligible for reelection, the coefficient would simply be the difference between the performance of first-term governors and second-term governors, that is, discipline – pure selection + mimicking. It should be clear that this difference in perfor-
mance by itself gives no information about either the absolute or the relative sizes of the three channels, information that structural estimation will allow us to identify.

Table 1 reports our replication of a typical reduced-form analysis using the data we use subsequently to estimate our model. It uses a governor’s job approval ratings (JAR) as a proxy for performance, denoted by \(y\). We discuss JAR as a performance measure in detail in section 4.3.1. We estimate

\[
y_{ist} = \mu_t + \mu_s + \gamma E_{ist} + \text{controls} + \nu_{ist}
\]

where an observation unit is a governor \(i\) in a state \(s\) in a period \(t\) where a period can be a month, a year or a term. In (1) \(E_{ist}\) is the dummy variable showing the governor is reelection eligible and the regression also includes state and time fixed effects and controls. Here \(y_{ist}\) is the average of JAR surveys conducted in a month, a year, or a governor’s entire term. As should be clear, the coefficient \(\gamma\) captures the effect of being reelection eligible and will contain the combination of the three effects as explained above.

Turning to the results in Table 1, the first three columns show that when we consider all governors then there is no significant difference between the performance of reelection-eligible governors and those that are not. When we restrict the sample to only those governors who subsequently win reelection (columns four to six), then we get a positive coefficient which is statistically significant for the monthly and annual analysis but not the term-level analysis. That is, we find that performance is higher when governors are in their first term, but that this depends on the level of survey aggregation used.

Given the results in the first three columns, a typical reduced-form analysis would have concluded that there is no significant effect of electoral accountability on performance. Turning to the results in columns 4 and 5, these show that once we restrict the sample to governors who subsequently win their reelection bid, performance is higher for governors in their first term. Our estimates are similar to the results in Tables 4 and 7 of Ferraz and Finan (2011), who find that in a sample of mayors serving in a two-term limit regime the effect of being reelection eligible is larger for winners than for the full sample. In the winner subsample performance differences cannot reflect selection, as all first-termers become lame ducks, thus the coefficient measures a mixture of discipline and mimicking. Once again, reduced-form analysis cannot distinguish these two effects further, while a structural analysis can.
2.3 Structural Estimation

There are very few papers that use a structural rather than reduced form approach to estimating electoral accountability. Sieg and Yoon (2014) ask whether electoral competition leads U.S. governors to moderate their fiscal policies: Democratic incumbents to act more fiscally conservative, Republican incumbents to act more fiscally liberal. They find this is the case for about 1/5 of Democratic incumbents and 1/3 of Republican incumbents. Our paper differs in two key respects. We study the effects of reelection on governor effort and overall performance in office rather than on their fiscal stance. The papers are thus complementary in focusing on different outcome aspects. Second, our model considers both the moral hazard and adverse selection problems of the electoral agency, whereas their model focuses on adverse selection abstracting from governors’ effort decision.

Finan and Mazzocco (2016) consider how electoral incentives affect the allocation of spending on public goods in the Brazilian federal legislature. They structurally estimate a model emphasizing the interaction among multiple representatives, as well as their decisions to run for office, paying special attention to inefficiencies due to electoral motivations and to corruption. They find that 26 percent of funds are misallocated relative to the social optimum, and study the welfare effects of alternative electoral institutions such as approval voting. While the paper is also concerned with the effect of electoral accountability on outcomes, the mechanisms it highlights – interaction among legislators and their decisions to run for office – are quite different than ours, as is the outcome studied (efficient versus inefficient allocation of public spending), and the level of government.

Finally, Avis, Ferraz, and Finan (2016) study the effects of random audits of Brazilian municipalities in their use of federal funds. In addition to electoral discipline and selection effects, they consider what they term a non-electoral discipline effect, whereby the finding of corruption may lead to judicial punishment or reputation costs. They argue that there is minimal support in their data for electoral effects of audits, with the non-electoral explaining 94 percent of the reduction in local corruption from the audit program. Hence, their paper not only looks at a very different measure of performance at a different level of government than ours, but also finds that for that measure at the municipal level, the mechanism by which information disciplines incumbents is overwhelmingly non-electoral rather than electoral.

---

6Structural estimation is relatively rare in political economy. Some examples are Merlo (1997), Diermeier, Eraslan, and Merlo (2003), and Strömberg (2008). Two other recent papers, Gowrisankaran et al. (2008) and DeBacker (2011) focus on the voter decision problem as a dynamic optimization problem, similar to our approach, but in their model politician’s actions are probabilistic and not strategic as in our model.
3 Model

As our benchmark model, we start with a simple political agency model with both moral hazard (unobserved politician effort) and adverse selection (unobserved politician preferences) that can generate stochastic policy outcomes and reelection rules. (In section 5.1.2 on identification we discuss why a model without moral hazard, that is, with only adverse selection, would not be consistent with all the findings of the paper.) Subsequent versions of the model relax some of the benchmark model’s assumptions. All voters are assumed to have the same information set and preferences, allowing modeling of a single representative voter. A governor may serve a maximum of two terms. After a governor’s first term, voters may choose to replace her with a randomly drawn challenger. If a governor has served two terms, the election is between two randomly drawn challengers. The equilibrium concept we use is Perfect Bayesian Equilibrium, which will be defined formally below.

3.1 Governor Types

All governors enjoy rents of \( r > 0 \) in each term they are in office. A governor is one of two types, either “good” \((\theta = G)\) or “bad” \((\theta = B)\), where the probability that a governor is good is \( \pi \equiv \mathbb{P}\{\theta = G\} \), where \( 0 < \pi < 1 \). Governors choose the level of their effort. The cost of exerting low effort \((e = L)\) is normalized to be zero. The difference between good and bad governors is in the cost they assign to exerting high effort \((e = H)\). In any term of office good governors have no cost of exerting high effort, while bad governors have a positive utility cost \( c \), which is expressed as a fraction of the rents \( r \) of office.\(^7\) For ease of exposition, we define \( c(e; \theta) \) \( r \) the cost of effort level \( e \) for a governor of type \( \theta \), where

\[
c(H; G) = c(L; G) = c(L; B) = 0 \quad \text{and} \quad c(H; B) = c
\]

(2)

We assume that, like the governor’s type \( \theta \), the cost \( c \) is observed by the governor but unobserved by the electorate. A bad governor draws \( c \) from a uniform distribution on the unit interval \([0,1]\) when first elected, where \( c \) remains the same in all terms while in office.\(^8\)

---

\(^7\)Note that the two types and their levels of effort should not be interpreted too literally. A bad governor can be one who is rent-seeking or otherwise not “congruent” with the voters; for example, leaders may differ in their inherent degree of “other-regarding” preferences towards voters, as discussed in Drazen and Ozbay (2015). Alternatively, a bad governor can be one who is low competence (and thus finds it very costly to exert sufficient effort to produce good outcomes), or otherwise a poor fit for the executive duties of a governor.

\(^8\)We also considered more general specifications, including a Beta \((a, b)\) distribution, where the uniform distribution we use is a special case with \( a = b = 1 \). However, \( a \) and \( b \) were not separately identified in our estimation.
The governor understands that her chance of winning reelection is \( \rho_H \) if she exerts high effort and \( \rho_L \) if she exerts low effort, where in equilibrium \( \rho_L < \rho_H \). Different levels of effort lead to different distributions of observed performance (as specified in equations (6) below). Hence, the reelection probabilities \( \rho_L \) and \( \rho_H \) are a combination of the performance of the governor given her effort and the probability of reelection given her performance, and they will be determined in equilibrium.

### 3.2 Governors’ Effort Choice

The problem of a governor of type \( \theta \) is

\[
\max_{e_1, e_2} \left[ 1 - c(e_1; \theta) \right] r + \left[ 1 - c(e_2; \theta) \right] r 
\]

where \( e_i \) is effort in term \( i \) and \( 1_H \) is an indicator which equals 1 if \( e_1 = H \) and 0 otherwise.

The actions of a good governor are trivial – she exerts high effort in the first term (\( e_1 = H \)) since it is costless and strictly increases her chances of reelection. Since effort is costless and she is indifferent over effort levels in the second term, we simply assume that \( e_2 = H \) as well.\(^9\)

For a bad governor it is clear that the optimal choice for the second term is \( e_2 = L \) since exerting high effort in the second term is costly and has no benefit.\(^10\) To derive \( e_1 \), note that if a bad governor exerts high effort in her first term, her payoff is \( (1 - c + \rho_H) r \), and if she exerts low effort, her payoff is \( (1 + \rho_L) r \). In words, by exerting high effort the governor would forego some of the first-term rent but would increase her chances of reelection, thus enjoying the rent for an extra term. She would therefore find it optimal to exert high effort if and only if

\[
c < \rho_H - \rho_L
\]

The voter does not observe \( c \), but understands the maximization problem that governors face. He therefore can calculate the probability \( \delta \) that a bad governor exerts high effort in her first term, that is, \( \delta \equiv \mathbb{P} \{ e_1 = H \mid \theta = B \} \). Given the assumption of a uniform distribution for \( c \),

\(^9\)If we assumed that good types like exerting high effort, i.e. \( c(H; G) < 0 \), she would strictly prefer \( e_2 = H \). This would also follow if, consistent with what we argue below about the relation between effort and expected performance, the good type preferred higher performance.

\(^10\)In reality, good last-term performance may of course improve opportunities after the governor leaves office. The basic point however is that for bad governors the impossibility of another term reduces a key incentive to perform well, so that they will put in less effort than good governors and perform less well, a phenomenon that we observe in the data.
we may then write

\[ \delta = \mathbb{P}(c < \rho_H - \rho_L) = \rho_H - \rho_L \]  

(5)

### 3.3 Voter’s Problem

The voter lives forever and prefers higher to lower \( y \), where \( y \) is the performance of the governor in office. The voter’s utility is linear in \( y \). We assume that this performance variable is in part influenced by the effort choice of the governor according to the rule

\[
\begin{align*}
\frac{y_i}{e_i} | (e_i = H) & \sim N \left(Y_H, \sigma_y^2\right) \\
\frac{y_i}{e_i} | (e_i = L) & \sim N \left(Y_L, \sigma_y^2\right)
\end{align*}
\]

(6a) \hspace{1cm} (6b)

for term \( i = 1, 2 \), where \( Y_H > Y_L \). Since the variance of the two distributions is the same, if the governor exerts high effort, the outcome will be drawn from a distribution that first-order stochastically dominates the one with low effort. Note that we also assume that the relationship between effort and performance is independent of the governor’s type or the term she is in.

We further assume probabilistic voting in that the utility of the voter is affected by a shock \( \varepsilon \sim N(\mu, \sigma_\varepsilon^2) \) occurring right before the election (that is, after \( e_1 \) is chosen). This “electoral” shock may reflect last-minute news about either the incumbent or the challenger, an exogenous preference for one of the candidates, or anything that affects election outcomes that is unrelated to the performance of the governor. Hence, the existence of the election shock makes elections uncertain events given the performance of incumbents. The mean of this shock, \( \mu \), jointly with other parameters capture how attractive the incumbent is relative to the challenger. We turn to the details of what \( \mu \) exactly captures in Section 5.1.2.

Define \( W(y_1, \varepsilon) \) as the voter’s life-time expected utility after observing the first-term performance of a governor and the election shock. It can be expressed recursively as

\[
W(y_1, \varepsilon) = y_1 + \beta \max_{R \in \{0, 1\}} \mathbb{E} \left\{ R \left[ y_2 + \varepsilon + \beta W(y_1', \varepsilon') \right] + (1 - R) W(y_1', \varepsilon') | y_1, \varepsilon \right\} 
\]

(7)

where \( \beta \) is the voter’s discount factor between electoral terms, and \( R \) is the decision to re-elect. After observing the performance of the incumbent governor and the election shock, the voter makes his reelection choice. If he reelects the governor, he will enjoy her second term performance as well as the election shock, which shows up as an additive term to the utility of the voter. Note that \( \varepsilon \) does not affect the type or actions of the challenger that the incumbent faces. Once the incumbent’s second term is over, a new governor drawn from
the pool of candidates will come to office. The successor governor will deliver a first-term performance \( y_1 \) and face a reelection shock of \( \varepsilon' \), giving \( W(y_1, \varepsilon') \) utility to the voter. If the voter does not reelect the incumbent, then a fresh draw from the pool of candidates occurs.

It is important to note that the voter realizes that he may have arrived at this node with \((y_1, \varepsilon)\) in one of three ways: a good governor, a bad governor who exerted high effort, or a bad governor who exerted low effort. The voter, of course, does not know which of these is the case, but has beliefs about them.

We can rewrite the voter’s problem as

\[
W(y_1, \varepsilon) = y_1 + \beta \max_{R \in (0,1)} \{ R \left[ \mathbb{E}(y_2 | y_1) + \varepsilon + \beta \mathbb{V} \right] + (1 - R) \mathbb{V} \} \tag{8}
\]

where we use \( \mathbb{V} \) to denote \( \mathbb{E}[W(y_1', \varepsilon')] \), namely the voter’s expected lifetime utility at the beginning of a two-period term. This is a constant since none of the stochastic variables are persistent. It can be written as

\[
\mathbb{V} = [\pi + (1 - \pi) \delta] \frac{1}{\sigma_y \sigma_\varepsilon} \int \int W(y_1', \varepsilon') \phi \left( \frac{y_1' - Y_H}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy_1' d\varepsilon' \tag{9}
\]

\[
+ (1 - \pi)(1 - \delta) \frac{1}{\sigma_y \sigma_\varepsilon} \int \int W(y_1', \varepsilon') \phi \left( \frac{y_1' - Y_L}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy_1' d\varepsilon'
\]

where \( \phi(\cdot) \) represents the standard normal PDF. Equation (9) makes it explicit that there is uncertainty with respect to the type of the governor, her effort and performance in the first term, as well as the election shock that will be drawn before the election at the end of the first term. In what follows, we proceed as if \( \mathbb{V} \) is a known constant, and it will be solved as a part of the equilibrium. Note further that

\[
\mathbb{E}(y_2 | y_1) = \hat{\pi}(y_1) Y_H + [1 - \hat{\pi}(y_1)] Y_L \tag{10}
\]

where \( \hat{\pi}(y_1) \equiv \mathbb{P}(\theta = G | y_1) \), that is, the voter’s posterior probability that the incumbent is good after observing first-term performance. Using (10) we can write \( W(y_1, \varepsilon) \) as

\[
W(y_1, \varepsilon) = y_1 + \beta \max_{R \in (0,1)} \{ R \left[ \hat{\pi}(y_1) Y_H + [1 - \hat{\pi}(y_1)] Y_L + \varepsilon + \beta \mathbb{V} \right] + (1 - R) \mathbb{V} \} \tag{11}
\]

### 3.4 Election

If types were observable, the voter would reelect only good governors since they would exert high effort in their second term while bad governors would not. Since voters only observe
performance $y_1$, not type or effort, their reelection rule is related to performance. However, the relationship is probabilistic, not deterministic, because the election shock might change the voter’s performance-based assessment of the incumbent. Solving the discrete choice problem in (11), the incumbent would win reelection, i.e. $R = 1$, if and only if

$$\hat{\pi}(y_1) > \frac{(1 - \beta) V - Y_L - \varepsilon}{Y_H - Y_L}$$

(12)

which shows that the incumbent will win reelection if the first-term outcome $y_1$ is sufficiently good (so that the voter has a high posterior probability of the incumbent being good) or if the election shock $\varepsilon$ is not too small or too negative. We can summarize the voting rule $R(y_1, \varepsilon)$ with the following

$$R(y_1, \varepsilon) = \begin{cases} 0 & \text{if } \varepsilon \leq \hat{\varepsilon}(y_1) \\ 1 & \text{if } \varepsilon > \hat{\varepsilon}(y_1) \end{cases}$$

(13)

where $\varepsilon = \hat{\varepsilon}(y_1)$ characterizes the points $(y_1, \varepsilon)$ for which (12) holds with equality with

$$\hat{\varepsilon}(y_1) = (1 - \beta) V - \hat{\pi}(y_1) (Y_H - Y_L) - Y_L$$

(14)

The voter uses the following Bayesian updating rule to infer the type of an incumbent

$$\hat{\pi}(y_1) \equiv \mathbb{P}(\theta = G|y_1) = \frac{\mathbb{P}(\theta = G) p(y_1|\theta = G)}{p(y_1)}$$

$$= \frac{\pi \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)}{[\pi + (1 - \pi) \delta] \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) + (1 - \pi) (1 - \delta) \phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)}$$

(15)

where $\delta$, as defined in (5), is the voter’s (correct) assessment about the probability that a bad governor will exert high effort in her first term, and $p(.)$ represents a generic density.

