A Structural Model of Electoral Accountability*

S. Boragan Aruoba† Allan Drazen‡ Razvan Vlaicu§

This Draft: April 16, 2015
First Draft: January 31, 2015

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 for 1982-2012. We find that the possibility of reelection provides a significant incentive for incumbents to exert effort. We also find a selection effect, although it is weaker in terms of its effect on average governor performance. According to our model, the widely-used two-term regime improves voter welfare by 4.2% compared to a one-term regime, and find that a three-term regime may improve voter welfare even further.

JEL Classification: D72, D73, H70, C57

Keywords: discipline, selection, political agency, elections, structural estimation, maximum likelihood

---

*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, and École Polytechnique for useful comments, and Seth Wechsler, Pablo Cuba-Borda and Camilo Morales-Jimenez for research assistance at various stages of the project.

†University of Maryland, Department of Economics, College Park, MD 20742, United States. E-mail: aruoba@econ.umd.edu.

‡University of Maryland, Department of Economics, College Park, MD 20742, United States, NBER, and CEPR. (corresponding author) E-mail: drazen@econ.umd.edu.

§Northwestern University, Kellogg School of Management, 2001 Sheridan Road, MEDS 5th floor, Evanston, IL 60208, United States. E-mail: r-vlaicu@kellogg.northwestern.edu.

1 Introduction

A key aspect of a well-functioning democracy is the accountability of officials via elections. Elections may improve outcomes by giving incumbents incentives to exert effort, thus *disciplining* poor performance (Barro [1973], Ferejohn [1986]). They may serve a *selection* function by screening out low performers (Banks and Sundaram [1993], Fearon [1999], Smart and Sturm [2013], Duggan and Martinelli [2015]), but may also lead incumbents to *pander* to voters with policies that improve their chances of reelection even if they are not socially beneficial.[1]

One may thus ask what are the effects of barring reelection at some point by introducing limits on the number of terms that a politician may serve. Though term limits may reduce electoral pandering and prevent politicians from becoming too "entrenched" in office and thus unresponsive to voter concerns, they may also reduce the incentives for incumbents to exert effort. They may imply a loss of the benefits of the experience gained by veteran lawmakers. Term limits may also reduce the information that voters have about candidates, negatively impacting the screening function of elections.[2] Separating and quantifying these various effects provides a significant challenge to assessing both the positive and the normative effects of imposing or changing term limits. Examining the effect of term limits further addresses the larger issue of electoral accountability in the political agency model. The wide application of this model in political economy suggests the importance of assessing its empirical relevance.

Many papers have used a reduced-form approach to try to estimate the effects of term limits. We discuss these papers in greater detail in Section 2. By its very nature, reduced-form estimation faces the difficulty of disentangling the importance of various factors – such as discipline versus selection – on the net effect of term limits. Nor can such an approach be used to consider counterfactual experiments central to assessing the welfare impact of term limits.

This paper proposes a structural approach to measuring the effects of term limits and thus electoral accountability. We set out a political agency model with adverse selection and moral hazard. In the model politicians are of two types: "good", who are always willing to exert high effort; and "bad", who would exert low effort in the absence of incentives, such as the possibility of another term in the office. Neither the effort level chosen by politicians

[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.

[2] These effects may also characterize indirectly-elected policymakers, as in Vlaicu and Whalley (2014).

nor their type are observable to voters. Instead, they observe incumbent performance, an outcome that partially depends on the effort. Voters use observed performance to decide whether or not to reelect the politician. We design our model to mimic those U.S. states where governors have a two-term limit in office, currently the most prevalent regime. As such, in their second term politicians choose their effort level, understanding that the reelection incentive is no longer relevant.

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. We consider 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. The structural model also allows us to run experiments to assess the welfare effects of changing term limits, where the invariance of structural parameters to the term-limit regime is critical in avoiding the Lucas (1976) critique.[3]

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 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 choose 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 is not stronger because of a stochastic relation between effort and 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 lifetime welfare of about 4.2% relative to the case of a one-term limit. More than half of this gain in welfare comes from the disciplining effect 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 selection effects would be roughly the same size, but mimicking reduces the latter by about

[3]Parameters are estimated using data from governors limited to two terms. Holding these parameters constant, we consider alternative regimes (such as a one-term or a three-term limit) where both the politicians and voters in the economy optimally respond to the changed incentives implied by the different electoral regimes.

40%.

We then consider a version of the model where effort is at least partially observable. This leads to increased discipline, but this effect is imperfect 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 large incentive to exert high effort. Even if effort were fully observable, only 43% of bad governors are disciplined, leading to a modest 1.6% increase in welfare relative to the case of unobservable effort.

Finally, we consider an electoral reform where we increase term limits from two to three terms. This reform leads to a large increase in discipline (61% of bad governors exert high effort in their first term and 26% of them do so in their second term), leading to an increase in the welfare of voters of 2.8% relative to the two-term benchmark.

The plan of the paper is as follows. In the next section we briefly review the literature on empirical estimation of the effects of electoral accountability. In section 3 we present our political agency model with a two-term limit. Section 4 describes our solution and estimation methods and data. We then present and discuss our estimates and their implications in Section 5. The final section presents conclusions. An appendix presents technical details.

2 Literature

As indicated above, there have been a number of papers using reduced-form estimation to test the effects of term limits on politician performance. For example, Besley and Case (1995, 2003), Besley (2006), and Alt, Bueno de Mesquita and Rose (2007) 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 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 performance, but this comparison cannot in itself reveal the relative strengths of discipline and selection effects in generating these outcomes. Disciplining governors eligible for reelection to perform better makes it harder to distinguish good and bad governors on the basis of outcomes, thus weakening the selection effect. Hence, the comparison of performance of reelection-eligible governors and lame-duck governors can only reveal a net effect.[4]

Some of the above research makes further assumptions to try to disentangle the effects.

[4] Ashworth (2012) makes a similar point in his excellent survey of research on electoral accountability.

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 strong 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 (2007) 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 different number of times but cannot be reelected again).[5]

As suggestive as these arguments are, they often rely on special 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 (2007) cannot reject the hypothesis that the discipline and selection effects are almost equal in magnitude. Structural estimation allows us to do that and, moreover, does not rely on comparison of outcomes across regimes (for example number of terms to which reelection is possible), where arguably other things have changed.

To our knowledge, the only structural approach to the effects of reelection has been that of Sieg and Yoon (2014).[6] They ask whether the mechanism of reelection gives an incentive for incumbents 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 some important respects. A key one is a different focus, where our paper looks at the incentive effects of reelection on governor effort and overall performance rather than on the stance of fiscal policies as in their paper. As such, the papers are complementary. Second, neither the moral hazard problem of low effort, central to our analysis, nor selection

[5] 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 does so disproportionately through attracting more competent types.

[6] Structural estimation is relatively rare in political economy. Some examples are Merlo (1997), Diermeier, Eraslan, and Merlo (2003), and Strömberg (2008).

over non-partisan characteristics, such as competence, play a role in Sieg and Yoon (2014), where competence is assumed to be fully observed. Hence, given the aim of their paper, there is no attempt to measure the contribution of selection versus discipline on improving outcomes, a focus of much of the earlier literature and of our paper.

3 Model

As our benchmark model, we start with a simple political agency model with voters and politicians that can generate stochastic policy outcomes and reelection rules. Subsequent versions of the model relax some of this 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 $0 \leq \pi \leq 1$. 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 θ, where

$$c(H;G) = c(L;G) = c(L;B) = 0 \text{ and } c(H;B) = c \tag{1}$$

We assume that, like the governor's type θ, the cost c is observed by the governor but unobserved by the electorate. A bad governor draws c from a uniform distribution on the

[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.

unit interval $[0,1]$ when first elected, where c remains the same in all terms while in office.[8] The governor understands that her chance of winning reelection is ρ_H if she exerts high effort and ρ_L if she exerts low effort, where in equilibrium $\rho_L < \rho_H$. Different levels of effort lead to different distributions of observed possible outcomes (as specified in equations (5) below). Hence, these probabilities 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 θ is

$$\max_{e_1,e_2} \left[1 - c\left(e_1;\theta\right)\right] r + \left[\mathbf{1}_H \rho_H + \left(1 - \mathbf{1}_H\right)\rho_L\right]\left[1 - c\left(e_2;\theta\right)\right] r \tag{2}$$

where e_i is effort in term i and $\mathbf{1}_H$ is an index 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 $\left(1 - c + \rho_H\right) r$, and if she exerts low effort, her payoff is $\left(1 + \rho_L\right) r$. 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

$$c < \rho_H - \rho_L \tag{3}$$

The voter does not observe c, but understands the maximization problem that governors face.

[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.

[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.

He therefore can calculate the probability δ that a bad governor exerts high effort in her first term, that is, $\delta \equiv \mathbb{P}\{e_1 = H | \theta = B\}$. Given the assumption of a uniform distribution for c, we may then write

$$\delta = \mathbb{P}(c < \rho_H - \rho_L) = \rho_H - \rho_L \tag{4}$$

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. For simplicity, we assume 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

$$y_i | (e_i = H) \sim N\left(Y_H, \sigma_y^2\right) \tag{5a}$$
$$y_i | (e_i = L) \sim N\left(Y_L, \sigma_y^2\right) \tag{5b}$$

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. Furthermore $\mu > 0$ will capture an incumbency advantage, as will be clear below.

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

$$W(y_1, \varepsilon) = y_1 + \beta \max_{R \in \{0,1\}} \mathbb{E}\{R[y_2 + \varepsilon + \beta W(y_1', \varepsilon')] + (1 - R) W(y_1', \varepsilon') | y_1, \varepsilon\} \tag{6}$$

where β is the voter's discount rate between electoral terms, and R is the decision to re-elect. After observing the performance of the incumbent governor, the voter makes his reelection choice. If he reelects the governor, he will enjoy her second term performance, which will be followed by the election of a new governor drawn from the pool of candidates. The successor governor will deliver a first-term performance y_1' and face a reelection shock of ε'. If the voter does not reelect the incumbent, then a fresh draw from the pool of candidates occurs.

