Electoral Accountability and Control in U.S. Cities

Holger Sieg
University of Pennsylvania and NBER

Chamna Yoon
Korea Advanced Institute of Science and Technology

November 8, 2021
Optimal Retention and Dynamic Agency

- We study 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.
- The principal can influence the agents’ behavior only through her choice of a retention rule which is required to be sequentially rational. No pre-commitment is allowed.
- We consider an environment in which the principal has only access to an imperfect monitoring technology.
- These types of problems commonly arise in many important relationships: tenure decisions in academics, partnership decisions in law firms, and reelection of politicians.
- We provide some new identification and estimation results for this class of dynamic games.
- We apply these methods to study political accountability of mayors of large U.S. cities.
Political Accountability

- A main concern for representative democracy is whether elections can serve as mechanisms of accountability.
- Can elections successfully align the incentives of politicians with voters’ preferences?
- Politicians are citizen candidates who cannot credibly commit themselves to policies prior to an election (Besley & Coate, 1997).
- Repeated elections mitigate the commitment problem of office holders 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 by choosing policies that are more desirable for voters.
A Dynamic Rent-Seeking Model

- We consider a rent-seeking model in the tradition of Barro (1973) and Ferejohn (1986).
- This model is more appropriate than a spatial model to study local electoral competition in cities since ideology is less important in city politics (Gyourko and Ferreira, 2009).
- In a rent-seeking environment, politicians have a short-run incentive to shirk from effort while in office, or equivalently to engage in rent-seeking activities that hurt other citizens.
- We consider a dynamic version of the rent-seeking game with imperfect monitoring that builds on Banks and Sundaram (1993, 1998) and Duggan (2017).
The Baseline Model

- We consider an infinite-horizon rent-seeking game with imperfect monitoring and a binding two-term limit.
- There is a continuum of citizen candidates that is partitioned into a finite set of types $j \in \{1, \ldots, n\}$ with $n \geq 2$.
- The probability of each type $j$ is given by $p_j$.
- The elected politician chooses a policy $x_t \in X$.
- In period $t + 1$, the incumbent faces a randomly drawn challenger with each type having probability $p_j$. 
Costly Effort

- Effort is costly, which gives rise to a moral hazard problem.
- Each politician type is characterized by a zero-effort policy denoted by $\hat{x}_j$.
- Order types such that $\hat{x}_1 < ... < \hat{x}_n$.
- Note that all types will play $\hat{x}_j$ is the second term because of the binding term limit.
- First-term incumbents may extra effort in an attempt to increase their chances of reelection.
Imperfect Monitoring

- The policy choice, $x$, generates a noisy outcome, $y = x + \epsilon$ which is observed by the voters.
- In contrast, the politicians’ types and their policy choices are not directly observable by the voters.
- The distribution function of $\epsilon = y - x$ is denoted by $F(\cdot)$ with continuous density $f(\cdot)$.
- Voters observe first term-policy outcomes and update beliefs.
- A belief system for voters is a probability distribution, denoted by $\mu(j|y)$, as a function of the observed signal.
Flow Pay-offs

- Given a policy choice $x$ and an outcome $y$ citizens obtain a pay-off given by $u(y)$, which increasing in $y$, if not in office.
- The politician’s pay-off in office is given by $w_j(x) + \beta_j$.
- $\beta_j$ measures the benefits for holding office of type $j$.
- $w_j(x)$ incorporates both the benefits from the policy choice and the costs from exerting effort for type $j$.
- Voters and politicians maximize inter-temporal expected utility with a common discount factor $\delta$. 
Cut-off Strategies

- Voters use cut-off rules as equilibrium strategies.
- Let $\bar{y}$ denote the cut-off point.
- Voters must be indifferent between reelecting the incumbent and electing the challenger if they observe outcome $\bar{y}$.
- Let $V^C$ be the continuation value of electing a challenger.
- The voter’s indifference condition is given by:

$$V^I(\bar{y}) \equiv \sum_{j} \mu(j|\bar{y}) \left[ E[u(y)|\hat{x}_j] + \delta V^C \right] = V^C$$

where $V^I(\bar{y})$ is the value function associated with an incumbent with observed policy outcome $\bar{y}$. 
Reelection Probabilities

▶ Each politician type knows that she is re-elected to a second term if and only if

\[ y = x + \epsilon \geq \bar{y} \]

or

\[ \epsilon \geq \bar{y} - x \]

▶ For any arbitrary policy choice \( x \), the probability of reelection is, therefore, given by \( 1 - F(\bar{y} - x) \).