Denoting the reelection probability conditional on first-term performance by $\psi(y_1)$, we have

$$\psi(y_1) \equiv \mathbb{P}(R = 1|y_1) = \mathbb{P}[\varepsilon > \hat{\varepsilon}(y_1)]$$

$$= 1 - \Phi\left[\frac{\hat{\varepsilon}(y_1) - \mu}{\sigma_\varepsilon}\right]$$

(16)

where $\Phi(.)$ denotes the CDF of a standard normal random variable.

Finally, the last piece we need is the probabilities $\rho_L$ and $\rho_H$ that the governor was
taking as given. These can be obtained by integrating $\psi(y_1)$ with respect to the performance distributions as in

$$
\rho_H = \frac{1}{\sigma_y} \int \psi(y_1) \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) dy_1 \quad (17)
$$

$$
\rho_L = \frac{1}{\sigma_y} \int \psi(y_1) \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) dy_1 \quad (18)
$$

To summarize the events, Figure 1 shows a game tree of the interaction between a governor and the voter. The sequence of actions and the information structure are as follows:

1. In her first term, a good governor ($\theta = G$) chooses $e_1 = H$. A bad governor ($\theta = B$) privately observes her cost $c$ and she chooses effort $e_1 \in \{L, H\}$. As a result of this choice, first-term performance $y_1$ is realized.

2. The voter observes the incumbent’s performance $y_1$ (which determines his current period utility) but not her effort $e_1$ or type $\theta$. He updates the probability that the incumbent is type $G$ using $\hat{\pi}(y_1)$.

3. An election shock $\varepsilon$ is realized.

4. An election is held between the incumbent and a randomly-drawn challenger. Based on his beliefs about the type of the incumbent $\hat{\pi}(y_1)$ and the election shock $\varepsilon$, the voter decides whether to retain the incumbent or replace her with the challenger. If the incumbent is not reelected, then the game restarts.

5. If the incumbent is reelected, a bad incumbent chooses $e_2 = L$ and a good incumbent chooses $e_2 = H$. Based on $e_2$, a performance $y_2$ is drawn by nature giving the utility of the voter in that term.

6. At the end of the term, a new election is held between two randomly-drawn candidates and the game restarts.

### 3.5 Equilibrium

A strategy for a governor is a choice of whether or not to exert high effort, i.e. $e_i(c) \in \{H, L\}$, in each period that she is in office, $i = 1, 2$, conditional on her (privately observed) cost of effort realization $c$. A strategy for the voter is a choice of whether or not to reelect the
incumbent, i.e. \( R(y_1, \varepsilon) \in \{0, 1\} \), given the observed incumbent’s first-term performance \( y_1 \), and an electoral shock realization \( \varepsilon \). The voter updates his beliefs about the incumbent’s type according to \( \hat{\pi}(y_1) \).

A perfect Bayesian equilibrium is a sequence of governor and voter strategies, and voter beliefs, such that in every period: the governor maximizes her future expected payoff, given the voter’s strategy, the voter maximizes his future expected payoff given the governor’s strategy, and the voter’s beliefs are consistent with governor’s strategy on the equilibrium path. As the environment is stationary, equilibrium outcomes will be a collection of equilibrium objects \( (\rho_H, \rho_L, \delta, V) \), where \( \delta \) is the probability that a bad governor exerts first-term effort (equivalently, the fraction of disciplined reelection-eligible bad governors), \( V \) is the voter’s life-time discounted utility, and \( \rho_H, \rho_L \) are reelection probabilities following, respectively, high and low first-term governor effort. Formally, we have the following definition.

**Definition** The outcome of a Perfect Bayesian Equilibrium of the game between a governor and the voter is a collection of scalars \( (\rho_L, \rho_H, \delta, V) \) where:

1. Given \( \rho_L \) and \( \rho_H \), a bad governor’s effort strategy \( e_1 \) leads to \( \delta \) and indirectly to \( V \).
2. Given \( \delta \) and \( V \), the voter’s reelection strategy leads to \( \rho_L \) and \( \rho_H \).

**Proposition 1** The Perfect Bayesian Equilibrium defined above exists and is unique.

**Proof.** See Appendix.

To understand the uniqueness result intuitively, consider first the decision of a bad governor in her first term. Her effort choice depends on the cost of high effort \( c \) relative to the increase in the reelection probability \( \rho \). Her maximization problem (3) implies that her decision will be to put in high effort \( e_1 = H \) if her cost \( c \) is no greater than \( \rho_H - \rho_L \), and to put in low effort \( e_1 = L \) otherwise. Hence, her decision may be described by a cutoff \( c^* = \rho_H - \rho_L \), which will be unique if the difference \( \rho_H - \rho_L \) (which is obviously between 0 and 1) is unique. The nature of the representative voter’s problem in (11) will clearly have a unique cutoff level in \( y_1 \) for each realization of \( \varepsilon \) as well.

Since the probability of reelection \( \psi(y_1) \) is monotonically increasing in first-term performance \( y_1 \) and the distribution of \( y_1 \) under high effort \( e_1 = H \) first-order stochastically dominates the distribution of \( y_1 \) under low effort \( e_1 = L \), the difference \( \rho_H - \rho_L \) is unique, so that \( \delta \) is as well. Finally, the voter’s life-time expected utility will obviously be unique as well in the equilibrium.
**Proposition 2** In equilibrium a good incumbent always exerts high effort; a bad incumbent exerts high effort if and only if (4) holds; the voter reelects the incumbent if and only if (12) holds; and voter beliefs about the incumbent’s type are given by (15).

**Proof.** Follows from the discussion above. ■

### 3.6 Model with Effort Signal

In this version of the model we allow the voter to observe a noisy signal about the effort level of the governor in the first term. We denote this signal by $z_1$ and assume that it is symmetric and correct with probability $\zeta$, that is

$$\zeta \equiv \mathbb{P} \{ z_1 = H | e_1 = H \} = \mathbb{P} \{ z_1 = L | e_1 = L \}$$

where $\frac{1}{2} \leq \zeta \leq 1$. The parameter $\zeta$ thus measures the informativeness of the signal. If $\zeta = \frac{1}{2}$ then the signal has no content, and the model is identical to the benchmark model. If $\zeta = 1$ then the signal fully reveals the incumbent’s effort level, and performance is no longer an informative signal.

The signal will only be relevant in the first term because once an incumbent is reelected, the voter has no more actions that may be informed by the signal. Thus, the only point where the signal is useful is when the voter updates his prior $\pi$ that the incumbent is good. The posterior is now defined by

$$\hat{\pi}(y_1, z_1) \equiv \mathbb{P}(\theta = G | y_1, z_1) = \frac{\pi p(y_1, z_1 | \theta = G) (1 - \pi) p(y_1, z_1 | \theta = B)}{\pi p(y_1, z_1 | \theta = G) + (1 - \pi) p(y_1, z_1 | \theta = B)}$$

which would then be used in calculating the voter’s expected utility from reelecting the incumbent and hence his reelection rule. Note that $\hat{\pi}(y_1, z_1)$ and $\psi(y_1, z_1)$ also have $z_1$ as an argument since they depend on $\hat{\pi}(y_1, z_1)$.

The incumbent understands that there will be a noisy signal about her first-term effort, which will affect her chances of reelection and uses the following expected reelection
probabilities in choosing her effort decision.

\[
\rho_H = \frac{1}{\sigma_y} \int [\zeta \psi (y_1, H) + (1 - \zeta) \psi (y_1, L)] \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) dy_1
\]

(21)

\[
\rho_L = \frac{1}{\sigma_y} \int [(1 - \zeta) \psi (y_1, H) + \zeta \psi (y_1, L)] \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) dy_1
\]

(22)

Further details are presented in the Appendix.

4 Solution, Estimation, and Data

In this section we discuss our strategy for solving and estimating the benchmark model. We also present our data. The details for the extension with an effort signal are presented in the Appendix.

4.1 Solution

The model has seven structural parameters: \( \pi, \beta, Y_H, Y_L, \sigma_y, \mu, \) and \( \sigma_\varepsilon \). As the definition of perfect Bayesian equilibrium shows, given the structural parameters, finding the equilibrium amounts to finding values for \( \rho_H, \rho_L, \delta \) and \( \mathbb{V} \). In the process of doing so, we need to evaluate five equilibrium mappings, \( \hat{\pi} (y_1), \hat{\varepsilon} (y_1), R (y_1, \varepsilon), W (y_1, \varepsilon) \) and \( \psi (y_1) \). We solve for the equilibrium as follows.

The first thing to notice is that once \( \mathbb{V} \) and \( \delta \) are known, \( \rho_H \) and \( \rho_L \) follow from (17) and (18), with \( W (y_1, \varepsilon), \hat{\pi} (y_1), R (y_1, \varepsilon), \hat{\varepsilon} (y_1), \) and \( \psi (y_1) \) obtained using (11), (12), (13), (14), and (16), respectively. Thus solving for the equilibrium amounts to satisfying (5) and (9). Define two residuals \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \) as the differences between conjectures for \( \mathbb{V} \) and \( \delta \) and the model-implied values from (9) and (5), respectively

\[
\mathcal{R}_1 \equiv \mathbb{V} - [\pi + (1 - \pi) \delta] \frac{1}{\sigma_y \sigma_\varepsilon} \int \int W (y_1', \varepsilon') \phi \left( \frac{y_1' - Y_H}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy'_1 d\varepsilon' \tag{23}
\]

\[
- (1 - \pi) (1 - \delta) \frac{1}{\sigma_y \sigma_\varepsilon} \int \int W (y_1', \varepsilon') \phi \left( \frac{y_1' - Y_L}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy'_1 d\varepsilon'
\]

\[
\mathcal{R}_2 \equiv \delta - \rho_H + \rho_L
\]

(24)

where equilibrium requires \( \mathcal{R}_1 = \mathcal{R}_2 = 0 \). This yields a nonlinear system of two equations in two unknowns, which we solve numerically. Consistent with our equilibrium uniqueness result, we are able to find a single solution to this system of equations given any set of structural parameters.
4.2 Estimation

We estimate the structural parameters using Maximum Likelihood. Our data set will consist of a measure of performance (for one or two terms) and reelection outcomes for a set of governors. As such, the unit of observation will be a governor stint. This can be either one or two terms, depending on whether the incumbent was reelected. Given the structure of the model, we can define the likelihood function analytically. For a governor who wins reelection, we observe the triplet \((y_1, R = 1, y_2)\). For a governor who loses reelection, we observe the pair \((y_1, R = 0)\). Each of these outcomes might come from different combinations of governor types, effort choices and reelection shocks. The density of a generic governor winning reelection while producing performance of \(y_1\) and \(y_2\) can be obtained as

\[
p_W(y_1, y_2) \equiv \frac{1}{\sigma_y^2} \left[ \pi \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) \psi(y_1) \phi \left( \frac{y_2 - Y_H}{\sigma_y} \right) \right. \\
+ (1 - \pi) \delta \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) \psi(y_1) \phi \left( \frac{y_2 - Y_L}{\sigma_y} \right) \\
\left. + (1 - \pi) (1 - \delta) \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) \psi(y_1) \phi \left( \frac{y_2 - Y_L}{\sigma_y} \right) \right] \\
(25)
\]

The three terms capture the cases where the governor is good, bad but disciplined, and bad and not disciplined, respectively. Similarly, the density of a governor of unspecified type losing reelection with first-term performance of \(y_1\) is given by

\[
p_L(y_1) \equiv \frac{1}{\sigma_y} \left[ \pi \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) [1 - \psi(y_1)] \right. \\
+ (1 - \pi) \delta \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) [1 - \psi(y_1)] \\
\left. + (1 - \pi) (1 - \delta) \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) [1 - \psi(y_1)] \right] \\
(26)
\]

For a governor \(k\) with \((y_{1k}, R_k, y_{2k})\), we compute her contribution to the log-likelihood using

\[
\log L_k = R_k \log [p_W(y_{1k}, y_{2k})] + (1 - R_k) \log [p_L(y_{1k})] \\
(27)
\]

and the log-likelihood is simply given by

\[
\log \mathcal{L} = \sum_{k=1}^{n} \log L_k \\
(28)
\]

Estimating the structural parameters requires maximizing \(\log \mathcal{L}\), which we do using standard
numerical optimization routines. We estimate six structural parameters \((\pi, Y_L, Y_H, \sigma_y, \mu, \sigma_z)\) and fix \(\beta = 0.85\), which represents roughly a 4% annual discount rate over a four-year term. Once estimates for the structural parameters are obtained, estimates for equilibrium outcomes \((\rho_H, \rho_L, \delta, \nu)\) can be directly obtained using the invariance property of Maximum Likelihood estimation. Standard errors are computed using the White correction for heteroskedasticity for the structural parameters, and the delta method for the equilibrium outcomes.

4.3 Data Description

4.3.1 Measuring Governor Performance

In order to estimate our model, we use data for U.S. governors. The key choice we need to make is the variable that proxies for performance \(y\) in the data. In the model \(y\) represents something that depends on governor effort, affects voters’ utility directly, and is observable to voters. Given our assumption of linear utility, it can be a measure of utility as well. Since reelection decisions depend on the performance in the first term, the measure we use needs to be a good predictor of reelection outcomes as well.

