Structural Estimation of Political Accountability

S. Borağan Aruoba* Allan Drazen† Razvan Vlaicu‡

This Draft: January 31, 2015

Abstract

This paper proposes a structural approach to the estimation of the effect of political accountability. We identify and quantify discipline and selection effects using U.S. gubernatorial elections for 1982-2012. We find that the possibility of re-election provides a significant incentive for incumbents to exert effort. We also find a selection effect, though it is weaker in terms of its effect on average governor performance. A structural approach also allows us to measure the welfare effects of different term limit lengths.

JEL Classification: D72, D73, H70

Keywords: discipline, moral hazard, political agency, elections, structural models

*University of Maryland, Department of Economics, 3105 Tydings Hall, College Park, MD 20742, United States. (corresponding author) E-mail: aruoba@econ.umd.edu.

†University of Maryland, Department of Economics, NBER, and CEPR. E-mail: drazen@econ.umd.edu.

‡Northwestern University, Kellogg School of Management, 2001 Sheridan Road, 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 by *disciplining* poor performance (Barro [1973], Ferejohn [1986]). They may serve a *selection* function by screening out low performers (Banks and Sundaram [1993], Fearon [1999]), 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 net effects of eliminating the possibility of reelection via limiting the number of terms a politician may serve. Though term limits may reduce electoral pandering and prevent politicians becoming too "entrenched" in office and thus unresponsive to voter concerns, the 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 wider issue of political accountability in the political agency model. The agency approach has found wide application in political economy, suggesting 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 below. 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 estimating the effects of term limits. We set out a simple political agency model with adverse selection and moral hazard. By structurally estimating the parameters of the model, we can quantify discipline and selection effects and assess their importance without relying on strong identification assumptions. We identify a benchmark of no electoral accountability (that is, where there is no possibility of re-election), on the basis of which we can measure how much electoral accountability improves outcomes,

[1] There is a large literature on political cycles in economic policy, with formal models going back at least to Nordhaus (1975). Brender and Drazen (2005, 2008) summarize key empirical findings. Welfare implications of such opportunistic behavior are studied by Maskin and Tirole (2004), among others.

[2] These effects may also affect indirectly-elected policymakers, as in Vlaicu and Whalley (2011).

as well as whether improvements come mainly through discipline or through selection. The structural model also allows us to run policy experiments on the welfare effects of changing term limits, where structural magnitudes are critical for making welfare comparisons.

Our main empirical findings are as follows. Under the assumption that only outcomes but not governor effort are observable (but that higher effort increases expected outcome measures), the possibility of re-election provides a significant incentive for incumbents to exert effort in order to increase their chances of re-election. Compared to the (counterfactual) benchmark case in which governors cannot be re-elected (a one-term limit, so there is no discipline effect on their first term behavior), allowing a second term leads to a 13 percentage point increase in the fraction of governors who exert high "effort" in their first term of office. Average performance (as measured by voter surveys of how well a governor is performing) rises by about 5%. Discipline is not stronger because 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.

We also find a selection effect, where the fraction of "good" governors (those willing to exert high "effort" even without the discipline of losing office) rises by 8 percentage point from the first to the second term of office. Partial observability of effort on the part of governors leads to increased discipline (the fraction of governors who exert effort), but this effect is imperfect because of the possibility of "bad" governors mimicking good ones in their effort choices, as well as the stochastic nature of election outcomes. Even if effort were fully observable, the latter implies that 43% of bad governors are disciplined. Were there a perfect relation between effort supplied and election outcomes, all governors would supply effort and be re-elected, but there would then be no selection effect.

The plan of the paper is as follows. In the next section we briefly review the literature on empirical estimation of the effect of term limits. In section 3 we present our basic political agency model with two-term limits. Section 4 first describes our method of structural estimation and the data. We then present and discuss our estimates and their implications in Section 5. The final section presents conclusions.

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) and List and Sturm (2006) compare the performance of reelection-eligible governors to that of lame-duck governors i.e. governors that are in their last legal term in office. They find significant differences in fiscal policy and environmental policy respectively. As the authors admit, the estimated difference has a limited interpretation. According to the logic of the model, even if average outcomes were the same, this does not mean elections have no effect, it just means that the discipline and selection effects cancel each other out. This identification strategy in effect measures the difference between the discipline and the selection effects of elections, both predicted by the canonical model. Discipline improves average outcomes in the first term, while selection improves them in the last term, because better-than-average politicians are reelected to a lame-duck term.

Ferraz and Finan (2011) apply the Besley and Case (1995, 2003) identification strategy to a cross-section of all Brazilian mayors in office between 2000-2004, some of which are in their second (and last) term and some of which are in their first term. Besley (2006) provides indirect evidence of a selection effect. He finds that in the U.S. lame-duck governors are more in tune with voter preferences, as measured by interest group ideological rankings.