▶ With probability \( F(\bar{y} - x) \) the challenger wins the election.
The Politician’s Decisions Problem

We can express the politician’s decision problem as a constrained optimization problem:

$$\max_{(x, r)} U_j(x, r)$$
$$s.t. \ g(x, r) \leq 0$$

where \( r \) is the reelection probability. Note that \( U_j(x, r) \) and \( g(x, r) \) are defined as:

$$U_j(x, r) = w_j(x) + \delta \left\{ r \left[ w_j(\hat{x}_j) + \beta + \delta V^C \right] + (1 - r) V^C \right\}$$
$$g(x, r) = r - (1 - F(\bar{y} - x))$$
Optimality Conditions

- The FOC of this problem is:

\[ w_j'(x) = -\delta f(\bar{y} - x) [w_j(\hat{x}_j) + \beta - (1 - \delta)V^C] \]

- The SOC of this problem is given by:

\[ w_j''(x) - \delta f'(\bar{y} - x) [w_j(\hat{x}_j) + \beta - (1 - \delta)V^C] \leq 0 \]

- It is well-known that the decision problem of the politician is not necessarily convex, which can give rise to mixed strategy equilibria.

- Generically, at most two strategies will have positive mass in a mixed strategy equilibrium.
For simplicity the notation, let’s focus on pure strategy equilibria. Everything can be extended to account for mixed strategy equilibria.

Updating of voter beliefs follows Bayes’ Rule. Conditional on observing outcome $y$, the posterior probability that the politician is type $j$ is:

$$
\mu(j|y) = \frac{p_j f(y - x_j)}{\sum_k p_k f(y - x_k)}
$$

Duggan (2017) then proves existence of equilibrium in mixed strategies.
Extension: Reelection or Incumbency Shocks

- Voter utility is given by: \( u(y) + d_I \kappa \).
- The parameter \( \kappa \) captures voters’ preferences for incumbents. \( \kappa \) is a discrete random variable, i.e. the distribution is given by \((\kappa_k, g_k), k=1,\ldots,K\).
- Timing: The reelection is realized after incumbents make effort choices, but before voters decide to retain the incumbent.
- The model then has cut-off values \( \bar{y}_k, k=1, \ldots,K \), which correspond to the \( K \) possible realizations of the reelection shock.
The Decision Problem in the Extended Model
A Parametrization

- Utility for each type is quadratic: $w_j(x) = -(x - \hat{x}_j)^2$.
- Voters' utility is linear: $u(y) = y$.
- $f(y - x)$ is normal with mean zero and constant variance $\sigma^2$. 
- Structural parameters of the model are the following:
  - the discount factor: $\delta$
  - the variance of monitoring technology: $\sigma^2$
  - the benefits of holding office: $\{\beta_j\}_{j=1}^n$
  - the parameters of the type distribution: $\{\hat{x}_j, p_j\}_{j=1}^n$
  - the parameters of the incumbency shock: $\{\kappa_k, g_k\}_{k=1}^K$.
- Nuisance Parameters:
  - the first-term effort levels: $\{x_j\}_{j=1}^n$
  - the cut-off values: $\{\bar{y}_k\}_{k=1}^K$. 
Our initial sample consists of all mayoral elections in the U.S. between 1990 and 2017.

We restrict attention to the 100 largest cities in U.S.

We also impose the sample restriction that the city had a binding two-term limit.

With these sample restrictions, our final sample consists of 135 mayors that served, at least, one term in office.

We find that 79 of the 111 mayors were reelected to the 2nd term (72 percent).

The remaining 32 mayors were not reelected.
There is not a single obvious performance measure for mayors.

We use the following three outcome measures:

- Employment rate.
- Expenditures per capita on education and welfare.
- Violent crime rate.

Differences in job performance are driven in our model by, at least, four factors: skill (type), effort, selection, and luck.