Existing empirical tests of the effect of reelection on governor performance (as discussed in section 2 above) use either economic variables, such as state unemployment rate or real income per capita growth, or fiscal variables, such as the growth in taxes, to measure governor performance. Such variables may be indicators of governor performance, but arguably governor performance reflects a larger set of variables, only some of which are quantifiable by (or even observable to) an outside observer. Corruption, which is shown to be important by Avis et al. (2016) or Finan and Mazzocco (2016) but difficult to measure, would be such a measure. See, for example, the discussion of Guy Hunt in section 5.1.3. We therefore want a broader measure of performance that might possibly capture the multi-faceted nature of performance.\(^{11}\) Nevertheless, for completeness sake, in Section 5.3, as a part of our robustness checks, we show our estimation results using two popular economic variables.

Theoretically, governor performance in this broader sense could be captured by expert evaluations (analogous to evaluations of U.S. presidents by historians), but such ratings are

\(^{11}\)Incidentally, the JAR-based performance measure is a better predictor of election outcomes than individual economic variables. In simple probit regressions, real income per capita growth and state unemployment rate have some limited success in predicting reelection outcomes. For example state unemployment rate has a significant coefficient and the probit regression has a McFadden \(R^2\) of 0.05. But once JAR is included in the same regression, unemployment rate loses its significance and the McFadden \(R^2\) almost quadruples to 0.19.
We chose to use job approval ratings (JAR) of governors from surveys of voters taken at various points during a governor’s term(s). A large fraction of the JAR data come from Beyle, Niemi, and Sigelman (2002), and we update their dataset through the end of 2014 using various online resources. Potential voters are asked to rate the governor as “excellent”, “good”, “fair” and “poor” or to say that they are “undecided”. As a measure of performance, we calculate for each governor the fraction of respondents who classify the governor as excellent or good out of those who express an opinion, eliminating the undecided respondents. We explain more precisely below how we convert this measure based on a survey taken at a specific point into a performance measure over the governor’s term.

The key question is then whether such approval ratings – and thus the resulting performance indicator – are a good measure of actual governor performance. There are two basic aspects of this question. First, does the JAR measure described above capture things believed to reflect true performance, such as economic variables that other studies used? Second, to what extent is this measure contaminated by things that do not reflect performance due to governor effort, such as various biases of survey respondents or pandering to voters?

To address these questions we regress our JAR-based performance measure on three sets of variables, and the results are reported in Table 2. The first set contains measures of state economic performance, such as state unemployment rate, growth of state per capita personal income, and state population growth. The second set are variables measure partisanship in the state: state’s population (possibly capturing homogeneity of preferences in smaller states), whether the governor is of the same party as the U.S. president, “partisan fit” which shows match between the party of the governor and how the state voted in the previous presidential election and the percent of voters of the governor’s or the opposing party. The third set of variables are governor characteristics: age, gender, years of education, whether the governor is a lawyer, or served in the military. The regressions also include state and time fixed effects.

The first column in Table 2 uses only measures of state economic performance. The

---

12It is also important to point out that JAR is not a relative rating, based on a comparison with a challenger, but it is an absolute evaluation of the governor’s performance in office, because the vast majority of JAR surveys are taken long before a challenger is identified. In our model, the challenger’s qualities enter through the election shock.

13We did not include fiscal outcomes sometimes used in some of the earlier literature (for example, higher government spending) in this regression because they are viewed differently by different groups of voters.

14The “partisan fit” variable follows Jacobson (2006) where it is 1 if governor’s party’s presidential candidate got more than 52% vote share in the state, it is -1, if governor’s party’s presidential candidate got less than 48% vote share in the state and 0 otherwise.
Second and third columns add each of the other sets of variables, one at a time, where the fourth column uses all variables. A number of things become clear. First, as the first two lines of the table across all columns make clear, the JAR-based performance measure is highly correlated in the correct direction with measures of state economic performance used in other studies. Looking at the $R^2$’s reported in the last row, a large fraction of the explanatory power in these regressions come from the first two variables, in addition to the fixed effects. Even without fixed effects the $R^2$ of the regression that includes only the first three variables (not shown) is 0.15, though in this case only unemployment rate is significant. Hence our JAR-based performance measure is indeed capturing governor performance in terms of the macroeconomic performance of the state.

Second, while most of the partisanship variables have no statistically significant effect on our governor performance measure, the measure is significantly related to both whether the governor is of the same party as the president and partisan fit. Interestingly, both variables have a negative effect on our JAR-based measure, that is, congruence of the governor’s party affiliation with either the president’s or with the voter’s preference in the most recent presidential election lowers the JAR rating. That is, partisan effects go in the opposite direction of naive conventional wisdom, where “partisan bias” is that a party’s adherents overrate a governor from the same party and underrate one from the opposing party. Jacobson (2006), who found an analogous negative correlation between a governor’s approval rating and his or her party being in the majority, explained it as such governors needing considerable cross-party appeal to win office. Similarly, we argue that the sign of the coefficients may be reflecting a performance effect – the bar for a Democrat governor, for example, to succeed in a Republican state is higher and thus they perform better; or alternatively, more “good” Democrat governors run for election in Republican states than “bad” Democrat governors. Thus, we think the case for arguing that JAR contains a partisan bias is weak, at best. Nevertheless, in Section 5.3 we consider an “adjusted” JAR measure where we strip our benchmark JAR measure from the effects of the two partisanship variables that are significant in Table 2.

Finally, among governor characteristics, only age has a significant (negative) effect on surveyed performance. This is consistent with the results of the JAR surveys being a good measure of governor performance to the extent that age (or any trait) is correlated with effort or with the relation between effort and performance. In any case, the coefficient is small: it shows that a one standard deviation difference (8 years) in age between two otherwise identical governors creates a JAR difference of 2.2 points.

One final question is whether a JAR-based performance measure as a proxy for $y$ might
reflect pandering to voters. Note first that if actions seen as pandering entered job approval ratings, our measure would still satisfy the conditions listed in the first paragraph as a proxy for $y$, and hence would allow us to estimate the model and measure the effect of reelection incentives on incumbent governor behavior. More specifically, the things the governor does that can be considered pandering, would still take effort, so that in order to pander and increase their chance of reelection, bad governors might choose to exert effort. However, the link between JAR and actual (not perceived) voter welfare would be weaker in this case and welfare statements would be problematic. On a conceptual level, arguing that JAR is so dominated by pandering and that it contains little information about a governor’s true performance is essentially arguing that voters (or individuals more generally) are simply unable to assess their own well-being. Though some take this extreme position, our comparison of the JAR data to narrative accounts of governor performance suggests that JAR does reflect a reasonably accurate assessment of voter welfare.¹⁵ Some of these narratives are presented in Section 5.1.3. On the more specific question of whether JAR is dominated by pandering, it may be argued that pandering is most likely in the election year. Our results are robust to dropping JAR surveys taken in the election year as we discuss in Section 5.3.¹⁶

To summarize the discussion so far, we think JAR captures what most people would consider the performance of governors. Through our regression, we were able to show that at least some key macroeconomic indicators significantly influence JAR. However, JAR is much more than just one or two indicators. The $R^2$ of the regression in Table 2, even after many controls including state and time fixed effects, is 0.44, which means we do not capture more than half of what is in JAR with this regression. Given the sensibility checks we have done, we are confident that it is the multi-faceted performance measure we need for our analysis. JAR clearly measures factors likely reflecting performance not captured by a univariate measure and does not show measurable evidence of being contaminated by non-performance-related factors.

Finally, we explain how we convert individual JAR survey results to measures that cover terms of governors. In order to eliminate effects of the governor’s reelection campaign, we use JAR up to and including June of the final year of the incumbent’s first term, i.e., the election year, given that U.S. gubernatorial elections take place in November. We do not restrict the

¹⁵There is some evidence that events not under a governor’s control happening right before elections do affect voters at the margin (Wolfers [2007], Healy, Malhotra, and Mo [2010]), but this is consistent with our modeling of the election shock $\varepsilon$.

¹⁶The view that voter assessments of government performance are not dominated by pandering is consistent with the findings of Brender and Drazen (2008) in a large cross-country sample at the national level, who find that voters punish rather than reward fiscal manipulation in election years.
second-term JAR-based measure. We take the simple average of the JAR numbers—that is, the fraction of those describing the governor’s performance as “excellent” or “good”, as described above—averaged over the entire term and use them as $y_1$ and $y_2$. From here on we use “JAR” to refer to the measure described in this paragraph. We revisit some of the choices we make in this section and consider alternatives in Section 5.3.

4.3.2 Governor Stints

Our model places some important constraints on the types of governor stints we can use in the estimation. We start with the universe of all governors that served from 1950 to the present, where we have collected basic information about the governor, some of which comes from Besley (2006), including the outcomes of their reelection bids.¹⁷ We then apply the following filters to eliminate governors who do not fit our model of a limit of two terms of equal length across governors.

- Drop governors who did not have any term limits, or had a one-term limit or a three-term limit.
- Drop governors during whose stints state election laws regarding term limits changed.
- Drop governor stints (not just the terms) where the governor was appointed, completed someone else’s term, or was elected through a special election outside the state’s regular electoral cycle.
- Drop governors who did not complete at least three years of their first term or at least two years of their second term (for example due to resignation, passing away, or being recalled).

These filters yield 169 governor stints.¹⁸ Combining this with the JAR data we compiled, a total of 5,549 surveys, yields 93 governor stints. Due to data availability and/or absence of term limits early on in our sample, except for one governor from the 1960s, our data

---

¹⁷We consider any governor that is eligible for reelection as having run for reelection, that is, we consider the choice of not running as losing. This is justified by our review of such cases where a reasonable interpretation of the events suggests that the governor decided that he or she would not be able to win reelection and either resigned or sought other alternatives. Perhaps not surprisingly, many of these governors perform quite badly in their first term, which results in being predicted by the model as “bad” governors who did not exert effort (see Table A1).

¹⁸A handful of governors serve multiple stints by being elected after some period following a completed term-limited stint. We treat each stint as a separate governor. Eliminating these governors from our sample does not change our results.
covers governor terms from 1978 to 2014. There are 26 election years from 32 states in our sample. The average age of a governor is 56, with 19 years of education on average, 91% of the governors in our sample are male, 55% of them are from the Democratic Party, 39% have served in the military and 46% of them are lawyers. Comparing these numbers with the population of all governors over this period, there does not seem to be a major bias in our sample.\textsuperscript{19} We provide the basic data that we use for estimation, namely \( (y_1, R, y_2) \) in Table A1 in the Appendix.

5 Estimation Results

5.1 Benchmark Model

5.1.1 Basic Results

The estimates of the six structural parameters and the four equilibrium outcomes are given in Table 3. Several things can be noted. 52% of governors in our sample are good and, based on the standard error, we strongly reject the two extremes, all governors being good or all governors being bad. Of the bad governors, 27% of them exert high effort in their first term and thus are disciplined. This is also highly statistically significant. Exerting high effort (for any governor) leads to an average increase in performance of over 20 JAR points, which is highly significant, both statistically and economically. High effort increases the probability of reelection from 45% to 72%. The mean of the election shock is 25.5 and it is highly significant. The election shock threshold \( \hat{\varepsilon}(y_1) \), the posterior probability that a type is good \( \hat{\pi}(y_1) \), and the reelection probability \( \psi(y_1) \), all conditional on observed \( y_1 \), are illustrated in Figure 2. The shapes of all these mappings originate from the shape of the \( \hat{\pi}(y_1) \) mapping, which in turn uses the normality of the process that determines \( y_1 \). A small first-term JAR, for example 25, signals to the voter that the governor did not exert high effort; as a result he assigns a near-zero probability of the governor being the good type. Then, for this governor to win reelection she needs an election shock of around 32 or larger. Since this is quite reasonable given the estimated values of \( \mu = 25.5 \) and \( \sigma_\varepsilon = 13.1 \), there is about a 30% probability for this governor to win reelection, despite poor first-term performance. As \( y_1 \) increases so does \( \hat{\pi}(y_1) \), until \( y_1 \) hits 80, after which the reelection probability remains constant at around 80%, reflecting the possibility of an unfavorable election shock after a

\textsuperscript{19}Our largest sample of 588 governor stints, which includes governors we dropped with the filters above show 96% governors as male, 55% were from the Democratic Party, 53% served in the military and 54% are lawyers. The average age of a governor is 53 and they have 19 years of education on average.
very strong performance in the first term.

In the specification of our model we assumed the variance of outcomes was independent of effort. When we allow the $y|e$ distribution to have a variance that depends on $e$, this unrestricted model has a log-likelihood which is only 0.23 log-points higher and there are no substantive change in the results. We conclude that the restriction we place by assuming a single variance is not rejected by the data, and given the small improvement in goodness of fit, a model selection criterion that favors parsimony (e.g. Schwarz Information Criterion) chooses our restricted model.

5.1.2 Identification

Before we turn to the implications of our model with the particular parameter estimates, it will be useful to discuss their identification. There are three key features of the parameter estimates we discuss in turn. First, we find very strong evidence that there are two types of governors ($0 < \pi < 1$). Second, we also find very strong evidence that some fraction of bad governors are disciplined ($\delta > 0$). Third, we find the presence of large election shocks. We demonstrate how our model identifies these three features by considering counterfactuals.

First, consider the possibility that there was only one type of governor. If $\pi = 0$ then there are only bad governors, while if $\pi = 1$ there are only good governors. In the latter case, since all governors are good, all governors would exert high effort in both terms, which means $\text{corr}(y_1, y_2) = 0$ as each term’s performance for a reelected governor is drawn independently from the same distribution. On the other hand, if all governors were bad then $\text{corr}(y_1, y_2)$ would be zero or even negative. \footnote{It is important to note that the cases where $\pi = 0$ and $\pi = 1$ are not covered by our model since updating of types as in (15) cannot occur. When $\pi = 1$ the reelection rule does not matter as all governors are good and they always exert effort. When $\pi = 0$, one can consider various other election schemes, under some of which, we may have some discipline. If this is the case then for these disciplined governors, $y_1$ will be drawn from the high-effort distribution and $y_2$ will be drawn from the low-effort distribution, which creates a negative correlation. Alternatively, if there are no disciplined bad governors, then to the extent any governor is reelected, $\text{corr}(y_1, y_2) = 0$ for the same reason as we explained for good governors.}

In our data, however, this correlation is 0.36 with a standard deviation of 0.14 (significant at the 5\% level) and our model delivers 0.39 over the long simulation explained in Section 5.1.3, which is very close.

Next, suppose that none of the bad governors exerted high effort so that $\delta = 0$. This would be a case of only adverse selection with no moral hazard, since types determine effort deterministically. If this were the case, then the difference between the performance of a reelected governor across terms would display two properties: it would have a zero mean and it would be a normally distributed variable with no skewness. \footnote{To see this first note that for a bad governor both $y_1$ and $y_2$ are drawn from $N(Y_L, \sigma^2_y)$ and for a good}
a mean of $-1.98$, though it is not significantly different from zero, and the Jarque-Bera test of normality is overwhelmingly rejected, primarily because of negative skewness. Our model delivers a mean of $-2.68$ and a negative skewness.