Alt, Bueno de Mesquita and Rose (2007) also use gubernatorial elections data from the U.S. with the goal to disentangle the discipline and selection effects of elections, while also, in some specifications, distinguishing between changes in outcomes due to selection versus changes due to experience in office. They cannot reject the hypothesis that the discipline and selection effects are almost equal in magnitude. Their identification strategy is based on comparisons across term limit regimes as well as switches between term limit regimes.[3]

To our knowledge, the only structural approach has been that of Sieg and Yoon (2014). They ask whether the mechanism of reelection gives an incentive for incumbents to moderate their fiscal policies – Democratic incumbents act more fiscally conservative, Republican incumbents 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 key respects. Rather than measuring governor performance by the economic variables used by Sieg and Yoon and the reduced-form estimation papers reviewed in this section, we look at voter ratings of a governor's job performance, which we find does a better job of explaining election outcomes. Second, Sieg and Yoon do not consider the moral hazard problem of low effort, which is a focus of our paper, nor do they consider selection over non-partisan characteristics, such as competence (which is assumed to be fully observed) or preference for rent seeking.

[3]For instance, by comparing the outcomes of second-term lame-duck governors (in states with two-term limits) to the outcomes of first-term lame-duck governors (in states with a one term limit), the authors claim to identify a selection effect, since the governors in the first category were screened while those in the second were not. The identification assumption, of course, needs to be that neither group has further reelection incentives, or that those incentives are exactly the same.

There is thus no attempt to the contribution of selection versus discipline on improving outcomes, a focus of much of the earlier literature and of our paper.

3 Model

We start with our benchmark model, which is the simplest political agency model that can generate stochastic policy outcomes and reelection rules. Subsequent versions of the model relax some of this model's assumptions.

3.1 Basic Setup

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. All voters have the same information set and preferences, thus, we simply consider a single representative voter.

The equilibrium concept we use is Perfect Bayesian Equilibrium, defined as follows. First, given her private type and her belief about the voter's strategy, a governor chooses effort to maximize her expected utility. Next, given his belief about the governor's strategy, at every possible node the voter chooses the action that maximizes his lifetime utility. Beliefs are consistent with strategies on the equilibrium path.

3.2 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 $\pi = \mathbb{P}\{\theta = G\}$, 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.[4] For ease of exposition,

[4] 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 governo 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.

we define $c(e;\theta)\, r$ the cost of effort level e for a governor of type θ, where

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

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 unit interval $[0,1]$ when first elected, and it remains the same in all terms while in office.[5] The governor understands that his chance of winning re-election is ρ_H if he exerts high effort and ρ_L if he 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 (4) below). Hence, these probabilities are a combination of the performance of the governor given his effort and the probability of re-election given his performance, and they will be determined in equilibrium.

3.3 Governors' Effort Choice

The problem of a governor of type θ is

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

where e_i is effort in term i, $\mathbf{1}_H$ is an index which equals 1 if $e_1 = H$ and 0 otherwise.

The actions of a good governor are trivial – he exerts high effort in the first term ($e_1 = H$) since it is costless and strictly increases his chances of re-election. Since effort is costless and he is indifferent over effort levels in the second term, we simply assume that $e_2 = H$ as well.[6]

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 it does not have any benefits. To derive e_1, note that if the bad governor exerts high effort in the first term, his payoff is $(1 - c + \rho_H)\, r$, and if she exerts low effort, her payoff is $(1 + \rho_L)\, r$. In words, by exerting high effort the governor would lose some of the first-term rent but increase the chances of reelection and thus enjoying the rent for an extra term. She would therefore find it optimal

[5] 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.

[6] If we assumed that good types like exerting effort (a negative cost), she would strictly prefer $e_2 = 1$. This would also follow if, consistent with what we argue below about the relation of effort to expected performance, the good type preferred higher performance.

to exert high effort if and only if

$$c < \rho_H - \rho_L \tag{2}$$

The voter does not observe c, but understands the maximization problem that governors face. He therefore can calculate the probability δ that a bad governor exerts high effort in her first term, that is, $\delta \equiv \mathbb{P}\left\{e_1 = H|\theta = B\right\}$. Given the assumptions of a uniform distribution for c, we may then write

$$\delta = \mathbb{P}\left(c < \rho_H - \rho_L\right) = \rho_H - \rho_L \tag{3}$$

3.4 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

$$y_i|\left(e_i = H\right) \sim N\left(Y_H, \sigma_y^2\right) \tag{4a}$$
$$y_i|\left(e_i = L\right) \sim N\left(Y_L, \sigma_y^2\right) \tag{4b}$$

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 which term she is in.

We further assume probabilistic voting in that the utility of the voter is affected by a shock $\varepsilon \sim N\left(\mu, \sigma_\varepsilon^2\right)$ 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 governors. Furthermore $\mu < 0$ will capture an incumbent bias.

Define $W\left(y_1, \varepsilon\right)$ as the voter's expected two-term utility after observing the first-term performance of a governor and the election shock

$$W\left(y_1, \varepsilon\right) = y_1 + \beta \max_{R \in \{0,1\}} \mathbb{E}\left\{R\left[y_2 + \beta W\left(y_1', \varepsilon'\right)\right] + (1 - R)\left[W\left(y_1', \varepsilon'\right) + \varepsilon\right]|y_1, \varepsilon\right\} \tag{5}$$

where β is the voter's discount rate between terms of governors 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 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 challenger more appealing. Note that ε does not affect the type or actions of the challenger in her stint should she get elected. 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 the probabilities of each.

