Electoral Accountability and Control in U.S. Cities*

Holger Sieg
University of Pennsylvania and NBER

Chamna Yoon
Korea Advanced Institute of Science and Technology

April 11, 2022

*We would like to thank two anonymous referees, Greg Crawford, Dennis Epple, John Ferejohn, Matias Iaryczower, Nolan McCarty, Jean Marc Robin, Thomas Romer, Howard Rosenthal, Bernard Salanié, Al Slivinski, Frank Wolak, as well as seminar participants at the Bank of Portugal, the Federal Reserve Bank in Cleveland, the University of Michigan, the University of Pennsylvania, Princeton University, Sciences Po, Stanford University, Western Ontario, and the University of Zürich for comments and suggestions. Sieg would like to thank the National Science Foundation for financial support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Yoon acknowledges support from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A8033628). The paper was edited by James J. Heckman.

Abstract

We consider a dynamic game of electoral competition with adverse selection, moral hazard, and imperfect monitoring. We show that this dynamic game can be estimated using a flexible maximum likelihood estimator. We implement the estimator using data from recent mayoral elections in large U.S. cities with binding two-term limits. Our empirical findings suggest that there are large differences in performance among different types of mayors. We find an economically important degree of policy responsiveness, with effort accounting for a larger fraction of the total effect than selection. Finally, we evaluate several institutional reforms that promise to increase policy responsiveness.

1 Introduction

An important concern for any representative democracy is whether elections can serve as an efficient mechanism of political accountability. Recent theoretical research has focused on determining under what conditions elections can successfully provide incentives for politicians to generate outcomes desired by the majority of voters. This question is particularly relevant if politicians are viewed as citizen candidates who cannot credibly commit themselves to policies and have, therefore, strong incentives to pursue their own objectives once in office.[1] Repeated elections can mitigate the commitment problems of officeholders whose ideal policies are different from those desired by the majority of voters. Short-run incentives may be tempered by the desire to be re-elected, inducing politicians to compromise or exert more effort by choosing policies that are more desirable for voters.[2] In the presence of term limits, however, the possibilities for policy responsiveness are attenuated. From an empirical perspective, the key question is then how much does policy respond to voters' preferences? The main objective of this paper is to show that we can answer this question by estimating a class of dynamic electoral games with adverse selection, moral hazard, and imperfect information. In particular, we estimate the degree of policy responsiveness in competitive U.S. mayoral elections and study the impact of institutional reforms on policy responsiveness.

The starting point of our analysis is a class of rent-seeking models proposed by Barro (1973) and Ferejohn (1986). In a rent-seeking environment, politicians have a short-run incentive to shirk while in office or equivalently to engage in rent-seeking activities that are not in the voters' best interests. We consider a dynamic version of the rent-seeking game with adverse selection, moral hazard, imperfect monitoring, and a binding two-term limit which is based on Banks and Sundaram (1998) and Duggan (2017). We allow for

[1]The citizen candidate approach is due to Osborne and Slivinski (1996) and Besley and Coate (1997).

[2]See Duggan and Martinelli (2017) for a recent survey of the electoral accountability literature.

stochastic reelection shocks which are essential to generate a model that can fit all key features of the data.

Since the incumbent type is unobserved and outcomes are imperfect measures of effort, voters use threshold strategies to effectively screen out underperforming incumbents. Voters thus based their decisions primarily on the observed performance of incumbents while in office, and not on political promises during campaigns or announced platforms. In equilibrium, electoral competition generates policy responsiveness via two separate channels. First, there is a selection effect since second-term incumbents tend to be of higher quality than first-term incumbents. Second, incumbents exert effort as long as they are not term-limited. One main empirical objective of this paper is then to quantify the importance of these two channels in determining the magnitude of policy responsiveness.

Estimating dynamic principal-agent models with imperfect monitoring is challenging and, to our knowledge, no general solutions to this problem have been proposed in the literature thus far. One problem stems from the fact that there are well-known non-convexities that arise in the politician's effort decision problem when there is imperfect monitoring. As a consequence, the solution to the effort choice may not be unique. When there are multiple solutions to the effort choice problem there is scope for the existence of mixed-strategy equilibria. The estimation of these models then needs to account for the fact that equilibria may be in pure or mixed strategies.

Since there is a finite number of politician types, the model generates distributions of observed policy outcomes that can be characterized by a mixture of normal distributions, assuming that the noise of the monitoring technology is normally distributed. If the equilibrium is in pure strategies the number of mixture components is equal to the number of types of incumbents. If equilibrium is in mixed strategies there can be more mixture components since some types will randomize among the optimal pure strategies. The model also implies that the distribution of observed second-term policies is another mixture of normal distributions. However, the model predicts an important difference. All incumbents play pure strategies in the second term since they face a binding two-term limit. Therefore, the number of components of the mixture distribution of second-term policies is given by the number of types that are reelected in equilibrium.

To generate a well-behaved likelihood function, we assume that there is a difference in the information set between voters and the econometrician. Voters are better informed than the econometrician. Hence, we only observe a noisy measure of the performance measure observed by voters in our model. We use latent factor methods, along the lines suggested by Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Schennach (2010), to model the underlying process that generates observed political performance measures. This allows us to use more than one performance measure in estimation. Moreover, we introduce election shocks into the model which implies that voter's cut-off strategies depend on the realization of the election shock. This approach then guarantees that the parameters of the model can be estimated using a flexible maximum likelihood estimator.

Our empirical analysis focuses on mayoral elections in large U.S. cities. We view a rent-seeking model as more appropriate than a spatial model.[3] Ideology is less important in city politics than in state and federal politics.[4] Our empirical analysis covers all mayoral elections in large U.S. cities between 1990 and 2017. We focus on the subsample

[3]As we discuss in detail below, a spatial model in which behavior is largely driven by differences in ideology may be more appropriate to understand electoral competition at the federal or state level. We have considered how to estimate these models in Sieg and Yoon (2017).

[4]There is some compelling evidence that suggests that ideological differences are not important at the local level of government. Ferreira and Gyourko (2009) use a regression discontinuity design and find that there are no differences in policy outcome between Democratic and Republican mayors in closely contested elections. Of course, this does not mean that there are no differences in ideology among U.S. cities, but it suggests that heterogeneity in ideology can be captured by a simple fixed-effects specification.

of cities that have a binding two-term limit for mayors. Approximately, half of all large cities in the U.S. have these term limits. We consider a variety of different observed outcomes such as expenditures on education and welfare, employment rates, and crime rates that can serve as noisy measures of political performance.

We start our empirical analysis by providing a careful non-parametric analysis to test for the presence of adverse selection and moral hazard. This analysis is in the spirit of the seminal paper by Chiappori and Salanie (2000). Our model has two predictions that can be directly tested without estimating the structural parameters. First, comparing first-term policies of politicians who are not reelected to first-term policies of those who won reelection, the model predicts that there should be a significant positive difference due to screening or selection. Second, comparing second-term policies with first-term policies of reelected incumbents, the difference in mean outcomes should be negative due to moral hazard and mean reversion. Using our three outcome measures, we show that the data are broadly consistent with these two predictions. We then show that we can use latent factor models to reduce the dimensionality of the policy space to a single dimension. These findings provide strong empirical support for our modeling approach.

We use standard model selection criteria to determine our preferred model specification. We find that a parsimonious model with three types of politicians and three election shocks fits the observed data well. Our empirical findings suggest that low-quality politicians account for 14 percent of all politicians while medium-quality politicians account for 62 percent. The remaining 24 percent are high-quality types. Hence, there is much scope for adverse selection. High-quality mayors lower the crime rate by more than 79 violent crimes per 100,000 individuals, increase the employment rate by 1.01 percentage points, and increase expenditures per capita on education and welfare by $220 compared to low-quality mayors. These differences in outcomes are not only statistically significant but also economically important. Nevertheless, we find that a significant fraction of low-and medium-quality types are reelected to a second term since voters only have access to a fairly noisy monitoring technology and election shocks may favor incumbents. However, the estimated benefits of holding office can be large which provides strong incentives to exert effort. In particular, all types exert substantial additional effort in the first term to win election to a second term. These findings suggest that there is much scope for moral hazard. We can decompose the policy responsiveness in our model into an effort effect, which arises since some first-term politicians exert substantial effort to gain reelection, and a selection effect, which results from the fact that voters use threshold strategies to screen out underperforming incumbents. We find that the effort effect is approximately three times as large as the selection effect.

Finally, we study three institutional changes that have been suggested to improve the degree of policy responsiveness. Investing in a better monitoring technology increases accountability allowing voters to improve the screening of poorly performing incumbents. Equally important, it provides incentives for medium- and high-quality types to exert additional effort. As a consequence, a better monitoring technology improves both dimensions of policy responsiveness. In contrast, increasing the benefits of holding office, primarily affects the effort margin, but has negligible effects on selection. Finally, we study policies that aim to increase the overall quality of the candidate pool. We find that increasing the fraction of high-quality types increases selection, but has an ambiguous effect on effort. While medium types tend to increase their efforts, high-quality types tend to decrease effort generating a non-monotonic overall aggregate effort effect.

Our paper is related to, at least, five areas in econometrics, empirical microeconomics, and political economy. First, our paper is related to the recent methodological literature on the identification and estimation of dynamic games. Some notable recent papers include Bajari, Benkard, and Levin (2007), Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008), Merlo and Tang (2012), and Hu and Shum (2013). In contrast to those papers, we focus on a new class of games with adverse selection and moral hazard that requires a different identification and estimation strategy than those considered in the previous papers.

Second, our paper builds on the econometric literature that has devised tests for moral hazard and adverse selection in a variety of different markets. Chiappori and Salanie (2000) and Chiappori, Jullien, Salanie, and Salanie (2006) test for asymmetric information in insurance markets. More recent papers have focused on revealing borrowers' private information through signaling devices in credit markets. Examples are Adams, Einav, and Levin (2009), Einav, Jenkins, and Levin (2012), Kawai, Onishi, and Uetake (2018). Xin (2019) estimates a model with moral hazard and adverse selection using data from online loans.

Third, our paper is related to the recent literature on the identification and estimation of optimal contract models such as Perrigne and Vuong (2011). Closely related to our study are recent papers by Gayle and Miller (2015) and Gayle, Golan, and Miller (2015) who estimate models of managerial compensation in the presence of adverse selection and/or moral hazard. Theses papers consider optimal contract models in which the principal can offer the agent a smooth earnings contract that is contingent on a continuous performance measure such as the stock price. In contrast, we focus on the case in which an infinitely-lived principal can only incentivize a sequence of shorter-lived agents using a discrete retention decision.

Fourth, our paper adds to the recent literature on estimating game-theoretic models in political economy.[5] Our paper is most closely related to recent research on estimating

[5]Other examples of theory-based estimation in political economy are Merlo (1997), who estimates a dynamic bargaining model of government formation, and Diermeier, Eraslan, and Merlo (2003), who extend that framework and provide additional evidence in support of the bargaining approach using data from a variety of European countries. Coate and Conlin (2004) and Coate, Conlin, and Moro (2008)

dynamic games of electoral competition. An important paper in the literature is Aruoba, Drazen, and Vlaicu (2019) who estimate a dynamic moral hazard model with two types that play a pure strategy equilibrium. Our model is richer and more closely related to the original model by Banks and Sundaram (1998) than the model estimated in that paper. For example, our estimation strategy allows for an arbitrary number of types, not just two, and fully endogenizes effort choices of all types. This feature of the model also allows us to study how the entry of new types can affect the degree of policy responsiveness along the equilibrium path. Moreover, our analysis accounts for the possibility of mixed-strategy equilibria, which potentially arise in these types of models. From an empirical perspective, we provide new non-parametric tests of the key predictions of our model which are missing from Aruoba et al. (2019). Finally, we measure policies using standard proxies such as expenditures, crime, and employment while Aruoba et al. (2019) measure outcomes using job approval ratings. As consequence, their analysis has little to say about policy responsiveness which is the main focus of our empirical analysis.

This paper also significantly differs from Sieg and Yoon (2017) who estimate a pure adverse selection model, in which politicians differ by ideology and ability, using data on gubernatorial elections. Here we estimate a rent-seeking model using a new data set that focuses on mayoral elections in the U.S. We assume that voters have only access to an imperfect monitoring technology while our previous paper assumed perfect monitoring. As a consequence, the equilibria may be in mixed strategies and we need a very different

estimate models of voter turn-out using data from Texas liquor referenda. Degan and Merlo (2011) also estimate a model of turn-out in multiple elections. Iaryczower and Shum (2012) estimate a game with asymmetric information to describe the voting behavior of judges in appellate courts. Myatt (2007) and Kawai and Watanabe (2013) consider models of strategic voting. Knight and Schiff (2010) estimate a model of social learning in presidential primaries. Lim (2013) studies differences between appointed and elected public officials. Garcia-Jimeno (2016) estimates a model of moral conflict to explain policies under prohibition.

strategy to identify and estimate this model.