Finally, our estimation results indicate the presence of large election shocks with a positive mean – the 95% confidence interval for election shocks is $-0.1$ to $51.1$. To put this in perspective, the 95% confidence intervals for performance conditional on high and low effort are 24.0 to 62.6 and 44.7 to 83.3, respectively. Note further that the effect of the mean election shock $\mu$ on welfare exceeds the mean effect of high versus low effort, that is $Y_H - Y_L$, which is 20.7. We explain how $\sigma_\varepsilon$ and $\mu$ are identified in turn. If $\sigma_\varepsilon$ were equal to 0, but we kept the estimated value for $\mu$, then any governor with $\hat{\varepsilon}(y_1) < \mu$ would have to win reelection, and all others would have to lose reelection; reelection would become a deterministic function of $y_1$. However in our sample there are many governors who lose reelection despite good first-term performance, as well as those who win despite bad performance, implying the presence of election shocks. More specifically, if we used a rule of predicting reelection when $\hat{\varepsilon}(y_1) < \mu$, out of the 93 governor stints in our sample, we would correctly predict the reelection outcome of 70 of them. We would, however, incorrectly predict reelection for 15 of them and incorrectly predict losing reelection for 8 of them. This shows that there need to be large election shocks beyond what is simply implied by $\mu$.

Turning to $\mu$, we note first that though one may be tempted to think of *incumbency advantage* as being characterized by $\mu > 0$, this is not the case. Even if $\mu = 0$ there could be an electoral advantage or disadvantage associated with incumbency: election outcomes are determined as in (15) and $V$ depends in complicated ways on all structural parameters and the behavioral response of governors, i.e. whether or not bad governors exert high effort.

We define incumbency advantage as the unconditional probability of winning reelection for an incumbent, $\mathbb{P}(R = 1)$, being greater than 0.5 and explore the dependence of reelection probability on various factors in Appendix D where we show how $\mathbb{P}(R = 1)$ changes under different parameter configurations. Thus, rather than focusing on the identification of $\mu$, a useful discussion here is to demonstrate the degree of incumbency advantage in our data, as $\mu$ will adjust in conjunction with other parameters to match this. There are two ways of measuring the incumbency advantage in our data. First, 61.2% of incumbents (57 out of 93) in our data win reelection. Second, reelection surprises in our data favor incumbents: if we look at governors who have $y_1 > Y_H$ but lose reelection and those with $y_1 < Y_L$ but win reelection, the former is 15% while the latter is 30%.

---

Governor they are drawn from $N \left( Y_H, \sigma_y^2 \right)$. Given that they are *iid*, the difference $(y_2 - y_1) \sim N \left( 0, 2\sigma_y^2 \right)$ in both cases, and given the properties of the normal distribution it would have no skewness.
Once these key parameters or equilibrium objects are pinned down, the other three structural parameters, that is, $Y_H$, $Y_L$ and $\sigma_y$, follow from matching other properties of the JAR data. These include the mean and variance of $y_1$ and $y_2$.

5.1.3 Measures of interest

To understand what the parameter estimates imply for our model with a two-term limit, we report results including outcomes in a (counterfactual) world with a one-term limit in Table 4. While some of the measures can be computed analytically, many cannot, and thus we resort to simulations where we simulate the model for 1,000,000 hypothetical governors. The first panel of the table illustrates the case where a governor was restricted to one term of office, in which case only good governors would exert high effort, leading to an average performance of 54 JAR points. Lifetime welfare for the voter is 360.3 in this case.22

In the second panel we look at summary measures for two-term limited governors. Having a two-term rather than a one-term limit is unambiguously better for the voter. First of all, more governors exert effort in their first terms, leading to a higher average JAR. This is because 27% of the bad governors exert effort in addition to all the good governors, leading to high effort 64.7% of the time in the first term, compared with 51.8% in the one-term case. This increases average JAR in the first term from 54 to 56.7. Second, because a higher fraction of bad than good governors are screened out in elections, more governors are good in the second term: 59.6% relative to the unconditional probability of 51.8%. Since these governors always exert effort, the average JAR in the second term is 55.6, compared to 54 in the one-term case. Putting these together, the life-time welfare of the voter goes up from 360.3 to 375.3, which is a 4.2% increase. Put differently, a voter in a two-term regime would be willing to give up about 2.3 JAR points every term ad infinitum in order to remain in that regime and not switch to a regime of one-term limits. Looking at Table 4, it is clear that the voter is better off because the governors’ performance in both terms is higher relative to the case of a one-term limit. In the next section we will further explore the different effects of electoral accountability in order to better understand them. Also, as a reference, the “first-best” in this economy, one where every term high effort is exerted, leads to a lifetime expected welfare of $(1 - \beta)^{-1}Y_H = 426.6$. This shows that our benchmark economy has

22Here and throughout the paper life-time welfare is computed as $[1/(1 - \beta)] \mathbb{E}(y)$ since given linear utility this is equivalent to the obvious definition of welfare $\mathbb{E}\left[\sum_{t=0}^{T} \beta^t y_t^s\right]$. Here the $\mathbb{E}(\cdot)$ operator is over all possible random events including the type of governors. In practice we compute $\mathbb{E}(y)$ via a simulation with 1,000,000 governors. Note that our definition of welfare excludes $\epsilon$ and since the definition of $V$ in (9) includes $\epsilon$ in the utility of the voters, these numbers are not identical.
about 12\% less welfare than this ideal one.

Table A1 lists the individual governors in our sample, their performance, and some potentially interesting statistics that can be computed from our estimation. In particular we show the performance measures $y_1$ and $y_2$ that go into the estimation, as well as $\hat{\pi} (y_1)$, the updated probability that the governor is a good type after observing $y_1$, $\psi (y_1)$, the probability that the governor will win reelection given her first-term performance, as well as a new statistic

$$
\pi (y_1, R, y_2) \equiv \mathbb{P} (\theta = G | y_1, R, y_2) = \begin{cases} \frac{\pi \frac{1}{\sigma_y} \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) \psi (y_1) \phi \left( \frac{y_2 - Y_L}{\sigma_y} \right)}{p_W (y_1, y_2)} & \text{if } R = 1 \\ \frac{\pi \frac{1}{\sigma_y} \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) [1 - \psi (y_1)]}{p_L (y_1)} & \text{if } R = 0 \end{cases}
$$

which shows the ex-post assessment of a governor’s type that one could calculate after having observed her performance in both terms and the reelection outcome.

We present three examples to illustrate how our model works. The first is Guy Hunt, who was the governor of Alabama between 1987 and 1993. His first-term performance is 60.1, which is slightly lower than $Y_H$ but sufficiently far away from $Y_L$ for voters to believe that he was a good type with 70\% probability just prior to his reelection bid. This implies a 74\% chance that he would win the election. He won a second term, but had performance of only 38.2 in the second term. As a result his $\pi (y_1, R, y_2)$ is only 8\%. According to our model, he was therefore probably a bad governor who exerted high effort in the first term and low effort in his second term. Incidentally, he was forced to resign towards the end of his second term because he was convicted of theft, conspiracy and ethics violations, including taking money from his first-term inaugural fund for personal use. Our second example is Mitch Daniels, who was the governor of Indiana between 2005 and 2013. His first-term performance was only 48.7, leading the voters to think that he was a “good” type with only 31\% probability. Despite this, he won reelection, evidently due to a large positive realization of the election shock.\textsuperscript{23} His second-term performance was 65.2, above the $Y_H = 64.0$ mean of the high-effort performance distribution. As a result, ex post it seems like he was a good governor who exerted high effort in his first term but had an unlucky performance draw, on which our model puts 84\% probability. Our third example is David Beasley, who was the Republican governor of South Carolina from 1995 to 1998. His first term performance

\textsuperscript{23}The \textit{Washington Post} named his reelection campaign “The Best Gubernatorial Campaign of 2008” where he won with more votes than any candidate in the state’s history. We can consider this as evidence of a large positive election shock.
was 60.8, which would give him a 75% probability of winning his reelection bid. However, he lost reelection to the Democrat challenger Jim Hodges. This is likely to be a case where a good governor was unlucky in drawing a negative election shock and losing his reelection bid. Indeed his loss was considered to be a surprise given how heavily Republican the state had been at the time.\footnote{Owners of video poker machines spent very heavily on advertisements attacking Beasley who had worked to ban them in the state during his first term. It was also argued that many conservative Republicans did not turn out to vote on election day because of what they perceived as Beasley’s supposed “flip-flops” on moving the Confederate flag from on top of the Capitol.}

Finally, we can also talk about how good a fit our model provides to the data. In Table A1 we report $\psi(y_{1k})$, the model’s implied probability that an incumbent $k$ will win reelection after observing $y_{1k}$. If we select a rule that predicts reelection whenever $\psi(y_{1k}) > 0.5$, then we can correctly predict the reelection outcomes for 75% of the governors (49 wins and 21 losses) in our sample, incorrectly predicting only 15 wins and 8 losses. One way to assess the performance of a probability forecast such as $\psi(y_{1})$ is to use the Brier (1950) score, which is defined as $(1/n) \sum_{k=1}^{n} [\psi(y_{1k}) - R_k]^2$ where $R_k \in \{0, 1\}$ is the election outcome. The Brier score is between zero (a perfect prediction) and one, with smaller numbers indicating a better forecast. Our model gets a Brier score of 0.195. For comparison, a naive forecast that uses the overall fraction of governors who win in our sample (61.2%) for each governor instead of the $\psi(y_{1k})$ measure gets a Brier score of 0.237.\footnote{As a point of comparison, a naive forecast that the incumbent wins 50% of the time would lead to a Brier score of 0.25, regardless of the outcome.}

A Diebold-Mariano (1995) test, as in Lahiri and Yang (2013) rejects equal accuracy between our model’s forecast and the naive forecast with a $p$-value of 0.01. Our model places quite a bit of structure on the relationship between the observable variables (reelection outcomes and first-term JAR in this case), which in principle puts it at a disadvantage against a reduced-form model like a probit. However, an estimated probit model that uses JAR in the first term as a predictor (i.e., a reduced-form model using the same observables as our structural model) yields lower prediction accuracy, that is, a higher Brier score (though the difference is no longer statistically significant). Therefore, we believe that the performance of our model being at least as good as an alternative reduced-form model with the same information and significantly better than a naive forecast is quite impressive.

\section{Measuring the Effects of Elections}

Elections have three consequences in our model: discipline (bad governors exert high effort to secure reelection), selection (more good than bad governors are reelected), and mimicking
In order to measure the first two effects, we compare the outcomes in the benchmark model with a counterfactual model where governors can only serve one term. Having more disciplined governors improves first-term outcomes relative to the one-term case since more governors overall will exert high effort. In turn, when there is a second term, the selection effect can be measured as the improvement in outcomes in the second term of the benchmark model relative to the one-term counterfactual model, as good governors have a higher reelection rate than bad governors and then they exert high effort in their second term.

These effects are not independent of each other. To see this, consider the case where all bad governors are disciplined, which means that all governors, good or bad, exert high effort in their first term. As a result, there will be no information voters can use to screen governors, thus leading to identical fractions of each type of governor across the first and second terms, so that the percentages in each term would be identical to the one-term counterfactual. This means the outcome in the second term will be identical to the one-term outcome, that is, there is no selection effect. It is important to realize that the lack of selection is a negative consequence of having more disciplined governors in the first term. We call this third effect “mimicking” and to remember that it makes outcomes worse use a negative sign. Thus, we distinguish between “pure selection”, which is the screening effect of elections were there no mimicking, and selection as defined above. Naturally, selection is equal to pure selection plus the (negative) effect of mimicking.

In order to identify mimicking, we consider a second counterfactual, one where there is no discipline as an equilibrium outcome. To obtain this, we assume that the cost of exerting high effort for bad governors is \( c = 1 \), which means none of them exerts high effort. This ensures that \( \delta = 0 \) in equilibrium and (5) no longer is a part of the description of equilibrium. Naturally, the voter solves his problem taking into account that \( \delta = 0 \) and this influences all equilibrium mappings including, for example, the reelection rule and thus the equilibrium outcomes \( \rho_L, \rho_H \) and \( V \). We solve this equilibrium using the structural parameters in Table 3. Some details of outcomes in this counterfactual case are presented in Table 5 and we use values from this table in what follows.

Table 6 shows two different approaches to computing the three effects in question: discipline, selection and mimicking. The first approach, labeled A, uses the change in the fraction of high-effort governors, measured in percentage points, while the second approach, labeled B, uses the change in performance, both as absolute change in performance and also as relative to the counterfactual as we explain now. Comparing the benchmark version with the
one-term case, we find a 12.9 percentage point increase in the fraction of governors exerting high effort in their first term, which leads to an increase of 2.7 JAR points in performance, or a 4.9% increase. These are our measures of discipline. The effect of selection is lower in magnitude, namely a 7.8 percentage point increase in the fraction of high-effort governors in the second term, leading to a 1.6 JAR point or 2.9% increase in performance. However, the improvement in the second term due to selection is partially cancelled due to mimicking – 3.4 percentage point decline in fraction of high-effort governors in the second term, leading to a 0.7 JAR point or 1.2% decline in performance. We use bootstrapping methods to compute confidence intervals and all the estimates in panel (a) of Table 6 are significant at the 5% level.\footnote{The changes using JAR points seem smaller than those that use change in the fraction of governors exerting high effort. This is because of the stochastic relationship between effort and performance.}

In Panel (b) we compare the measures of discipline and selection. When we allow mimicking to reduce the gains due to selection, both types of measures, A and B, show that the effect of discipline is significantly different from selection – almost 5 percentage point difference in fraction of good governors in term 1 versus term 2 which leads to about a 2 percentage points increase in performance in term 1 versus in term 2. However, when we strip mimicking from selection and compute pure selection, the differences are not significant at the 10% level.

As we discussed in the previous section, voters are better off in the two-term-limit world relative to a one-term-limit one by about 4.2% in welfare. The decomposition in this section suggests that about 2/3 of this is due to the disciplining effect of elections: going from the one-term-limit counterfactual to the no-discipline counterfactual, welfare goes up from 360.3 to 365.5 while from the no-discipline counterfactual to the benchmark model welfare goes up from 365.5 to 375.3.

Returning to the discussion of the reduced-form estimates for discipline and selection in Section 2, results in this section clearly demonstrate the advantages of using a structural approach. In Section 2, a simple regression of JAR-based performance on an election-eligible dummy would lead one to conclude that there was no significant effect on governor performance of being election-eligible. We also argued that this coefficient will be the combination of the three effects we measured in this section following discipline−pure selection+mimicking. Table 6 shows that both discipline and pure selection are large and are highly statistically significant and that selection is reduced by mimicking, leading to the difference between discipline and selection (pure selection net of mimicking) to be statistically significant.
5.3 Robustness

In this section, we consider three types of exercises concerning the robustness of our parameter estimation. First, as explained in Section 4.3, we made some choices in preparing the JAR data for estimation. Here we consider some alternative choices. Second, we use the “adjusted JAR”, which strips the JAR data from what may be considered as partisan effects, as we also explain in Section 4.3. Third, we use two state-wide macroeconomic indicators as performance measures instead of JAR. The results are reported in Table 7, where in the interest of space, we only report our new estimates for $\pi$ and $\delta$ as well as one measure for discipline, selection and welfare.