We can write the voter's problem as

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

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

$$\begin{aligned} \mathbb{V} \;=\; & [\pi + (1 - \pi)\delta] \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_H}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \\ & + (1 - \pi)(1 - \delta) \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_L}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \end{aligned} \tag{7}$$

where $\phi(\cdot)$ represents the standard normal PDF. (7) makes it explicit that there is uncertainty with respect to the type of the governor, his effort and performance in the first term, as well as the ε 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

$$\mathbb{E}(y_2|y_1) = \hat{\pi}(y_1) Y_H + [1 - \hat{\pi}(y_1)] Y_L \tag{8}$$

where $\hat{\pi}(y_1) \equiv \mathbb{P}(\theta = G|y_1)$, that is, the posterior of the voter on π after observing first term performance. Using (8) we can write $W(y_1, \varepsilon)$ as

$$W(y_1, \varepsilon) = y_1 + \beta \max_{R \in \{0,1\}} \left[R \left\{ \hat{\pi}(y_1) Y_H + [1 - \hat{\pi}(y_1)] Y_L + \beta \mathbb{V} \right\} + (1 - R)(\mathbb{V} + \varepsilon) \right] \tag{9}$$

3.5 Election

If types were observable, the voter would reelect only good governors since they would exert high effort in their second term and bad governors would not. Since neither types nor effort are not observable, and due to the existence of the election shock, the reelection problem becomes an imperfect one. Solving the discrete choice problem in (9), the incumbent would win reelection, i.e. $R = 1$ if and only if

$$\hat{\pi}(y_1) > \frac{(1 - \beta)\mathbb{V} - Y_L + \varepsilon}{Y_H - Y_L} \tag{10}$$

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 large (so that the challenger does not have a large candidate-specific advantage). We can summarize the decision rule $R(y_1, \varepsilon)$ with the following

$$R(y_1, \varepsilon) = \begin{cases} 0 & \text{if } \varepsilon \geq \hat{\varepsilon}(y_1) \\ 1 & \text{if } \varepsilon < \hat{\varepsilon}(y_1) \end{cases} \tag{11}$$

where $\varepsilon = \hat{\varepsilon}(y_1)$ characterizes the points (y_1, ε) for which (10) holds with equality with

$$\hat{\varepsilon}(y_1) = \hat{\pi}(y_1)(Y_H - Y_L) - (1 - \beta)\mathbb{V} + Y_L \tag{12}$$

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

$$\begin{aligned} \hat{\pi}(y_1) &\equiv \mathbb{P}(\theta = G | y_1) = \frac{\mathbb{P}(\theta = G)\mathbb{P}(y_1 | \theta = G)}{\mathbb{P}(y_1)} \\ &= \frac{\pi \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)}{[\pi + (1 - \pi)\delta]\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) + (1 - \pi)(1 - \delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)} \end{aligned} \tag{13}$$

where δ, as defined in (3), is the voter's belief about the probability that a bad governor will exert high effort in her first term.

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

$$\begin{aligned}
\psi\left(y_{1}\right) &\equiv \mathbb{P}\left(R=1 \mid y_{1}\right)=\mathbb{P}\left[\varepsilon<\hat{\varepsilon}\left(y_{1}\right)\right] \\
&= \Phi\left[\frac{\hat{\varepsilon}\left(y_{1}\right)-\mu}{\sigma_{\varepsilon}}\right]
\end{aligned} \tag{14}$$

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

$$\rho_H = \int \psi\left(y_{1}\right) \phi\left(\frac{y_{1}-Y_{H}}{\sigma_{y}}\right) d y_{1} \tag{15}$$

$$\rho_L = \int \psi\left(y_{1}\right) \phi\left(\frac{y_{1}-Y_{L}}{\sigma_{y}}\right) d y_{1} \tag{16}$$

3.6 Equilibrium

A Perfect Bayesian Equilibrium of the game between the 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.

Figure 1 shows a game tree that summarizes the game between the governor and the voter. The sequence of actions and the information structure can be summarized 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, the election shock, 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 $e_2 = H$.

6. Based on e_2, a performance y_2 is drawn by nature giving the utility of the voter in that term

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

3.7 Discussion

Key magnitudes can be summarized as follows:

- Fraction of bad governors who supply effort in their first term (% of bad governors who are disciplined) : $\delta \in [0, 1]$.

- Increase in average first-term performance (discipline effect) : $(1 - \pi)\,\delta\,(Y_H - Y_L) \geq 0$, which is strictly positive if $\delta > 0$.

- Increase in fraction of good governors among second-term governors (selection effect) :

$$\frac{(\pi - \pi^2)\,(1 - \delta)\,(\rho_H - \rho_L)}{\pi\rho_H + (1 - \pi)\,\delta\rho_H + (1 - \pi)\,(1 - \delta)\,\rho_L} \geq 0 \tag{17}$$

  which is strictly positive if $\delta < 1$.

To summarize, the baseline model with a two-term limit generates a discipline and a selection effect. Discipline increases first-term average outcomes, while selection increases final-term average outcomes because more good governors survive reelection and exert high effort. There is a trade-off between the two, however: higher discipline reduces the fraction of bad incumbents that are screened out. The former is strictly positive as long as some bad governors choose to exert effort in their first term; the latter is strictly positive if not all of them do.