Finally, our empirical analysis is also related to previous studies that have tested predictions of accountability models using linear regression models. One important paper is Alt, Bueno de Mesquita, and Rose (2011) who also study gubernatorial election. They identify the effort or accountability effect by comparing first-term outcomes between governors that face a binding one-term limit and those of governors that face a binding two-term limit. They identify the competence or selection effect by comparing second-term policies of two-term-limit governors with first-term policies of one-term-limit governors. To our knowledge, no large U.S. city uses a one-term limit. Instead, we follow Besley and Case (1995) and others who study elections with incumbents who face a binding two-term limit.[6]

The rest of the paper is organized as follows. Section 2 provides a dynamic game of electoral competition. Section 3 introduces our data set. Section 4 provides some new non-parametric tests of the main predictions of these types of accountability models. Section 5 derives our strategy to estimate the parameters of the dynamic game. Section 6 reports the estimation results. Section 7 reports our measures of the degree of policy responsiveness that arise in equilibrium and discusses potential reforms that can be implemented to increase the degree of policy responsiveness. Section 8 offers some conclusions.

[6] Approximately 50 percent of all large U.S. cities have adopted some type of term limit in the past decades.

2 The Model

2.1 A Dynamic Game of Electoral Competition

We consider a discrete-time, infinite-horizon electoral game with periods indexed by $t = 1, 2, ...$ In each period, an incumbent mayor chooses a policy $x_t \in X$. Infinitely-lived voters observe a noisy signal of the policy $y_t \in Y$, which is a scalar. If the incumbent is in the first term, then an election is held in which the incumbent is matched against a randomly drawn challenger. If the incumbent is in the second term, a challenger is picked randomly in an open election as the new mayor, since the incumbent faces a binding two-term limit.[7]

There is a continuum of politicians that is partitioned into a finite set of types $T = \{1, ...n\}$ with $n \geq 2$. Types of politicians are independent and identically distributed. Let p_j denote the probability of each type in the population. A politician type j is characterized by its preferences. In particular, the flow-payoff of type j when in office for an arbitrary policy x is given by

$$w_j(x) + \beta_j \quad = \quad -(x - \hat{x}_j)^2 + \beta_j \tag{1}$$

where $\hat{x}_j$ is the policy choice of type j that maximizes $w_j(x)$ (or minimizes effort costs); and β_j captures the office benefit.[8] Without loss of generality, let us order the types such that $\hat{x}_1 < ... < \hat{x}_n$. Hence, $\hat{x}_j$ can be interpreted as the quality of type j.

[7]This model was developed by Banks & Sundaram (1993,1998) and recently studied by Duggan (2017). We extend the baseline model to allow for reelection shocks and type-specific benefits of holding office. The exposition of our model largely follows Duggan (2017) which should be consulted for additional discussions of the key assumptions and concepts.

[8]Note that along the equilibrium path no politician will ever choose $x < \hat{x}_j$ as long as the benefits of holding office are sufficiently high. As a consequence, this objective function is equivalent to a quadratic effort cost function that is commonly used in applied contract theory.

Preferences are private information; voters do not observe the politicians' types. The policy choice x_t is also not directly observed by voters, but the outcome y_t is publicly observed. Hence, we consider a game of imperfect monitoring in which the observed policy outcome y_t depends stochastically on the policy choice x_t:

$$y_t = x_t + \epsilon_t \tag{2}$$

where the shocks ϵ_t are independently distributed with common density $f(\cdot)$. We assume that the distribution of y_t given x_t satisfies the standard monotone likelihood ratio property. Greater policy outcomes induce the voter to favorably update her beliefs about the policy adopted by the incumbent in the first period. In the empirical model, we assume that the ϵ_t are normally distributed with mean zero and constant variance σ_ϵ^2.

Following the citizen-candidate approach of Osborne and Slivinski (1996) and Besley and Coate (1997), politicians and voters cannot make binding commitments regarding future actions. Voters, including politicians who are out of office, receive a payoff:

$$u(y, d, \kappa) = u(y) + d\,\kappa \tag{3}$$

where $u(y)$ is monotonically increasing in y, d is an indicator variable that is equal to one if the incumbent is in the second term and zero otherwise, and $\kappa \in \mathcal{K}$ is the random utility shock that voters get from reelecting an incumbent to a second term. We assume that κ is a discrete random variable. Hence, the preference shock κ_k is realized with probability g_k for $k = 1, .., K$. Moreover, κ is realized before the voters make reelection decisions, but after politicians make effort decisions. These shocks capture random events that affect election outcomes, but are not related to policy choices made by the mayor. The timing of the election implies that voter's cut-off strategies depend on the realization of the shock while politicians' effort decisions do not. As we discuss in detail below, this shock, therefore, generates a model that is sufficiently flexible to fit the data in our application.[9]

[9]Note that this specification abstracts from heterogeneity among voters since the policy space is one-

In our empirical model we assume that voters' utility is linear, i.e. $u(y, d, \kappa) = y + d\ \kappa$.

Citizens – voters and politicians – have a common discount factor $\delta \in [0, 1]$. Given sequences of policy choices $\{x_t\}_{t=1}^{\infty}$, policy outcomes $\{y_t\}_{t=1}^{\infty}$, the incumbency status of the mayor in office $\{d_t\}_{t=1}^{\infty}$, and election shocks $\{\kappa_t\}_{t=1}^{\infty}$, the total payoff of a citizen is the discounted sum of per-period payoffs,

$$\sum_{t=1}^{\infty} \delta^{t-1} \{I_t\ (w_j(x_t) + \beta_j)\ +\ (1 - I_t)\ u(y_t, d_t, \kappa_t)\} \tag{4}$$

where $I_t \in \{0, 1\}$ is an indicator variable that is equal to one when the citizen holds office in any period and zero otherwise.

2.2 Stationary Electoral Equilibrium

Following Duggan (2017), we restrict attention to a stationary electoral equilibrium which is a refinement of a stationary perfect Bayesian equilibrium. Strategies in this dynamic game are potentially complex, as voters' and politicians' strategies could conceivably depend on observed histories of policy outcomes and electoral outcomes. To rule out implausible behavior by voters and politicians, the literature imposes refinements that strengthen the concept of perfect Bayesian equilibrium.

A stationary strategy of a type j politician is a pair $\pi_j = (\pi_j^1, \pi_j^2)$, where π_j^1 (π_j^2) denotes a mixed strategy in term 1 (2). It is obvious that a second-term incumbent will implement $\hat{x}_j$ as a policy due to the existence of a binding two-term limit. Hence, each type of politician has a dominant pure strategy in the second term. As we will discuss in detail below, first-term strategies can be pure or mixed depending on the solution of the politician's effort decision problem.

A stationary strategy for the voter is a mapping $\rho : Y \times \mathcal{K} \rightarrow \{0, 1\}$. Note that

dimensional and preferences are monotonic. We discuss relaxing these assumptions in the conclusions.

the voters' strategy depends on both the observed signal y and the reelection shock κ while the politicians' strategy is not conditional on the reelection shock due to the timing assumption made above.

A belief system for the voter is a probability distribution $\mu(\cdot|y)$ defined on $T \times X$ as a function of the observed signal. It represents the voter's posterior beliefs about an incumbent's type and second-term policy choice conditional on the observed first-term policy outcome. The marginal $\mu_T(j|y)$ gives the voter's posterior beliefs about the incumbent's type.

A strategy profile, denoted by $(\pi_1, ..., \pi_n, \rho)$, consists of strategies for voters and politician types. It is sequentially rational if given beliefs a) the incumbent and challenger cannot gain by deviating from their strategies; and b) voters of each type vote sincerely for the candidate that maximizes their expected lifetime utility for any realization of y. Beliefs are consistent with the strategy profile if for every first-term policy y on the path of play implied by $(\pi_1^1, ..\pi_n^1)$, the distribution of beliefs is derived from these strategies via Bayes' rule. A stationary perfect Bayesian Equilibrium is a pair of stationary strategy profiles and beliefs such that the strategies are sequentially rational given the beliefs, and the beliefs are consistent with the strategies.[10]

In addition, the literature imposes two other refinements on the equilibrium. First, the literature assumes that the equilibrium is deferential, i.e. voters favor the incumbent when indifferent. Second, the literature assumes that the voting strategies are monotonic which implies that each voter uses a cut-off strategy. Hence, the voter reelects the incumbent if and only if the payoff from y meets or exceeds a cutoff that depends in our model on the realization of the election shock. A voter's strategy profile that is deferential and monotonic implies that the incumbent is reelected if and only if $y \geq \bar{y}_k$, where $\bar{y}_k$ is

[10]See Duggan (2017) for a more detailed discussion of the equilibrium concept used in these types of games.

the election threshold that corresponds to the realized election shock κ_k. In summary, a stationary electoral equilibrium is a stationary perfect Bayesian equilibrium that is deferential and monotonic. A stationary electoral equilibrium is then characterized by five conditions.

First, voters must be indifferent between reelecting the incumbent and electing the challenger if they observe outcome $\bar{y}_k$, for each for each realization of the election shock κ_k. Formally, let V^C be the continuation value of electing a challenger, then the indifference condition is given by:

$$V_k^I(\bar{y}_k) \equiv \sum_j \mu_T(j|\bar{y}_k) \left[E[u(y)|\hat{x}_j] + \kappa_k + \delta V^C\right] = V^C \quad k = 1, .., K \tag{5}$$

where $V_k^I(y)$ is the value function associated with an incumbent with observed policy outcome y when the election shock is κ_k. Note that we have already imposed the restriction that each incumbent type implements $\hat{x}_j$ in the second term.

Second, each type j knows that she is re-elected to a second term if and only if $y \geq \bar{y}_k$ for each $k = 1, ..., K$. For an arbitrary policy choice x, the reelection probability $r(x)$ is thus given by:

$$r(x) \quad = \quad \sum_{k=1}^{K} (1 - F(\bar{y}_k - x))\; g_k \tag{6}$$

With probability $\sum_k F(\bar{y}_k - x)\; g_k$ the challenger will be elected. The incumbent, therefore, solves the following constrained decision problem:

$$\max_x \; U_j(x) \quad = \quad w_j(x) + \delta \left\{r(x)\; [w_j(\hat{x}_j) + \beta_j - (1-\delta)V^C] + V^C\right\} \tag{7}$$

subject to the reelection constraint in equation (6). The FOC of this problem is:

$$w_j'(x) + \delta \Big[\sum_k \; f(\bar{y}_k - x) g_k\Big]\; [w_j(\hat{x}_j) + \beta_j - (1-\delta)V^C] \quad = \quad 0 \tag{8}$$

The SOC of this problem is given by:

$$w_j''(x) - \delta\Big[\sum_k f'(\bar{y}_k - x)\; g_k\Big]\; [w_j(\hat{x}_j) + \beta_j - (1-\delta)V^C] \;\; \leq \;\; 0 \tag{9}$$

It is well-known that this problem is not necessarily convex. If the decision problem has multiple solutions, the politician will be indifferent among them and, therefore, may play a mixed strategy in the first term. Let us assume for notational simplicity that agents mix over at most I strategies.[11]

Third, updating of voters' beliefs follows Bayes' Rule. Conditional on observing the outcome y the posterior probability that the politician is of type j is:

$$\mu_T(j|y) \;\; = \;\; \frac{p_j \sum_{i=1}^{I} f(y|x_{ij})\; \pi_{ij}^1}{\sum_{k=1}^{n} p_k \sum_{i=1}^{I} f(y|x_{ik})\; \pi_{ik}^1} \tag{10}$$

Fourth, V^C is recursively defined by the following equation:

$$\begin{aligned} V^C \;\; = \;\; & \sum_{j=1}^{n} p_j \sum_{i=1}^{I} \pi_{ij}^1 \Bigg\{ E[u(y)|x_{ij}] \\ & + \delta \sum_{k=1}^{K} \Bigg[(1 - F(\bar{y}_k - x_{ij}))\, \Big(E[u(y)|\hat{x}_j] + \kappa_k + \delta V^C \Big) + F(\bar{y}_k - x_{ij}) V^C \Bigg] g_k \Bigg\} \end{aligned} \tag{11}$$

Note that this equation can be solved analytically for V^C.

Fifth, for each k, equilibrium requires that for $y \geq \bar{y}_\kappa$, voters prefer the incumbent:

$$V_k^I(y) \equiv \sum_{j=1}^{n} \mu_T(j|y) \left[E[u(y)|\hat{x}_j] + \kappa_k + \delta V^C \right] \geq V^C \tag{12}$$

and for $y \leq \bar{y}_\kappa$ voters prefer the challenger:

$$V_k^I(y) \equiv \sum_{j=1}^{n} \mu_T(j|y) \left[E[u(y)|\hat{x}_j] + \kappa_k + \delta V^C \right] \leq V^C \tag{13}$$

Duggan (2017) then proves the existence of a stationary electoral equilibrium in mixed strategies in a model without election shocks and common benefits of holding office. The

[11] See Appendix A for an example and some additional discussions regarding mixed strategy equilibria.

key finding of that paper is that policy responsiveness in dynamic elections is subject to a strict bound, owing to the commitment problem of voters: if first-term office holders generated utility greater than the ideal point of the highest quality type of politicians, then voters would have an incentive to replace office holders after their first term to reap the benefit from the effort of newly elected politicians; but then first-term office holders would shirk, instead. One of the key empirical objectives of this paper is to estimate the magnitude of policy responsiveness.