The election shock ε shows up as an additive term to the utility of the voter, where a positive ε makes the incumbent more appealing relative to the challenger. Note that ε does not affect the type or actions of the challenger that the incumbent faces. It is also important to note that the voter realizes that he may arrive at this node with (y_1, ε) in one of three ways: a good governor, a bad governor who exerted high effort, and 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

$$W(y_1, \varepsilon) = y_1 + \beta \max_{R \in \{0,1\}} \{R[\mathbb{E}(y_2|y_1) + \varepsilon + \beta \mathbb{V}] + (1-R)\mathbb{V}\} \tag{7}$$

where we use $\mathbb{V}$ to denote $\mathbb{E}[W(y_1', \varepsilon')]$ which is a constant since none of the stochastic variables are persistent. It can be written as

$$\begin{aligned} \mathbb{V} \;=\; & [\pi + (1-\pi)\delta] \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_H}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \\ & + (1-\pi)(1-\delta) \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_L}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \end{aligned} \tag{8}$$

where $\phi(\cdot)$ represents the standard normal PDF. Equation (8) 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

$$\mathbb{E}(y_2|y_1) = \hat{\pi}(y_1) Y_H + [1 - \hat{\pi}(y_1)] Y_L \tag{9}$$

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 (9) we can write $W(y_1, \varepsilon)$ as

$$W(y_1, \varepsilon) = y_1 + \beta \max_{R \in \{0,1\}} [R\{\hat{\pi}(y_1) Y_H + [1 - \hat{\pi}(y_1)] Y_L + \varepsilon + \beta \mathbb{V}\} + (1-R)\mathbb{V}] \tag{10}$$

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 neither type nor effort is observable, and due to the existence of the election shock, the reelection decision is not linked deterministically to y_1. Solving the discrete choice problem in (10), the incumbent would win reelection, i.e. $R = 1$, if and only if

$$\hat{\pi}(y_1) > \frac{(1-\beta)\mathbb{V} - Y_L - \varepsilon}{Y_H - Y_L} \tag{11}$$

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 ε is not too small or too negative (so that the incumbent does not have too small an incumbency advantage). We can summarize the decision rule $R(y_1, \varepsilon)$ with the following

$$R(y_1, \varepsilon) = \begin{cases} 0 & \text{if } \varepsilon \leq \hat{\varepsilon}(y_1) \\ 1 & \text{if } \varepsilon > \hat{\varepsilon}(y_1) \end{cases} \tag{12}$$

where $\varepsilon = \hat{\varepsilon}(y_1)$ characterizes the points (y_1, ε) for which (11) holds with equality with

$$\hat{\varepsilon}(y_1) = (1-\beta)\mathbb{V} - \hat{\pi}(y_1)(Y_H - Y_L) - Y_L \tag{13}$$

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

$$\begin{aligned} \hat{\pi}(y_1) &\equiv \mathbb{P}(\theta = G|y_1) = \frac{\mathbb{P}(\theta = G)\, p(y_1|\theta = G)}{p(y_1)} \\ &= \frac{\pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)}{[\pi + (1-\pi)\delta]\,\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) + (1-\pi)(1-\delta)\,\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)} \end{aligned} \tag{14}$$

where δ, as defined in (4), 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

$$\begin{aligned} \psi(y_1) &\equiv \mathbb{P}(R = 1|y_1) = \mathbb{P}[\varepsilon > \hat{\varepsilon}(y_1)] \\ &= 1 - \Phi\left[\frac{\hat{\varepsilon}(y_1) - \mu}{\sigma_\varepsilon}\right] \end{aligned} \tag{15}$$

Finally, the last piece we need is the probabilities ρ_L and ρ_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

$$\rho_H = \int \psi(y_1)\,\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 \tag{16}$$

$$\rho_L = \int \psi(y_1)\,\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \tag{17}$$

3.5 Equilibrium

Good governors always exert effort. A strategy for a bad governor is a choice of whether or not to exert effort, i.e. $e_i(B, c) \in \{H, L\}$, 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 ε. 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 $(\rho_H, \rho_L, \delta, \mathbb{V})$, where δ is the probability that a bad governor exerts first-term effort (equivalently, the fraction of disciplined reelection-eligible bad governors), $\mathbb{V}$ is the voter's life-time discounted utility, and ρ_H, ρ_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* $(\rho_L, \rho_H, \delta, \mathbb{V})$ *where*

1. *Given* δ*, the voter's choices lead to* ρ_L, ρ_H *and* $\mathbb{V}$*.*

2. *Given* ρ_L*,* ρ_H *and* $\mathbb{V}$*, a bad governor's choice of* e_1 *leads to* δ*.*

To summarize, 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. 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 θ. He updates the probability that the incumbent is type G using $\hat{\pi}(y_1)$.

3. An election shock ε 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 $\hat{\pi}(y_1)$, the election shock ε, and her performance y_1, 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.

3.6 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. We denote this signal by z and assume that it is symmetric and correct with probability ζ, that is

$$\zeta \equiv \mathbb{P}\{z = H|e = H\} = \mathbb{P}\{z = L|e = L\} \tag{18}$$

where $\frac{1}{2} \leq \zeta \leq 1$. The parameter ζ 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 π that the incumbent is good.

Using z_1 to denote the signal regarding e_1, the posterior would be defined by

$$\begin{aligned}
\hat{\pi}(y_1, z_1) &\equiv \mathbb{P}(\theta = G | y_1, z_1) = \frac{\pi p(y_1, z_1 | \theta = G)}{\pi p(y_1, z_1 | \theta = G) + (1-\pi) p(y_1, z_1 | \theta = B)} \\
&= \begin{cases} \dfrac{\pi \zeta \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)}{[\pi + (1-\pi)\delta] \zeta \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) + (1-\pi)(1-\delta)(1-\zeta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)} & \text{if } z_1 = H \\ \dfrac{\pi (1-\zeta) \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)}{[\pi + (1-\pi)\delta](1-\zeta)\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) + (1-\pi)(1-\delta)\zeta\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)} & \text{if } z_1 = L \end{cases} \qquad (19)
\end{aligned}$$

which would then be used in calculating the voter's expected utility from reelecting the incumbent and hence his reelection rule. Note that $\hat{\varepsilon}(y_1, z_1)$ 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

$$\rho_H = \int [\zeta \psi(y_1, H) + (1-\zeta)\psi(y_1, L)] \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 \qquad (20)$$

$$\rho_L = \int [(1-\zeta)\psi(y_1, H) + \zeta \psi(y_1, L)] \phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \qquad (21)$$

Further details are presented in the Appendix.

3.7 Model with a Three-Term Limit

In this version we allow for a governor to remain in office for a maximum of three terms. The model is a straightforward extension of the two-term benchmark, but outcomes in the first term now have an effect on *both* the voter's and the governor's decisions in the second term. The voter now uses not only y_2 but also y_1 to decide whether to reelect a second-term incumbent to a third term. This, in turn, makes a bad governor's effort choice in the second term contingent not only on the cost of exerting effort which she drew when she came to office, but also on y_1. That is, suppose a bad governor exerted high effort in her first term but drew a performance outcome y_1 that was low, nonetheless being reelected to a second term due to an election shock. She may decide not to exert (costly) high effort in her second term since the low y_1 makes it more difficult to convince the voter to reelect her to a third term. We present the full model 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 two extensions – the version with an effort signal and the version with a three-term limit – are presented in the Appendix.

4.1 Solution

The model has seven structural parameters: $\pi, \beta, Y_H, Y_L, \sigma_y, \mu$, and σ_ε. As the definition of perfect Bayesian equilibrium shows, given the structural parameters, finding the equilibrium amounts to finding values for ρ_H, ρ_L, δ 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 the equilibrium as follows.

The first thing to notice is that once $\mathbb{V}$ and δ are known, ρ_H and ρ_L follow from (16) and (17), derived from (11), (13), and (15). Thus solving the equilibrium amounts to satisfying (8) and (4). Define two residuals $\mathcal{R}_1$ and $\mathcal{R}_2$ as the differences between conjectures for $\mathbb{V}$ and δ and the model-implied values from (8) and (4), respectively

$$\begin{aligned} \mathcal{R}_1 &\equiv \mathbb{V} - [\pi + (1-\pi)\delta] \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_H}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \qquad (22) \\ &\quad - (1-\pi)(1-\delta) \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_L}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \\ \mathcal{R}_2 &\equiv \delta - \int \psi(y_1) \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 + \int \psi(y_1) \phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \qquad (23) \end{aligned}$$

where equilibrium requires $\mathcal{R}_1 = \mathcal{R}_2 = 0$. This can be solved easily using a nonlinear equation solver. Note that for given values for δ and $\mathbb{V}$ (and thus ρ_H and ρ_L), $\hat{\pi}(y_1)$ follows from (14), $\hat{\varepsilon}(y_1)$ follows from (13), $R(y_1, \varepsilon)$ follows from (12), $W(y_1, \varepsilon)$ follows from (10) and $\psi(y_1)$ follows from (15).

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 of one or two terms. 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

$$\begin{aligned} p_W(y_1, y_2) \equiv\ & \pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_H}{\sigma_y}\right) \\ & +(1-\pi)\delta\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_L}{\sigma_y}\right) \\ & +(1-\pi)(1-\delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_L}{\sigma_y}\right) \end{aligned} \tag{24}$$

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

$$\begin{aligned} p_L(y_1) \equiv\ & \pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)[1-\psi(y_1)] \\ & +(1-\pi)\delta\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)[1-\psi(y_1)] \\ & +(1-\pi)(1-\delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)[1-\psi(y_1)] \end{aligned} \tag{25}$$

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

$$L_k = R_k \log\left[p_W(y_{1k}, y_{2k})\right] + (1 - R_k)\log\left[p_L(y_{1k})\right] \tag{26}$$

and the log-likelihood is simply given by

$$\log \mathcal{L} = \sum_{k=1}^{n} L_k \tag{27}$$

Estimating the structural parameters requires maximizing $\log \mathcal{L}$, which we do using standard numerical optimization routines. We estimate six structural parameters $(\pi, Y_L, Y_H, \sigma_y, \mu, \sigma_\varepsilon)$ and fix $\beta = 0.85$, which represents roughly a 4% annual discounting over a four-year term. Once estimates for the structural parameters are obtained, estimates for equilibrium outcomes $(\rho_H, \rho_L, \delta, \mathbb{V})$ 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. Further details of the computational method are given in the Appendix.

4.3 Data Description

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 enters voters' utility directly (and thus is observable to them) and at least in part affected by the effort of the governor. We choose job approval ratings (JAR) for this purpose because relative to alternatives such as economic, environmental or fiscal outcomes, it seems to best fit our criteria.[11] Our implicit assumption in treating the JARs as measuring voter welfare is that voter approval ratings represent an accurate assessment by voters of their own welfare and not pandering by the governor. The robustness of our results to dropping the results of JAR surveys in the election year discussed in Section 5.3, where it may be argued that pandering would be most likely, supports this assumption.