3.8 Model with Effort Signal

In this version of the model we allow the voters to observe a noisy signal about the effort level of the governor. We denote this signal with z and assume that it is symmetric and correct with probability ζ, that is

$$\mathbb{P}\left\{z = H | e = H\right\} = \mathbb{P}\left\{z = L | e = L\right\} = \zeta,$$

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 affected 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

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

which would then be used in calculating his expected utility from reelecting the incumbent and hence his reelection rule. $\hat{\varepsilon}\left(y_1, z_1\right)$ and $\psi\left(y_1, z_1\right)$ also have z_1 as an argument since they use $\hat{\pi}\left(y_1, z_1\right)$.

The incumbent understands that there will be a noisy signal about his performance, which will affect her chances of reelection and uses

$$\begin{aligned}
\rho_H &= \int \left[\zeta\psi\left(y_1, H\right) + (1-\zeta)\psi\left(y_1, L\right)\right]\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 \\
\rho_L &= \int \left[(1-\zeta)\psi\left(y_1, H\right) + \zeta\psi\left(y_1, L\right)\right]\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1
\end{aligned}$$

Further details are presented in the Appendix.

4 Solution and Estimation

In this section we discuss our strategy for solving and estimating the benchmark model.

4.1 Solution

The model has seven structural parameters: $\pi, \beta, Y_H, Y_L, \sigma_y, \mu$, and σ_ε. As (3.6) 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 four equilibrium mappings, $\hat{\pi}(y_1)$, $\hat{\varepsilon}(y_1)$, $W(y_1, \varepsilon)$ and $\psi(y_1)$. We solve the equilibrium as follows.

Define two residuals $\mathcal{R}_1$ and $\mathcal{R}_2$ between conjectures for $\mathbb{V}$ and δ and the model-implied values from (7) and (3)

$$
\begin{aligned}
\mathcal{R}_1 \equiv\ & \mathbb{V} - [\pi + (1-\pi)\delta] \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_H}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \qquad (18)\\
& - (1-\pi)(1-\delta) \int\int W(y_1', \varepsilon') \phi\left(\frac{y_1' - Y_L}{\sigma_y}\right) \phi\left(\frac{\varepsilon' - \mu}{\sigma_\varepsilon}\right) dy_1' d\varepsilon' \\
\mathcal{R}_2 \equiv\ & \delta - \int \psi(y_1) \phi\left(\frac{y_1 - Y_H}{\sigma_y}\right) dy_1 + \int \psi(y_1) \phi\left(\frac{y_1 - Y_L}{\sigma_y}\right) dy_1 \qquad (19)
\end{aligned}
$$

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 $\mathbb{V}$, $\hat{\pi}(y_1)$ follows from (13), $\hat{\varepsilon}(y_1)$ follows from (12), $R(y_1, \varepsilon)$ follows from (11), $W(y_1, \varepsilon)$ follows from (9) and $\psi(y_1)$ follows from (14).

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, an observation unit 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 probability of a generic governor winning reelection while producing performance of y_1 and y_2 can be obtained as

$$
\begin{aligned}
p_W(y_1, y_2) \equiv\ & \pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_H}{\sigma_y}\right) \\
& +(1-\pi)\delta\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_L}{\sigma_y}\right) \qquad (20)\\
& +(1-\pi)(1-\delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)\psi(y_1)\phi\left(\frac{y_2 - Y_L}{\sigma_y}\right)
\end{aligned}
$$

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

$$\begin{aligned} p_L(y_1) \equiv \; & \pi\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)[1-\psi(y_1)] \\ & +(1-\pi)\delta\phi\left(\frac{y_1 - Y_H}{\sigma_y}\right)[1-\psi(y_1)] \\ & +(1-\pi)(1-\delta)\phi\left(\frac{y_1 - Y_L}{\sigma_y}\right)[1-\psi(y_1)] \end{aligned} \tag{21}$$

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

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

and the log-likelihood is simply given by

$$\log \mathcal{L} = \sum_{k=1}^{n} L_k \tag{22}$$

Estimating the structural parameters requires maximizing $\log \mathcal{L}$, which we do using standard numerical optimization routines. 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.

5 Results

5.1 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 y in the data is. 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.[7] A large fraction of the 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. For each governor we use 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 and the spillover from the challengers' evaluation of the electorate to the incumbents' JAR, we use JAR up to and including June of an election year at the end of the first term of the incumbent. 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.

Our model places some important constraints on the type of governor we can use in the estimation. We start with the universe of all governors that served from 1950 to present, where we have collected some basic information about the governor, some of which comes from Besley (2006). We also know the outcomes of their reelection efforts.[8] We then apply the following filters to eliminate governors who do not fit our model:

- 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 elections rule in the state changed.
- Drop governors with 2-year terms.
- Drop governor stints (not just the terms) where the governor was appointed, filled in someone else's term or elected through a special election.
- Drop governors who did not complete at least three years in each term in office (for example due to resignation, passing away or being recalled).

These filters yield 149 governors stints.[9] Combining this with the JAR data we compiled

[7] 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.

[8] 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. This analyis is available upon request.