3 Data

We focus on mayoral elections in the U.S. between 1990 and 2017. We restrict our sample to the 100 largest cities. We also impose the sample restriction that the city had a binding two-term limit during the terms that mayors in the sample served in office.[12] The vast majority of cities started to adopt term limits in the late 1980s and early 1990s. With these sample restrictions, our final sample consists of 111 mayors that served, at least, one term in office. 79 of them reelected to the 2nd term. The remaining 32 mayors were not reelected.[13] The reelection probability for incumbents is, therefore, 71.2 percent in the sample. Note that this is low compared to members of the legislature and comparable to the reelection rate of governors.

There is not a single obvious performance measure for political executives such as mayors and governors. Two approaches have been explored in the empirical literature. Most of the previous studies have followed Besley and Case (1995) and focused on taxes, expenditures, debt service, unemployment rates, and minimum wage policies to study

[12] An online appendix reports all large cities in the U.S. with a two-term limit.

[13] We also drop mayors who are serving his first term in 2017 because his reelection status is not determined. If the term limit was adopted in the middle of the mayor's last term, we drop those observations.

the performance of governors. Alternatively, some studies have used survey data or job approval ratings to measure performance. Our performance measurements are in the spirit of Besley and Case (1995) but adapted to the city context. We use the following three outcomes to measure the performance of mayors: employment rates, expenditures on education and welfare, and the violent crime rate. The use of employment and crime rates is fairly common and does not need much justification.[14] We use expenditures for education to proxy school quality. Similarly, welfare spending serves as a proxy for the quality of services for the old and poor. Note that most large U.S. cities obtain a significant fraction of their school and welfare budgets from state and federal transfers. Hence, higher expenditures often reflect the ability of mayors to negotiate a better deal with governors and state legislatures.

The employment rates are provided by the BLS. The violent crime rate – defined as the number of violent crimes per 100,000 persons – is obtained from the FBI uniform crime reports and available until 2014. City expenditures are obtained from the Lincoln Land Institute. Table 1 summarizes the main economic outcome variables.[15]

Our dynamic game is stationary. Moreover, we need to estimate the model by pooling among cities. To account for heterogeneity among cities as well as time fixed effects, we use a procedure, which was suggested by Besley and Case (1995) and has been employed in most empirical studies since then. First, we regress our outcome measures on time dummies and demographics such as population using a balanced panel. Second, we

[14] For example, Arnold and Carnes (2012) argue that crime played a decisive role in opinion polls in NYC.

[15] Housing prices may also be used as performance measures since amenities and public goods tend to be capitalized into housing prices (Rosen, 1979). However, housing price increases are not necessarily good for renters, i.e voter preferences are not necessarily monotonically increasing in housing prices. We, therefore, do not rely on housing prices as performance measures in this paper. Nevertheless, it is useful to note that the predictions of our model are consistent with the observed housing price patterns.

regress the residuals from the first regression on city dummies for the time periods when the two-term limit was in place. The remainder of the empirical analysis is done based on the residuals from the second regression.

4 Nonparametric Tests

In this section, we provide several non-parametric tests of the presence of adverse selection and moral hazard that are in the spirit of the seminal paper by Chiappori and Salanie (2000). Let us denote the vectors of policy outcomes observed by the econometrician by z_1 and z_2 where the subscript indicates the first or the second term. The starting point of our empirical analysis are nonparametric estimates of the densities of the three outcomes of interest conditional on term and reelection status. Let us denote these densities by $f(z_1|W = 0)$, and $f(z_1|W = 1)$ and $f(z_2)$. Figures 1-3 plot the estimated densities using annual data of the three outcomes during our sample period.

The plots suggest that the distributions of first-term outcomes conditional on winning stochastically dominate the distributions of first-term outcomes conditional on losing. This finding is consistent with one of the main prediction of our model which suggests that voters should use threshold rules to screen out poorly performing candidates in the first term. Moreover, the distributions of first-term outcomes conditional on winning also stochastically dominate the distributions of second-term outcomes which is consistent with the notion that the reduction in effort due to the term limit is stronger than the selection effect.

We have also conducted several formal statistical tests to investigate the differences among these distributions using data at the annual frequency. Table 2 reports the results from pairwise difference-in-means tests and Kolmogorov-Smirnov tests. All pairwise tests suggest that we can reject the null hypothesis that the distributions are the same. Moreover, the differences in means tests tend to reject the null hypothesis that the means are the same. These findings, therefore, confirm the predictions of our model.

We also find that other popular outcome measures such as total own-source revenues or total expenditures are not strategic, i.e. they do not differ by term, reelection status, or party affiliation.[16]

The next step of our analysis is to determine whether it is possible to reduce the dimensionality of the outcome space to a one-dimensional space using techniques from factor analysis. While the observed outcome variables are likely to be correlated with the mayor's performance in office, they are imperfect measures subject to measurement error. Our estimation approach, therefore, assumes that the agents in our model are better informed than the econometricians.[17] Following Carneiro, Hansen, and Heckman (2003), we estimate a variety of different latent factor models. A commonly used test determines the number of latent factors based on the number of eigenvalues of the covariance matrix of the outcomes that are larger than one. This criterion suggests that a one-dimensional latent factor model is appropriate in our application, i.e. we find that only one of three estimated eigenvalues is larger than one. Hence, the lth measurement of outcome y_t, denoted by z_{lt}, in term t can be written as:

$$z_{lt} \;=\; \lambda_l \, y_t + u_{lt} \quad l = 1, 2, .., L \tag{14}$$

[16]Besley and Case (1995), Alt et al. (2011) and Sieg and Yoon (2017) all have used total own-source revenue and total expenditures to measure the ideology of governors. These papers report some evidence that suggests that governors moderate their ideological stand in the first period to get reelected, and then pursue more extreme policies in the second term when they are term-limited. We have tested this hypothesis for mayors in our sample. We find no evidence of that type of strategic behavior for the mayors in our sample. As a consequence, mayors seem to behave differently than governors which may reflect differences in the importance of ideology among federal, state, and local governments in the U.S.

[17]This is a common assumption in the relevant literature since the pioneering work of McFadden (1973).

The factor loadings λ_l are identified using the observed covariance matrix of the measures.[18]

Note that our model generates one policy outcome for each mayoral four-year term. We, therefore, use four-term averages as outcome measures for the remaining analysis. We find that the estimated factor loadings all have the expected sign. Not surprisingly, we find that employment and spending are positively correlated with the latent factor while crime is negatively correlated with the latent factor.

Cunha, Heckman, and Schennach (2010) provide conditions that guarantee that the distributions of the latent factors are non-parametrically identified. Hence, we can obtain non-parametric estimates of the conditional densities of y, denoted by $f(y_1|W = 0)$, $f(y_1|W = 1)$, and $f(y_2)$. Two approaches are commonly used in the literature to estimate these densities. First, we can obtain regression-scored factors which are unbiased estimators of y_1 and y_2 (Thomson, 1951).[19] We then use a kernel estimator to obtain the density of the estimated factors conditional on term and reelection status. Figure 4 plots the estimated conditional densities of the regression-scored common factor.[20]

Not surprisingly, the conditional distributions of the latent factor are similar to the conditional distributions of the three measurements. Again, we conduct pairwise difference-in-means tests and Kolmogorov-Smirnov tests. These tests confirm that the distribution of first-term outcomes conditional on winning stochastically dominates the

[18]We discuss suitable normalizations below. For a recent survey of the measurement error literature see Schennach (2016).

[19]Regression-scored factors tend to perform better than Bartlett-scored factors and other alternatives proposed in the literature. For a detailed discussion of these issues see, for example, Mulaik (2009). With one latent factor, the estimated factors are similar to the first principal component of the data matrix.

[20]The factor loadings in the Figure 4 are 0.60, 0.61, and -0.30, respectively. The scale is set so that the variance of the latent factor is equal to one.

two other distributions.

Second, we can apply Kotlarski's theorem to directly estimate the conditional densities without having to impute the values of y. To see how that works, rewrite the measurement system for two outcomes as:

$$z_l/\lambda_l \quad = \quad y + \epsilon_l/\lambda_l \qquad l = 1, 2 \tag{15}$$

Kotlarski's Theorem implies that the characteristic function of y is given by:

$$\phi_y(t) \quad = \quad \exp\left(\int_0^t \frac{\phi(0,u)}{\phi_1(0,u)} du\right) \tag{16}$$

where $\phi(.,.)$ is the characteristic function of $(z_1/\lambda_1, z_2/\lambda_2)$ with first partial derivative $\phi_1(.,.)$. The density of the latent factor y can then be estimated by

$$f(y) \quad = \quad \frac{1}{2\pi} \int_{-T}^{T} \left(1 - \frac{t}{T}\right) \exp\left(-ity\right) \hat{\phi}_y(t) dt \tag{17}$$

where T is a smoothing parameter and $\hat{\phi}_y(t)$ is an estimator of the characteristic function.[21] The results of this approach are illustrated in Figure 5.[22]

The estimated conditional densities plotted in Figure 5 are smoother versions of the ones shown in Figure 4. Table 3 compares the first two moments of the three conditional densities for the two approaches. We find that there are only small differences in these moments. The main difference is that the densities based on Kotlarski's Theorem have a slightly larger standard deviation than the estimated densities based on the estimated factors.

Summarizing, the empirical evidence provided above suggests that the key predictions of our dynamic game are borne out in the data. There is clear evidence of strategic effort

[21] See Krasnokutskaya (2011) for a detailed discussion of how to implement this estimator.

[22] The factor loadings in the Figure 5 are 0.60, 0.69 and -0.18 respectively. We fixed the first factor loading at 0.60 to compare results between the two approaches. We estimate the remaining factor loadings using the covariance matrix of the observed outcomes.

in the first term relative to the second term and selection between incumbents that win and lose elections. Finally, a one-dimensional latent factor model appears to be sufficient to explain the three outcomes that are subject to strategic behavior. As a consequence, we next discuss how to estimate the dynamic game with moral hazard and adverse selection introduced above.

5 Estimation of the Dynamic Game

We show in this section that we can estimate the parameters of the model using a maximum likelihood estimator. The basic intuition behind our estimation strategy is that our model generates parametric mixture models to explain the conditional densities of the observed outcomes estimated in the previous section. The number of mixtures primarily depends on the number of types of politicians.[23] In addition, we impose the restrictions that the conditional distributions of the latent factor predicted by our model are similar to the non-parametrically estimated distributions shown in Figure 5. Hence, we add a penalty function that matches the deciles of these distributions to the likelihood function.

We start by defining the information set of the econometrician. Let W denote an observed indicator which is equal to one if the mayor is reelected to a second term and zero otherwise. The key informational assumption is that we as econometricians do not have the same information set as the voters in our model. In particular, we only observe a vector of measurements of first-term outcomes z_1. In contrast to the voters in our model, we do not observe the scalar y_1. Moreover, we also observe a vector of measurements

[23] It also depends on whether some types play pure or mixed strategies in equilibrium. In that sense our approach is closely related to the recent econometric literature on mixture models. For more details, see, for example, the survey by Compiani and Kitamura (2016).

of second-term outcomes z_2 if the mayor is reelected to a second term ($W = 1$), but we do not observe y_2. We observe these data for a random sample of mayors in cities with binding two-term limits. The sample size is given by N.

For computational reasons, it is useful to consider a sequential two-stage estimator.[24]

5.1 First Stage

To derive the likelihood function, let us initially assume that all n politician types play pure strategies in equilibrium. We discuss how to extend the estimator to account for mixed strategies and how to test for the existence of mixed-strategy equilibria below.

Recall that the structural parameters of the model are the variance of the monitoring technology σ_ϵ^2, the benefits of holding office $\{\beta_j\}_{j=1}^n$, the parameters of the type distribution $\{\hat{x}_j, p_j\}_{j=1}^n$, and the parameters of the incumbency shock $\{\kappa_k, g_k\}_{k=1}^K$.[25] In the first stage, we treat the first-term effort levels $\{x_j\}_{j=1}^n$ and the cut-off values $\{\bar{y}_k\}_{k=1}^K$ as additional parameters. Hence, define the first-stage parameters when all agents play pure strategies:

$$\theta_1 \;=\; (\{\hat{x}_j, p_j, x_j\}_{j=1}^n, \{g_k, \bar{y}_k\}_{k=1}^K, \sigma_\epsilon) \tag{18}$$

Note that we have replaced the structural parameters $\{\beta_j\}_{j=1}^n$ by $\{x_j\}_{j=1}^n$ and $\{\kappa_k\}_{k=1}^K$ by $\{\bar{y}_k\}_{k=1}^K$. This turns out to be a convenient re-parametrization of the likelihood function and simplifies some of the computational burden associated with the analysis.[26]

Recall that the lth measurement of outcome y_t, denoted by z_{lt}, in term t can be

[24] Note the model can also be estimated using a nested-fixed point maximum likelihood estimator that directly imposes all equilibrium conditions.