A large fraction of the JAR data come from Beyle, Niemi, and Sigelman (2002), and we update their dataset through the 2012 election using various online resources. The underlying data come from surveys of voters at various points of each governor's term, where they are asked to rate the governor as excellent, good, fair and poor (or "undecided"). For each governor we measure performance as the fraction of respondents who classify the governor as excellent or good out of those who express an opinion, eliminating the undecided respondents. 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. We do not restrict the second-term JAR. We take the simple average of the JAR numbers over a term of the governor and use them as y_1 and y_2. From here on we use JAR to refer to the adjusted measures described in this paragraph.

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 present, where we have collected basic information about the governor, some of which comes from Besley (2006). We also know the outcomes of their reelection bids.[12] We then apply

[11] We also tried real income per capita growth, unemployment and change in unemployment. The former variable produced some significant effect on election outcomes but it was tiny in size which meant that much was "explained" by the election shock. As such, our model was not very informative. Nor did these economic variables have a high correlation with JAR. See Stein (1990) for an argument on why governors may be held less accountable than national leaders for economic conditions.

[12] We 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 A2). Dropping these governors would lead to a high value for π and discipline / selection effects become very small. This analyis is available upon request.

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 governors with 2-year terms.
- 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 149 governor stints.[13] Combining this with the JAR data we compiled yields 93 governor stints. Due to data availability and the prevalence of 2-year terms and/or absence of term limits early on in our sample, except for one governor from the 1960s, our data covers elections from 1982 to 2012. There are 26 election years from 32 states in our sample. 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.

Our model assumes that all governors are identical, except for their types. In order to conform to this assumption, our measures of performance need to be uncorrelated with any observable feature of the governor. This is indeed the case. Our measures of y_1 and y_2 have negligible correlations with characteristics of governors such as age, party, whether or not they are in the same party as the U.S. president, education level and gender, as well as characteristics of the states in which they serve, such as the Census division to which they belong.

We provide the basic data that we use for estimation, namely (y_1, R, y_2) in Table A1 in the Appendix. We also revisit some of the choices we make in this section and consider alternatives in Section 5.3.

[13] A 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.

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 1. Several things can be noted. 52% of governors in our sample are good and we strongly reject the two extremes, all governors being good or bad. Of the bad governors, 27% of them exert high effort in their first term and thus are disciplined. 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. Figure 2 shows the distribution of JAR for the 57 reelected incumbents in our sample. The red and blue normal distributions show the estimated performance distributions with the dashed lines showing their means. It is clear from this figure that it is unlikely that the outcome distributions come from a single normal distribution as there are two peaks near the two estimated means of the distributions.

High effort increases the probability of reelection from 45% to 72%. There is also a significant incumbency advantage: an incumbent enters the reelection with an advantage that is equivalent to having 9.34 JAR points more than his actual JAR. The election shock has a large standard deviation, which shows that there are many elections in which a governor with a low JAR is nonetheless reelected or a governor with a high JAR is not reelected. 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 3. 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. Small first-term JAR, for example 25, signal 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 15 or larger. Since this is quite reasonable given the estimated values of $\mu = 9.34$ and $\sigma_\varepsilon = 13.07$, there is about a 30% probability for this governor to win reelection, despite bad first-term performance. As y_1 increases so does $\hat{\pi}(y_1)$, until y_1 hits 70 after which the reelection probability remains constant at around 80%, reflecting the possibility of an unfavorable election shock after a very strong performance in the first term.

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. First, we find very strong evidence that there are two types of governors (that is, $0 < \pi < 1$). To understand why this is the case, assume to the contrary that there was only one type of governor.[14] In this case y_1 and y_2 for a governor would be *uncorrelated* because they would be *iid* draws from the same distribution. In our data, however, this correlation is 0.36. With two types of governors, and assuming for the moment that $\delta = 0$ (that is, bad governors never exert high effort), we can generate a positive correlation from the variation due to differences in governor types.

Second, we find $\delta > 0$. If on the contrary $\delta = 0$, this would imply that $(y_2 - y_1)$ would be zero-mean and symmetric because every governor draws her performance outcomes from the same distribution across her terms (with mean Y_H for good governors and with mean Y_L for bad governors), and any differences in the performance between terms is completely due to luck. In our data $y_2 - y_1$ has a negative skewness, indicating that it has a distribution with a thicker left tail. Our model with $\delta > 0$ is able to match this because the disciplined bad governors will have $y_2 - y_1$ that is negative since they would be drawing from two different distributions in their two terms.

Third, we find $\sigma_\varepsilon > 0$ and $\mu > 0$, indicating the presence of sizable election shocks that, on average, favor incumbents. If σ_ε were equal to 0, then any governor with $\hat{\pi}(y_1) > \pi$ would have to win reelection and all others would have to lose. In our sample however there are many governors who lose reelections despite good first-term performance, implying the presence of election shocks. Furthermore, defining $\chi_W \equiv \mathbb{P}(R = 1 | \hat{\pi}(y_1) < \pi)$ and $\chi_L \equiv \mathbb{P}(R = 0 | \hat{\pi}(y_1) > \pi)$ as the two surprises (winning elections despite bad performance and losing elections despite good performance), the election shock having a zero mean would imply $\chi_W = \chi_L$. However, in our sample $\chi_W > \chi_L$, which requires that $\mu > 0$, that is, an incumbency advantage.

Finally, once these key parameters or equilibrium objects are pinned down, the other three structual parameters, that is, Y_H, Y_L and σ_y, follow from matching some of the other properties of the JAR data. These include the mean and variance of y_1 and y_2.

[14] It should be clear that if there was only one type of governor, it is irrelevant if we called them good or bad. The bad governors would never exert high effort because there is no benefit in doing so.

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) one-term world in Table 2. 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 governor-terms. 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 a voter $\mathbb{V}$ is 360.2 in this case.

In the second panel we look at summary measures for two-term limited governors. The two-term world is unambiguously better for the voter. First, more governors exert effort in the first term, 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 51.7% 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 governors than good governors are screened out in elections, more governors are good in the second term at 59.6% relative to the unconditional probability of 51.7%. 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 360.2 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 one-term regime.[15] Looking at Table 2, it is clear that the voter is better off because the governors' performance in both terms is higher relative to the one-term case. We turn to understanding the different effects of electoral accountability in next section.

Given our estimated model, we can compute a few interesting magnitudes about the governors and their performance in our sample, which are reported in Table A2. 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

[15] Relative to the one-term outcome of 54, this 2.3 point increase is also a 4.2% increase. Since the utility of the voter is linear in JAR, "consumption-equivalent" welfare is equivalent to simply comparing lifetime values.

a new object

$$\bar{\pi}(y_1, R, y_2) \equiv \mathbb{P}(\theta = G | y_1, R, y_2) = \begin{cases} \dfrac{\pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_H}{\sigma_y}\right)}{p_W(y_1, y_2)} & \text{if } R = 1 \\ \dfrac{\pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)[1 - \psi(y_1)]}{p_L(y_1)} & \text{if } R = 0 \end{cases} \tag{28}$$

which shows the ex-post assessment of a governor's type, 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 is a good type with 70% probability just prior to his reelection bid. This implies a 74% chance that he will win the election. He wins a second term and has performance of only 38.2 in the second term. As result $\bar{\pi}(y_1, R, y_2)$ 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. Our second example is Mitch Daniels, who was the governor of Indiana between 2005 and 2013. His first-term performance is only 48.7, leading the voters to think that he is a good type with only 31% probability. Despite this, he wins reelection.[16] In his second term his performance is 65.2. As a result, *ex post* it seems like he was a good governor who exerted effort in his first term but had an unlucky performance draw. Our third example is David Beasley, who was the Republican governor of South Carolina from 1995 to 1999. His first term performance is 60.8, which gives him a 75% probability of winning his reelection bid. However, he loses reelection to the Democrat challenger Jim Hodges. This is likely to be a case where a good governor was unlucky to draw a negative election shock and lose the reelection. Indeed his loss is considered to be a surprise given how heavily Republican the state had been at the time.[17]

Finally, we can also talk about how good a fit our model provides to the data. In Table A2

[16] The *Washington Post* names 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.

[17] Owners of video poker machines spent very heavily on advertisements attacking Beasley who had worked to ban it in the state during his first term. It is 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.

we report $\psi(y_{1k})$, the model's implied probability that an incumbent k will win reelection. 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.[18] A Diebold-Mariano (1995) test, as described by 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 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.

5.2 Measuring the Effects of Elections

One of the key advantages of our structural approach is the ability to conduct counterfactual exercises to see how outcomes would differ if we changed various aspects of the environment such that the governors and voters behaved differently than they do in the data. We use this advantage for two important purposes. First, in this section we measure the magnitudes of the effects of having the possibility of reelection. Second, in Section 5.5, we consider a change in term limits.

Elections have three consequences in our model: discipline (bad governors exert high effort to secure reelection), selection (more good governors are reelected since bad governors who exerted low effort are identified), and mimicking (bad governors who are disciplined look like good governors). 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.

[18] As a point of comparison, a naive forecast that the incumbent wins 50% of the time would lead to a Bier score of 0.25, regardless of the outcome.

Having more disciplined governors increases first-term outcomes relative to the one-term case since more governors overall will exert high effort. In turn, when there is a second election, the selection effect can be measured as the increase in outcomes in the second term of the benchmark model relative to the one-term counterfactual model, as there will be more good governors than bad governors who survive reelection and 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". 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 minus the 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 (4) no longer is a part of the description of equilibrium. Naturally the voters take this into account and adjust their behavior accordingly in solving their problem. In other words, the voter solves his problem taking in to account that $\delta = 0$ and this influences all equilibrium mappings including, for example, the reelection rule and thus the equilibrium outcomes ρ_L, ρ_H and $\mathbb{V}$. We solve this equilibrium using the structural parameters in Table 1. Some details of outcomes in this counterfactual case are presented in Table A1 in the Appendix.