The objective of the structural analysis is to determine the relative importance of each factor.
Accounting for Heterogeneity Among Cities and Time Series Effects

- To account for heterogeneity among cities and time series effects, we use the procedure proposed by Besley and Case (1995).
- First, we regress our outcome measures on time dummies and population using a balanced panel.
- Second, we regress the residuals from the first regression on city dummies for the time periods when the two term limit was adopted.
- The residuals from the second regression are used as outcome measures for the structural estimation.
Employment Rate

Kernel density estimate

- lost
- reelected 1st
- reelected 2nd

kernel = epanechnikov, bandwidth = 0.3646
Kernel density estimate

Kernel = epanechnikov, bandwidth = 0.3495
Violent Crime Rate

Kernel density estimate

-4 -3 -2 -1 0 1 2 3 4
violent_rate

Density

Kernel = epanechnikov, bandwidth = 0.3373
### Formal Tests

<table>
<thead>
<tr>
<th>Mayor Type</th>
<th>Employment Rate</th>
<th>Spending Rate</th>
<th>Crime Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) First term, lost</td>
<td>-0.157</td>
<td>-0.067</td>
<td>0.383</td>
</tr>
<tr>
<td>(2) First term, win</td>
<td>0.149</td>
<td>0.123</td>
<td>-0.143</td>
</tr>
<tr>
<td>(3) Second term</td>
<td>-0.088</td>
<td>-0.103</td>
<td>-0.040</td>
</tr>
</tbody>
</table>

#### Difference in Means Test

<table>
<thead>
<tr>
<th>Test</th>
<th>t-statistic</th>
<th>One sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test (1) vs (2)</td>
<td>-2.767</td>
<td>0.003</td>
</tr>
<tr>
<td>(one sided) p-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test (2) vs (3)</td>
<td>2.820</td>
<td>0.003</td>
</tr>
<tr>
<td>(one sided) p-value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Kolmogorov-Smirnov Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>One sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>first term, lost vs first term, win</td>
<td>0.192</td>
<td>0.002</td>
</tr>
<tr>
<td>(one sided) p-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>first term, win vs second term</td>
<td>-0.178</td>
<td>0.000</td>
</tr>
<tr>
<td>(one sided) p-value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We observe a number of variables that can potentially be used as noisy performance measurements of our latent outcome variable.

Following Carneiro, Hansen and Heckman (2003), we assume that the $k$th measurement of outcome $y_t$, denoted by $z^k_t$, in term $t$ satisfies:

$$z^k_t = \mu^k y_t + u^k_t = \mu^k (x_t + \epsilon_t) + u^k_t$$

The parameters of the measurement model and the distribution of $y$ are identified regardless of the structural model.
The estimates of the factor loadings have the expected signs.

We find that only the largest eigenvalue of the outcome matrix is larger than one.

We thus conclude that a measurement model with one latent factor is sufficient.

Other outcome measures (with the exception of housing prices) do not seem to be “strategic”, i.e. do not differ by term status.
Estimated Density of the Latent Factor: Imputation
### Formal Tests for Imputed Latent Factor

<table>
<thead>
<tr>
<th></th>
<th>Sample Averages</th>
<th>Komogorov-Smirnow Test</th>
<th>Difference-in-Mean Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>First term, lost</td>
<td>-0.256</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First term, win</td>
<td>0.257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second Term</td>
<td>-0.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>first term, lost vs first term, win</td>
<td>0.358</td>
<td>(one sided) p-value 0.005</td>
<td></td>
</tr>
<tr>
<td>first term, win vs second term</td>
<td>-0.222</td>
<td>(one sided) p-value 0.045</td>
<td></td>
</tr>
<tr>
<td>first term, lost vs first term, win</td>
<td>-2.264</td>
<td>(one sided) p-value 0.013</td>
<td></td>
</tr>
<tr>
<td>first term, win vs second term</td>
<td>2.238</td>
<td>(one sided) p-value 0.014</td>
<td></td>
</tr>
</tbody>
</table>
Kotlarski’s Theorem

- Rewrite the measurement system for two outcomes as:

\[ z_1 = y + \epsilon_1 \]
\[ z_2/\mu_2 = y + \epsilon_2/\mu_2 \]

- Kotlarski’s Theorem implies that the characteristic function of \( y \) is given by:

\[ \phi_y(t) = \exp \left( \int_0^t \frac{\phi(0, u)}{\phi_1(0, u)} du \right) \]

where \( \phi(., .) \) is the characteristic function of \((z_1, z_2/\mu_2)\) with first partial derivative \( \phi_1(., .) \).