[25] The discount factor δ is normalized to 0.8 which corresponds to an annual discount factor of 0.95.

[26] $\{\beta_j\}_{j=1}^n$ and $\{\kappa_k\}_{k=1}^K$ are identified and estimated in the second stage as discussed below.

written as:

$$z_{lt} \quad = \quad \lambda_l \, y_t + u_{lt} \qquad l = 1, 2, .., L, \quad t = 1, 2 \tag{19}$$

We assume that errors in the measurement error model are normally distributed, i.e. $u_{lt} \sim N(0, \sigma_l^2)$ for $l = 1, 2, 3, ..., L$. Recall that the error in the monitoring technology satisfies $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. We also assume that the u_{lt}'s and ϵ_t are mutually independent of each other.[27] As discussed above, the factor loading and the variances of the measurement model are identified from the measurement model and are, therefore, treated as known. Define

$$\eta_t \quad = \quad \begin{bmatrix} \eta_{1t} \\ ... \\ \eta_{Lt} \end{bmatrix} = \begin{bmatrix} z_{1t} - \lambda_1 x_t \\ ... \\ z_{Lt} - \lambda_L x_t \end{bmatrix} = \begin{bmatrix} \lambda_1 \epsilon_t + u_{1t} \\ ... \\ \lambda_L \epsilon_t + u_{Lt} \end{bmatrix} \tag{20}$$

Then the joint distribution of ϵ_t and η_t is given by

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{11}\Sigma_{12} \\ \Sigma_{21}\Sigma_{22} \end{bmatrix}\right) \tag{21}$$

where the components of the covariance matrix are defined as $\Sigma_{11} = \sigma_\epsilon^2$, $\Sigma_{12} = [\lambda_1 \sigma_\epsilon^2, ..., \lambda_L \sigma_\epsilon^2]$, and

$$\Sigma_{22} \quad = \quad \begin{bmatrix} \lambda_1^2 \sigma_\epsilon^2 + \sigma_1^2 & ... & \lambda_1 \lambda_L \sigma_\epsilon^2 \\ ... & ... & ... \\ \lambda_L \lambda_1 \sigma_\epsilon^2 & ... & \lambda_L^2 \sigma_\epsilon^2 + \sigma_L^2 \end{bmatrix} \tag{22}$$

The density of ϵ_t conditional on η_t satisfies:

$$\epsilon_t \mid \eta_t \sim N\left(\Sigma_{12}\Sigma_{22}^{-1}\eta, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \tag{23}$$

[27]Note that the normality assumption is not essential for identification as discussed in detail in Cunha et al. (2010).

Thus, the joint density of performance measures $(z_{11}, ..., z_{L1}, z_{12}, ..., z_{L2})$ of a mayor who wins reelection can be written as:

$$
\begin{aligned}
f_{\theta_1}(z_{11}, ..., z_{L1}, z_{12}, ..., z_{L2}) &= \sum_{j=1}^{n} p_j \left\{ \phi_\eta (z_{11} - \lambda_1 x_j, ..., z_{L1} - \lambda_L x_j) \right. \\
&\times \left[\sum_{k=1}^{K} \left[1 - \Phi \left(\frac{\bar{y}_k - x_j - \Sigma_{12}\Sigma_{22}^{-1}\eta_j}{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}} \right) \right] g_k \right] \Bigg\} \\
&\times \phi_\eta (z_{12} - \lambda_1 \hat{x}_j, ..., z_{L2} - \lambda_L \hat{x}_j) \qquad (24)
\end{aligned}
$$

where $\phi_\eta()$ is a density of η. Similarly, the joint density of the first-term performance $(z_{11}, ..., z_{1L})$ of a mayor who is not reelected to a second term can be written as:

$$
\begin{aligned}
f_{\theta_1}(z_{11}, ..., z_{L1}) &= \sum_{j=1}^{n} p_j \left\{ \phi_\eta (z_{11} - \lambda_1 x_j, ..., z_{L1} - \lambda_L x_j) \right. \qquad (25) \\
&\times \left[\sum_{k=1}^{K} \Phi \left(\frac{\bar{y}_k - x_j - \Sigma_{12}\Sigma_{22}^{-1}\eta_j}{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}} \right) g_k \right] \Bigg\}
\end{aligned}
$$

Note that our model explains the observed distributions of z_1 and z_2 as mixtures of normals. The number of components of each mixture is equal to n, the number of types of politicians in the model. The relative weights of the mixtures depend on the type probabilities p_j. The means of the mixtures are $\lambda_l x_j$ for the first-term policy measurements. Hence, we can treat x_j as an additional parameter. Conditional on x_j the density does not depend on β_j. Also note that $\lambda_l \hat{x}_j$ are the means for the observed second-term policy measurements. Hence, the difference between first-term and second-term policies identifies the degree of effort.

The reelection probability – given by the term in brackets $[\cdot]$ – also depends the distribution of the preference shocks g_k, and the truncation parameters $\bar{y}_k$. Conditional on $\bar{y}_k$, the reelection probability does not depend on κ_k. Hence we treat $\bar{y}_k$ as an additional parameter in the first stage of the estimation algorithm. As we discuss in detail below, we find that the thresholds need to be sufficiently spaced apart to make sure that, the estimated model is consistent with the observation some high-performing mayors are not reelected to a second term while some low-performing mayors are retained.

Mayor i's contribution to the log-likelihood is then given by:

$$\log L_i(\theta_1) = W_i \log[f_{\theta_1}(z_{11i}, ..., z_{L1i}, z_{12i}, ..., z_{L2i})] + (1 - W_i) \log[f_{\theta_1}(z_{11i}, ..., z_{L1i})] \quad (26)$$

The log-likelihood for the sample of size N is given by

$$\log L(\theta_1) = \frac{1}{N} \sum_{i=1}^{N} \log L_i(\theta_1) \quad (27)$$

Additionally, we need to make sure that the predicted conditional distributions of the latent factor y are similar to the non-parametrically estimated conditional distributions shown in Figure 5. We can accomplish that objective by adding a penalty function to the likelihood. This penalty function consists of M functions, denoted by $g_m(\theta_1)$, which are based on the difference between the estimated and predicted deciles of the conditional distributions of y_1 and y_2. The penalized likelihood function can then be written as

$$\log L_p(\theta_1) = \log L(\theta_1) - \xi \sum_{m=1}^{M} g_m(\theta_1) \quad (28)$$

where ξ is a weighting parameter.

In principle, we can also estimate the model using a constrained maximum likelihood estimator. We prefer the penalized maximum likelihood estimator over a constrained maximum likelihood estimator since it is computationally challenging to find parameters that exactly satisfy the constraints. We explored a variety of different values for ξ and find that the main results are quite robust as long as ξ is sufficiently high to force the model to capture the nine deciles of the conditional distributions of the latent outcome y. We thus conclude that the first stage parameters of the model can be estimated using a penalized maximum likelihood estimator which treats the key policy outcome observed by the voters in our model as a latent variable.

5.2 Second Stage

Define the second-stage parameters as:

$$\theta_2 \;=\; (\{\beta_j\}_{j=1}^{n}, \{\kappa_k\}_{k=1}^{K}) \tag{29}$$

Given a first-stage estimator of θ_1, we then estimate the remaining structural parameters in the second stage by using the two equilibrium conditions. First, voters must be indifferent between reelecting the incumbent and electing the challenger if they observe outcome $y_1 = \bar{y}_k$ for each $k = 1, ..., K$. Recall that this indifference condition is given by:

$$V_k^I(\bar{y}_k) \equiv \sum_j \mu(j|\bar{y}_k)\left[E[u(y)|\hat{x}_j] + \kappa_k + \delta V^C\right] = V^C \quad k = 1, ..., K \tag{30}$$

The reelection shocks κ_k are, therefore, primarily identified from the cut-off levels $\bar{y}_k$. An increase in the value of κ_k makes retaining incumbents more attractive which tends to decrease the election threshold $\bar{y}_k$.

Second, recall that the first order condition of the effort problem is given by:

$$w_j'(x_j) + \delta\Big[\sum_{k=1}^{K} f(\bar{y}_k - x_j)g_k\Big]\left[w_j(\hat{x}_j) + \beta_j - (1-\delta)V^C\right] \;=\; 0 \quad j = 1, ..., n \tag{31}$$

Hence, the benefits of holding office β_j are primarily identified from the first-term effort levels x_j. As the benefits of holding office increase, effort also tends to increase.

The sequential estimator is consistent and asymptotically normally distributed under the standard regularity assumptions. When we implement the sequential estimator we use a bootstrap algorithm to compute standard errors to account for the sequential nature of the estimation strategy.

5.3 Mixed Strategies

We can extend the first- and the second-stage estimators to allow for mixed strategies. Define the first-stage parameters when all agents play mixed strategies as:

$$\theta_1 \;=\; ((\hat{x}_j, p_j, \pi_j^1, x_j)_{j=1}^n, (g_k, \bar{y}_k)_{k=1}^K, \sigma_\epsilon) \tag{32}$$

and note that π_j^1 and x_j are now I-dimensional vectors. In a mixed strategy equilibrium the joint density of performance measures $(z_{11}, ..., z_{L1}, z_{12}, ..., z_{L2})$ of a mayor who wins reelection can be written as:

$$\begin{aligned} f_{\theta_1}(z_{11}, ..., z_{L1}, z_{12}, ..., z_{L2}) \;&=\; \sum_{j=1}^n p_j \left\{ \sum_{i=1}^I \pi_{ij}^1 \; \phi_\eta (z_{11} - \lambda_1 x_{ij}, ..., z_{L1} - \lambda_L x_{ij}) \right. \\ &\times \; \left[\sum_{k=1}^K \left[1 - \Phi \left(\frac{\bar{y}_k - x_{ij} - \Sigma_{12}\Sigma_{22}^{-1}\eta_{ij}}{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}} \right) \right] g_k \right] \Bigg\} \\ &\times \; \phi_\eta (z_{12} - \lambda_1 \hat{x}_j, ..., z_{L2} - \lambda_L \hat{x}_j) \end{aligned} \tag{33}$$

Similarly, the density of the first-term performance of a losing mayor can be written as:

$$\begin{aligned} f_{\theta_1}(z_{11}, ..., z_{L1}) \;&=\; \sum_{j=1}^n p_j \left\{ \sum_{i=1}^I \pi_{ij}^1 \; \phi_\eta (z_{11} - \lambda_1 x_{ij}, ..., z_{L1} - \lambda_L x_{ij}) \right. \\ &\times \; \left[\sum_{k=1}^K \Phi \left(\frac{\bar{y}_k - x_{ij} - \Sigma_{12}\Sigma_{22}^{-1}\eta_{ij}}{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}} \right) \; g_k \right] \Bigg\} \end{aligned} \tag{34}$$

Let's assume for simplicity that $I = 2$ which is typically the case as discussed in Appendix A. Note that the mixed strategy likelihood function with n types is nested by the likelihood function of a pure strategy equilibrium with $2n$ types. Assume, for simplicity, that effort levels in the mixed strategy equilibrium are strictly monotonic in type, i.e. that the highest possible effort of type j is lower than the lowest level of type $j+1$. Then the mixed strategy likelihood function imposes the constraint on the likelihood with pure strategies that zero-effort policies are the same for consecutive types, i.e. that $\hat{x}_j = \hat{x}_{j+1}$

for $j = 1, 3, ..., n-1$. Hence, the mixed strategy likelihood function for n types is a restricted version of the pure strategy likelihood function with $2n$ types.[28]

Mixed strategies also impose two parameter restrictions for the second-stage parameters, namely that benefits of "consecutive pure" types are the same $\beta_j = \beta_{j+1}$ for $j = 1, 3, ..., n-1$, and that the "adjacent pure" types are indifferent between the effort levels:

$$\begin{aligned} 0 &= w_j'(x_j) + \delta \Big[\sum_{k=1}^{K} f(\bar{y}_k - x_j) g_k \Big] \left[w_j(\hat{x}_j) + \beta_j - (1-\delta) V^C \right] \\ &= w_j'(x_{j+1}) + \delta \Big[\sum_{k=1}^{K} f(\bar{y}_k - x_{j+1}) g_k \Big] \left[w_j(\hat{x}_j) + \beta_j - (1-\delta) V^C \right] \end{aligned} \tag{35}$$

In our application, we can reject the hypothesis that players use mixed strategies in equilibrium as we discuss in more detail in the next section.