The bottom panel of Table 2 shows two different approaches to computing these three effects. The first approach uses the change in the number of good governors, measured in percentage points, while the second approach uses the change in performance. Comparing the benchmark version with the one-term case, we find that there is a 13 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 5% increase. These are our measures of discipline. The effect of selection is lower in magnitude, namely a 7.9 percentage point increase in the fraction of good governors (or equivalently governors that exert high effort) in the second term, leading to a 1.6 JAR point or 3.0% increase in performance. However, the improvement in the second term due to selection is partially cancelled due to mimicking – the pure selection effect is 2.6 JAR points, almost identical to discipline – but second-term performance is 1 JAR point lower due to mimicking.

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 60% 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 by 5.8 points while from the no-discipline counterfactual to the benchmark model welfare goes up by 9.3 points.

To highlight the significance of our structural approach in measuring the effect of elections, we close with a comparison to a reduced-form approach. A typical analysis of the effect of elections, for example Besley and Case (1995), compares the performance of politicians who are in their last terms in office (lame ducks) with those eligible for reelection, controlling for various observable characteristics of politicians. According to our model, the first and second (last) term outcomes are given by the stylized equations

$$y_1 = \textit{baseline} + \textit{discipline} \tag{29}$$

$$y_2 = \textit{baseline} + \textit{pure selection} - \textit{mimicking} \tag{30}$$

where "baseline" captures the level of performance that would be observed in the absence of electoral accountability, that is, independent of the effect of elections.[19] If we compute the performance of lame ducks, relative to all others in our model, or, equivalently, regress the performance of all governors on a lame-duck dummy, we get $y_2 - y_1 = \textit{pure selection} - \textit{mimicking} - \textit{discipline}$. Using the numbers in Table 2 for the second approach that uses JAR, with selection as the measure of *pure selection−mimicking*, we get -1.1. It is not clear how to interpret this number in isolation since by itself it gives us information about neither the absolute nor the relative sizes of the three channels we are able to separately identify. For example, the common finding that performance falls in a governor's last term is often interpreted as reflecting simply the lack of discipline, suggesting that removing term limits would increase voter welfare. To understand why this conclusion does not follow,

[19] It should be obvious that the three terms other than "baseline" are complicated functions of the equilibrium outcomes (i.e. behaviors of governors and voters), as well as of the structural parameters of the model.

consider our two-term-limit results relative to the counterfactual of having a one-term limit. First-term performance is indeed higher due to discipline, while second-term performance is higher due to selection (which itself is mitigated by mimicking). The fall in average performance in the last term is simply due to the relative strengths of the various effects, which reduced-form estimation cannot separate.

5.3 Robustness

As we explained in Section 4.3, we made some choices in preparing the JAR data for estimation. 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). In Table 3 we consider the robustness of our results to making different 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 π and δ is not affected by these variations.

We also considered allowing the distributions of Y_H and Y_L to have different variances (Free σ_y^H). This change also produces little substantive changes in the results. It is also useful to note that the log-likelihood of the restricted model (our benchmark) is only 0.23 log-points smaller than the likelihood of this unrestricted model and thus the restriction we place is not rejected by the data.

5.4 Noisy Effort Signal

The implications of a noisy effort signal 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 4 reports discipline and selection measures (analogous to Table 2) 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. (See the Appendix for the details of how this version is solved.) Throughout this section, we assume the structural parameters shown in Table 1 are unchanged but solve for the equilibrium objects for every ζ considered. We show the re-computed δ 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 3.0%. 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.

To better understand the magnitude of these effects, we also considered the case of $\zeta = 1$, that is, perfect observability of effort, as shown in the third column of Table 4. (This, of course, is not equivalent to perfect observability of type, since bad governors 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 43%, an increase by more than half of the 27% when effort was unobservable, but not by more as one would be perhaps 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 ε (reelection probability $\rho_L = 0.45$); conversely, even if a bad type exerts high effort, she is not guaranteed reelection (reelection probability $\rho_H = 0.72$) if the realization of ε is sufficiently negative. So, 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.

To confirm this, we simulated the model with full observability of effort ($\zeta = 1$) but with the election shock turned off ($\sigma_\varepsilon = 0$), so that the election shock is constrained to take its mean value $\mu = 9.34$. There is no possibility of a very positive realization of ε to "save" a low-effort incumbent, though the average incumbency advantage is still present. 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 component. 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 4 shows how the welfare of the voter changes in each case. Having a moderately informative effort signal is worth 0.3% of welfare to the voter while making effort fully observable leads to an improvement of 1.6%, which is sizable. Much of the increase in these cases come from the higher discipline. When selection is absent due to shutting down election shocks, the welfare gain falls to 1.1%.

5.5 Electoral Reform: A Limit of Three Terms

In this section we consider an electoral reform where governors are allowed to stay in office for three terms. To do so we use the estimated parameters from Table 1. We realize that of the structural parameters, π, μ and σ_ε may depend on the term-limit regime in place. For example, changing term limits may change incentives for people to enter politics and thus may change the composition of the pool of candidates. Since we do not model the choice of running for office, we cannot capture this effect in our model.[20] For example π may be lower under a three-term electoral regime since expected lifetime rents are now higher, reducing welfare and causing us to overstate the gain from the change in term limits. Also, the parameters that govern the election shock process could possibly depend on which term the incumbent is in; for example having more information about a governor may increase incumbency advantage so that a governor who has served two terms may be reelected more easily than a governor who has served only one.[21] Similarly, perhaps election shocks may not be as large in a second reelection bid. A change in the incumbency advantage μ would in turn change reelection probabilities and thus the discipline effect. As such, if μ did indeed go up, keeping the election shock process unchanged across terms would likely understate the welfare gains of the electoral reform we consider.

[20] Thus, perhaps one way to interpret our counterfactual experiment is one where a governor who thought she was running for office in a two-term limit regime is told that she has a limit of three terms on her first day in office.

[21] Rogers (2014) finds that incumbents in state legislatures face weaker challengers as they near their final term, indicating an increasing incumbency advantage.

Finally, we assume that the mapping from effort to outcomes is invariant to the term of the governor. One can argue that this can indeed be affected by the term of the governor in more than one way. For example a third-term governor may have more fatigue and thus even with effort outcomes may be worse. Alternatively, since a third-term governor knows the working of the system much better, perhaps things get done more easily, thus increasing performance conditional on effort. These factors would have complicated effects on welfare; the possibility of fatigue would reduce the gains from the reform while the idea that a third-term governor will be more efficient would increase the gains from reform. Thus the results in this section should be read with these caveats in mind.

Table 5 summarizes our results, where key statistics, obtained via simulating the model with a three-term limit, are contrasted with those from the benchmark model with a two-term limit. With a higher "prize" (staying in office for an extra term), bad governors now exert effort with 61.3% probability in the first term and 25.9% probability in the second term, compared with 26.8% probability in the first term for the two-term regime.[22] As a result average performance in the first term goes up to 60.1 from 56.7. Even though in their third term all bad governors exert low effort, governors in office exert high effort 69.7% of the time, compared with 62.7% in the two-term case, showing that there is a large increase in discipline. The flip side of this, though, is decreased selection since now more bad governors "slip through" elections and survive to their third term. The 7.9 percentage point increase in the fraction of good governors in the two-term regime (relative to a one-term regime, our first measure of selection) goes down to 4.5 for term three in the three-term regime.[23] Overall, the voters benefit from the reform, with welfare rising by 2.8%.

While tedious, in principle one can continue this exercise and increase the term limit to four, five or more terms. We think, however, that the problems we explain at the beginning of this section would become more and more significant as we increase the number of terms a governor can serve and would make our assumption of treating parameters as structural less defendable. As such we do not undertake this exercise. Nevertheless, we think it is useful to know that an electoral reform in the U.S. allowing governors to serve more than two terms

[22] One may question if having 61.3% of bad governors exerting high effort is too high. While it is difficult to answer this question exactly, we can get an idea by looking at the data for unconstrained governors. There are 137 governors in no-term-limit states who are eligible for reelection after their first term and 88% of them choose to run for a second term. Of those who win the second term, 54% of them run for a third term. The percentage seeking reelection to a fourth term, is 45%, still quite high. Hence, it would appear that the option of staying in office for an extra term is indeed valuable because ex post (i.e. at end of the second term) many governors take advantage of it.

[23] We can compute our measures of discipline only for the first term and selection only for the third term in the three-term regime since discipline and selection are combined in the second term.

may lead to a significant increase in voter welfare, if our assumption of holding parameters constant across regimes is not very egregious.

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.

Many papers have used a reduced-form approach to try to estimate the effects of electoral accountability on discipline and selection, but this approach faces the difficulty of disentangling the importance of these two effects on 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 estimation 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.

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 of course related. Understanding why some governors don't perform well should help explain why the threat of not being reelected may not induce them to perform better.

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

[1] Alt, J., E. Bueno de Mesquita and S. Rose (2009) "Disentangling Accountability and Competence in Elections: Evidence from U.S. Term Limits," *Journal of Politics* 73(1): 171-186.

[2] Ashworth, S. (2012), "Electoral Accountability: Recent Theoretical and Empirical Work," *Annual Review of Political Science* 15: 183–201.

[3] Banks, J. and R. Sundaram (1993) "Adverse Selection and Moral Hazard in a Repeated Elections Model," in W. Barnett et al. (eds.) *Political Economy: Institutions, Information, Competition and Representation*, New York, NY: Cambridge University Press.

[4] Barro, R. (1973), "The Control of Politicians: An Economic Model," *Public Choice* 14: 19-42.

[5] Besley, T. (2006) *Principled Agents? The Political Economy of Good Government*, Oxford University Press.

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

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

[8] Beyle, T., R. Niemi, and L. Sigelman (2002), "Gubernatorial, Senatorial, and State-level Presidential Job Approval Ratings: A Compilation of Data," *State Politics and Policy Quarterly* 2(3): 215-229.

[9] Brender, A. and A. Drazen (2005), "Political Budget Cycles in New Versus Established Democracies," *Journal of Monetary Economics* 52(7): 1271-1295.

[10] Brender, A. and A. Drazen (2008), "How Do Budget Deficits and Economic Growth Affect Reelection Prospects? – Evidence from a Large Panel of Countries," *American Economic Review* 98(5): 2203-2220.

[11] Brier, G.W. (1950), "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review,* 78: 1–3.

[12] Diebold, F.X. and R. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253-263.

[13] Diermeier, D., H. Eraslan, and A. Merlo (2003),"A Structural Model of Government Formation," *Econometrica* 71: 27-70.

[14] Drazen, A. and E.Y. Ozbay (2015), "Does 'Being Chosen to Lead' Induce Non-Selfish Behavior? Experimental Evidence on Reciprocity," working paper.