Our benchmark measure of governor performance averaged the results of all JAR surveys over a governor’s first term up to and including June of the election year, where we used the fraction of respondents who classify the governor as excellent or good out of those who express an opinion (that is, eliminating the undecided respondents). The first six columns in Table 7 following the Benchmark, report the results from three alternative choices: using all surveys in the first term up to the election (All Surveys); dropping all surveys taken in the election year (No Election Year); taking the average JAR in each year of the term and then taking the year-by-year average so that respondent sentiment in a year with many surveys would not be overweighted (Year-by-Year Average); using the median (Median JAR) or the minimum JAR (Minimum JAR) rather than the average; and, taking the fraction of respondents who classified the governor as excellent or good out of all respondents including the undecided (Keep Undecideds), which essentially classifies the undecided as expressing low approval. As the estimates make clear, the results are robust to all of these alternative performance calculations. The key is that the identification of $\pi$ and $\delta$ is not affected by these variations – roughly half of all governors are bad and of those about a quarter of them are disciplined.

The next column shows the estimation results with the “adjusted JAR”. Results are virtually identical to the benchmark results. This is not surprising because the adjustments tend to be small – the median adjustment is about 2 JAR points. The last two columns show the estimation results with macroeconomic variables rather than JAR as measures of governor performance. First, we should note that the state unemployment rate is by far the most important determinant of JAR (that we could identify) and it is a reasonable predictor of reelection performance as we explained above. State per-capita personal income growth has a smaller correlation with both JAR and reelection outcomes. Both versions show smaller discipline and selection effects relative to our benchmark (about a third) but discipline is
still larger than selection. The estimated value of $\delta$ in the version with unemployment rate is similar to the benchmark value at 0.23 but in the version with income growth $\delta$ is small and statistically indistinguishable from zero at 5% significance. When we inspect the detailed results when using state-level macroeconomic indicators, we see however, that stochastic outcomes play a much larger role. (See also footnote 11) To see this, consider the ratio $(Y_H - Y_L)/\sigma_y$, as a measure of the importance of luck. This ratio is about 2.1 in our benchmark model, and it is 2.2 in the version with the unemployment rate. This means exerting high effort is expected to create about twice the increase in performance relative to one standard deviation increase in luck. In the income growth growth version, this ratio is only 0.4, indicating that luck is much, much more important. We find this an unattractive feature of using these narrower measures of performance, further strengthening the view presented in section 4.3 of the preferability of using a multi-faceted measure of performance such as JAR.

5.4 Noisy Effort Signal

The implications of a noisy effort signal counterfactual discussed in section 3.6 help to understand the importance of the election shock for the strength of discipline effects, as well as the trade-off between discipline and selection. Table 8 reports discipline and selection measures (analogous to Table 6) for different values of the partially and fully informative signals of governor effort, the latter both in the presence and absence of an election shock. Throughout this section, we assume the structural parameters shown in Table 3 are unchanged but solve for the equilibrium objects for every $\zeta$ considered. We show the re-computed $\delta$ in the table.

The first column shows the benchmark results presented above, which correspond to $\zeta = 0.5$ in this version. The second column shows the effect of a partially informative signal of effort, $\zeta = 0.75$. Relative to case of an uninformative (or no) signal, the fraction of bad governors disciplined rises from 27% to 30%. This is consistent with what theory would lead us to expect: a higher probability of observing “shirking” leads to more bad types exerting high effort. We also find a stronger selection effect, although the change is small; 3.2% instead of 2.9%. Hence, the higher selection effect due to observability is present as theory would suggest, but is small. The reason for this will become clear shortly.

The next column shows the effects of an increase in $\zeta$ to 0.9, which means the signal is correct 90% of the time. Now 35% of bad governors are disciplined, while selection is as large as it was with no signal, indicating a decline relative to $\zeta = 0.75$. The former result is once again obvious. The latter follows from the mimicking effect we discussed in Section 5.2
where more disciplined governors means more bad governors getting reelected because they look like good governors.

To better understand the magnitudes of these effects, we also considered the case of $\zeta = 1$, that is, perfect observability of effort, as shown in the fourth column of Table 8. (This, of course, is not equivalent to perfect observability of type, since bad governors still can, and do, mimic the effort levels of good governors.) We see that the fraction of bad governors disciplined in their first term rises to 42%, an increase by more than half of the 27% when effort was unobservable, but not by more as one might be inclined to expect. The reason why full observability of effort does not lead to all bad types exerting high effort in their first term is the existence of the election shock. Even if a governor is known to be of bad type – perfectly indicated in this case by low effort – she can still win reelection with a sufficiently positive realization of $\varepsilon$ (her reelection probability is $\rho_L = 0.23$); conversely, even if a bad type exerts high effort, she is not guaranteed reelection (her reelection probability is $\rho_H = 0.66$) if the realization of $\varepsilon$ is sufficiently negative. Therefore, bad types with a sufficiently high draw of $c$ will still find it optimal to exert low effort, even though it will be fully apparent to the voter that they did so. Hence, discipline is mitigated by the randomness of reelection outcomes due to reasons unrelated to performance, as theory once again would suggest. Turning to selection, what we called “pure selection” is virtually fully cancelled by mimicking and there is a negligible selection effect (zero up to rounding error). This shows that as a function of $\zeta$ selection has a hump shape. Here, increased precision of the reelection rule creates a tighter connection between effort and reelection.

To confirm our conjecture that the lack of full discipline is due to the presence of the election shock, we solve the model with full observability of effort ($\zeta = 1$) and with $\sigma_\varepsilon \approx 0$, so that the election shock is constrained to take its mean value $\mu = 25.5$. There is no longer the possibility of a very positive realization of $\varepsilon$ to “save” a low-effort incumbent. Now all bad governors exert high effort, and all are reelected. Mimicking of good governors by all bad governors implies there is no selection effect, and the fraction of good governors in the second term is identical to the fraction in the first term.

We can now see why partial observability of effort implied such a small increase in the selection effect relative to the case of no observability. As we discussed in Section 5.2, the mimicking by bad governors in the first term reduces the effect of selection by making it more difficult to distinguish types based on the performance signal. Perfect observability of effort (and hence low effort making it unambiguous that a governor is bad) does not induce perfect discipline on governors when reelection has a significant exogenous random compo-
nent. In the limit, when effort is perfectly observable and low effort guarantees electoral
defeat, discipline is perfect (that is, there will be no governors in the third group), but the
selection effect goes to zero precisely because of full mimicking by bad governors.

The last row in Table 8 shows how the welfare of the voter changes in each case. Having a
moderately informative effort signal is worth 0.5% of welfare to the voter while making effort
fully observable leads to an improvement of 2.1%, which is sizable. Much of the increase
in these cases come from the higher discipline. But these gains pale in comparison to the
one in the last column where in addition to making effort fully observable, we eliminate the
uncertainty of elections. This achieves a welfare improvement of 10.4% over the two-term
benchmark.

Our results show that greater transparency would not in itself significantly increase effort
especially due to the randomness of election outcomes (that is, their dependence on other
factors). If greater transparency made election outcomes themselves less stochastic, they
could increase effort and thus welfare significantly, as suggested by our final exercise.

6 Conclusions

In this paper we constructed a political agency model with adverse selection and moral
hazard, and we structurally estimated the model. The aim was to disentangle the various
effects that electoral accountability has on policymaker performance – specifically discipline
and selection effects – and, more generally, to assess the empirical relevance of the widely-
used political agency model.

Previous research has used a reduced-form approach to try to estimate the effects of elec-
torial accountability on discipline and selection, but this approach faces the difficulty of disen-
tangling the importance of these two determinants of policymaker performance. Structural
estimation allows us to separate empirically the discipline and selection effects of elections.
We estimated the effects on the performance of U.S. governors of the common two-term limit
regime relative to the counterfactual case where reelection is not allowed, so that elections
can neither discipline nor allow selection based on performance. A crucial advantage of a
structural model is the possibility of estimating specific parameters representing discipline
effects and the relative prevalence of governor types, a possibility that reduced-form estima-
tion does not allow. This is what allows counterfactual experiments to assess the welfare
effects of electoral accountability under different informational and electoral regimes.

We found a significant discipline effect of reelection incentives, as well as a somewhat
weaker selection effect. Quantifying these effects allows us to assess their relative importance. More generally, our results indicate that a formal political agency model stressing the role of accountability finds support in the data, an important point given the widespread use of the political agency approach in theoretical political economy models.

We further found that electoral accountability has important welfare implications, where the possibility of reelection induces a significant increase in welfare due to its inducing higher effort. We should note, as we made clear in the paper, that we consider only the discipline and selection effects of elections and the implied utility effects. There may be other effects – loss of experience induced by term limits versus allowing “new blood” to be injected into politics – that may have significant utility implications. However, a tractable structural model requires focusing on a limited number of issues, and we believe that those we chose are first order.\footnote{An obvious direction to take is extending the model to have a term limit of }$n > 2$ terms. Doing so takes some additional work – for example now governors will take in to account the outcome $y_{t-1}$ in making their effort decision $e_t$ in period $t$ since the former influences elections outcomes as well. We worked out the version with $n = 3$. However we choose not to pursue this further because we do not think it is reasonable to assume some of the parameter estimates remain unchanged in this new regime with a different term limit. For example the change in the term limit may change self-selection into politics and thus may alter $\pi$ or an incumbent facing reelection for a second time will likely draw election shocks from a different distribution than the one she drew from in the first reelection. This extension may thus require a richer model.

Further research may help address some basic questions raised by these results. Why is there such a large fraction of “bad” governors in the data? Why don’t reelection incentives discipline a larger fraction of them? Arguing that there is a large stochastic element to elections doesn’t really answer the second question. These two questions are likely related. For example, it is reasonable to think about the link between the importance of rents in politics and self-selection into politics.

In our opinion, structural estimation can be quite helpful in gaining a deeper econometric understanding of issues of politician performance and electoral accountability. We believe this paper is a useful step in that direction.
References


Table 1: Reduced-Form Analysis of Electoral Accountability

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Approval Rating (JAR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reelection Eligible</td>
<td>0.71</td>
<td>0.63</td>
<td>-1.72</td>
<td>5.76**</td>
<td>5.26**</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td>(2.22)</td>
<td>(2.12)</td>
<td>(2.74)</td>
<td>(2.44)</td>
<td>(1.88)</td>
</tr>
<tr>
<td>Survey Aggregation</td>
<td>Month</td>
<td>Year</td>
<td>Term</td>
<td>Month</td>
<td>Year</td>
<td>Term</td>
</tr>
<tr>
<td>Governors</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>Winners</td>
<td>Winners</td>
<td>Winners</td>
</tr>
<tr>
<td>States</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Observations</td>
<td>2,378</td>
<td>357</td>
<td>150</td>
<td>1,638</td>
<td>357</td>
<td>114</td>
</tr>
</tbody>
</table>

Note: Estimation is done via OLS. The unit of observation is an opinion survey on a given governor in a month (columns 1 and 4), year (columns 2 and 5) or a term (columns 3 and 6). Columns 1-3 use all governors while columns 4-6 restrict the sample to those who win reelection. The sample consists of states with a two-term limit. All regressions include year fixed effects (column 1 also includes month fixed effects) and state fixed effects, as well as governor controls (age, age squared, gender, years of education), political controls (party of governor, same party as president). Standard errors clustered at the state level. ** denotes significance at the 5% level.
Table 2: Determinants of Job Approval Ratings

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable = JAR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unemployment</strong></td>
<td>−2.93***</td>
<td>−3.09***</td>
<td>−2.80***</td>
<td>−2.94***</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.66)</td>
<td>(0.77)</td>
<td>(0.73)</td>
</tr>
<tr>
<td><strong>Income Growth</strong></td>
<td>0.72***</td>
<td>0.59**</td>
<td>0.75***</td>
<td>0.66***</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.24)</td>
<td>(0.25)</td>
<td>(0.24)</td>
</tr>
<tr>
<td><strong>Population Growth</strong></td>
<td>1.45</td>
<td>1.40*</td>
<td>1.56</td>
<td>1.57*</td>
</tr>
<tr>
<td></td>
<td>(0.90)</td>
<td>(0.82)</td>
<td>(0.96)</td>
<td>(0.85)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>−0.28**</td>
<td></td>
<td>−0.28**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td></td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td><strong>Lawyer</strong></td>
<td>2.58</td>
<td>2.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.86)</td>
<td>(1.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>−0.83</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.75)</td>
<td>(3.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Served in Military</strong></td>
<td>−1.66</td>
<td>−1.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.02)</td>
<td>(2.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Years of Education</strong></td>
<td>−0.02</td>
<td>−0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td></td>
<td>0.75</td>
<td>1.00*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.46)</td>
<td>(0.52)</td>
<td></td>
</tr>
<tr>
<td><strong>Same Party President</strong></td>
<td>−2.16*</td>
<td>−2.17**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.24)</td>
<td>(1.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Partisan Fit</strong></td>
<td>−2.28**</td>
<td>−1.96**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% Voters Governor Party</strong></td>
<td>−0.20</td>
<td>−0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% Voters Other Party</strong></td>
<td>0.13</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.39</td>
<td>0.41</td>
<td>0.42</td>
<td>0.44</td>
</tr>
</tbody>
</table>

*Note:* The regression include state and year fixed effects. It uses 1,020 governor-year observations over the period 1976-2010 and across 50 U.S. states. Second block of variables are characteristics of the governor. “Same party president” shows if the governor is from the same party as the U.S. President. See the text for the definition of “Partisan Fit”. “% Voters” variables shows the voters affiliated with the governor’s party and the other party, respectively. *, ** and *** represent significance at 10%, 5% and 1% levels, respectively, using standard errors that are clustered at the state level.
Table 3: Parameter Estimates

<table>
<thead>
<tr>
<th>Structural Parameters</th>
<th>Equilibrium Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$ 0.52</td>
<td>$\delta$ 0.27</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>$Y_L$ 43.33</td>
<td>$\rho_L$ 0.45</td>
</tr>
<tr>
<td>(2.49)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>$Y_H$ 63.99</td>
<td>$\rho_H$ 0.72</td>
</tr>
<tr>
<td>(1.67)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>$\sigma_y$ 9.84</td>
<td>$\Sigma$ 499.72</td>
</tr>
<tr>
<td>(0.80)</td>
<td>(31.95)</td>
</tr>
<tr>
<td>$\mu$ 25.53</td>
<td></td>
</tr>
<tr>
<td>(5.81)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_\epsilon$ 13.07</td>
<td></td>
</tr>
<tr>
<td>(4.32)</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* White standard errors are below estimates. Standard errors for the equilibrium objects are computed using the delta method. $\beta$ is fixed at 0.85.
Table 4: Some Properties of the Estimated Model

