Estimating a Dynamic Game with Moral Hazard and Imperfect Monitoring: Evidence from U.S. Mayors

Holger Sieg
University of Pennsylvania and NBER

Chamna Yoon
Korea Advanced Institute of Science and Technology

July 8, 2019
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: managers & owners or investors, tenure decisions in academics, partnership decisions in law firms, and politicians & voters.

There are no general identification and estimation results for this class of dynamic games.
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.

Repeated elections mitigate the commitment problems 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.
We consider a rent-seeking model in the tradition of Barro (1973) and Ferejohn (1986).

This model appears to be 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-Sundaram (1993, 1998).
Data

- Our 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.
Political Performance Measures

- There is not a single obvious performance measure for mayors.

- We use the following four “noisy” outcome measures:
  - Employment rate.
  - Housing price index.
  - 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 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.
## Political Performance Measures: Evidence

<table>
<thead>
<tr>
<th>Mayor Type</th>
<th>Employment Rate</th>
<th>Housing Price</th>
<th>Spending Rate</th>
<th>Crime Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) One term mayor</td>
<td>-0.097</td>
<td>-0.337</td>
<td>-0.161</td>
<td>0.507</td>
</tr>
<tr>
<td>(2) Reelected (1st)</td>
<td>0.196</td>
<td>0.228</td>
<td>0.218</td>
<td>-0.235</td>
</tr>
<tr>
<td>(3) Reelected (2nd)</td>
<td>-0.154</td>
<td>-0.086</td>
<td>-0.146</td>
<td>-0.027</td>
</tr>
<tr>
<td>t-test (1) vs (2)</td>
<td>-1.325</td>
<td>-2.262</td>
<td>-1.735</td>
<td>3.270</td>
</tr>
<tr>
<td>(one sided) p-value</td>
<td>0.094</td>
<td>0.013</td>
<td>0.043</td>
<td>0.001</td>
</tr>
<tr>
<td>t-test (2) vs (3)</td>
<td>2.191</td>
<td>1.552</td>
<td>2.225</td>
<td>-1.239</td>
</tr>
<tr>
<td>(one sided) p-value</td>
<td>0.015</td>
<td>0.062</td>
<td>0.014</td>
<td>0.109</td>
</tr>
</tbody>
</table>
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 has a bliss point 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 exert extra effort in an attempt to increase their chances of reelection.
Imperfect Monitoring

- The policy choice, $x$, generates a noisy outcome, $y = x + \epsilon$.
- Neither the politicians’ types nor the policy choices are directly observable by the voters, but the policy outcomes are.
- 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 $u(y)$ if not in office.
- The politician’s pay-off in office is given by $w_j(x) + \beta$.
- $\beta$ measures the benefits for holding office.
- $w_j(x)$ incorporates both the benefits from the policy choice and the costs from exerting effort for type $j$.
- Voters and politicians maximize intertemporal expected utility with a common discount factor, $\delta$. 
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 given by \( 1 - F(\bar{y} - x) \).
- With probability \( F(\bar{y} - x) \) the challenger will be elected.
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 $$U_j(x, r)$$ and $$g(x, r)$$ are defined as:

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

$$g(x, r) = r - (1 - F(\bar{y} - x))$$

It is well-known that the decision problem of the politician is not necessarily convex.
If the decision problem has multiple solutions, the politician will be indifferent among them.
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:
  \[ w_j''(x) - \delta f'(\bar{y} - x) [w_j(\hat{x}_j) + \beta - (1 - \delta)V^C] \leq 0 \]
Updating of Beliefs

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

$$\mu(j|y) = \frac{p_j \sum_x f(y - x) \pi_j(x)}{\sum_k p_k \sum_x f(y - x) \pi_k(x)}$$

- Banks and Sundaram (1998) then prove 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 that can take on two values:

\[
\kappa = \begin{cases} 
\kappa_l, & \text{with probability } \lambda \\
\kappa_h, & \text{with probability } 1 - \lambda 
\end{cases}
\]

- Timing: The reelection is realized after incumbents make effort choices, but before voters decide to retain the incumbent.
- The model then has two cut-off values \( \bar{y}_l \) and \( \bar{y}_h \) which correspond to the two possible realization of the reelection shock.
- The previous literature has typically assumed that \( \kappa_l = -\infty \), i.e. that the reelection shock randomly kills incumbents with probability \( \lambda \).
The Decision Problem in the Extended Model

The vertical lines are re-election cutoffs.
A Parametrization

- Utility for politicians are quadratic:
  \[ w_j(x) = -(x - \hat{x}_j)^2 \quad j = 1, \ldots, n \]

- Voter utility is linear \( u(y) = y \).