[15] Duggan, J. and C. Martinelli (2015), "Electoral Accountability and Responsive Democracy," working paper.

[16] Fearon, J. (1999), "Electoral Accountability and Control of Politicians: Selecting Good Types Versus Sanctioning Poor Performance," in Przeworski, Stokes, and Manin (eds.) *Democracy, Accountability, and* R*epresentation*, Cambridge, UK: Cambridge University Press.

[17] Ferejohn, J. (1986) "Incumbent Performance and Electoral Control," *Public Choice* 50: 5-26.

[18] Ferraz, C. and F. Finan (2011), "Electoral Accountability and Corruption in Local Governments: Evidence from Audit Reports," *American Economic Review* 101(4): 1274-1311.

[19] Gagliarducci, S. and T. Nannicini (2013), "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection," *Journal of the European Economic Association,* 11, p. 369–398.

[20] Lahiri, K. and L. Yang (2013), "Forecasting Binary Outcomes," *Handbook of Economic Forecasting*, Vol. 2B, (eds. A. Timmermann and G. Elliott), 1025-1106.

[21] List, J. and M. Sturm (2006), "How Elections Matter: Theory and Evidence from Environmental Policy," *Quarterly Journal of Economics* 121(4): 1249-1281.

[22] Lucas, R. (1976). "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets*, K. Brunner and A. Meltzer eds. Carnegie-Rochester Conference Series on Public Policy 1, 19–46

[23] Maskin, E. and J. Tirole (2004), "The Politician and the Judge: Accountability in Government," *American Economic Review* 94(4): 1034-1054.

[24] Merlo, A. (1997), "Bargaining over Governments in a Stochastic Environment," *Journal of Political Economy,* 105: 101-131.

[25] Rogers, S. (2014), "Term Limits: Keeping Incumbents in Office," working paper, Vanderbilt University.

[26] Sieg, H. and C. Yoon (2014), "Estimating a Dynamic Game of Gubernatorial Elections to Evaluate the Impact of Term Limits," working paper, University of Pennsylvania.

[27] Smart, M. and D.M. Sturm (2013) "Term Limits and Electoral Accountability," *Journal of Public Economics* 107: 93-102.

[28] Stein, R. (1990) "Economic Voting for Governor and U. S. Senator: The Electoral Consequences of Federalism," *Journal of Politics* 52(1): 29-53.

[29] Strömberg, D. (2008), "How the Electoral College Influences Campaigns and Policy: The Probability of Being Florida," *American Economic Review* 98(3): 769-807.

[30] Vlaicu, R. and A. Whalley (2014) "Hierarchical Accountability in Government," working paper, University of Maryland.

Table 1: Parameter Estimates

Structural Parameters		Equilibrium Objects	
π	0.52	δ	0.27
	(0.08)		(0.06)
Y_L	43.33	ρ_L	0.45
	(2.48)		(0.07)
Y_H	63.99	ρ_H	0.72
	(1.67)		(0.05)
σ_y	9.84	$\mathbb{V}$	391.78
	(0.80)		(11.48)
μ	9.34		
	(2.62)		
σ_ϵ	13.07		
	(4.31)		

Note: White standard errors are below estimates. Standard errors for the equilibrium objects are computed using the delta method. β is fixed at 0.85.

Table 2: Some Properties of the Estimated Model

One-Term Limit

Good governors	51.7%
High Effort	51.7%
Average Performance (JAR Points)	54.0
Life-time Discounted Welfare for Voter	360.2

Two-Term Limit (Benchmark)

Good governors in Term 1	51.7%
Good governors in Term 2	59.6%
Good governors Overall	54.7%
High effort in Term 1	64.7%
High effort in Term 2	59.6%
High effort Overall	62.7%
Average Performance in Term 1 (JAR Points)	56.7
Average Performance in Term 2 (JAR Points)	55.6
Average Performance Overall (JAR Points)	56.3
Life-time Discounted Welfare for Voter	375.3

Measures of Interest

Discipline A : Change in Fraction of High-Effort Governors in Term 1 (Benchmark vs. 1-Term)	13.0
Discipline B : Change in Performance In Term 1 (Benchmark vs. 1-Term)	2.7
% Change in Performance in Term 1 (Benchmark vs. 1-Term)	5.0%
Selection A : Change in Fraction of Good Governors in Term 2 (Benchmark vs. 1-Term)	7.9
Selection B : Change in Performance In Term 2 (Benchmark vs. 1-Term)	1.6
% Change in Performance In Term 2 (Benchmark vs. 1-Term)	3.0%
Mimicking A : Change in Fraction of Good Governors in Term 2 (Benchmark vs. $\delta = 0$)	-4.6
Mimicking B : Change in Performance In Term 2 (Benchmark vs. $\delta = 0$)	-1.0
% Change in Performance In Term 2 (Benchmark vs. $\delta = 0$)	-1.7%

Note: The numbers on this table are obtained by simulating the model for 1,000,000 terms, given the structural parameters in Table 1. The one-term limit assumes governors are not eligible to run for reelection. The $\delta = 0$ version is solved assuming $c = 1$. All changes in fractions (such as the ones for Discipline A, Selection A and Mimicking A measures) reported as percentage point changes.

Table 3: Robustness of Estimation Results

	Benchmark	**All Surveys**	**No Election Year**	**Year-by-Year Average**
π	0.52	0.50	0.53	0.53
	(0.08)	(0.10)	(0.09)	(0.09)
δ	0.27	0.26	0.26	0.29
	(0.06)	(0.06)	(0.07)	(0.07)
Discipline B	5.0%	4.8%	4.7%	5.4%
Selection B	3.0%	2.8%	2.8%	3.2%
Welfare Gain	4.7%	4.1%	3.9%	4.5%

	Median JAR	**Minimum JAR**	**Keep Undecideds**	**Free σ_y^H**
π	0.54	0.45	0.46	0.48
	(0.09)	(0.12)	(0.12)	(0.08)
δ	0.25	0.23	0.23	0.26
	(0.07)	(0.06)	(0.06)	(0.08)
Discipline B	4.6%	6.5%	4.3%	5.1%
Selection B	2.9%	3.5%	2.3%	2.8%
Welfare Gain	3.9%	5.3%	3.5%	4.2%

Notes: The top of each panel show the re-estimated π and δ for each case with standard errors in parentheses. See Table 2 for the definitions of the discipline and selection measures. Reported welfare gains are relative to the one-term regime.

Table 4: Results from the Version with Effort Signal

	$\zeta = 0.5$	$\zeta = 0.75$	$\zeta = 1$	$\zeta = 1$ **and** $\sigma_\epsilon = 0$
δ	0.27	0.30	0.43	1.00
Discipline B	5.0%	5.5%	7.9%	18.4%
Selection B	3.0%	3.2%	3.8%	0.0%
Welfare Gain Relative to Benchmark	-	0.3%	1.6%	1.1%

Notes: The first column ($\zeta = 0.5$) shows the benchmark results from Tables 1 and 2. Structural parameters are kept as in Table 1. See Table 2 for the definitions of the discipline and selection measures.

Table 5: Electoral Reform: Three-Term Limit

	Two-Term	**Three-Term**
Good governors in Term 1	51.7%	51.7%
Good governors in Term 2	59.6%	55.6%
Good governors in Term 3	-	56.2%
Good governors Overall	54.7%	54.2%
High effort in Term 1	64.7%	81.3%
High effort in Term 2	59.6%	67.1%
High effort in Term 3	-	56.2%
High effort Overall	62.7%	69.7%
Average Performance in Term 1 (JAR points)	56.7	60.1
Average Performance in Term 2 (JAR points)	55.6	57.2
Average Performance in Term 3 (JAR points)	-	54.9
Average Performance Overall (JAR points)	56.3	57.7
Measures of Interest		
Fraction of Disciplined Bad Governors (Term 1)	26.8%	61.3%
Fraction of Disciplined Bad Governors (Term 2)	-	25.9%
Discipline A (Term 1)	13.0	29.6
Discipline B (Term 1)	5.0%	11.3%
Selection A (Term 2)	7.9	-
Selection B (Term 2)	3.0%	-
Selection A (Term 3)	-	4.5
Selection B (Term 3)	-	1.7%
Welfare Gain Relative to Two-Term Benchmark	-	2.8%

Notes: Structural parameters are kept as in Table 1. See Table 2 for the definitions of the discipline and selection measures.