<table>
<thead>
<tr>
<th>One-Term Limit</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Good governors</td>
<td>51.8%</td>
</tr>
<tr>
<td>High Effort</td>
<td>51.8%</td>
</tr>
<tr>
<td>Average Performance (JAR Points)</td>
<td>54.0</td>
</tr>
<tr>
<td>Life-time Discounted Welfare for Voter</td>
<td>360.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Two-Term Limit (Benchmark)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Good governors in Term 1</td>
<td>51.8%</td>
</tr>
<tr>
<td>Good governors in Term 2</td>
<td>59.6%</td>
</tr>
<tr>
<td>Good governors Overall</td>
<td>54.8%</td>
</tr>
<tr>
<td>High effort in Term 1</td>
<td>64.7%</td>
</tr>
<tr>
<td>High effort in Term 2</td>
<td>59.6%</td>
</tr>
<tr>
<td>High effort Overall</td>
<td>62.8%</td>
</tr>
<tr>
<td>Average Performance in Term 1 (JAR Points)</td>
<td>56.7</td>
</tr>
<tr>
<td>Average Performance in Term 2 (JAR Points)</td>
<td>55.6</td>
</tr>
<tr>
<td>Average Performance Overall (JAR Points)</td>
<td>56.3</td>
</tr>
<tr>
<td>Life-time Discounted Welfare for Voter</td>
<td>375.3</td>
</tr>
</tbody>
</table>

*Note:* The numbers on this table are obtained by simulating the model for 1,000,000 governors, given the structural parameters in Table 3. The one-term limit assumes governors are not eligible to run for reelection.
Table 5: No-Discipline ($\delta = 0$) Counterfactual

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Good governors in Term 1</td>
<td>51.8%</td>
</tr>
<tr>
<td>Good governors in Term 2</td>
<td>63.0%</td>
</tr>
<tr>
<td>Good governors Overall</td>
<td>56.3%</td>
</tr>
<tr>
<td>High effort in Term 1</td>
<td>51.8%</td>
</tr>
<tr>
<td>High effort in Term 2</td>
<td>63.0%</td>
</tr>
<tr>
<td>High effort Overall</td>
<td>56.3%</td>
</tr>
<tr>
<td>Average Performance in Term 1 (JAR Points)</td>
<td>54.0</td>
</tr>
<tr>
<td>Average Performance in Term 2 (JAR Points)</td>
<td>56.3</td>
</tr>
<tr>
<td>Average Performance Overall (JAR Points)</td>
<td>55.0</td>
</tr>
<tr>
<td>Life-time Discounted Welfare for Voter</td>
<td>365.5</td>
</tr>
</tbody>
</table>

Notes: The no-discipline counterfactual is obtained by setting the cost of exerting high effort for bad governors to $c = 1$ so that they never exert high effort, and re-solving the model so that the voters optimally react to this. The solution uses the estimates of the structural parameters reported in Table 3.
Table 6: Discipline and Selection

(a) Measures of Discipline and Selection

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Change in Fraction of High-Effort Governors in Term 1 (Benchmark vs. 1-Term)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discipline A</td>
<td></td>
<td>12.9**</td>
</tr>
<tr>
<td>Discipline B</td>
<td>Change in Performance In Term 1 (Benchmark vs. 1-Term)</td>
<td>2.7 [4.9%]**</td>
</tr>
<tr>
<td>Selection A</td>
<td>Change in Fraction of High-Effort Governors in Term 2 (Benchmark vs. 1-Term)</td>
<td>7.8**</td>
</tr>
<tr>
<td>Selection B</td>
<td>Change in Performance In Term 2 (Benchmark vs. 1-Term)</td>
<td>1.6 [2.9%]**</td>
</tr>
<tr>
<td>Mimicking A</td>
<td>Change in Fraction of High-Effort Governors in Term 2 (Benchmark vs. δ = 0)</td>
<td>−3.4**</td>
</tr>
<tr>
<td>Mimicking B</td>
<td>Change in Performance In Term 2 (Benchmark vs. δ = 0)</td>
<td>−0.7 [−1.2%]**</td>
</tr>
</tbody>
</table>

Welfare Gain Relative to One-Term Limit 4.2%**

(b) Comparison

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discipline A - Selection A</td>
</tr>
<tr>
<td>Discipline B - Selection B</td>
</tr>
<tr>
<td>Discipline A - Pure Selection A</td>
</tr>
<tr>
<td>Discipline B - Pure Selection B</td>
</tr>
</tbody>
</table>

**Notes:** The numbers on this table are obtained by simulating the model for 1,000,000 governors, given the structural parameters in Table 3. See the notes to Table 4. The δ = 0 version is solved assuming c = 1. In panel (a) all changes in fractions (such as the ones for Discipline A, Selection A and Mimicking A measures) are reported as percentage point changes. In panel (b) Pure Selection is defined as Selection minus Mimicking, where Mimicking is negative to emphasize its welfare-reducing nature. Numbers in square brackets are in percentage point units and show the difference between the corresponding term in square brackets in panel (a). Numbers in panel (b) may not exactly correspond to the differences in the numbers in panel (a) due to rounding. In both panels (**) denotes significance at 5% level and (*) denotes significance at 10% level.
Table 7: Robustness of Estimation Results

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>All Surveys</th>
<th>No Election Year</th>
<th>Year-by-Year Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi)</td>
<td>0.52</td>
<td>0.50</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.10)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>(\delta)</td>
<td>0.27</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Discipline B</td>
<td>4.9%</td>
<td>4.8%</td>
<td>4.6%</td>
</tr>
<tr>
<td>Selection B</td>
<td>2.9%</td>
<td>2.8%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Welfare Gain</td>
<td>4.2%</td>
<td>4.0%</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Median JAR</th>
<th>Minimum JAR</th>
<th>Keep Undecideds</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi)</td>
<td>0.54</td>
<td>0.45</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.12)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>(\delta)</td>
<td>0.25</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Discipline B</td>
<td>4.5%</td>
<td>6.4%</td>
<td>4.3%</td>
</tr>
<tr>
<td>Selection B</td>
<td>2.9%</td>
<td>3.5%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Welfare Gain</td>
<td>3.9%</td>
<td>5.3%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adjusted JAR</th>
<th>Unemployment Rate</th>
<th>Income Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi)</td>
<td>0.52</td>
<td>0.42</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>(\delta)</td>
<td>0.26</td>
<td>0.23</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Discipline B</td>
<td>4.6%</td>
<td>1.9%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Selection B</td>
<td>2.8%</td>
<td>0.9%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Welfare Gain</td>
<td>3.9%</td>
<td>1.5%</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

Notes: The top of each panel show the re-estimated \(\pi\) and \(\delta\) for each case with standard errors in parentheses. See Table 6 for the definitions of the discipline and selection measures. Reported welfare gains are relative to the one-term regime.
Table 8: Results from the Version with An Effort Signal

<table>
<thead>
<tr>
<th>$\zeta = 0.5$</th>
<th>$\zeta = 0.75$</th>
<th>$\zeta = 0.9$</th>
<th>$\zeta = 1$</th>
<th>$\zeta = 1$ and $\sigma_c = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>0.27</td>
<td>0.30</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>Discipline B</td>
<td>4.9%</td>
<td>5.5%</td>
<td>6.5%</td>
<td>7.8%</td>
</tr>
<tr>
<td>Selection B</td>
<td>2.9%</td>
<td>3.2%</td>
<td>3.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Welfare Gain</td>
<td>-</td>
<td>0.5%</td>
<td>1.2%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

Notes: The first column ($\zeta = 0.5$) shows the benchmark results from Tables 2, 3 and 4. Structural parameters are kept as in Table 2. See Table 4 for the definitions of the discipline and selection measures. The “B” measures are reported in percent of the value in a world with a one-term limit. Welfare gain is relative to the benchmark with no signal (or equivalently with $\zeta = 0.5$).
Figure 1: Game Tree

\[
\begin{align*}
\text{Nature} & \quad \quad \pi \ (\text{Good}) \\
& \quad \quad 1 - \pi \ (\text{Bad}) \\
\text{Governor} & \quad e_1 = 1 \\
& \quad (y_1, \varepsilon) \\
\text{Voter} & \quad R = 1 \\
& \quad R = 0 \\
\text{Governor} & \quad e_1 = 0 \\
& \quad (y_1, \varepsilon) \\
\text{Nature} & \quad e_2 = 0 \\
\text{Voter} & \quad R = 1 \\
& \quad R = 0 \\
\text{Governor} & \quad e_2 = 0 \\
\text{Game Restarts} & \quad R = 1 \\
\text{Nature} & \quad e_2 = 1 \\
\text{Voter} & \quad R = 0 \\
\text{Governor} & \quad e_2 = 0 \\
\text{Game Restarts} & \quad R = 0 \\
\end{align*}
\]
Figure 2: Equilibrium Mappings

(a) Election shock threshold $\hat{\varepsilon}(y_1)$

(b) Posterior Probability $\hat{\pi}(y_1)$

(c) Reelection Probability $\psi(y_1)$

Notes: The red horizontal lines in the first and second panels are $\mu$ (the mean of the election shock process) and $\pi$ (the unconditional probability of a governor being good), respectively.
Appendix (For Online Publication)

A Existence and Uniqueness of Equilibrium

We begin with the following lemma that higher first-term performance increases the probability the voter assigns to the incumbent being good.

**Lemma 1** The posterior probability that a governor is good, \( \hat{\pi}(y_1) \), is a unique function that is monotonically increasing in \( y_1 \).

**Proof.** The definition of \( \hat{\pi}(y_1) \) as a posterior probability clearly implies that it is unique. To show that \( \hat{\pi} \) is increasing in \( y_1 \), note that all good governors exert high effort in the first term, while bad governors exert high effort in the first term only if \( c < \rho_H - \rho_L \). Since both \( \rho_H \) and \( \rho_L \in (0, 1) \) and since \( c = 1 \) for some governors, in any equilibrium some bad governors are exerting low effort in the first term. Since according to (6a) and (6b), the distribution of \( y_1 \) when \( \epsilon_1 = H \) first-order stochastically dominates that of \( y_1 \) when \( \epsilon_1 = L \), a higher value of \( y_1 \) must therefore raise the posterior probability that a governor is good. Formally, one can differentiate (15) with respect to \( y_1 \) to obtain

\[
\frac{d}{dy_1} \hat{\pi}(y_1) = \frac{\pi (1 - \pi) (1 - \delta)}{2\pi} \frac{(Y_H - Y_L) \exp \left\{ -\frac{1}{2} \left[ \left( \frac{y_1 - Y_H}{\sigma_y} \right)^2 + \left( \frac{y_1 - Y_L}{\sigma_y} \right)^2 \right] \right\}}{\sigma_y \left\{ \left[ \pi + (1 - \pi) \delta \right] \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) + (1 - \pi) (1 - \delta) \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) \right\}} > 0
\]

since \( Y_H > Y_L \) (where \( \pi \) is the prior probability that the governor is good and \( \pi \) is the mathematical constant “pi”).

**Proof of Proposition 1.** Consider first the voter’s problem. The solution to the voter’s problem (11) given in (12) yields a unique critical value for \( \hat{\pi}(y_1) \) for any \( \varepsilon \). This implies, given the above lemma, that there exists a unique cut-off value of \( y_1 \) for each \( \varepsilon \), below which the voter does not reelect the incumbent, at or above which he does.

The voter’s decision yields the reelection probability (16) that may be written

\[
\psi(y_1) = 1 - \Phi \left[ \frac{(1 - \beta) \nu - \hat{\pi}(y_1)(Y_H - Y_L) - Y_L - \mu}{\sigma_y} \right]
\]

where \( \Phi(.) \) denotes the CDF of a standard normal random variable. \( \psi(y_1) \) is unique, so that the reelection probabilities defined by (17) and (18) are unique and contained in \((0, 1)\).
lemma and $Y_H > Y_L$ further implies that $\psi(y_1)$ is monotonically increasing in $y_1$, so that first-order stochastic dominance implies not only that $\rho_H > \rho_L$, but also that the difference

$$
\rho_H - \rho_L = \frac{1}{\sigma_y} \int \psi(y_1) \left[ \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) - \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) \right] dy_1 \tag{A-2}
$$

is contained in $(0, 1)$ and is unique.

Now, consider governors. By assumption, good governors always supply high effort. The decision problem of a bad governor in (3) implies that her effort decision in the first term of office will be described by a single cutoff $c^* = \rho_H - \rho_L \in (0, 1)$, where $c^*$ is unique given that $\rho_H - \rho_L$ is unique in equilibrium. This also implies that $\delta = \rho_H - \rho_L \in (0, 1)$ is unique. Finally, $V$ – the voter’s life-time expected utility one period in the future – will obviously be unique as well in the equilibrium.

Hence, taking the decision problems of voters and governors together, the equilibrium described in the paper exists and is unique.

\section{Two-Term Model with a Noisy Effort Signal}

We here set out some of the key equations that would differ from the unobservable effort benchmark model to complement the discussion in the text. The voter’s value function, conditional on first-term observables would be:

$$
W(y_1, z_1, \varepsilon) = y_1 + \beta \max_{R \in \{0, 1\}} \mathbb{E} \left\{ R [y_2 + \varepsilon + \beta W(y'_1, z'_1, \varepsilon')] + (1 - R) W(y'_1, z'_1, \varepsilon') \bigg| y_1, z_1, \varepsilon \right\} \tag{A-3}
$$

which leads to (8) with the new definition of $V \equiv \mathbb{E}[W(y'_1, z'_1, \varepsilon')]$

$$
V = \left( \frac{\pi + (1 - \pi) \delta}{\sigma_y \sigma_\varepsilon} \right) \int \int \left[ \zeta W(y'_1, H, \varepsilon') + (1 - \zeta) W(y'_1, L, \varepsilon') \right] \phi \left( \frac{y'_1 - Y_H}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy'_1 d\varepsilon' \\
+ \left[ \frac{(1 - \pi) (1 - \delta)}{\sigma_y \sigma_\varepsilon} \right] \int \int \left[ \zeta W(y'_1, L, \varepsilon') + (1 - \zeta) W(y'_1, H, \varepsilon') \right] \phi \left( \frac{y'_1 - Y_L}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy'_1 d\varepsilon' 
$$
The incumbent’s posterior reputation becomes:

\[ \hat{\pi}(y_1, z_1) \equiv P(\theta = G|y_1, z_1) = \frac{p(y_1, z_1|\theta = G) P(\theta = G)}{p(y_1, z_1|\theta = G) P(\theta = G) + p(y_1, z_1|\theta = B) P(\theta = B)} \]

\[
= \begin{cases} 
\pi \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) & \text{if } z_1 = H \\
\pi \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) (1 - \zeta) + (1 - \pi) \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) & \text{if } z_1 = L
\end{cases}
\]

because

\[
p(y_1, z_1|\theta = G) = p(y_1, z_1|\theta = G, e_1 = H) P(e_1 = H|\theta = G) + p(y_1, z_1|\theta = G, e_1 = L) P(e_1 = L|\theta = G).
\]

and \(\hat{\pi}(y_1, z_1)\) replaces \(\hat{\pi}(y_1)\) in various equations such as (10) and (14).