[9] 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.

yield 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 a 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, 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 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 very small correlations with some characteristics of governors such as age, party, education and gender, as well as characteristics of the states they serve in, such as the Census region or division the state is located in.

5.2 Benchmark Model

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, though the estimate is not particularly sharp. 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 statistically, and economically significant. Figure 2 shows the distribution of JAR of 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.

High effort increases the probability of re-election from 45% to 72%. There is also a significant incumbent bias: an incumbent enters the reelection with an advantage that is equivalent to having 9 JAR points more than his actual JAR. The election shock has a very large standard deviation, which shows that there are many elections in which a governor with a low JAR is nonetheless 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 on observed y_1, are illustrated in Figure 3.

These parameter estimates imply the following measures in Table 2 of estimated governor types, governor performance, and the effect of having a two-term rather than a one-term limit (in which there would be neither discipline nor selection effects) on measures of accountability. These numbers come from simulating the model for 100,000 terms. While some of the measures can be computed analytically many cannot, which is why we use a simulation. The first block of the table illustrates the counterfactual case where a governor was restricted to one term of office, in which case only good governors would supply effort, leading to an average performance of 54 JAR points.

In the next block we then look at summary measures for two-term limited governors, which are summarized for discipline and selection in the third block of Table 2. While 52% of governors are good, the possibility of re-election based on performance induces effort on the part of 65% of first-term governors. That is, 13% of first-term governors (or 27% of bad governors) who exert effort in their first term do so because of the possibility of a second term. This implies average performance in term one of 56.7, an increase of 2.7 Jar points (or 5% higher performance). This can be thought of as the welfare gain due to discipline.

In terms of selection, our parameter estimates in Table 1 imply that 60% of second-term governors are good, an increase of 8% over the fraction of good first-term governors. The effect on selection of a higher fraction of good governors on performance cannot be measured simply by taking the difference in average performance among second-term versus first-term governors, since discipline effects, by making it harder to distinguish good from bad governors after the first term, by themselves would lower average second term performance. In fact we see that average second term performance over all re-elected governors is a bit lower than average first term performance. However, since only good governors exert effort in their final term of office, the effect on performance of the increase in the fraction of good governors can be found by taking the difference between average performance in the second of two terms to average performance in the case where governors are (counterfactually) limited to a single term, that is, the fourth line of the third block, showing a an increase in performance due to selection of 3%.We discuss the trade-off between discipline and selection in section 5.4.

5.3 Robustness

We looked at the effect on our results of alternative ways of calculating governor performance from the voter surveys, as shown in Table 3. As discussed in section 5.1, 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). Our alternatives included: using all survey in the first term up to the election (All Surveys); dropping all surveys 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 made clear, the results are robust to all of these alternative performance calculations.

We also considered allowing the distributions of Y_H and Y_L to have different variances (Free σ_y^H). This change also had no significant effect on the results.

5.4 Noisy Effort Signal

The estimated implications of a noisy effort signal discussed in section 3.8 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 gives 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 calculation of residuals analogous to (18) and (19).)

The first column shows the effect of a partially informative signal of effort, $\zeta = .75$. Relative to case of no signal shown in Table 2, the fraction of governors disciplined rises from 27% to 30%. This is as theory would lead us to expect, as a higher probability of observing "shirking" leads to more bad types exerting effort. We also find a stronger selection effect, though the change is small 8.5% instead of 7.8% when effort was unobservable. 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 second column of Table 4. (This, of course, is not equivalent to perfect observability of type, since bad governors can 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. The reason that full observability of effort does not lead to all bad types exerting effort in their first term is the existence of the election shock. Even if a type is known to be bad – perfectly indicated in this case by low effort – he can still win re-election with a sufficiently negative (i.e. challenger-favoring) realization of ε ($\rho_L = .45$); conversely, even if a type exerts effort, he is not guaranteed re-election ($\rho_H = .72$) if the realization of ε is sufficiently high. So, bad types with a high draw of c will find it optimal to exert low effort. Hence, discipline is mitigated by the randomness of re-election probability, as theory once again would suggest.

We therefore simulated the model with full observability of effort ($\zeta = 1$) but with the election shock very close to 0 ($\sigma_\varepsilon = 0$), so that the election shock is known to take its mean value $\mu = -9.3$. There is no possibility of a very negative realization of ε to "save" a low effort incumbent. Since c, the cost to a bad governor of exerting effort as a fraction of office rents, has a maximum of 1, all bad governors exert effort, and all are re-elected. 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. There are three groups of first-term governors: good; bad who exert high effort; and, bad who exert low effort. A more precise signal of effort makes it easier to separate the first group from the third, therefore improving selection. However, it also induces more bad governors to exert effort, making it more difficult to distinguish types on the basis of effort. Hence, a noisy signal of effort has effects on selection working in conflicting directions. 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 that the incumbent loses re-election, discipline is perfect (that is, there will be no governors in the third group), but the selection effect goes to zero precisely because of this.

6 Conclusions

In this paper we structurally estimated a political accountability model in order to separate empirically the discipline and selection effects of elections. We estimated the effects on the performance of U.S. governors of common two-term limits relative to the counterfactual case of where re-election 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 governor type, a possibility that reduced-form estimation does not allow. This is what allows counterfactual experiments to estimate the welfare effects of term limits.