Figure 1: Game Tree

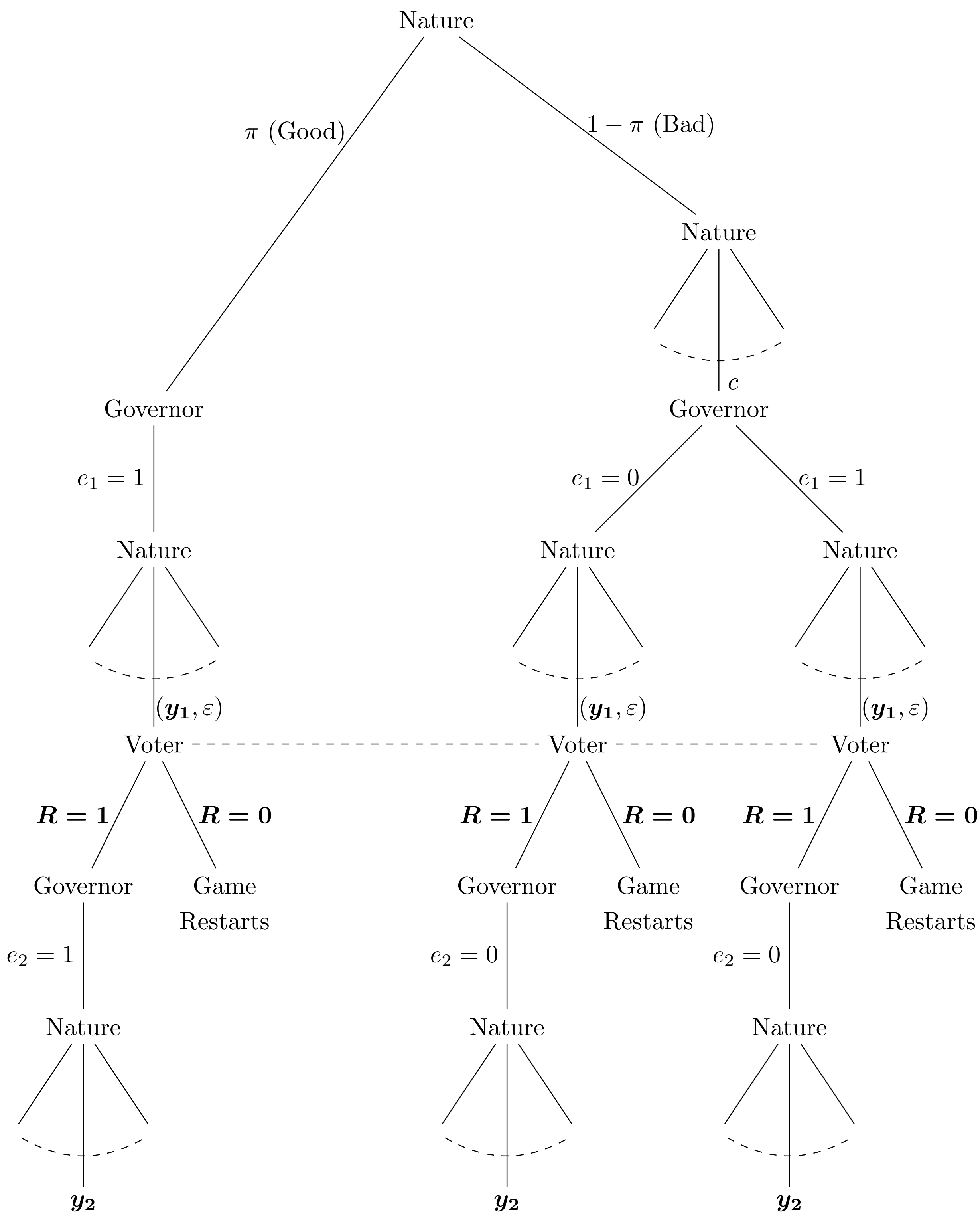
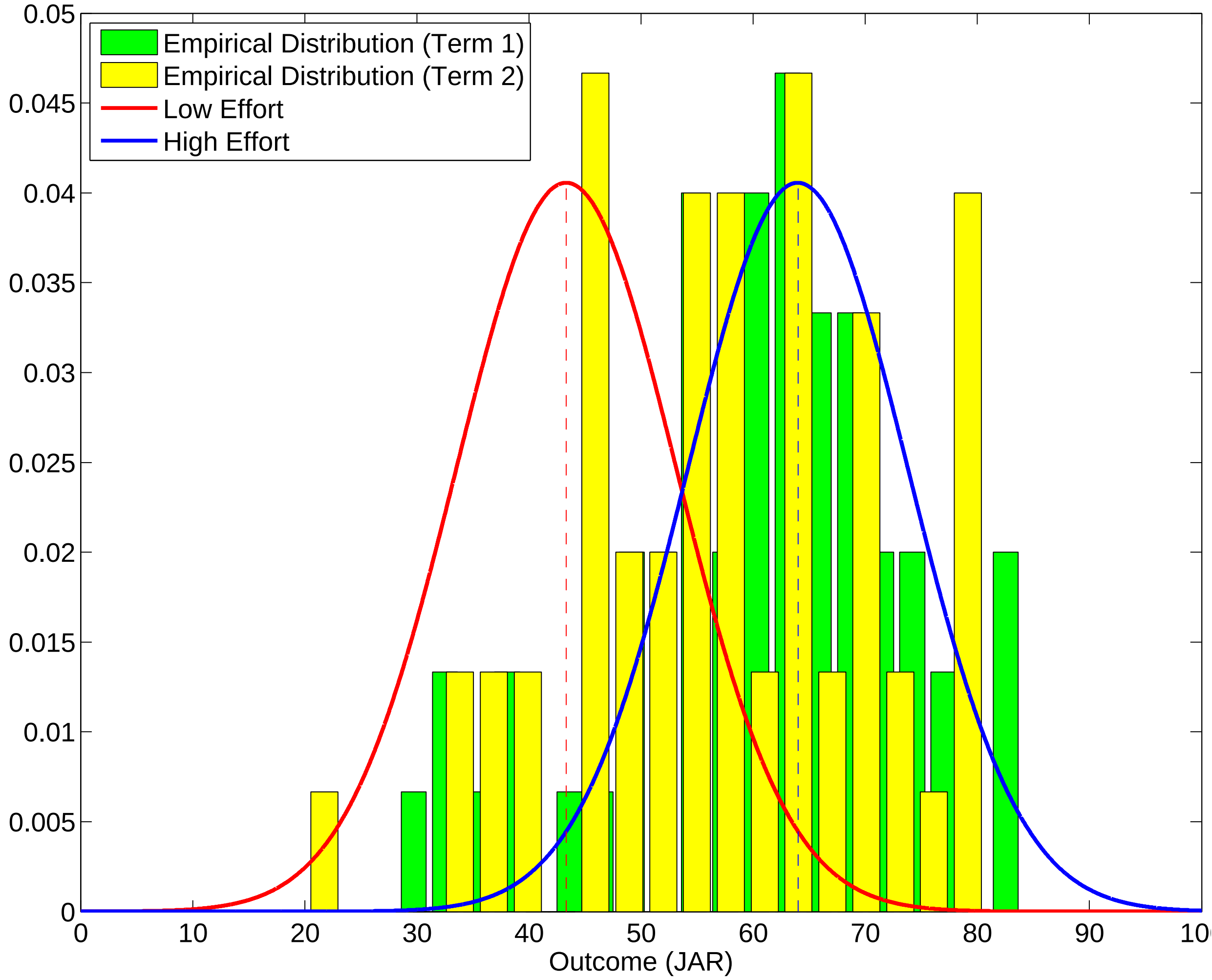
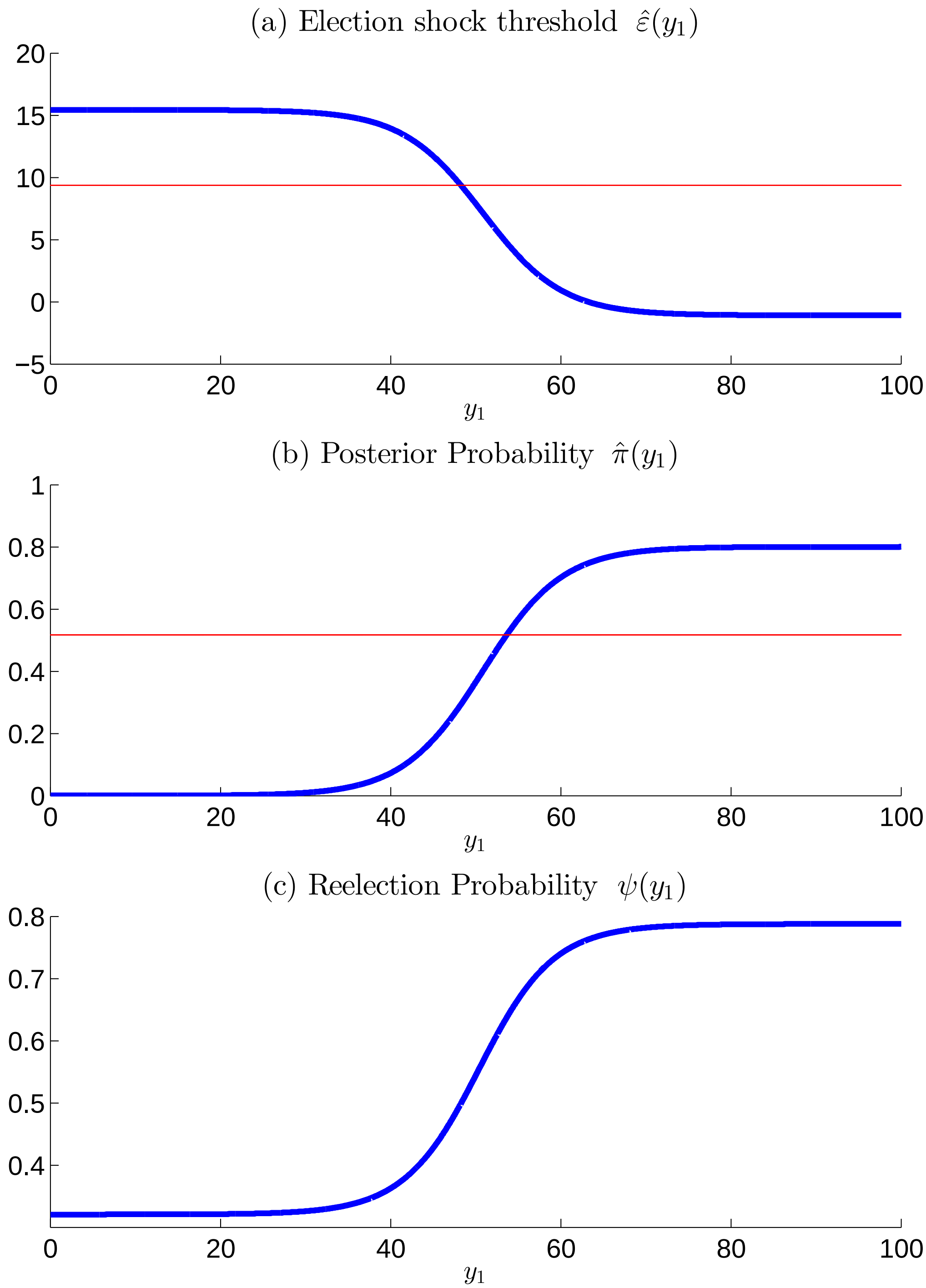


Figure 2: Outcome (JAR) Distributions (Only Reelected Incumbents)



Notes: The bars show the empirical distribution of JARs in the first and second terms of reelected incumbents. The blue and red lines show the estimated normal distribution of outcomes conditional on high and low effort, respectively, with the vertical dotted lines indicating the means of the distributions.

Figure 3: Equilibrium Mappings



Notes: The red horizontal lines in the first and second panels are μ (the mean of the election shock process) and π (the unconditional probability of a governor being good), respectively.