- The density of the latent factor \( y \) can then be estimated by

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

where \( T \) is a smoothing parameter and \( \hat{\phi}_y(t) \) is an estimator of the characteristic function. See Krasnokutskaya (2011) for details.
Estimated Density of the Latent Factor: Kotlarski
Comparison: Imputation vs. Kotlarski

- Conditional Moments:

<table>
<thead>
<tr>
<th></th>
<th>Imputation</th>
<th>Kotlarski</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(y_1</td>
<td>W = 0)$</td>
<td>-0.256</td>
</tr>
<tr>
<td>$E(y_1</td>
<td>W = 1)$</td>
<td>0.257</td>
</tr>
<tr>
<td>$E(y_2</td>
<td>W = 1)$</td>
<td>-0.131</td>
</tr>
<tr>
<td>$\text{Std}(y_1</td>
<td>W = 0)$</td>
<td>1.025</td>
</tr>
<tr>
<td>$\text{Std}(y_1</td>
<td>W = 1)$</td>
<td>1.037</td>
</tr>
<tr>
<td>$\text{Std}(y_2</td>
<td>W = 1)$</td>
<td>0.903</td>
</tr>
</tbody>
</table>

- The conditional means are almost the same while the conditional standard deviations are slightly larger when we use Kotlarski’s Theorem.
A Maximum Likelihood Estimator: Imputation

Let $W_i$ be an indicator that is equal to 1 if mayor $i$ is reelected and serves a second term and zero otherwise. The likelihood of observing $(y_1, y_2, W)$ can be written as:

$$L(y_1, y_2, W | (\hat{x}_j, p_j, x_j)_{j=1}^J, (g_k, \bar{y}_k)_{k=1}^K, \sigma_\epsilon) = f(y_1, y_2)^W f(y_1)^{1-W}$$

where

- $x_j$ are the endogenous first-term effort levels,
- $\hat{x}_j$ are the second-term effort levels,
- $\bar{y}_k$ is the cut-off points that are generated by the reelection shock $k$, and
- $\sigma_\epsilon$ is the standard deviation of monitoring technology of the voters.
A Maximum Likelihood Estimator (cont.)

The joint likelihood of a mayor winning reelection while producing performance of $y_1$ and $y_2$ can be written as:

$$f(y_1, y_2) = \sum_{j=1}^{J} p_j \frac{1}{\sigma_{\epsilon}} \phi \left( \frac{y_1 - x_j}{\sigma_{\epsilon}} \right) \left[ \sum_{k=1}^{K} g_k 1(y_1 \geq \bar{y}_k) \right] \frac{1}{\sigma_{\epsilon}} \phi \left( \frac{y_2 - \hat{x}_j}{\sigma_{\epsilon}} \right)$$

The likelihood of a mayor losing reelection with first term performance of $y_1$ can be written as:

$$f(y_1) = \sum_{j=1}^{J} p_j \frac{1}{\sigma_{\epsilon}} \phi \left( \frac{y_1 - x_j}{\sigma_{\epsilon}} \right) \left[ \sum_{k=1}^{K} g_k 1(y_1 < \bar{y}_k) \right]$$

Hence, we approximate the non-parametrically identified densities with mixtures of normals.
A Maximum Likelihood Estimator: Kotlarski

- We can express the likelihood in terms of the observed $z_1$ and $z_2$ instead of the unobserved $y_1$ and $y_2$.

- The likelihood function is then given by:

$$L(z_1, z_2, W|x_1, x_2, \hat{x}_1, \hat{x}_2, \bar{y}, \sigma_{\epsilon}) = \int f(y_1, y_2)^W f(y_1)^{1-W} g(y_1, y_2|z_1, z_2) \, dy_1 \, dy_2$$

where $g(\cdot)$ is identified from our measurement model.