6 Empirical Results

We have estimated the model using a sample of cities for which we do not have missing variables. Table 4 reports the parameter estimates and estimated standard errors for a three-type model with three election shocks. Columns I and II show the estimates of the penalized MLE using two different values of the penalty parameter. In Column I ξ equals 0.4, while in Column II ξ equals 1.0. Standard errors are computed using a sequential bootstrap algorithm.

Comparing Columns I and II in Table 4 shows that the estimates are remarkably similar across these two specifications of the estimator. We focus in the discussion below on our preferred model in Column I. Table 4 suggests that all types of politicians exert

[28]More generally, we need to perform all pairwise tests to determine whether there exits an i and a j such that $\hat{x}_j = \hat{x}_i$.

effort in the first term to get reelected. The differences between first-term policies x_j and second-term policies $\hat{x}_j$ are significant and large. As consequence, we find much evidence of moral hazard. Figure 6 illustrates the optimal effort choices of all types of politicians for the model estimated in Column I.

Our findings imply that all types play pure strategies in equilibrium in all specifications that we estimated. Figure 6 shows the constraint set, which contains the feasible combination of reelection probabilities and effort levels. It is fairly flat relative to the indifference curves. As a consequence, there is very little scope for mixed-strategy equilibria in the neighborhood of our preferred model estimates. We can formally test whether there exists a mixed strategy equilibrium that is consistent with our first-stage estimates. Here we can use Wald tests to test the null hypothesis that $\hat{x}_j = \hat{x}_{j+1}$ for $j = 1$ or $j = 2$ which is a necessary condition for the existence of a mixed strategy equilibrium. Given the large differences in the estimates and the small estimated standard errors, it is not surprising that we reject both null hypotheses at any reasonable level of confidence.

The parameter σ_ϵ is the standard deviation of the monitoring technology. The parameter estimate is approximately 0.97. This result suggests that the monitoring technology is not very precise which is consistent with the hypothesis that most citizens do not pay much attention to local politics and that the media coverage of local politics may not be comprehensive in many cities.

There are three election shocks in our preferred model. The probability distribution of the reelection shocks is given by g_k. These shocks translate into different cut-off thresholds denoted by $\bar{y}_k$ in Table 4. The estimated model must be consistent with the observation that some high-performing mayors are not reelected while some low-performing mayors are retained. We find that our model needs to have, at least, three shocks to fit this feature of the data.

Table 5 reports the parameter estimates and estimated standard errors of the second stage estimator. Recall that we find that the higher-quality types exert more effort along the equilibrium paths than low-quality types. This finding then implies that high-quality politicians have much higher benefits of holding office than low-quality types. This makes sense since a large fraction of these benefits may arise due to career concerns. These estimates may also reflect that high-quality types are more likely to be elected to higher office or obtain administrative positions at the state or federal level than low- and medium-quality types.

The heterogeneity in the cut-off threshold implies that there is much heterogeneity in the points in the support of the election shocks. This finding is consistent with the observation that some local elections may be determined by factors that are not directly under the control of the mayor. For example, the election shocks may capture the impact of differences in ideology or valence that are not explicitly captured by our model.

Our model implies that there are large and economically important differences among mayoral candidates. Hence, there is much scope for adverse selection. In Table 6 we report the differences in policies between low-quality and high-quality types using the model estimated in Column I. We convert the differences predicted by our models into differences in observed outcomes using our measurement model. For comparison purposes, we also report the standard deviations of the key outcome measures.

Table 6 suggests that high-quality types yield much better outcomes than low-quality types in both terms. The difference in first-term outcomes is approximately 1.01 percentage points of the employment rate, $220 additional expenditures per capita on education and welfare, and 79 fewer violent crimes per 100,000 individuals. The differences in second-term policies are slightly smaller and purely reflect differences in zero-effort policies. We, therefore, conclude that there is also much scope for adverse selection in mayoral elections in large U.S. cities.

Finally, we consider the goodness of fit of our model. Tables 7 and 8 report some statistics that measure the goodness of fit of our preferred model reported in Column I of Tables 4 and 5. First, consider the three observed outcome variables that enter into our likelihood function. Table 7 reports the estimated and predicted means and standard deviations of the three observed outcome variables conditional on term and reelection status. Overall, we find that our model captures the main features of the data reasonably well, despite the fact the distributions of the observed outcomes are generated by a single latent factor. To obtain a better fit of these moments, researchers probably need to consider models in which the policy space is multi-dimensional. As we discuss in the conclusions of this paper, it is rather difficult to characterize the equilibria that arise in these types of models.

Tables 8 reports the nine estimated and predicted deciles of the distributions of the latent factor conditional on term and reelection status. Note that we use these deciles in the penalty function. Overall, we find that our model fits these conditional distributions of the latent factor rather well. As we increase the value of the penalty parameter we improve the fit of the distribution of the latent factor and decrease the fit of the distributions of the observed measurements. Overall, we conclude that our model fits both the distributions of the observed measurements and the distribution of the latent factor well.

7 Policy Responsiveness

7.1 Measuring the Effort and Selection Effects

Our model predicts that policies respond to voters' preferences for two reasons. First, incumbents exert effort to win reelection to a second term. Consider for simplicity the case in which all types play a pure strategy in the first term. We can then measure the magnitude of this "effort effect" using the average difference between the optimal first-term policy and the zero-effort, second-term policies:

$$E = \sum_{j=1}^{n} p_j \left(x_j - \hat{x}_j \right) \tag{36}$$

Note that this difference in outcomes is also the expected difference between the first-term outcomes of a model with a two-term limit and the outcomes of a model with a one-term limit.

Second, voters use cut-off strategies to eliminate incumbents that produce undesirable outcomes. The ability of voters to perform this task is complicated by the fact that they only have access to an imperfect monitoring technology. Voters do not directly observe the quality of each incumbent. Instead, they update beliefs about the incumbent's type based on the observed first-term policy outcomes. In that sense, the model captures "retrospective" voting behavior. Voting is also "prospective" in this model since voters use the information of the first-term outcomes to predict the performance of the incumbent in a second term.

One of the key advantages of our modeling approach is that we can measure how successful voters are in eliminating low-quality types along the equilibrium path. To construct an appropriate measure of this "selection effect," note that the probability of type j in the second term can be written as:

$$s_j = \frac{p_j \, r_j}{\sum_{k=1}^{n} p_k \, r_k} \tag{37}$$

where r_j is the survival probability in equation (6). A natural measure of the selection effect is the differences between the expected zero-effort policy adopted by reelected incumbents in the second term and the expected zero-effort policy of politicians in the underlying population:

$$S \;=\; \sum_{j=1}^{n} (s_j - p_j)\; \hat{x}_j \tag{38}$$

The average quality of reelected incumbents is higher than the average quality of politicians in the population. Note that this term is also the expected difference between the second-term outcomes of a model with a two-term limit and the outcomes of a model with a one-term limit.

Table 9 summarizes our estimates of the effort and selection effects. Recall that our policy outcome measure y is normalized such that the mean is zero and the standard deviation is one. Hence, the effort effect is approximately 0.46 of a standard deviation. The selection effect is smaller and approximately 0.17 of standard deviation. We can also translate these effects into observed outcomes using our measurement error model. We find that the effort effect increases the employment rate by 0.23 percentage points, increases spending by $50, and lowers the violent crimes rate by 18 crimes per 100,000 individuals. The selection effect is approximately a third of the effort effect. The sum of the two effects is approximately a quarter of the difference between high- and low-quality politicians shown in Table 6.

7.2 Improving Policy Responsiveness

Here, we discuss institutional reforms that promise to improve the degree of policy responsiveness. Our model allows us to study the efficacy of several reform options that are commonly discussed in the literature. First, voters can improve the accountability of local politicians by investing in a better monitoring technology. The local media plays a curial role in informing voters about local politicians and providing a critical assessment of their performances. Similarly, citizens can spend more time and energy attending political meetings or spreading information through informal networks. To assess the impact of improving the monitoring technology, we plot the effort and the selection effect as a function of σ_ϵ. Figure 7 summarizes our main findings. Overall, we find that improving the monitoring technology (i.e. lowering the value of σ_ϵ) appears to be a promising option for increasing policy responsiveness. It allows voters to screen out a higher fraction of low-quality types and induces politicians to exert more effort along the equilibrium path. Hence, this reform affects both margins of policy responsiveness.

Alternatively, voters can increase the benefits of holding office. For example, they can increase the pay of local politicians or improve other fringe benefits associated with the job. Mayors tend to be underpaid compared to private-sector managers, especially once one accounts for the fact that most mayors tend to work seven days a week. When the office is more attractive, politicians have stronger incentives to exert effort. We implement this policy by considering a proportional increase in the benefits of holding office for all three types. We consider increases between 0 and 20 percent. Figure 8 illustrates the policy responsiveness by plotting the effort and the selection effect as a function of the policy parameter. We find that increasing the benefits of holding office has only a small effect on selection. However, increasing the benefits of holding office significantly improves effort. For example, a ten percent increase in the benefits for all types increases effort by almost ten percent. This policy may also have the additional benefit of increasing the overall quality pool of the candidates. We address this option directly in our next counterfactual.

Finally, voters can try to attract better candidates and improve the overall quality of the talent pool. Our findings suggest that the candidate pool has a large fraction of low- and medium-quality politicians. Anecdotal evidence suggests that mayors often lack qualifications that seem to be desirable for the job. This finding is consistent with the notion proposed by Diermeier, Keane, and Merlo (2005) that politics should be viewed as an occupational choice. As a consequence, the quality of individuals that choose politics as an occupation is largely a function of the available outside options in the private or public sectors of the economy. Local politicians typically have to survive an exhausting elimination process to become a viable candidate for mayor of a large city. As a consequence, it may not be surprising that many high-quality mayoral candidates do not pursue a career in local politics.

One nice feature of our model is that we can analyze how an increase in the quality of the candidate pool affects voters' welfare and overall policy responsiveness. To illustrate some of the basic mechanisms that are at work, we vary the fraction of high-quality types (p_3) around the estimated value adjusting the values of the p_1 and p_2 proportionally. Figure 9 summarizes our main findings.

Overall, we find that increasing the share of high-quality types has a small positive impact on selection. Since the distribution of types is more dispersed, this reform has a similar effect as an investment into a better monitoring technology. However, an increase in p_3 has a non-linear impact on effort choices. It induces the medium-quality type to exert more effort, while the high-quality type reduces its effort. Therefore, the net effect can be non-monotonic. We thus conclude that an increase in the share of high-quality politicians improves selection, but has an ambiguous effect on effort.

In summary, one key advantage of our model is that we can evaluate the impact of important institutional reforms that are aimed to increase the degree of policy responsiveness within an internally consistent model. Our model has not much to say about the costs associated with these reforms, and hence we cannot conduct a full welfare analysis associated with these reforms. Nevertheless, we conclude that our model provides some new quantitative insights into the effectiveness of several important institutional reforms that are discussed among policymakers and engaged citizens.

8 Conclusions

We have developed a new method for estimating dynamic electoral games with adverse selection, moral hazard, and imperfect monitoring. We have shown that these games can be used to measure the degree of policy responsiveness in competitive local elections with a potentially large number of politician types that endogenously pick effort levels. One key challenge posed by these types of models stems from the fact that there are well-known non-convexities that arise in the politician's effort decision problem when there is imperfect monitoring. As a consequence, there is scope for the existence of mixed-strategy equilibria. We have shown how to account for the potential existence of mixed-strategy equilibria in estimation. We have estimated the dynamic game using panel data from U.S. mayoral elections. We find that a parsimonious model with three types of politicians and three electoral shocks fits the observed outcomes well.

Our empirical results provide new insights into the nature of political competition in large U.S. cities. High-quality politicians are much more effective in reducing crime, increasing spending on education and welfare, and raising the level of employment than medium- and low-quality types. As a consequence, there are potentially large economic benefits from electing and retaining high-quality candidates. Nevertheless, there is a significant fraction of low- and medium-quality politicians that are reelected in equilibrium to a second term. Our analysis suggests that the monitoring technology used by voters is noisy which makes it difficult for voters to screen out low- and medium-quality incumbents. Hence, there is a limited degree of electoral accountability. We have quantified the effectiveness of a variety of institutional reforms that promise to increase policy responsiveness. Investing in a better monitoring technology is a promising policy change since it increases both effort and selection along the equilibrium path. Increasing the benefits of holding office primarily improves the overall effort of incumbents while increasing the quality of the candidate pool allows voters to retain a larger fraction of high-quality types.

Our paper provides ample scope for future research. In our application, there are only a small number of policy measures that indicate strategic behavior by incumbents in response to term limits. Hence, a model with a one-dimensional policy space explains the key regularities observed in the data well. Nevertheless, future research should focus on finding compelling models and estimation strategies that can handle strategic behavior in multi-dimensional spaces. Of course, dynamic voting games in multi-dimensional spaces are notoriously difficult to handle as discussed, for example, by Banks and Duggan (2008). It is fair to say that some theoretical progress will be needed before researchers can take these models to the data.