Appendix (For Online Publication)

A 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:

$$W(y_1, z_1, \varepsilon) = y_1 + \beta \max_{R \in \{0,1\}} \mathbb{E} \left\{ \begin{array}{l} R\left[y_2 + \varepsilon + \beta W(y_1', z_1', \varepsilon')\right] + \\ +(1-R) W(y_1', z_1', \varepsilon') \end{array} \middle| y_1, z_1, \varepsilon \right\} \tag{A-1}$$

which leads to (7) with the new definition of $\mathbb{V} \equiv \mathbb{E}\left[W(y_1', z_1', \varepsilon')\right]$

$$\begin{aligned} \mathbb{V} &= [\pi + (1-\pi)\delta] \int\int \left[\zeta W(y_1', H, \varepsilon') + (1-\zeta) W(y_1', L, \varepsilon')\right] \phi\left(\frac{y_1' - Y_H}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \\ &\quad + (1-\pi)(1-\delta) \int\int \left[\zeta W(y_1', L, \varepsilon') + (1-\zeta) W(y_1', H, \varepsilon')\right] \phi\left(\frac{y_1' - Y_L}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \end{aligned}$$

The incumbent's posterior reputation becomes:

$$\begin{aligned} \hat{\pi}(y_1, z_1) &\equiv \mathbb{P}(\theta = G | y_1, z_1) = \frac{p(y_1, z_1 | \theta = G) \mathbb{P}(\theta = G)}{p(y_1, z_1 | \theta = G) \mathbb{P}(\theta = G) + p(y_1, z_1 | \theta = B) \mathbb{P}(\theta = B)} \\ &= \begin{cases} \dfrac{\pi \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\zeta}{[\pi + (1-\pi)\delta]\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\zeta + (1-\pi)(1-\delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)(1-\zeta)} & \text{if } z_1 = H \\ \dfrac{\pi \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)(1-\zeta)}{[\pi + (1-\pi)\delta]\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)(1-\zeta) + (1-\pi)(1-\delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)\zeta} & \text{if } z_1 = L \end{cases} \end{aligned}$$

because

$$\begin{aligned} p(y_1, z_1 | \theta = G) &= p(y_1, z_1 | \theta = G, e_1 = H) \mathbb{P}(e_1 = H | \theta = G) + \\ &\quad + p(y_1, z_1 | \theta = G, e_1 = L) \mathbb{P}(e_1 = L | \theta = G). \end{aligned} \tag{A-2}$$

and $\hat{\pi}(y_1, z_1)$ replaces $\hat{\pi}(y_1)$ in various equations such as (9) and (13).

Reelection probabilities conditional on voter information are:

$$\begin{aligned} \psi(y_1, z_1) &= \mathbb{P}(R = 1 | y_1, z_1) = [\varepsilon > \hat{\varepsilon}(y_1, z_1)] \\ &= 1 - \Phi\left[\frac{\hat{\varepsilon}(y_1, z_1) - \mu}{\sigma_\varepsilon}\right] \end{aligned} \tag{A-3}$$

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

$$\rho_H = \int [\zeta\psi(y_1, z_1 = H) + (1-\zeta)\psi(y_1, z_1 = L)]\,\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 \tag{A-4}$$

$$\rho_L = \int [(1-\zeta)\psi(y_1, z_1 = H) + \zeta\psi(y_1, z_1 = L)]\,\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \tag{A-5}$$

B Model with a Three-Term Limit

Here we highlight some of the key changes in the model with a three-term limit. More details are available upon request. Define the vector $\mathbf{S}_t \equiv [y_t, \varepsilon_t, \hat{\pi}_{t-1}]$ for term t of a governor, $\hat{\pi}_t \equiv \hat{\pi}(\mathbf{S}_t)$ is the updated probability that a governor is of good type given the information in $\mathbf{S}_t$. The bad governors solve the following problems in the three periods they are potentially in office

$$V_1(c) = \max_{e_1} \left\{ \begin{array}{c} r(1-ce_1) + e_1 \int \psi(\mathbf{S}_1) V_2(\mathbf{S}_1, c)\,\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 \\ +(1-e_1)\int \psi(\mathbf{S}_1) V_2(\mathbf{S}_1, c)\,\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \end{array} \right\} \tag{A-6}$$

$$V_2(\mathbf{S}_1, c) = \max_{e_2} r(1-ce_2) + \left[e_2\rho_2^H(\mathbf{S}_1) + (1-e_2)\,\rho_2^L(\mathbf{S}_1)\right] V_3(c) \tag{A-7}$$

$$V_3(c) = \max_{e_3} r(1-ce_3) \tag{A-8}$$

where $\psi(\mathbf{S}_1)$ is the probability of reelection conditional on performance, and $\rho_2^H(\mathbf{S}_1)$ and $\rho_2^L(\mathbf{S}_1)$ are the reelection probabilities conditional on high and low effort in the first term. The bad governors choose $e_3 = 0$, as should be obvious and they choose $e_2 = 1$ with probability $\delta_2(\mathbf{S}_1) = \rho_2^H(\mathbf{S}_1) - \rho_2^L(\mathbf{S}_1)$, similar to the two-term model.

The voter solves the problem

$$W_1(\mathbf{S_1}) = y_1 + \beta \max_{R_1} [R_1 \{\mathbb{E}[W_2(\mathbf{S_2})|\mathbf{S_1}] + \varepsilon_1\} + (1-R_1)\mathbb{W}] \tag{A-9}$$

after the first term of the governor and

$$W_2(\mathbf{S_2}) = y_2 + \beta \max_{R_2} [R_2 \{\mathbb{E}[y_3|\mathbf{S_2}] + \varepsilon_2 + \beta\mathbb{W}\} + (1-R_2)\mathbb{W}] \tag{A-10}$$

after her second term, where $\mathbb{W} \equiv \mathbb{E}[W_1(\mathbf{S_1})]$. In this version some of the expectations are quite complicated (and we do not reproduce them here). The reelection rules are

$$R_1 = 1 \text{ iff } \mathbb{E}\left[W_2\left(\mathbf{S_2}\right)|\mathbf{S_1}\right] + \varepsilon_1 \geq \mathbb{W} \text{ or } \varepsilon_1 > \hat{\varepsilon}_1\left(\mathbf{S_1}\right) \tag{A-11}$$

$$R_2 = 1 \text{ iff } \mathbb{E}\left(y_3|\mathbf{S_2}\right) + \varepsilon_2 + \beta\mathbb{W} \geq \mathbb{W}\text{or } \varepsilon_2 > \hat{\varepsilon}_2\left(\mathbf{S_2}\right) \tag{A-12}$$

where

$$\hat{\varepsilon}_1\left(\mathbf{S_1}\right) \equiv \mathbb{W} - \mathbb{E}\left[W_2\left(\mathbf{S_2}\right)|\mathbf{S_1}\right] \tag{A-13}$$

$$\hat{\varepsilon}_2\left(\mathbf{S_2}\right) \equiv (1-\beta)\,\mathbb{W} - \hat{\pi}_2\left(Y_H - Y_L\right) - Y_L \tag{A-14}$$

The environment is complete with the election probability in term t

$$\psi(\mathbf{S_t}) = \mathbb{P}\left[\varepsilon_t > \hat{\varepsilon}_t\left(\mathbf{S_t}\right)\right] \tag{A-15}$$

and the updating by the voter

$$\hat{\pi}(\mathbf{S_t}) = \frac{\hat{\pi}_{t-1}\phi\left(\frac{y_t - Y_H}{\sigma_y}\right)}{\hat{\pi}_{t-1}\phi\left(\frac{y_t - Y_H}{\sigma_y}\right) + \left(1 - \hat{\pi}_{t-1}\right)\tilde{\phi}_t} \tag{A-16}$$

where $\hat{\pi}_0 = \pi$ and

$$\tilde{\phi}_t \equiv \delta_t \phi\left(\frac{y_t - Y_H}{\sigma_y}\right) + (1-\delta_t)\phi\left(\frac{y_t - Y_L}{\sigma_y}\right) \tag{A-17}$$

C Some Computational Details

We need to evaluate some integrals numerically to obtain (22) and (23) in the text. Note that all the integrals we deal with have the following general form

$$\int \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \xi\left(x\right) dx \tag{A-18}$$

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

$$\int \frac{1}{\sqrt{\pi}} \exp\left(-\hat{x}^2\right) \xi\left(\sqrt{2}\sigma\hat{x} + \mu\right) d\hat{x} \tag{A-19}$$

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

$$\frac{1}{\sqrt{\pi}}\sum_{i=1}^{m}\omega_i\xi\left(\sqrt{2}\sigma\hat{x}_i+\mu\right) \tag{A-20}$$

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

Turning to the integrals in (22) the first one is

$$\mathbb{A}_1=\int\int W\left(y_1',\varepsilon'\right)\phi\left(\frac{y_1'-Y_H}{\sigma_y}\right)\phi\left(\frac{\varepsilon'-\mu}{\sigma_\varepsilon}\right)dy_1'd\varepsilon' \tag{A-21}$$

which is double integral, but, this can actually be simplified to a single integral by realizing that the dependence of $W(.)$ on ε is through an indicator function. In particular, we can write

$$\mathbb{A}_1=\int\underbrace{\left\{\int W\left(y_1',\varepsilon'\right)\phi\left(\frac{\varepsilon'-\mu}{\sigma_\varepsilon}\right)d\varepsilon'\right\}}_{\xi_1\left(y_1'\right)}\phi\left(\frac{y_1'-Y_H}{\sigma_y}\right)dy_1' \tag{A-22}$$

where

$$\begin{aligned}\xi_1\left(y_1'\right) &= y_1'+\beta\int_{\hat{\varepsilon}\left(y_1'\right)}^{\infty}\left\{\hat{\pi}\left(y_1'\right)Y_H+\left[1-\hat{\pi}\left(y_1'\right)\right]Y_L+\beta\mathbb{V}\right\}\phi\left(\frac{\varepsilon'-\mu}{\sigma_\varepsilon}\right)d\varepsilon' &\text{(A-23)}\\ &\quad+\beta\mathbb{V}\int_{-\infty}^{\hat{\varepsilon}\left(y_1'\right)}\phi\left(\frac{\varepsilon'-\mu}{\sigma_\varepsilon}\right)d\varepsilon'+\beta\int_{-\infty}^{\hat{\varepsilon}\left(y_1'\right)}\varepsilon'\phi\left(\frac{\varepsilon'-\mu}{\sigma_\varepsilon}\right)d\varepsilon' \\ &= y_1'+\beta\left\{\hat{\pi}\left(y_1'\right)Y_H+\left[1-\hat{\pi}\left(y_1'\right)\right]Y_L+\beta\mathbb{V}\right\}\psi\left(y_1'\right) &\text{(A-24)}\\ &\quad+\beta\mathbb{V}\left[1-\psi\left(y_1'\right)\right]+\beta\left(\mu+\sigma_\varepsilon\frac{\phi\left(\frac{\hat{\varepsilon}\left(y_1'\right)-\mu}{\sigma_\varepsilon}\right)}{1-\psi\left(y_1'\right)}\right)\end{aligned}$$

and thus $\mathbb{A}_1$ can be computed using a Gauss-Hermite approximation using $\xi_1\left(y_1'\right)$.