- We also add a penalty function that uses the conditional moments of $y_1$ and $y_2$ reported above.
## Parameter Estimates: 1st Stage

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Imputation $J = 2$</th>
<th>Imputation $J = 3$</th>
<th>Kotlarski $J = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{x}_1$</td>
<td>-0.5507</td>
<td>-1.3861</td>
<td>-2.1994</td>
</tr>
<tr>
<td>$\hat{x}_2$</td>
<td>0.4213</td>
<td>-0.2497</td>
<td>-0.1875</td>
</tr>
<tr>
<td>$\hat{x}_3$</td>
<td>–</td>
<td>0.7402</td>
<td>1.0154</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.6197</td>
<td>0.1235</td>
<td>0.1200</td>
</tr>
<tr>
<td>$p_2$</td>
<td>0.3803</td>
<td>0.6896</td>
<td>0.6800</td>
</tr>
<tr>
<td>$p_3$</td>
<td>–</td>
<td>0.1869</td>
<td>0.2000</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.8020</td>
<td>0.6986</td>
<td>0.8342</td>
</tr>
<tr>
<td>$x_1$</td>
<td>-0.4143</td>
<td>-1.3871</td>
<td>-2.1789</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.9058</td>
<td>-0.0001</td>
<td>0.0648</td>
</tr>
<tr>
<td>$x_3$</td>
<td>–</td>
<td>1.4107</td>
<td>1.2432</td>
</tr>
<tr>
<td>$g_1$</td>
<td>0.4483</td>
<td>0.4483</td>
<td>0.3540</td>
</tr>
<tr>
<td>$g_2$</td>
<td>0.2964</td>
<td>0.2964</td>
<td>0.2344</td>
</tr>
<tr>
<td>$g_3$</td>
<td>0.1303</td>
<td>0.1303</td>
<td>0.1702</td>
</tr>
<tr>
<td>$g_4$</td>
<td>0.1250</td>
<td>0.1250</td>
<td>0.2414</td>
</tr>
<tr>
<td>$\bar{y}_1$</td>
<td>-2.3800</td>
<td>-2.3800</td>
<td>-4.3207</td>
</tr>
<tr>
<td>$\bar{y}_2$</td>
<td>-0.3408</td>
<td>-0.3408</td>
<td>-2.9182</td>
</tr>
<tr>
<td>$\bar{y}_3$</td>
<td>0.8859</td>
<td>0.8859</td>
<td>-0.104</td>
</tr>
<tr>
<td>$\bar{y}_4$</td>
<td>2.3313</td>
<td>2.3313</td>
<td>2.3528</td>
</tr>
</tbody>
</table>
A Minimum Distance Estimator

Given a first-stage ML estimator of \((\hat{x}_j, p_j, x_j)_{j=1}^{J}, (g_k, \bar{y}_k)_{k=1}^{K}\), we can estimate \((\beta_j)_{j=1}^{J}\) and \((\kappa_k)_{k=1}^{K}\) using a minimum distance estimator that exploits the following two equilibrium conditions:

- the \(K\) voter’s indifference conditions,

\[
V_k^l(\bar{y}_k) - V^C \equiv \sum_j \mu(j|\bar{y}_\kappa) \left[ E[u(y)|\hat{x}_j] + \kappa_k + \delta V^C \right] - V^C = 0
\]

- the FOCs of the decision problem of the \(n\) types of politicians.

\[
w_j^l(x_j) + \delta \left[ \sum_{k=1}^{K} f(\bar{y}_k - x_j)g_k \right] \left[ w_j(\hat{x}_j) + \beta_j - (1 - \delta) V^C \right] = 0
\]
### Parameter Estimate: 2nd Stage

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Imputation $J = 2$</th>
<th>Imputation $J = 3$</th>
<th>Kotlarski</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$ benefits of office</td>
<td>2.226</td>
<td>0.5396</td>
<td>1.5956</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>10.2323</td>
<td>3.9361</td>
<td>8.86332</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>−</td>
<td>18.5074</td>
<td>9.1615</td>
</tr>
<tr>
<td>$\kappa_1$ election</td>
<td>0.8303</td>
<td>1.8922</td>
<td>3.1968</td>
</tr>
<tr>
<td>$\kappa_2$ shock</td>
<td>0.6818</td>
<td>0.8632</td>
<td>3.1638</td>
</tr>
<tr>
<td>$\kappa_3$</td>
<td>0.1571</td>
<td>0.4977</td>
<td>1.1104</td>
</tr>
<tr>
<td>$\kappa_4$</td>
<td>-0.1176</td>
<td>-0.1517</td>
<td>0.1757</td>
</tr>
</tbody>
</table>
Some Comments on Identification and Estimation

- Instead of estimating the model sequentially, we can use a simultaneous nested fixed point estimator.
- Both the sequential and simultaneous estimator can be extended to allow for mixed strategy equilibria.
- Model selection is based on the Akaike Information Criterion:

\[
AIC = 2k - 2 \ln(\hat{L})
\]

where \(k\) is the number of 1st stage parameters and \(\hat{L}\) is the maximized likelihood value.