Reelection probabilities conditional on voter information are:

\[
\psi(y_1, z_1) = P(R = 1|y_1, z_1) = \left[ \varepsilon > \hat{\varepsilon}(y_1, z_1) \right] \]

\[
= 1 - \Phi \left[ \frac{\hat{\varepsilon}(y_1, z_1) - \mu}{\sigma_\varepsilon} \right]
\]

We may then write reelection probabilities, as perceived by the incumbent:

\[
\rho_H = \frac{1}{\sigma_y} \int \left[ \zeta \psi(y_1, z_1 = H) + (1 - \zeta) \psi(y_1, z_1 = L) \right] \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) dy_1
\]

\[
\rho_L = \frac{1}{\sigma_y} \int \left[ (1 - \zeta) \psi(y_1, z_1 = H) + \zeta \psi(y_1, z_1 = L) \right] \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) dy_1
\]

C Some Computational Details

We need to evaluate some integrals numerically to obtain (23) and (24) in the text. Note that all the integrals we deal with have the following general form

\[
\int \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right) \xi(x) dx
\]

where \(x \sim N(\mu, \sigma^2)\) is a generic normal random variable and \(\xi(x)\) is a known function. Let’s apply a change of variables \(\tilde{x} = \frac{(x - \mu)}{\sqrt{2\sigma}}, \) where \(\tilde{x} \sim N(0, 0.5).\) This also means \(x = \sqrt{2\sigma}\tilde{x} + \mu.\)
Then, written using the pdf of \( x \) the integral simplifies to

\[
\int \frac{1}{\sqrt{\pi}} \exp\left(-\hat{x}^2\right) \xi \left(\sqrt{2\sigma} \hat{x} + \mu\right) d\hat{x}
\]

(A-9)

Finally, using Gauss-Hermite quadrature we can approximate this integral using

\[
\frac{1}{\sqrt{\pi}} \sum_{i=1}^{m} \omega_i \xi \left(\sqrt{2\sigma} \hat{x}_i + \mu\right)
\]

(A-10)

where the \( \hat{x}_i \) and \( \omega_i \) are the Gauss-Hermite quadrature nodes and weights respectively and \( m \) is the order of integration.

Turning to integrals in (23), they can be computed by a Gauss-Hermite approximation with \( \xi_1 (y_1) = \psi (y_1) \).

\[
R_1 \equiv \delta - \frac{1}{\sigma_y} \int \xi_1 (y_1) \phi \left( \frac{y_1 - Y_H}{\sigma_y} \right) dy_1 + \frac{1}{\sigma_y} \int \xi_1 (y_1) \phi \left( \frac{y_1 - Y_L}{\sigma_y} \right) dy_1
\]

(A-11)

As for the integrals in (24), the first one is

\[
A_1 \equiv \frac{1}{\sigma_y \sigma_\varepsilon} \int \int W (y'_1, \varepsilon') \phi \left( \frac{y'_1 - Y_H}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) dy'_1 d\varepsilon'
\]

(A-12)

\[
= \frac{1}{\sigma_y} \int \left\{ \frac{1}{\sigma_\varepsilon} \int W (y'_1, \varepsilon') \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) d\varepsilon' \right\} \phi \left( \frac{y'_1 - Y_H}{\sigma_y} \right) dy'_1
\]

(A-13)

Using the definition of \( W(.) \) in (11), \( \xi_2 (y'_1) \) can be written as

\[
\xi_2 (y'_1) = y'_1 + \beta \left\{ \hat{\pi} (y'_1) Y_H + [1 - \hat{\pi} (y'_1)] Y_L + \beta \mathcal{V} \right\} \frac{1}{\sigma_\varepsilon} \int_{\varepsilon(y'_1)}^{\infty} \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) d\varepsilon'
\]

(A-14)

\[
+ \beta \frac{1}{\sigma_\varepsilon} \int_{\varepsilon(y'_1)}^{\infty} \varepsilon' \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) d\varepsilon'
\]

\[
+ \frac{1}{\sigma_\varepsilon} \mathcal{V} \int_{-\infty}^{\varepsilon(y'_1)} \phi \left( \frac{\varepsilon' - \mu}{\sigma_\varepsilon} \right) d\varepsilon'
\]

(A-15)

\[
= y'_1 + \beta \left\{ \hat{\pi} (y'_1) Y_H + [1 - \hat{\pi} (y'_1)] Y_L + \beta \mathcal{V} \right\} \psi (y'_1)
\]

\[
+ \beta \left( \mu + \sigma_\varepsilon \frac{\phi \left( \varepsilon(y'_1) - \mu \right)}{\psi (y'_1)} \right)
\]

\[
+ \beta \mathcal{V} [1 - \psi (y'_1)]
\]
where for the second term we use the formula for the expected value of a truncated normal distribution

\[ \mathbb{E}(x|x > a) = \mu + \sigma \frac{\phi \left( \frac{a - \mu}{\sigma} \right)}{1 - \Phi \left( \frac{a - \mu}{\sigma} \right)} \text{ where } x \sim N(\mu, \sigma^2) \]  

(A-16)

Thus \( A_1 \) can be computed using a Gauss-Hermite approximation using \( \xi_2(y'_1) \).

The second integral in (24) can be computed analogously

\[
A_2 \equiv \frac{1}{\sigma_y \sigma_{\varepsilon}} \int \int W(y'_1, \varepsilon') \phi \left( \frac{y'_1 - Y_L}{\sigma_y} \right) \phi \left( \frac{\varepsilon' - \mu}{\sigma_{\varepsilon}} \right) dy'_1 d\varepsilon' \quad (A-17)
\]

where the only difference relative to \( A_1 \) is that the integration is with respect to the low-effort distribution. Thus we have

\[
\mathcal{R}_2 \equiv \nabla \left[ \pi + (1 - \pi) \delta \right] A_1 - (1 - \pi) (1 - \delta) A_2 \quad (A-18)
\]

D Incumbency Advantage

In Table A2 we show how the unconditional probability of winning reelections for incumbents, \( \mathbb{P}(R = 1) \), depends on structural parameters. Recall that we define incumbency advantage as \( \mathbb{P}(R = 1) > 0.5 \) in the text and in our sample 61.2% of incumbents win reelection. The key message of this appendix will be that incumbency advantage reflected in this number does not simply depend on \( \mu \), but on all structural parameters in complicated ways, so that \( \mu > 0 \) is neither necessary nor sufficient in itself for the model to match the data.

In Table A2, we start by a very simple configuration of parameters: \( \pi = 0.5, \mu = 0, \sigma_y = 0, \sigma_{\varepsilon} = 0 \) and \( \delta \) is forced to be zero by assuming very large cost of exerting effort for bad governors. In this case, all good governors exert high effort, none of the bad governors exert high effort, and since \( \sigma_y = 0 \) effort can be perfectly observed by the voters. They reelect all good governors and replace bad governors, yielding a 50% reelection probability of incumbents. In column 2 we change \( \pi \) to the estimated value of 0.52 and this simply increases the reelection probability to 0.52. This shows that the fraction of good governors is important for the incumbency advantage.

In the remaining columns we allow bad governors to optimally choose their effort and thus \( \delta \in (0, 1) \) in all cases. Column 3 uses the same parameter configuration as column 2 where now 42% of bad governors are disciplined and the reelection probability is reduced to
This is because in this equilibrium voters reelect only 42% of governors that produce a high performance (remember that effort leads deterministically to performance since $\sigma_y = 0$) and none of the governors that exert low effort and get low performance are reelected. This shows that bad governors being disciplined reduces incumbency advantage as the voters become worried about reelecting a bad governor who will exert low effort. This has a close link to the freshman effect.

In the rest of the columns we turn on the stochastic process for outcomes so that effort can no longer be inferred from performance. In column 4 now $\rho_L > 0$ since a lucky bad governor that exerts low effort can obtain a draw for $y_1$ that would get her reelected. Turning on the stochastic process for $y$ increases the reelection probability since now the voters realize that they cannot be sure about the type of the governor based on performance.

In the rest of the table we show how $\mu$ and $\sigma_\varepsilon$ affect the reelection probability. The last column shows the results with our benchmark parameters where the reelection probability is 63%. First in column (5) we increase $\mu$ to 5 while keeping $\sigma_\varepsilon = 0$. This gives a reelection probability that is exactly 63% – same as our benchmark specification. This means we can match the degree of incumbency advantage with $\mu = 5$ and $\sigma_\varepsilon = 0$ but, of course, this misses the observation in the data that there are a number of surprise wins and losses. In column 6 we increase $\sigma_\varepsilon$ to 1 while keeping $\mu$ at 5. This reduces the reelection probability to 0.36. In columns 7 and 8 we keep increasing $\sigma_\varepsilon$ towards the benchmark value of 13.07 and in order to keep the reelection probability near 63%, something estimation would aim to do since reelection outcomes are part of the observed data, we need to keep increasing $\mu$. This means that the value of $\mu$ does not pin down the incumbency advantage but once the rest of the parameters are pinned down based on other observations, then $\mu$ adjusts to deliver the right reelection probability, and thus the right degree of incumbency advantage.
Table A1: Governors

<table>
<thead>
<tr>
<th>State</th>
<th>Name (Year Facing Reelection)</th>
<th>$y_1$</th>
<th>$\hat{\pi}(y_1)$</th>
<th>$\psi(y_1)$</th>
<th>$y_2$</th>
<th>$\bar{\pi}(y_1, R, y_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>Fob James Jr. (1982)</td>
<td>26.6</td>
<td>0%</td>
<td>32%</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>AL</td>
<td>George C. Wallace (1986)</td>
<td>40.8</td>
<td>8%</td>
<td>37%</td>
<td>-</td>
<td>8%</td>
</tr>
<tr>
<td>AL</td>
<td>Guy Hunt (1990)</td>
<td>60.1</td>
<td>70%</td>
<td>74%</td>
<td>38.2</td>
<td>8%</td>
</tr>
<tr>
<td>AL</td>
<td>Fob James Jr. (1998)</td>
<td>48.0</td>
<td>28%</td>
<td>49%</td>
<td>-</td>
<td>28%</td>
</tr>
<tr>
<td>AL</td>
<td>Don Siegelman (2002)</td>
<td>61.2</td>
<td>72%</td>
<td>75%</td>
<td>-</td>
<td>72%</td>
</tr>
<tr>
<td>AL</td>
<td>Bob Riley (2006)</td>
<td>54.0</td>
<td>53%</td>
<td>65%</td>
<td>63.6</td>
<td>90%</td>
</tr>
<tr>
<td>AR</td>
<td>Mike Beebe (2010)</td>
<td>81.4</td>
<td>80%</td>
<td>79%</td>
<td>80.4</td>
<td>100%</td>
</tr>
<tr>
<td>CA</td>
<td>Pete Wilson (1994)</td>
<td>33.0</td>
<td>2%</td>
<td>33%</td>
<td>41.4</td>
<td>0%</td>
</tr>
<tr>
<td>CO</td>
<td>Bill Owens (2002)</td>
<td>71.0</td>
<td>79%</td>
<td>78%</td>
<td>57.1</td>
<td>89%</td>
</tr>
<tr>
<td>CO</td>
<td>Bill Ritter (2010)</td>
<td>56.9</td>
<td>63%</td>
<td>70%</td>
<td>-</td>
<td>63%</td>
</tr>
<tr>
<td>DE</td>
<td>Thomas R. Carper (1996)</td>
<td>65.8</td>
<td>77%</td>
<td>77%</td>
<td>80.7</td>
<td>100%</td>
</tr>
<tr>
<td>FL</td>
<td>Robert Graham (1982)</td>
<td>68.2</td>
<td>78%</td>
<td>78%</td>
<td>80.6</td>
<td>100%</td>
</tr>
<tr>
<td>FL</td>
<td>Bob Martinez (1990)</td>
<td>44.3</td>
<td>16%</td>
<td>41%</td>
<td>-</td>
<td>16%</td>
</tr>
<tr>
<td>FL</td>
<td>Lawton Chiles (1994)</td>
<td>38.7</td>
<td>6%</td>
<td>35%</td>
<td>50.0</td>
<td>3%</td>
</tr>
<tr>
<td>FL</td>
<td>Jeb Bush (2002)</td>
<td>66.2</td>
<td>77%</td>
<td>77%</td>
<td>57.9</td>
<td>89%</td>
</tr>
<tr>
<td>FL</td>
<td>Charlie Crist (2010)</td>
<td>66.3</td>
<td>77%</td>
<td>77%</td>
<td>-</td>
<td>77%</td>
</tr>
<tr>
<td>GA</td>
<td>Zell Miller (1994)</td>
<td>56.8</td>
<td>63%</td>
<td>70%</td>
<td>70.1</td>
<td>98%</td>
</tr>
<tr>
<td>GA</td>
<td>Roy Barnes (2002)</td>
<td>77.8</td>
<td>80%</td>
<td>79%</td>
<td>-</td>
<td>80%</td>
</tr>
<tr>
<td>GA</td>
<td>Sonny Perdue (2006)</td>
<td>61.2</td>
<td>72%</td>
<td>75%</td>
<td>58.1</td>
<td>87%</td>
</tr>
<tr>
<td>IN</td>
<td>Robert D. Orr (1984)</td>
<td>37.0</td>
<td>4%</td>
<td>34%</td>
<td>46.0</td>
<td>1%</td>
</tr>
<tr>
<td>IN</td>
<td>Evan Bayh (1992)</td>
<td>69.9</td>
<td>79%</td>
<td>78%</td>
<td>57.0</td>
<td>88%</td>
</tr>
<tr>
<td>IN</td>
<td>Mitch Daniels (2008)</td>
<td>48.7</td>
<td>31%</td>
<td>51%</td>
<td>65.2</td>
<td>84%</td>
</tr>
<tr>
<td>KS</td>
<td>Mike Hayden (1990)</td>
<td>50.0</td>
<td>36%</td>
<td>54%</td>
<td>-</td>
<td>36%</td>
</tr>
<tr>
<td>KS</td>
<td>Joan Finney (1994)</td>
<td>33.8</td>
<td>2%</td>
<td>33%</td>
<td>-</td>
<td>2%</td>
</tr>
<tr>
<td>KS</td>
<td>Kathleen Sebelius (2006)</td>
<td>63.6</td>
<td>75%</td>
<td>76%</td>
<td>64.9</td>
<td>97%</td>
</tr>
<tr>
<td>KY</td>
<td>Paul E. Patton (1999)</td>
<td>60.3</td>
<td>71%</td>
<td>74%</td>
<td>51.9</td>
<td>62%</td>
</tr>
<tr>
<td>KY</td>
<td>Ernie Fletcher (2007)</td>
<td>39.6</td>
<td>7%</td>
<td>36%</td>
<td>-</td>
<td>7%</td>
</tr>
<tr>
<td>LA</td>
<td>Edwin W. Edwards (1987)</td>
<td>21.3</td>
<td>0%</td>
<td>32%</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>LA</td>
<td>Charles Roemer (1991)</td>
<td>56.9</td>
<td>63%</td>
<td>70%</td>
<td>-</td>
<td>63%</td>
</tr>
<tr>
<td>LA</td>
<td>Edwin W. Edwards (1995)</td>
<td>32.2</td>
<td>1%</td>
<td>33%</td>
<td>-</td>
<td>1%</td>
</tr>
<tr>
<td>LA</td>
<td>Mike Foster (1999)</td>
<td>77.7</td>
<td>80%</td>
<td>79%</td>
<td>70.4</td>
<td>99%</td>
</tr>
</tbody>
</table>