We found a significant discipline effect of re-election incentives, as well as a somewhat weaker selection effect. Perhaps this is not surprising, but quantifying these effects allow 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 political economy models.

Further research is aimed at addressing 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? These 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.

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] 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.

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

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

[5] 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.

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

[7] 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.

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

[9] 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.

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

[11] 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 Representation*, Cambridge, UK: Cambridge University Press.

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

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

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

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

[16] Nordhaus, W. (1975) "The Political Business Cycle," *Review of Economic Studies* 42: 169-190.

[17] 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.

[18] Smart and Sturm (2008) "Term Limits and Electoral Accountability," CEPR Discussion paper No. 4272.

[19] Vlaicu, R. and A. Whalley (2011) "Hierarchical Accountability in Government: Theory and Evidence," working paper, University of Maryland.

APPENDIX – 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 baseline model. Let z_1 be a pre-election signal about the incumbent's first-term effort e_1. Assume that the signal is symmetric:

$$\mathbb{P}\left\{z_1 = H | e_1 = H\right\} = \mathbb{P}\left\{z_1 = L | e_1 = L\right\} = \zeta \tag{23}$$

where $\frac{1}{2} \leq \zeta \leq 1$. The parameter ζ thus measures the informativeness of the signal.

The voter's value function, conditional on first-term observables would be:

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

and the voter's ex-ante value function would be:

$$\begin{aligned} \mathbb{V} &= \left[\pi + \left(1-\pi\right)\delta\right] \int\int \left[\zeta W\left(y_1, 1, \varepsilon\right) + \left(1-\zeta\right) W\left(y_1, 0, \varepsilon\right)\right] dF\left(y_1\right) dH\left(\varepsilon\right) + \\ &\quad + \left(1-\pi\right)\left(1-\delta\right) \int\int \left[\zeta W\left(y_1, 0, \varepsilon\right) + \left(1-\zeta\right) W\left(y_1, 1, \varepsilon\right)\right] dG\left(y_1\right) dH\left(\varepsilon\right) \end{aligned}$$

Expected second-term governor performance, given the electorate's observables, is:

$$\mathbb{E}\left(y_2 | y_1, z_1\right) = \hat{\pi}\left(y_1, z_1\right) Y_H + \left[1 - \hat{\pi}\left(y_1, z_1\right)\right] Y_L$$

The threshold election shock is:

$$\hat{\varepsilon}\left(y_1, z_1\right) = \hat{\pi}\left(y_1, z_1\right)\left(Y_H - Y_L\right) - \left(1-\beta\right)\mathbb{V} + Y_0$$

The incumbent's posterior reputation becomes:

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

because

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

Reelection probabilities conditional on voter information are:

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

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

$$\begin{aligned}\rho_H &= \int [\zeta\psi(y_1, z_1 = H) + (1-\zeta)\,\psi(y_1, z_1 = L)]\,dF(y_1) \\ \rho_L &= \int [(1-\zeta)\,\psi(y_1, z_1 = H) + \zeta\psi(y_1, z_1 = L)]\,dG(y_1)\end{aligned}$$

Finally, residuals, analogous to (18) and (19), are:

$$\begin{aligned}\mathcal{R}_1 &\equiv \mathbb{V} - [\pi + (1-\pi)\,\delta]\underbrace{\int\int [\zeta W(y_1, z_1 = H, \varepsilon) + (1-\zeta)\,W(y_1, z_1 = L, \varepsilon)]\,dJ(\varepsilon)\,dF(y_1)}_{\mathbb{A}_1} \\ &\quad - (1-\pi)(1-\delta)\underbrace{\int\int [\zeta W(y_1, z_1 = L, \varepsilon) + (1-\zeta)\,W(y_1, z_1 = H, \varepsilon)]\,dJ(\varepsilon)\,dG(y_1)}_{\mathbb{A}_2} \\ \mathcal{R}_2 &\equiv \delta - \int [\zeta\psi(y_1, z_1 = H) + (1-\zeta)\,\psi(y_1, z_1 = L)]\,dF(y_1) + \\ &\quad + \int [\zeta\psi(y_1, z_1 = L) + (1-\zeta)\,\psi(y_1, z_1 = H)]\,dG(y_1)\end{aligned}$$

Table 1: Parameter Estimates

π	0.52	δ	0.27
	(0.08)		(0.05)
Y_L	43.33	ρ_L	0.45
	(2.43)		(0.07)
Y_H	63.99	ρ_H	0.72
	(1.75)		(0.05)
σ_y	9.84	$\mathbb{V}$	391.78
	(0.77)		(9.57)
μ	-9.34	σ_ϵ	13.07
	(0.83)		(1.75)

Note: White standard errors are below estimates.