<table>
<thead>
<tr>
<th></th>
<th>(K = 2)</th>
<th>(K = 3)</th>
<th>(K = 4)</th>
<th>(K = 5)</th>
<th>(K = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(J = 2)</td>
<td>564.97</td>
<td>566.31</td>
<td>562.14</td>
<td>566.14</td>
<td>570.14</td>
</tr>
<tr>
<td>(J = 3)</td>
<td>562.18</td>
<td>563.52</td>
<td>559.36</td>
<td>563.36</td>
<td>567.36</td>
</tr>
<tr>
<td>(J = 4)</td>
<td>564.06</td>
<td>565.40</td>
<td>561.23</td>
<td>565.23</td>
<td>569.23</td>
</tr>
</tbody>
</table>

- AIC is minimized when \(J = 3\) and \(K = 4\).
Model Fit: First Term & Lost

![Graph showing data and model fit](image-url)
Model Fit: First Term & Reelected
Model Fit: Second Term

![Graph showing model fit with data and model curves.](image-url)
Consider our preferred model with three types. We compare outcomes for the best and the worst type to measure the potential gains from monitoring.

The next table reports the differences in outcomes by term:

<table>
<thead>
<tr>
<th></th>
<th>First Term</th>
<th>Second Term</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model outcome</td>
<td>1.41</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>Employment rate</td>
<td>0.70</td>
<td>0.49</td>
<td>0.83</td>
</tr>
<tr>
<td>Expenditures per capita</td>
<td>93</td>
<td>201</td>
<td>155</td>
</tr>
<tr>
<td>Crime rate</td>
<td>-90</td>
<td>-63</td>
<td>214</td>
</tr>
</tbody>
</table>

Hence, the potential benefits from monitoring are large, approximately 1 standard deviation of the performance measure.
Policy Responsiveness: Effort versus Selection

For each type, we can measure the impact of effort on policy as the difference between the policy adopted in the first period, and the bliss point. Averaging over types we obtain:

\[
E = \sum_{j=1}^{n} p_j (x_j - \hat{x}_j)
\]

A natural measure of the selection effect is then given by:

\[
S = \sum_{j=1}^{n} (s_j - p_j) \hat{x}_j
\]

This measures compares the average quality of politicians in the first term with the average quality of incumbents that are reelected and serve a second term.
Using our preferred model we obtain the following estimates of the effort and the selection effects:

<table>
<thead>
<tr>
<th>Model outcome</th>
<th>Effort</th>
<th>Selection</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment rate</td>
<td>0.15</td>
<td>0.05</td>
<td>0.83</td>
</tr>
<tr>
<td>Expenditures per capita</td>
<td>28</td>
<td>9</td>
<td>155</td>
</tr>
<tr>
<td>Crime rate</td>
<td>-19</td>
<td>-6</td>
<td>214</td>
</tr>
</tbody>
</table>

Note that improvements in the monitoring technology largely determine the relative magnitude of these effects.
Increasing the Benefit of Holding Office

\[ \beta \text{ relative to the baseline value} \]

- Effort: blue line
- Selection: red line
Varying the Share of High-Quality Politicians
Conclusions

- We have studied optimal retention of mayors when voters have only access to an imperfect monitoring technology using a sample of large U.S. cities with a two-term limit.
- Non-parametric tests indicate that our model is broadly consistent with the data.
- We have developed a new maximum likelihood estimator that allows for multiple types and mixed strategy equilibria.
- Our model estimates provide strong evidence that there are statistically significant and economically important differences among mayoral types. Moreover, the monitoring technology available to voters is noisy.
- Our analysis of policy responsiveness suggests that the effort effect is three times as large as the selection effect.
- Investing into a better monitoring technology or increasing the overall quality of the talent pool are likely to increase the degree of policy responsiveness in equilibrium.