Continued on next page
Table A1 – Governors (continued)

<table>
<thead>
<tr>
<th>State</th>
<th>Name (Year Facing Reelection)</th>
<th>$y_1$</th>
<th>$\hat{\pi}(y_1)$</th>
<th>$\psi(y_1)$</th>
<th>$y_2$</th>
<th>$\bar{\pi}(y_1, R, y_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>Kathleen Babineaux Blanco (2007)</td>
<td>46.1</td>
<td>21%</td>
<td>45%</td>
<td>-</td>
<td>21%</td>
</tr>
<tr>
<td>ME</td>
<td>John Baldacci (2006)</td>
<td>51.1</td>
<td>41%</td>
<td>57%</td>
<td>47.3</td>
<td>15%</td>
</tr>
<tr>
<td>MD</td>
<td>Harry Hughes (1982)</td>
<td>32.0</td>
<td>1%</td>
<td>33%</td>
<td>44.8</td>
<td>0%</td>
</tr>
<tr>
<td>MD</td>
<td>William D. Schaefer (1990)</td>
<td>61.9</td>
<td>73%</td>
<td>76%</td>
<td>33.6</td>
<td>4%</td>
</tr>
<tr>
<td>MD</td>
<td>Parris N. Glendening (1998)</td>
<td>36.5</td>
<td>4%</td>
<td>34%</td>
<td>55.4</td>
<td>5%</td>
</tr>
<tr>
<td>MD</td>
<td>Robert L. Ehrlich (2006)</td>
<td>57.1</td>
<td>63%</td>
<td>70%</td>
<td>-</td>
<td>63%</td>
</tr>
<tr>
<td>MD</td>
<td>Martin O’Malley (2010)</td>
<td>55.3</td>
<td>58%</td>
<td>67%</td>
<td>57.5</td>
<td>76%</td>
</tr>
<tr>
<td>MI</td>
<td>Jennifer Granholm (2006)</td>
<td>52.4</td>
<td>46%</td>
<td>61%</td>
<td>40.8</td>
<td>5%</td>
</tr>
<tr>
<td>MS</td>
<td>Ray Mabus (1991)</td>
<td>58.6</td>
<td>67%</td>
<td>72%</td>
<td>-</td>
<td>67%</td>
</tr>
<tr>
<td>MS</td>
<td>Kirk Fordice (1995)</td>
<td>50.1</td>
<td>37%</td>
<td>55%</td>
<td>53.8</td>
<td>38%</td>
</tr>
<tr>
<td>MS</td>
<td>Ronnie Musgrove (2003)</td>
<td>68.0</td>
<td>78%</td>
<td>78%</td>
<td>-</td>
<td>78%</td>
</tr>
<tr>
<td>MS</td>
<td>Haley Barbour (2007)</td>
<td>54.1</td>
<td>54%</td>
<td>65%</td>
<td>67.4</td>
<td>96%</td>
</tr>
<tr>
<td>MO</td>
<td>Bob Holden (2004)</td>
<td>40.2</td>
<td>8%</td>
<td>36%</td>
<td>-</td>
<td>8%</td>
</tr>
<tr>
<td>MO</td>
<td>Matt Blunt (2008)</td>
<td>41.3</td>
<td>9%</td>
<td>37%</td>
<td>-</td>
<td>9%</td>
</tr>
<tr>
<td>MT</td>
<td>Marc Racicot (1996)</td>
<td>77.4</td>
<td>80%</td>
<td>79%</td>
<td>79.9</td>
<td>100%</td>
</tr>
<tr>
<td>MT</td>
<td>Judy Martz (2004)</td>
<td>29.4</td>
<td>1%</td>
<td>33%</td>
<td>-</td>
<td>1%</td>
</tr>
<tr>
<td>MT</td>
<td>Brian Schweitzer (2008)</td>
<td>72.3</td>
<td>79%</td>
<td>78%</td>
<td>66.7</td>
<td>98%</td>
</tr>
<tr>
<td>NE</td>
<td>Bob Kerrey (1986)</td>
<td>84.0</td>
<td>80%</td>
<td>79%</td>
<td>-</td>
<td>80%</td>
</tr>
<tr>
<td>NE</td>
<td>Kay A. Orr (1990)</td>
<td>52.7</td>
<td>48%</td>
<td>61%</td>
<td>-</td>
<td>48%</td>
</tr>
<tr>
<td>NE</td>
<td>Ben Nelson (1994)</td>
<td>63.9</td>
<td>75%</td>
<td>77%</td>
<td>77.8</td>
<td>100%</td>
</tr>
<tr>
<td>NV</td>
<td>Kenny C. Guinn (2002)</td>
<td>59.7</td>
<td>70%</td>
<td>74%</td>
<td>62.6</td>
<td>94%</td>
</tr>
<tr>
<td>NV</td>
<td>Jim Gibbons (2010)</td>
<td>31.8</td>
<td>1%</td>
<td>33%</td>
<td>-</td>
<td>1%</td>
</tr>
<tr>
<td>NJ</td>
<td>Richard J. Hughes (1965)</td>
<td>58.1</td>
<td>66%</td>
<td>72%</td>
<td>64.1</td>
<td>95%</td>
</tr>
<tr>
<td>NJ</td>
<td>Thomas H. Kean (1985)</td>
<td>65.6</td>
<td>77%</td>
<td>77%</td>
<td>72.7</td>
<td>99%</td>
</tr>
<tr>
<td>NJ</td>
<td>James J. Florio (1993)</td>
<td>30.0</td>
<td>1%</td>
<td>33%</td>
<td>-</td>
<td>1%</td>
</tr>
<tr>
<td>NJ</td>
<td>Jon Corzine (2009)</td>
<td>48.1</td>
<td>29%</td>
<td>50%</td>
<td>-</td>
<td>29%</td>
</tr>
<tr>
<td>NM</td>
<td>Gary E. Johnson (1998)</td>
<td>51.2</td>
<td>42%</td>
<td>58%</td>
<td>53.7</td>
<td>42%</td>
</tr>
<tr>
<td>NM</td>
<td>Bill Richardson (2006)</td>
<td>62.2</td>
<td>74%</td>
<td>76%</td>
<td>55.3</td>
<td>80%</td>
</tr>
<tr>
<td>NC</td>
<td>James G. Martin (1988)</td>
<td>65.6</td>
<td>77%</td>
<td>77%</td>
<td>55.2</td>
<td>82%</td>
</tr>
<tr>
<td>NC</td>
<td>James B. Hunt Jr. (1996)</td>
<td>68.6</td>
<td>78%</td>
<td>78%</td>
<td>69.9</td>
<td>99%</td>
</tr>
<tr>
<td>NC</td>
<td>Mike Easley (2004)</td>
<td>56.9</td>
<td>63%</td>
<td>70%</td>
<td>59.3</td>
<td>85%</td>
</tr>
<tr>
<td>NC</td>
<td>Bev Perdue (2012)</td>
<td>44.3</td>
<td>16%</td>
<td>41%</td>
<td>-</td>
<td>16%</td>
</tr>
</tbody>
</table>

Continued on next page
Table A1 – Governors (continued)

<table>
<thead>
<tr>
<th>State</th>
<th>Name (Year Facing Reelection)</th>
<th>$y_1$</th>
<th>$\hat{\pi}(y_1)$</th>
<th>$\psi(y_1)$</th>
<th>$y_2$</th>
<th>$\bar{\pi}(y_1, R, y_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH</td>
<td>Richard F. Celeste (1986)</td>
<td>55.9</td>
<td>60%</td>
<td>68%</td>
<td>63.6</td>
<td>93%</td>
</tr>
<tr>
<td>OH</td>
<td>George V. Voinovich (1994)</td>
<td>64.3</td>
<td>76%</td>
<td>77%</td>
<td>74.7</td>
<td>100%</td>
</tr>
<tr>
<td>OH</td>
<td>Bob Taft (2002)</td>
<td>83.9</td>
<td>80%</td>
<td>79%</td>
<td>20.3</td>
<td>0%</td>
</tr>
<tr>
<td>OH</td>
<td>Ted Strickland (2010)</td>
<td>59.6</td>
<td>69%</td>
<td>74%</td>
<td>-</td>
<td>69%</td>
</tr>
<tr>
<td>OK</td>
<td>Henry L. Bellmon (1990)</td>
<td>45.8</td>
<td>20%</td>
<td>44%</td>
<td>-</td>
<td>20%</td>
</tr>
<tr>
<td>OK</td>
<td>David Walters (1994)</td>
<td>40.8</td>
<td>8%</td>
<td>37%</td>
<td>-</td>
<td>8%</td>
</tr>
<tr>
<td>OK</td>
<td>Frank Keating (1998)</td>
<td>68.2</td>
<td>78%</td>
<td>78%</td>
<td>60.5</td>
<td>94%</td>
</tr>
<tr>
<td>OK</td>
<td>Brad Henry (2006)</td>
<td>69.0</td>
<td>78%</td>
<td>78%</td>
<td>71.0</td>
<td>99%</td>
</tr>
<tr>
<td>OR</td>
<td>Barbara Roberts (1994)</td>
<td>28.6</td>
<td>1%</td>
<td>32%</td>
<td>-</td>
<td>1%</td>
</tr>
<tr>
<td>OR</td>
<td>John Kitzhaber (1998)</td>
<td>61.8</td>
<td>73%</td>
<td>75%</td>
<td>59.8</td>
<td>91%</td>
</tr>
<tr>
<td>OR</td>
<td>Ted Kulongoski (2006)</td>
<td>47.2</td>
<td>25%</td>
<td>47%</td>
<td>46.0</td>
<td>6%</td>
</tr>
<tr>
<td>PA</td>
<td>Richard L. Thornburgh (1982)</td>
<td>49.6</td>
<td>35%</td>
<td>53%</td>
<td>44.8</td>
<td>8%</td>
</tr>
<tr>
<td>PA</td>
<td>Robert P. Casey (1990)</td>
<td>81.3</td>
<td>80%</td>
<td>79%</td>
<td>33.2</td>
<td>5%</td>
</tr>
<tr>
<td>PA</td>
<td>Ed Rendell (2006)</td>
<td>55.0</td>
<td>57%</td>
<td>67%</td>
<td>52.3</td>
<td>50%</td>
</tr>
<tr>
<td>RI</td>
<td>Lincoln C. Almond (1998)</td>
<td>44.5</td>
<td>17%</td>
<td>42%</td>
<td>55.7</td>
<td>23%</td>
</tr>
<tr>
<td>RI</td>
<td>Donanld Carcieri (2006)</td>
<td>59.4</td>
<td>69%</td>
<td>73%</td>
<td>46.4</td>
<td>32%</td>
</tr>
<tr>
<td>SC</td>
<td>Carroll Campbell (1990)</td>
<td>73.3</td>
<td>79%</td>
<td>78%</td>
<td>70.6</td>
<td>99%</td>
</tr>
<tr>
<td>SC</td>
<td>David Beasley (1998)</td>
<td>60.8</td>
<td>71%</td>
<td>75%</td>
<td>-</td>
<td>71%</td>
</tr>
<tr>
<td>SC</td>
<td>Jim Hodges (2002)</td>
<td>46.7</td>
<td>23%</td>
<td>46%</td>
<td>-</td>
<td>23%</td>
</tr>
<tr>
<td>SC</td>
<td>Mark Sanford (2006)</td>
<td>54.9</td>
<td>56%</td>
<td>66%</td>
<td>47.4</td>
<td>25%</td>
</tr>
<tr>
<td>SD</td>
<td>M. Michael Rounds (2006)</td>
<td>75.0</td>
<td>80%</td>
<td>79%</td>
<td>65.1</td>
<td>98%</td>
</tr>
<tr>
<td>TN</td>
<td>Ned R. McWherter (1990)</td>
<td>70.3</td>
<td>79%</td>
<td>78%</td>
<td>51.4</td>
<td>70%</td>
</tr>
<tr>
<td>TN</td>
<td>Don Sundquist (1998)</td>
<td>61.6</td>
<td>73%</td>
<td>75%</td>
<td>47.7</td>
<td>43%</td>
</tr>
<tr>
<td>TN</td>
<td>Phil Bredesen (2006)</td>
<td>61.8</td>
<td>73%</td>
<td>75%</td>
<td>74.2</td>
<td>100%</td>
</tr>
<tr>
<td>WV</td>
<td>Arch A. Moore Jr. (1988)</td>
<td>42.7</td>
<td>12%</td>
<td>39%</td>
<td>-</td>
<td>12%</td>
</tr>
<tr>
<td>WV</td>
<td>Gaston Caperton (1992)</td>
<td>28.3</td>
<td>1%</td>
<td>32%</td>
<td>36.7</td>
<td>0%</td>
</tr>
<tr>
<td>WV</td>
<td>Cecil H. Underwood (2000)</td>
<td>59.7</td>
<td>70%</td>
<td>74%</td>
<td>-</td>
<td>70%</td>
</tr>
<tr>
<td>WV</td>
<td>Bob Wise (2004)</td>
<td>58.9</td>
<td>68%</td>
<td>73%</td>
<td>-</td>
<td>68%</td>
</tr>
<tr>
<td>WY</td>
<td>Jim Geringer (1998)</td>
<td>66.3</td>
<td>77%</td>
<td>77%</td>
<td>50.0</td>
<td>61%</td>
</tr>
<tr>
<td>WY</td>
<td>Dave Freudenthal (2006)</td>
<td>73.4</td>
<td>79%</td>
<td>78%</td>
<td>78.0</td>
<td>100%</td>
</tr>
</tbody>
</table>

Notes: $y_1$ and $y_2$, if the governor is reelected, show the JAR performance of the governor. $\hat{\pi}(y_1)$ is the updated probability of the governor being good, and $\psi(y_1)$ is the probability that the governor will win re-election, both conditional on first-term performance. $\bar{\pi}(y_1, R, y_2)$ is the probability that the governor is good, having observed both terms’ performance, where available.
Table A2: Incumbency Advantage and Structural Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$</td>
<td>0.50</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>$\sigma_y$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9.84</td>
<td>9.84</td>
<td>9.84</td>
<td>9.84</td>
<td>9.84</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>5.00</td>
<td>5.00</td>
<td>22.00</td>
<td>25.53</td>
<td></td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>10.00</td>
<td>13.07</td>
</tr>
</tbody>
</table>

Equilibrium Objects

| $\delta$ | 0 | 0 | 0.42 | 0.38 | 0.53 | 0.48 | 0.33 | 0.27 |
| $\rho_L$ | - | - | 0.00 | 0.01 | 0.23 | 0.03 | 0.40 | 0.45 |
| $\rho_H$ | - | - | 0.42 | 0.39 | 0.77 | 0.51 | 0.73 | 0.72 |

$\mathbb{P}(R = 1)$ | 0.50 | 0.52 | 0.31 | 0.36 | 0.63 | 0.36 | 0.63 | 0.63 |

Notes: The parameters not shown are kept fixed at their values in Table 2. In columns (1) and (2) $\delta$ is forced to be 0, which means the bad governors are assumed to have high enough cost of exerting effort so that none of them choose to exert high effort.