Alternatively, research such as Daley and Snowberg (2009) have maintained a one-dimensional policy space and focused on multi-tasking models that are in the spirit of Holmström and Milgrom (1991).[29] In these models politicians may have explicit incentives to engage in activities such as fund-raising or interacting with lobbyists instead of devoting themselves to improving outcomes that voters care about. But again, some additional theoretical development is needed to turn these models into fully dynamic voting models with imperfect monitoring before these models can be estimated.

It would also desirable to include heterogeneity among voters either in preferences for ideology or valence along the lines discussed, for example, in Bernhardt, Camara, and Squintani (2011) and Sieg and Yoon (2017). These studies consider dynamic spatial electoral games assuming perfect monitoring. Integrating a spatial dimension into the

[29]Another interesting branch of the electoral accountability literature examines the possibility of political inefficiencies due to pandering. For example, Maskin and Tirole (2004) develop a model in which politicians have private information about the state of the world and may take action contrary to their information to secure re-election.

rent-seeking game studied in this paper is promising, especially if the object is to study state or federal politics.

Finally, our methods can be applied to study other applications outside of political economy. We have studied the interaction between a single long-lived principal and a series of short-lived agents in the presence of both moral hazard and adverse selection. These types of agency problems commonly arise in many other important relationships. Prominent examples are managers and owners of firms, financial advisors and investors, tenure decisions in academics, or partnership decisions in law firms. There are no general identification and estimation results for these types of dynamic games. However, we are confident that the results presented in this paper will provide useful for studying these types of environments as well.

A Mixed Strategies

It is well-known that the decision problem of the agent is not necessarily convex when the principal has only access to an imperfect monitoring technology. We illustrate this result in Figure A.1 which is based on specification of a model with two types and an election shock that can take on two values. The figure shows the indifference curves of both types of politicians at the optimum. Note that the constraint set is not convex. In this example, the low-quality type plays a mixed strategy while the high-quality type plays a pure strategy in equilibrium. Note that the figure suggests that the number of pure strategies that are used in a mixed strategy equilibrium, denoted by I, is, at most, two. Duggan and Martinelli (2017) provide sufficient conditions that imply that result for a model without election shocks. Multiple global constrained maxima appear to be non-generic properties of the model.

References

Adams, W., Einav, L., and Levin, J. (2009). Liquidity Constraints and Imperfect Information in Subprime Lending. *The American Economic Review*, *99 (1)*, 49–84.

Aguirregabiria, V. and Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica*, *75*, 1–54.

Alt, J., Bueno de Mesquita, E., and Rose, S. (2011). Disentangling accountability and competence in elections: evidence from US term limits. *The Journal of Politics*, *73*(01), 171–186.

Arnold, R. and Carnes, N. (2012). Holding Mayors Accountable: New York's Executives from Koch to Bloomberg. *American Journal of Political Science*, *56*, 949–63.

Aruoba, B., Drazen, A., and Vlaicu, R. (2019). A Structural Model of Electoral Accountability. *International Economic Review*, *60 (2)*, 517–545.

Bajari, P., Benkard, L., and Levin, J. (2007). Estimating Dynamic Models of Imperfect Competition. *Econometrica*, *75*, 1331–71.

Banks, J. S. and Duggan, J. (2008). A Dynamic Model of Democratic Elections in Multidimensional Policy Spaces. *Quarterly Journal of Political Science*, *3*(3), 269–299.

Banks, J. S. and Sundaram, R. (1998). Optimal Retention in Agency Problems. *Journal of Economic Theory*, *82*, 293–323.

Barro, R. (1973). The Control of Politicians: An Economic Model. *Public Choice*, *14*, 19–42.

Bernhardt, D., Camara, O., and Squintani, F. (2011). Competence and Ideology. *The Review of Economic Studies*, *78*(2), 487–522.

Besley, T. and Case, A. (1995). Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits. *The Quarterly Journal of Economics*, *110*(3), 769–798.

Besley, T. and Coate, S. (1997). An Economic Model of Representative Democracy. *Quarterly Journal of Economics*, *112*, 85–114.

Carneiro, P., Hansen, K., and Heckman, J. (2003). Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College. *International Economic Review*, *44 (2)*(May), 361–422.

Chiappori, P.-A., Jullien, B., Salanie, B., and Salanie, F. (2006). Asymmetric Information in Insurance: General Testable Implications. *The Rand Journal of Economics*, *37 (4)*, 783–398.

Chiappori, P.-A. and Salanie, B. (2000). Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy*, *108 (1)*, 56–78.

Coate, S. and Conlin, M. (2004). A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence. *American Economic Review*, *94*, 1476–1504.

Coate, S., Conlin, M., and Moro, A. (2008). The Performance of Pivotal-Voter Models in Small-Scale Elections: Evidence from Texas Liquor Referenda. *Journal of Public Economics*, *92*, 582–96.

Compiani, G. and Kitamura, Y. (2016). Using Mixtures in Econometric Models: A Brief Review and Some New Results. *Econometrics Journal*, *19*, C95–127.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931.

Daley, B. and Snowberg, E. (2009). Even if it is not Bribery: The Case for Campaign Finance Reform. *Journal of Law, Economics, and Organization*, *27 (2)*, 324–349.

Degan, A. and Merlo, A. (2011). A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association*, *9*(2), 209–245.

Diermeier, D., Eraslan, H., and Merlo, A. (2003). A Structural Model of Government Formation. *Econometrica*, *71*, 27–70.

Diermeier, D., Keane, M., and Merlo, A. (2005). A Political Economy Model of Congressional Careers. *American Economic Review*, *95(1)*, 347–373.

Duggan, J. (2017). Term Limits and Bounds on Policy Responsiveness in Dynamic Elections. *Journal of Economic Theory*, *170*, 426–463.

Duggan, J. and Martinelli, C. (2017). The Political Economy of Dynamic Elections: A Survey and Some New Results. *Journal of Economic Literature*, *55*, 916–984.

Einav, L., Jenkins, M., and Levin, J. (2012). Contract Pricing in Consumer Credit Markets. *Econometrica*, *80 (4)*, 1387–1432.

Ferejohn, J. (1986). Incumbent Performance and Electoral Control. *Public Choice*, *50*, 5–26.

Ferreira, F. and Gyourko, J. (2009). Do Political Parties Matter? Evidence from U.S. Cities. *Quarterly Journal of Economics*, *123*, 399–422.

Garcia-Jimeno, C. (2016). The Political Economy of Moral Conflict: An Empirical Study of Learning and Law Enforcement under Prohibition. *Econometrica*, *84 (2)*, 511–70.

Gayle, G., Golan, L., and Miller, R. (2015). Promotion, Turnover, and Compensation in the Executive Labor Market. *Econometrica*, *83 (6)*, 2293–2369.

Gayle, G. and Miller, R. (2015). Identifying and Testing Models of Managerial Compensation. *Review of Economic Studies*, *82 (3)*, 1074–1118.

Holmström, B. and Milgrom, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, *7*, 24–52.

Hu, Y. and Shum, M. (2013). Identifying Dynamic Games with Serially-Correlated Unobservables. In *Advances in Econometrics (Volume 31)*. Emrald Publishing.

Iaryczower, M. and Shum, M. (2012). The Value of Information in the Court: Get it Right, Keep it Tight. *American Economic Review*, *102*, 202–37.

Kawai, K., Onishi, K., and Uetake, K. (2018). Signaling in Online Credit Markets. Working Paper.

Kawai, K. and Watanabe, Y. (2013). Inferring strategic voting. *The American Economic Review*, *103*(2), 624–662.

Knight, B. and Schiff, N. (2010). Momentum and Social Learning in Presidential Primaries. *Journal of Political Economy*, *118*(6), 1110–1150.

Krasnokutskaya, E. (2011). Identification and Estimation of Auction Models with Unobserved Heterogeneity. *The Review of Economic Studies*, *78*(1), 293–327.

Lim, C. (2013). Preferences and Incentives of Appointed and Elected Public Officials. *American Economic Review*, *103*(4), 1360–1397.

Maskin, E. and Tirole, J. (2004). The Politician and the Judge: Accountability in Government. *American Economic Review*, *94*(4), 1034–54.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*. Academic Press.

Merlo, A. (1997). Bargaining over Governments in a Stochastic Environment. *Journal of Political Economy*, *105*, 101–131.

Merlo, A. and Tang, X. (2012). Identification and Estimation of Stochastic Sequential Bargaining Games. *Econometrica*, *80*, 373–389.

Mulaik, S. A. (2009). *Foundations of factor analysis*. CRC Press.

Myatt, D. P. (2007). On the theory of strategic voting. *The Review of Economic Studies*, *74*(1), 255–281.

Osborne, M. and Slivinski, A. (1996). A Model of Political Competition with Citizen-Candidates. *Quarterly Journal of Economics*, *111*, 65–96.

Perrigne, I. and Vuong, Q. (2011). Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard. *Econometrica*, *79 (5)*, 1499–1539.

Pesendorfer, M. and Schmidt-Dengler, P. (2008). Asymptotic Least Squares Estimators for Dynamic Games. *Review of Economic Studies*, *75*, 901–28.

Rosen, S. (1979). *Wage-based Indexes of Urban Quality of Life*. Johns Hopkins University Press.

Schennach, S. (2016). Recent Advances in the Measurement Error Literature. *Annual Review of Economics*, *8(10*, 341–377.

Sieg, H. and Yoon, C. (2017). Estimating Dynamic Games of Electoral Competition to Evaluate Term Limits in U.S. Gubernatorial Elections. *American Economic Review*, *107 (7)*, 1824–57.

Thomson, G. H. (1951). *The Factorial Analysis of Human Ability*. University of London Press.

Xin, Y. (2019). Asymmetric Information, Reputation, and Welfare in Online Credit Markets. Working Paper.

Tables

Table 1: City Economic and Policy Variables : 1990–2016

	Sample Mean	Standard Deviation
Employment rate (%)	93.13	2.63
Spending on education and public welfare ($)	2030.79	1013.50
Violent Crime rate per 100,000 people	1093.21	615.09

Table 2: Formal Tests

Mayor Type	Employment Rate	Education Spending	Crime Rate
(1) First term, lost	-0.157	-0.067	0.383
(2) First term, win	0.149	0.123	-0.143
(3) Second term	-0.088	-0.103	-0.040
Difference in Means Test			
t-test (1) vs (2)	-2.767	-1.676	4.350
(one sided) p-value	0.003	0.047	0.000
t-test (2) vs (3)	2.820	2.605	-1.133
(one sided) p-value	0.003	0.005	0.129
Kolmogorov-Smirnov Tests			
first term, lost vs first term, win	0.192	0.146	-0.324
(one sided) p-value	0.002	0.033	0.000
first term, win vs second term	-0.178	-0.136	0.117
(one sided) p-value	0.000	0.008	0.043

Note: These tests use annual data of the three outcomes.

Table 3: Comparison of the Estimated Densities

	Density of estimated Factors	Density estimated via Kotlarski's Theorem
$E(y_1\|W=0)$	-0.256	-0.225
$E(y_1\|W=1)$	0.257	0.245
$E(y_2\|W=1)$	-0.131	-0.120
$Std(y_1\|W=0)$	1.151	1.286
$Std(y_1\|W=1)$	1.096	1.278
$Std(y_2\|W=1)$	0.990	1.204

Table 4: First Stage Parameter Estimates

		I		II	
		$\xi = 0.4$		$\xi = 1.0$	
2nd term effort	$\hat{x}_1$	-2.15	(0.171)	-1.81	(0.166)
	$\hat{x}_2$	-0.30	(0.047)	-0.29	(0.052)
	$\hat{x}_3$	1.00	(0.145)	0.87	(0.151)
type probability	p_1	0.14	(0.001)	0.18	(0.003)
	p_2	0.62	(0.018)	0.51	(0.019)
	p_3	0.24		0.31	
monitoring tech	σ_ϵ	0.97	(0.040)	0.98	(0.020)
1st term effort	x_1	-1.61	(0.259)	-1.16	(0.271)
	x_2	-0.06	(0.064)	-0.03	(0.084)
	x_3	1.98	(0.244)	1.65	(0.171)
probability	g_1	0.51	(0.036)	0.53	(0.022)
	g_2	0.11	(0.009)	0.11	(0.014)
	g_3	0.38		0.36	
voters' cut-off	$\bar{y}_1$	-2.40	(0.510)	-2.13	(0.450)
	$\bar{y}_2$	-0.45	(0.178)	-0.77	(0.218)
	$\bar{y}_3$	2.21	(0.123)	2.04	(0.173)

Note: standard errors are in parentheses.

Table 5: Second Stage Parameter Estimates

		I		II	
		$\xi = 0.4$		$\xi = 1.0$	
	β_1	8.61	(3.754)	10.04	(3.221)
benefits of holding office	β_2	10.50	(4.407)	9.66	(3.920)
	β_3	16.72	(3.128)	15.21	(3.210)
	κ_1	2.54	(0.426)	2.13	(0.389)
election	κ_2	1.33	(0.209)	1.52	(0.247)
shocks	κ_3	0.04	(0.147)	0.14	(0.213)

Note: standard errors are in parentheses.