Table 2: Comparison of 2-Term Benchmark with a 1-Term Limit

One-Term Limit	
Good governors	51.7%
High effort	51.7%
Average Performance (JAR Points)	54.0
Life-time Discounted Welfare for Voter	359.18

Two-Term Limit (Benchmark)	
Good governors in Term 1	51.8%
Good governors in Term 2	59.5%
Good governors overall	54.8%
High effort in Term 1	64.7%
High effort in Term 2	59.5%
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	374.44

Measures of Interest	
Discipline 1 : Fraction Disciplined	26.8%
Discipline 2 : Increase in Performance In Term 1 Relative to 1-Term	5.0%
Selection 1 : Increase in Good Governors in Term 2 Relative to 1-Term (in p.p.)	7.83
Selection 2 : Increase in Performance In Term 2 Relative to 1-Term	3.0%
Performance / Welfare Improvement Relative to 1-Term Limit	4.2%

Table 3: **Robustness of Estimation Results**

	Benchmark	**All Surveys**	**No Election Year**
π	0.52	0.50	0.53
	0.08	*0.10*	*0.09*
δ	0.27	0.26	0.26
	0.05	*0.06*	*0.07*
Discipline 1	26.8%	26.2%	26.3%
Discipline 2	5.0%	4.9%	4.7%
Selection 1	7.8%	7.8%	7.5%
Selection 2	2.9%	2.9%	2.8%
Welfare Gain	4.3%	4.1%	4.0%

	Year-by-Year Average	**Median JAR**	**Minimum JAR**
π	0.53	0.54	0.45
	0.09	*0.09*	*0.12*
δ	0.29	0.25	0.23
	0.07	*0.07*	*0.06*
Discipline 1	29.4%	24.7%	23.2%
Discipline 2	5.4%	4.6%	6.5%
Selection 1	8.4%	7.3%	7.0%
Selection 2	3.2%	2.9%	3.5%
Welfare Gain	4.6%	4.0%	5.4%

	Keep Undecideds	**Free σ_y^H**	
π	0.46	0.48	
	0.12	*0.08*	
δ	0.23	0.26	
	0.06	*0.08*	
Discipline 1	23.5%	26.0%	
Discipline 2	4.3%	5.1%	
Selection 1	7.0%	7.6%	
Selection 2	2.4%	2.8%	
Welfare Gain	3.6%	4.2%	

Notes: See Table 2 for the definitions of the discipline and selection variables.

Table 4: Results from the Version with Effort Signal

	$\zeta = 0.75$	$\zeta = 1$	$\zeta = 1$ **and** $\sigma_\epsilon = 0$
δ	0.30	0.43	1.00
Discipline 1	29.8%	42.8%	100.0%
Discipline 2	5.5%	7.9%	18.4%
Selection 1	8.5%	10.1%	0.0%
Selection 2	3.2%	3.8%	0.0%

Notes: See Table 2 for the definitions of the discipline and selection variables.