The second integral in (22) can be computed analogously

$$\begin{aligned}\mathbb{A}_2 &= \int\int W\left(y_1',\varepsilon'\right)\phi\left(\frac{y_1'-Y_L}{\sigma_y}\right)\phi\left(\frac{\varepsilon'-\mu}{\sigma_\varepsilon}\right)dy_1'd\varepsilon' &\text{(A-25)}\\ &= \int\xi_1\left(y_1'\right)\phi\left(\frac{y_1'-Y_L}{\sigma_y}\right)dy_1'\end{aligned}$$

and $\mathbb{A}_2$ can be computed using a Gauss-Hermite approximation using $\xi_1(y_1')$. Thus we have

$$\mathcal{R}_1 \equiv \mathbb{V} - [\pi + (1-\pi)\,\delta]\,\mathbb{A}_1 - (1-\pi)(1-\delta)\,\mathbb{A}_2 \tag{A-26}$$

Turning to (23), the integrals can be computed by a Gauss-Hermite approximation with $\xi_2(y_1) = \psi(y_1)$.

$$\mathcal{R}_2 \equiv \delta - \int \xi_2(y_1)\,\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 + \int \xi_2(y_1)\,\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \tag{A-27}$$

Table A1: No-Discipline ($\delta = 0$) Counterfactual

Good governors in Term 1	51.7%
Good governors in Term 2	64.1%
Good governors Overall	56.6%
High effort in Term 1	51.7%
High effort in Term 2	64.1%
High effort Overall	56.6%
Average Performance in Term 1 (JAR Points)	54.0
Average Performance in Term 2 (JAR Points)	56.5
Average Performance Overall (JAR Points)	55.0
Life-time Discounted Welfare for Voter	366.0

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 1.

Table A2: Governors

State	Name (Reelection Year)	y_1	$\hat{\pi}(y_1)$	$\psi(y_1)$	y_2	$\bar{\pi}(y_1, R, y_2)$
AL	Fob James Jr. (1982)	26.6	0%	32%	-	0%
AL	George C. Wallace (1986)	40.8	8%	37%	-	8%
AL	Guy Hunt (1990)	60.1	70%	74%	38.2	8%
AL	Fob James Jr. (1998)	48.0	28%	49%	-	28%
AL	Don Siegelman (2002)	61.2	72%	75%	-	72%
AL	Bob Riley (2006)	54.0	53%	65%	63.6	90%
AR	Mike Beebe (2010)	81.4	80%	79%	80.4	100%
CA	Pete Wilson (1994)	33.0	2%	33%	41.4	0%
CO	Bill Owens (2002)	71.0	79%	78%	57.1	89%
CO	Bill Ritter (2010)	56.9	63%	70%	-	63%
DE	Thomas R. Carper (1996)	65.8	77%	77%	80.7	100%
FL	Robert Graham (1982)	68.2	78%	78%	80.6	100%
FL	Bob Martinez (1990)	44.3	16%	41%	-	16%
FL	Lawton Chiles (1994)	38.7	6%	35%	50.0	3%
FL	Jeb Bush (2002)	66.2	77%	77%	57.9	89%
FL	Charlie Crist (2010)	66.3	77%	77%	-	77%
GA	Zell Miller (1994)	56.8	63%	70%	70.1	98%
GA	Roy Barnes (2002)	77.8	80%	79%	-	80%
GA	Sonny Perdue (2006)	61.2	72%	75%	58.1	87%
IN	Robert D. Orr (1984)	37.0	4%	34%	46.0	1%
IN	Evan Bayh (1992)	69.9	79%	78%	57.0	88%
IN	Mitch Daniels (2008)	48.7	31%	51%	65.2	84%
KS	Mike Hayden (1990)	50.0	36%	54%	-	36%
KS	Joan Finney (1994)	33.8	2%	33%	-	2%
KS	Kathleen Sebelius (2006)	63.6	75%	76%	64.9	97%
KY	Paul E. Patton (1999)	60.3	71%	74%	51.9	62%
KY	Ernie Fletcher (2007)	39.6	7%	36%	-	7%
LA	Edwin W. Edwards (1987)	21.3	0%	32%	-	0%
LA	Charles Roemer (1991)	56.9	63%	70%	-	63%
LA	Edwin W. Edwards (1995)	32.2	1%	33%	-	1%
LA	Mike Foster (1999)	77.7	80%	79%	70.4	99%
LA	Kathleen Babineaux Blanco (2007)	46.1	21%	45%	-	21%
ME	John Baldacci (2006)	51.1	41%	57%	47.3	15%
MD	Harry Hughes (1982)	32.0	1%	33%	44.8	0%
MD	William D. Schaefer (1990)	61.9	73%	76%	33.6	4%
MD	Parris N. Glendening (1998)	36.5	4%	34%	55.4	5%
MD	Robert L. Ehrlich (2006)	57.1	63%	70%	-	63%
MD	Martin O'Malley (2010)	55.3	58%	67%	57.5	76%
MI	Jennifer Granholm (2006)	52.4	46%	61%	40.8	5%
MS	Ray Mabus (1991)	58.6	67%	72%	-	67%

Continued on next page

Table A2 – Governors *(continued)*

State	Name (Reelection Year)	y_1	$\hat{\pi}(y_1)$	$\psi(y_1)$	y_2	$\bar{\pi}(y_1, R, y_2)$
MS	Kirk Fordice (1995)	50.1	37%	55%	53.8	38%
MS	Ronnie Musgrove (2003)	68.0	78%	78%	-	78%
MS	Haley Barbour (2007)	54.1	54%	65%	67.4	96%
MO	Bob Holden (2004)	40.2	8%	36%	-	8%
MO	Matt Blunt (2008)	41.3	9%	37%	-	9%
MT	Marc Racicot (1996)	77.4	80%	79%	79.9	100%
MT	Judy Martz (2004)	29.4	1%	33%	-	1%
MT	Brian Schweitzer (2008)	72.3	79%	78%	66.7	98%
NE	Bob Kerrey (1986)	84.0	80%	79%	-	80%
NE	Kay A. Orr (1990)	52.7	48%	61%	-	48%
NE	Ben Nelson (1994)	63.9	75%	77%	77.8	100%
NV	Kenny C. Guinn (2002)	59.7	70%	74%	62.6	94%
NV	Jim Gibbons (2010)	31.8	1%	33%	-	1%
NJ	Richard J. Hughes (1965)	58.1	66%	72%	64.1	95%
NJ	Thomas H. Kean (1985)	65.6	77%	77%	72.7	99%
NJ	James J. Florio (1993)	30.0	1%	33%	-	1%
NJ	Jon Corzine (2009)	48.1	29%	50%	-	29%
NM	Gary E. Johnson (1998)	51.2	42%	58%	53.7	42%
NM	Bill Richardson (2006)	62.2	74%	76%	55.3	80%
NC	James G. Martin (1988)	65.6	77%	77%	55.2	82%
NC	James B. Hunt Jr. (1996)	68.6	78%	78%	69.9	99%
NC	Mike Easley (2004)	56.9	63%	70%	59.3	85%
NC	Bev Perdue (2012)	44.3	16%	41%	-	16%
OH	Richard F. Celeste (1986)	55.9	60%	68%	63.6	93%
OH	George V. Voinovich (1994)	64.3	76%	77%	74.7	100%
OH	Bob Taft (2002)	83.9	80%	79%	20.3	0%
OH	Ted Strickland (2010)	59.6	69%	74%	-	69%
OK	Henry L. Bellmon (1990)	45.8	20%	44%	-	20%
OK	David Walters (1994)	40.8	8%	37%	-	8%
OK	Frank Keating (1998)	68.2	78%	78%	60.5	94%
OK	Brad Henry (2006)	69.0	78%	78%	71.0	99%
OR	Barbara Roberts (1994)	28.6	1%	32%	-	1%
OR	John Kitzhaber (1998)	61.8	73%	75%	59.8	91%
OR	Ted Kulongoski (2006)	47.2	25%	47%	46.0	6%
PA	Richard L. Thornburgh (1982)	49.6	35%	53%	44.8	8%
PA	Robert P. Casey (1990)	81.3	80%	79%	33.2	5%
PA	Ed Rendell (2006)	55.0	57%	67%	52.3	50%
RI	Lincoln C. Almond (1998)	44.5	17%	42%	55.7	23%
RI	Donanld Carcieri (2006)	59.4	69%	73%	46.4	32%
SC	Carroll Campbell (1990)	73.3	79%	78%	70.6	99%

Continued on next page

Table A2 – Governors *(continued)*

State	Name (Reelection Year)	y_1	$\hat{\pi}(y_1)$	$\psi(y_1)$	y_2	$\bar{\pi}(y_1, R, y_2)$
SC	David Beasley (1998)	60.8	71%	75%	-	71%
SC	Jim Hodges (2002)	46.7	23%	46%	-	23%
SC	Mark Sanford (2006)	54.9	56%	66%	47.4	25%
SD	M. Michael Rounds (2006)	75.0	80%	79%	65.1	98%
TN	Ned R. McWherter (1990)	70.3	79%	78%	51.4	70%
TN	Don Sundquist (1998)	61.6	73%	75%	47.7	43%
TN	Phil Bredesen (2006)	61.8	73%	75%	74.2	100%
WV	Arch A. Moore Jr. (1988)	42.7	12%	39%	-	12%
WV	Gaston Caperton (1992)	28.3	1%	32%	36.7	0%
WV	Cecil H. Underwood (2000)	59.7	70%	74%	-	70%
WV	Bob Wise (2004)	58.9	68%	73%	-	68%
WY	Jim Geringer (1998)	66.3	77%	77%	50.0	61%
WY	Dave Freudenthal (2006)	73.4	79%	78%	78.0	100%

Notes: y_1 and y_2, when available, show the JAR performance of the governor. The absence of y_2 indicate that the governor did not win a second term. $\hat{\pi}(y_1)$ is the updated probability of the governor being good, conditional on first-term performance and $\psi(y_1)$ is the probability that the governor will win re-election based 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.