Table 6: Difference between High- and Low-quality Politicians

	First Term	Second Term	Standard Deviation
Model outcome	2.04	1.30	1.00
Employment rate	1.01	0.65	0.83
Expenditures per capita	220	141	156
Crime rate	-79	-50	215

Table 7: Model Fit of Observed Measures

	Employment Rate		Expenditures		Crime Rate	
	data	model	data	model	data	model
$E(z_1 \vert W=0)$	-0.17	-0.20	-0.06	-0.23	0.39	0.06
$E(z_1 \vert W=1)$	0.17	0.35	0.19	0.40	-0.15	-0.10
$E(z_2 \vert W=1)$	-0.08	-0.05	-0.16	-0.05	-0.05	0.01
$Std(z_1 \vert W=0)$	1.03	1.21	1.09	1.18	0.93	1.02
$Std(z_1 \vert W=1)$	0.92	1.27	1.01	1.25	1.10	1.02
$Std(z_2 \vert W=1)$	1.03	1.21	0.90	1.17	0.86	1.01

Note: the data moments are based on four-year averages for each term.

Table 8: Fit of Distribution of the Latent Factor

	Lost		Reelected 1st		Reelected 2nd	
percentile	data	model	data	model	data	model
10	-2.24	-2.21	-1.40	-1.28	-1.84	-1.78
20	-1.50	-1.48	-0.76	-0.71	-1.16	-1.14
30	-1.01	-0.99	-0.31	-0.29	-0.71	-0.71
40	-0.61	-0.64	0.08	0.06	-0.35	-0.36
50	-0.25	-0.30	0.45	0.42	-0.02	-0.03
60	0.10	0.07	0.83	0.83	0.30	0.29
70	0.46	0.45	1.25	1.33	0.64	0.64
80	0.87	0.89	1.82	1.99	1.04	1.04
90	1.40	1.42	3.28	2.99	1.62	1.59

Table 9: Policy Responsiveness: Effort versus Selection

	Effort Effect	Selection Effect	Standard Deviation
Model outcome	0.46	0.17	1.00
Employment rate	0.23	0.08	0.83
Expenditures per capita	50	18	156
Crime rate	-18	-7	215

Figures

Figure 1 plot the estimated density of spending on education and welfare using annual data during our sample period.

Figure 2 plot the estimated density of the employment rate using annual data during our sample period.

Figure 3 plot the estimated density of the violent crime rate using annual data during our sample period.

Figure 4 plots the estimated conditional densities of the regression-scored common factor.

Figure 5 plots the estimated conditional densities of the common factor using Kotlarski's Theorem.

Figure 6 illustrates the optimal effort choices of all types of politicians.

Figure 7 plots the effort and the selection effect as a function of the precision of the monitoring technology.

Figure 8 plots the effort and the selection effect as a function of the office benefits.

Figure 9 plots the effort and the selection effect as a function of the share of high-quality politicians.

Figure A.1 illustrates a case in which there are two global optima for type 1.