Figure 1: Game Tree

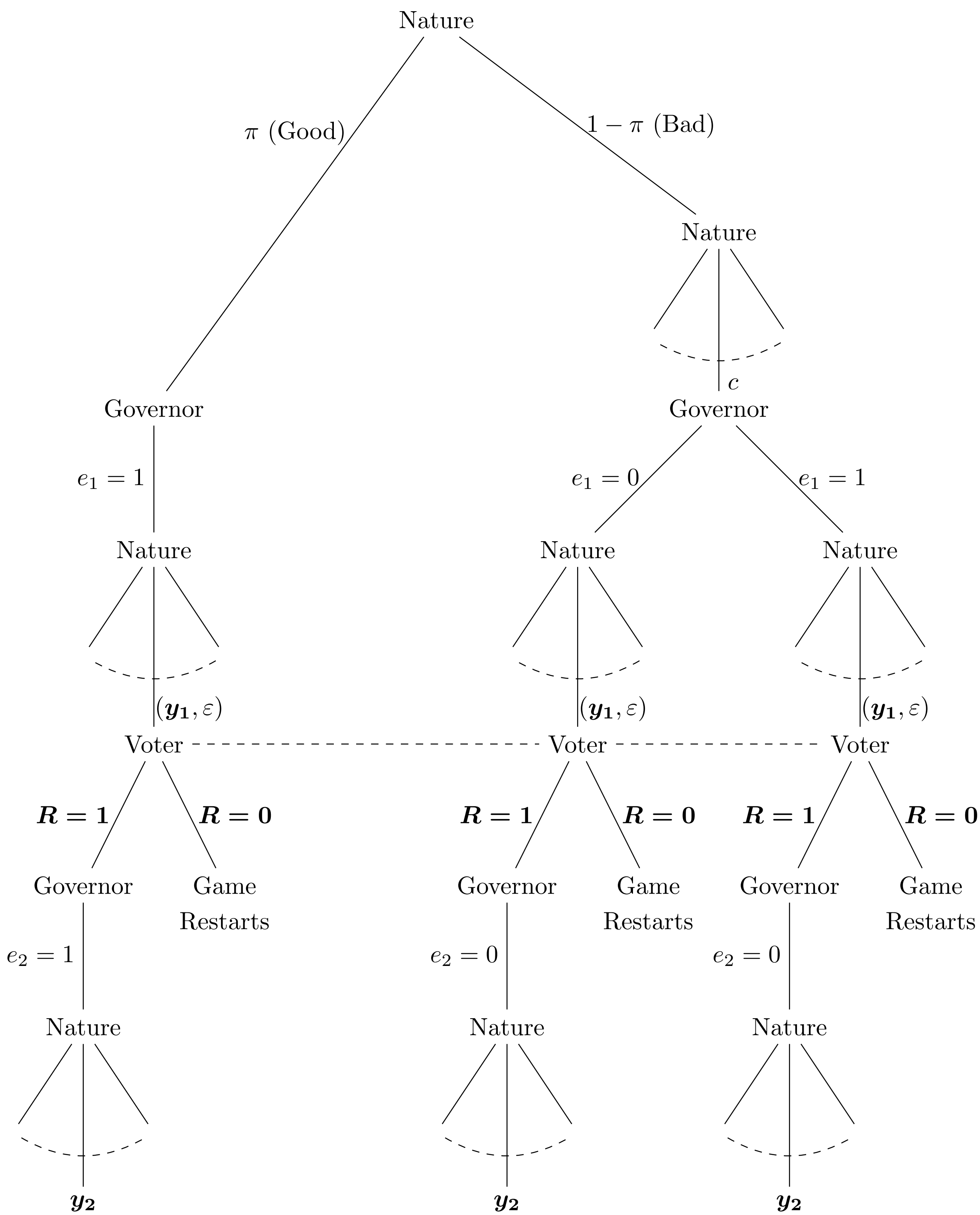

Figure 2: Outcome (JAR) Distributions (Only Reelected Incumbents)

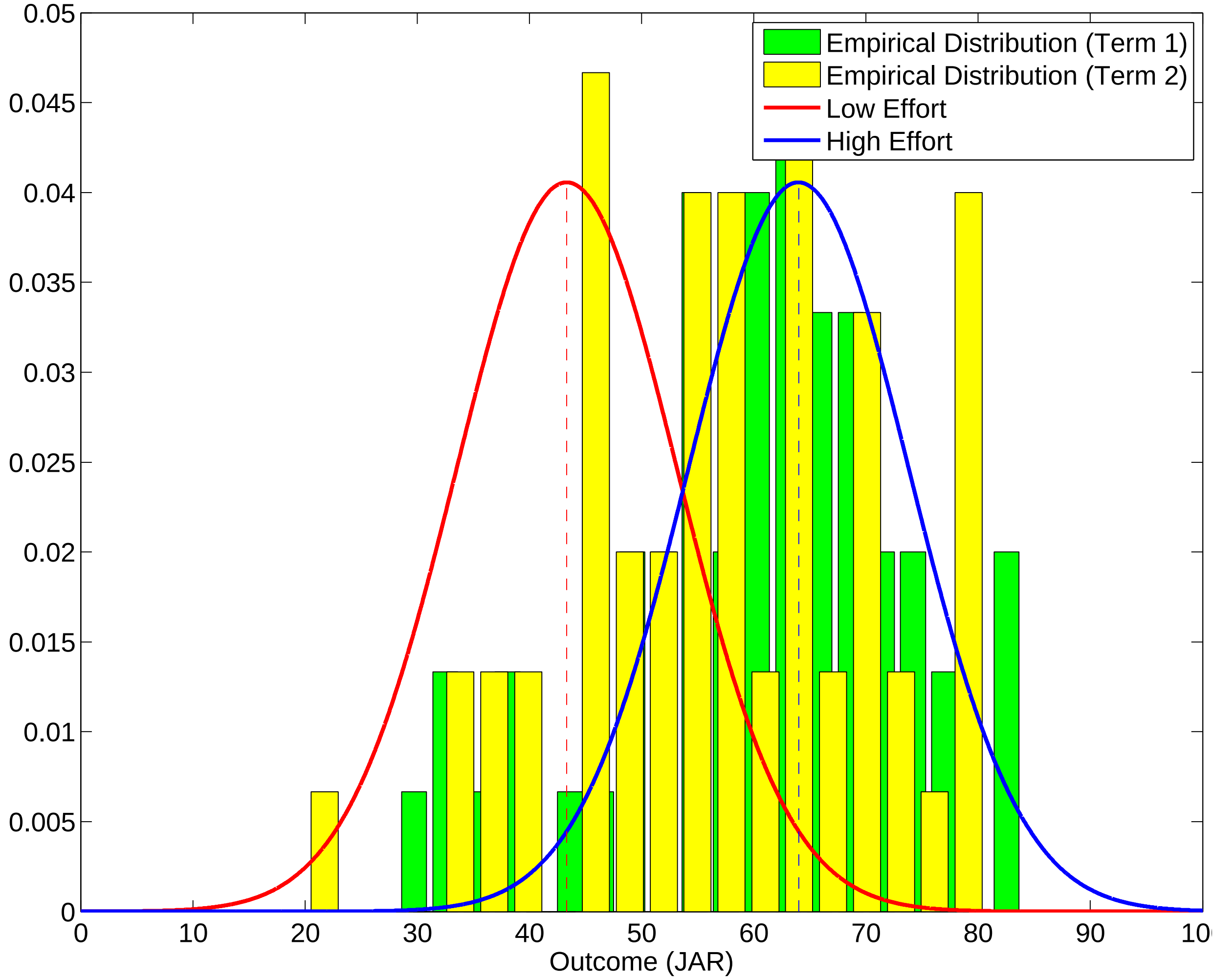

Figure 3: Equilibrium Mappings