Figure 1: Spending on Education and Welfare

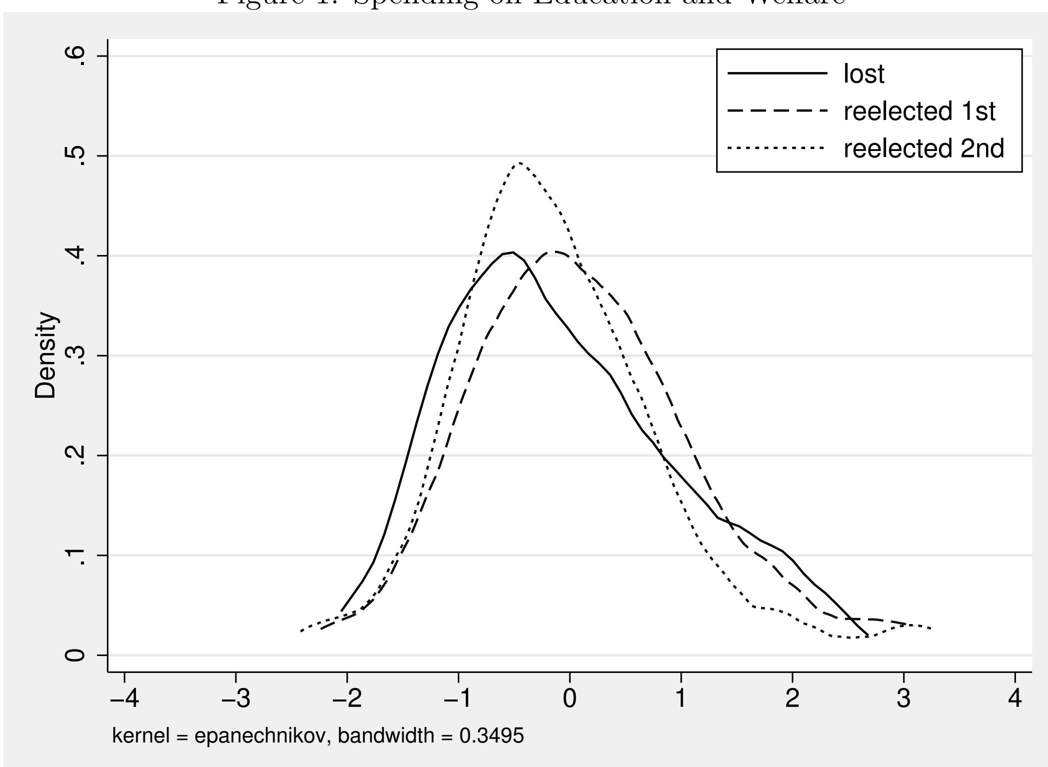

Figure 2: Employment Rate

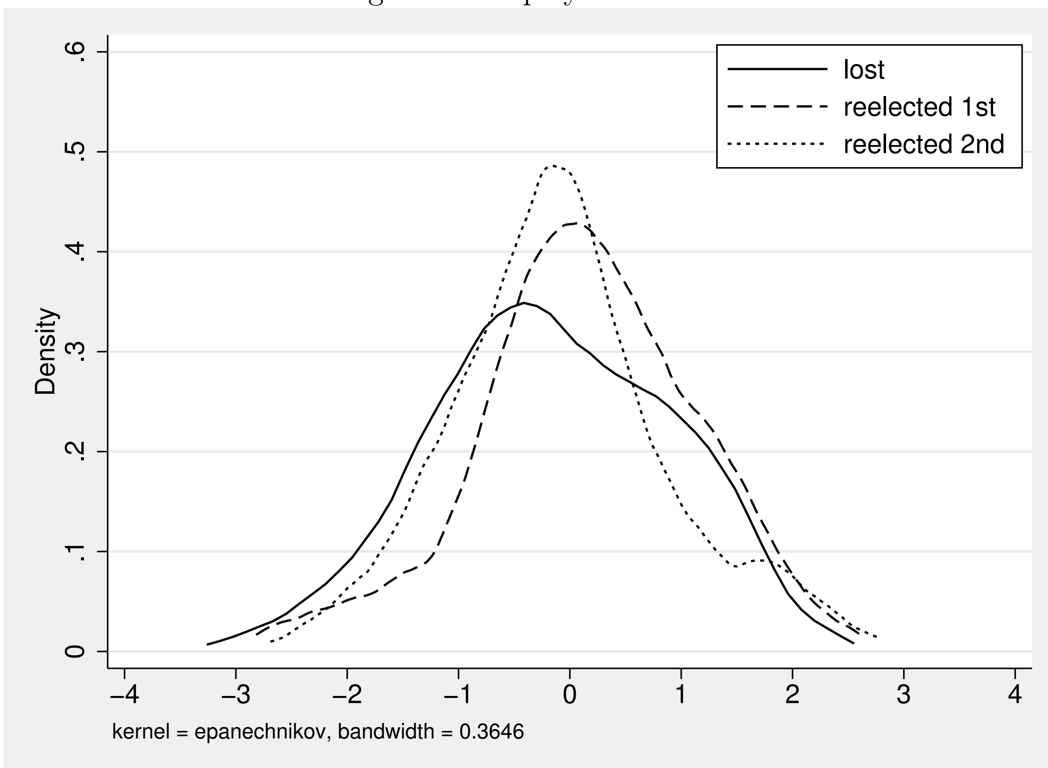

Figure 3: Violent Crime Rate

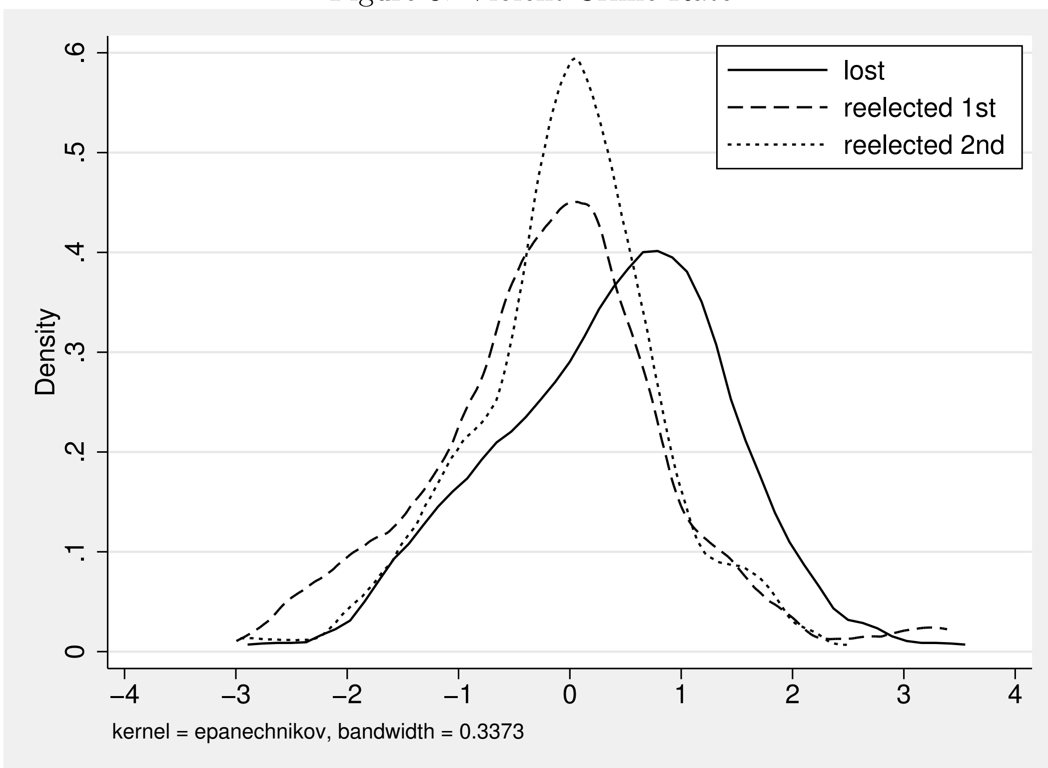

Figure 4: Kernel Estimates of the Density of Common Factor

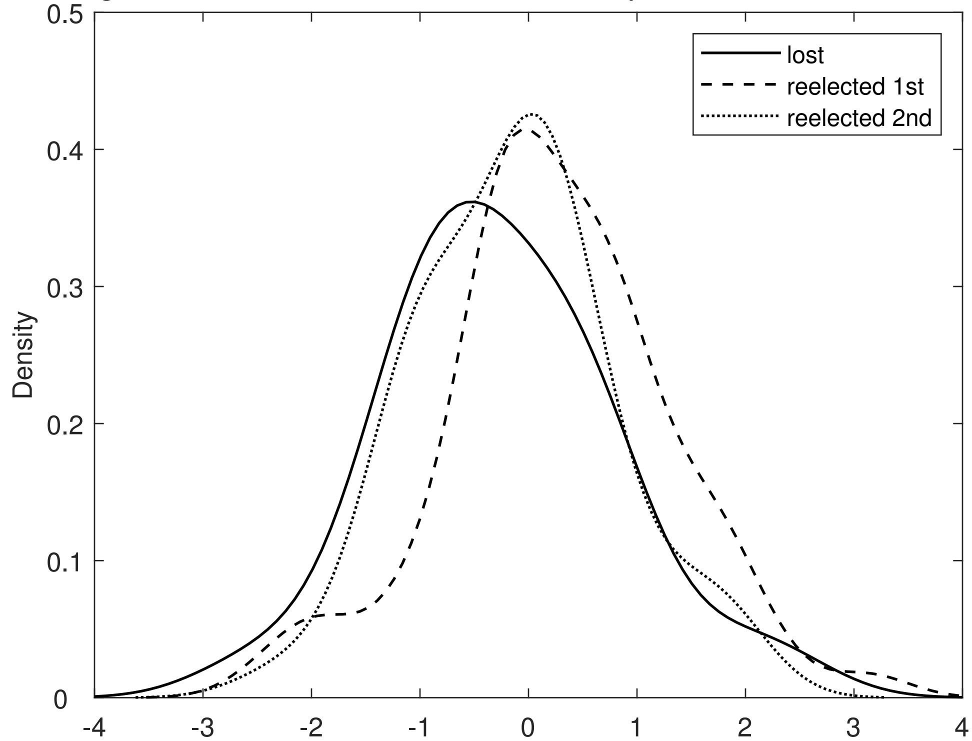

Figure 5: The Density of the Common Factor Estimated using Kotlarski's Theorem

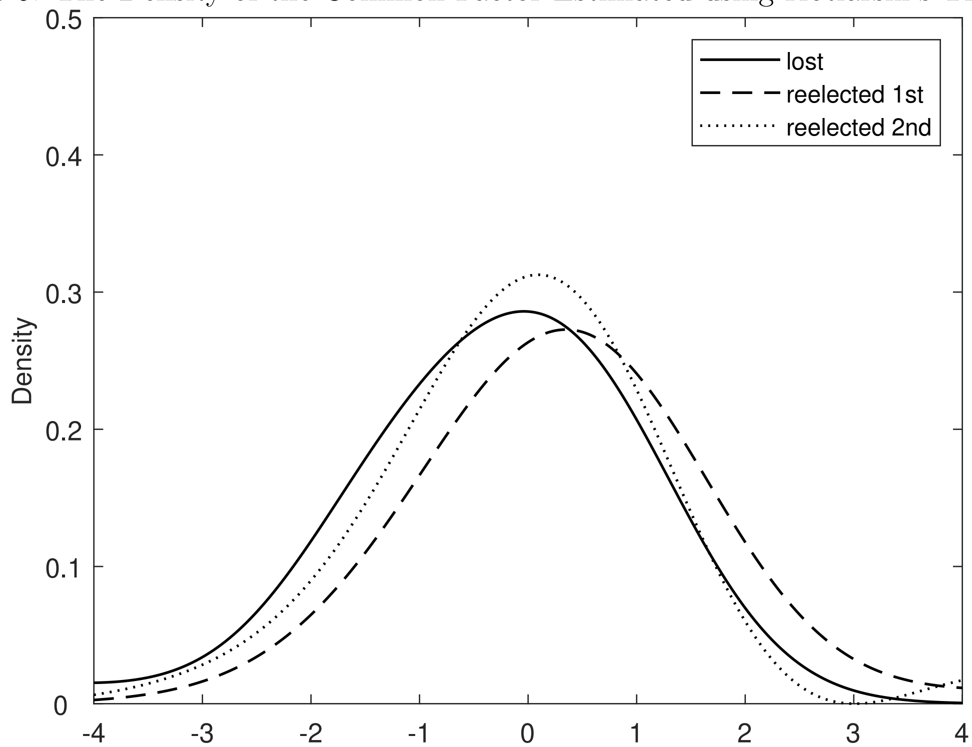

Figure 6: The Effort Choice Problem

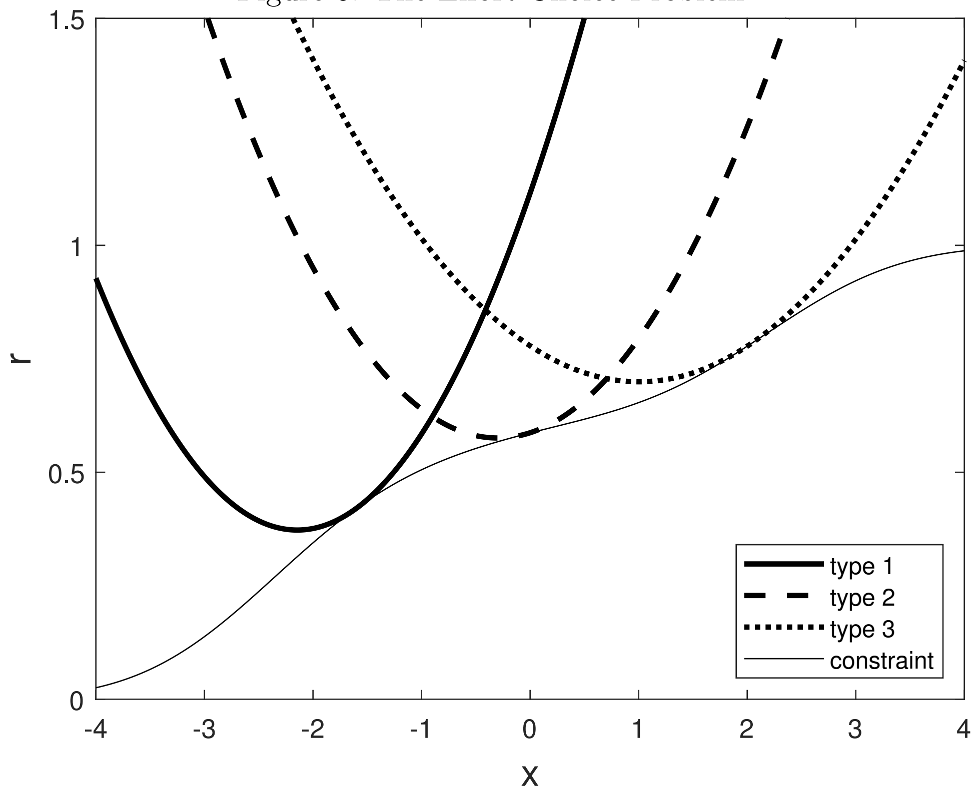

Figure 7: Improving the Monitoring Technology

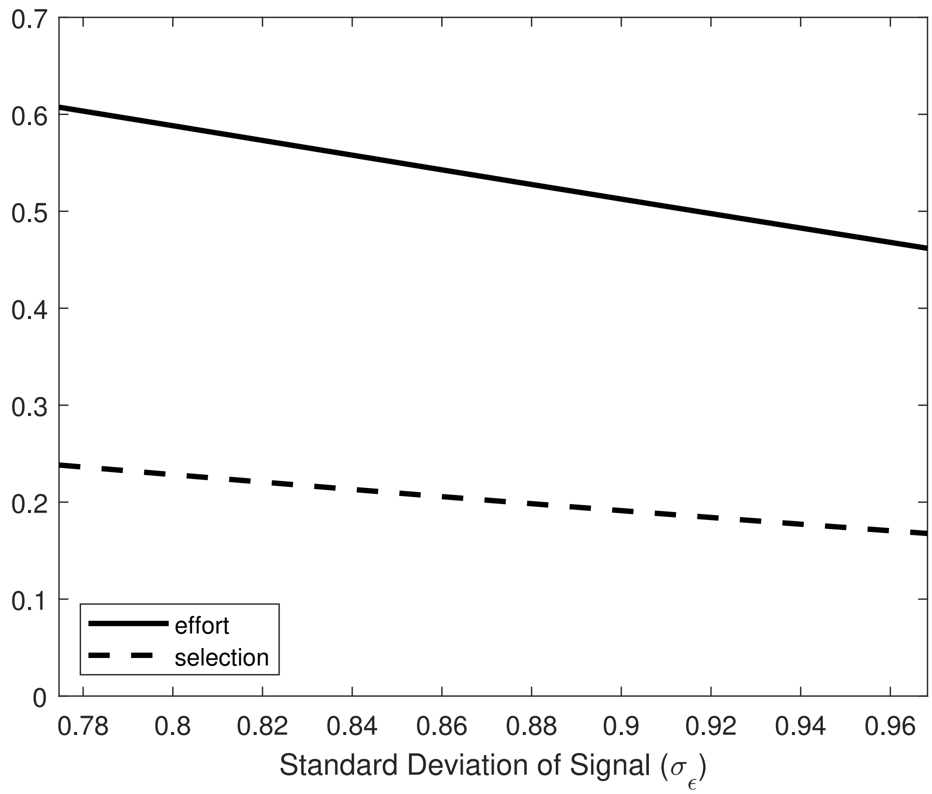

Figure 8: Increasing the Benefit of Holding Office

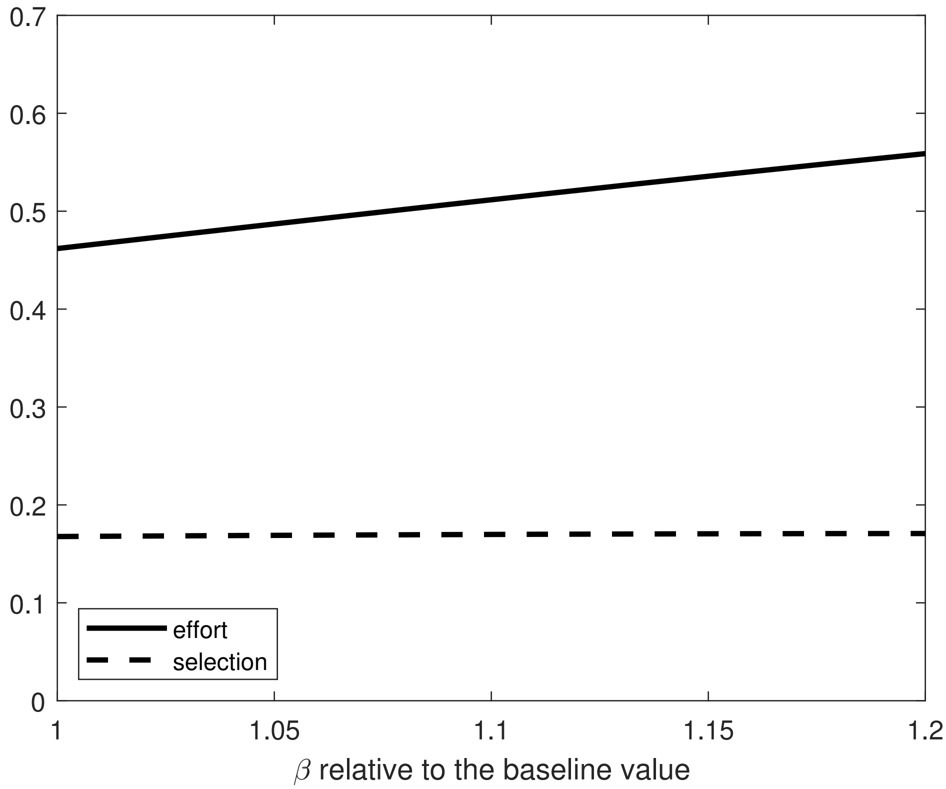

Figure 9: Varying the Share of High-Quality Politicians

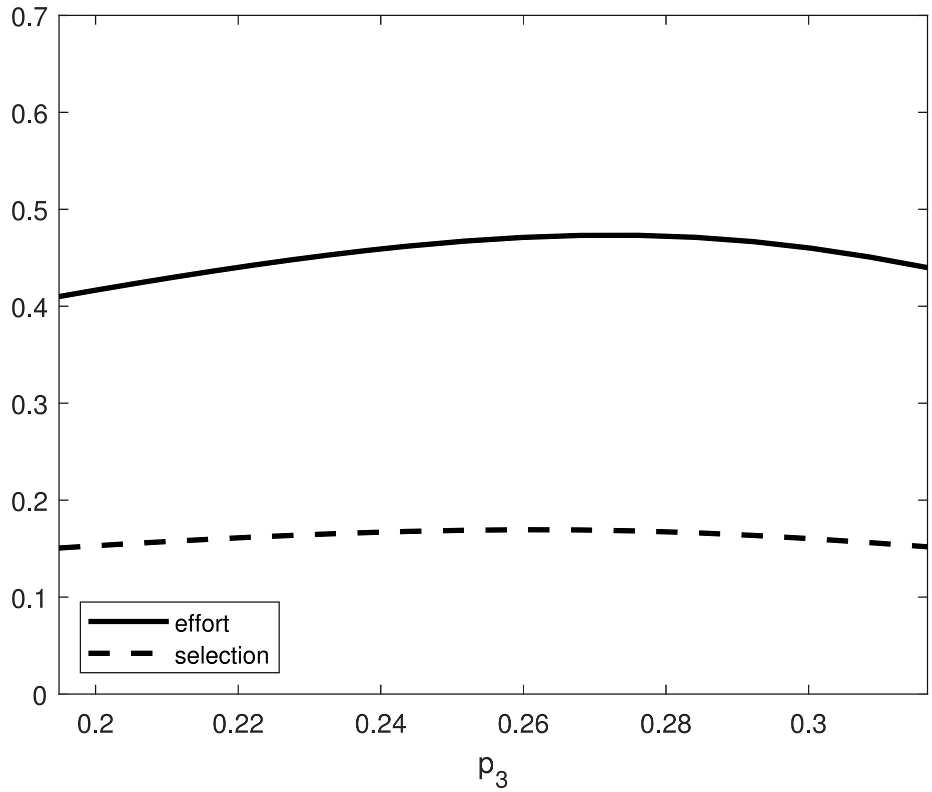

Figure A.1: Multiplicity of Global Optima