- \( f(y - x) \) is normal with mean zero and constant variance \( \sigma^2_\epsilon \).

- Structural parameters of the model are the following:
  - the discount factor: \( \delta \)
  - the benefits of holding office: \( \beta \)
  - the variance of monitoring technology: \( \sigma^2_\epsilon \)
  - the parameters of the type distribution: \( \{\hat{x}_j, p_j\}_{j=1}^n \)
  - the parameters of the incumbency shock: \( \lambda, \kappa_l, \kappa_h \)
Measurement Error in Outcomes

- We need to assume that we observe, at least, one noisy measure of the outcome variable $y$.
- If there is no measurement error, we obtain a degenerate likelihood function.
- We observe a number of variables that can potentially be used as noisy performance measurements.
- 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.
It is fairly straightforward to see that the magnitude of $\sigma_\epsilon$ determines the shape of the outcome distribution $y$.

If $\sigma_\epsilon$ is small, the distribution of $y$ is clearly multi-modal in both periods. As we increase the value of $\sigma_\epsilon$, the distribution becomes uni-modal and fairly flat.
Identification: Benefits of Holding Office

- In contrast, a change in the benefits of holding office, $\beta$, primarily affects the willingness of politicians to exert effort in the first period.
- Hence, it primarily affects the distribution of first-term outcomes by shifting this distribution to the right.
- Moreover, these incentives primarily affect lower effort types, since high effort types are likely to get reelected even if they just chose their bliss points.
The reelection shock is necessary to fit the distribution of first-term losers, i.e., we observe some "good" mayors that are not reelected.

Voters’ preferences for incumbents ($\kappa_h$) shifts $\bar{y}_h$ and primarily affects low types.

Voters’ preferences for incumbents ($\kappa_l$) shifts $\bar{y}_l$ and primarily affects high types.

The extended model provides a better explanation of why both low and high types exert effort in the first period.
A Maximum Likelihood Estimator

▶ Let $W_i$ be an indicator that is equal to 1 if mayor $i$ is reelected and serves a second term and zero otherwise.
▶ We compute her contribution to the log-likelihood as:

$$
\log L_i = W_i \log \left[ f \left( z_1^1, \ldots, z^K_1, z_1^1, \ldots, z^K_2 \right) \right] + (1 - W_i) \log \left[ f \left. \left( z_1^1, \ldots, z^K_1 \right) \right] \right)
$$

▶ We show that these densities are mixtures of normals.
▶ The structural parameters of the model can, therefore, be estimated using a full information maximum likelihood estimator.
▶ Note that we need to compute the equilibrium of the game to evaluate the log-likelihood function, i.e. we use a nested fixed point algorithm.
## Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Model I</th>
<th>Model II</th>
</tr>
</thead>
<tbody>
<tr>
<td>type distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{x}_1$</td>
<td>-0.418 (0.012)</td>
<td>-0.386 (0.007)</td>
</tr>
<tr>
<td>$\hat{x}_2$</td>
<td>0.226 (0.006)</td>
<td>0.181 (0.004)</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.571 (0.005)</td>
<td>0.508 (0.006)</td>
</tr>
<tr>
<td>monitoring technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.105 (0.003)</td>
<td>0.110 (0.003)</td>
</tr>
<tr>
<td>benefits of office</td>
<td>$\beta$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.730 (0.018)</td>
<td>0.567 (0.014)</td>
</tr>
<tr>
<td>election shock</td>
<td>$\kappa_I$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001 (0.000)</td>
<td>0.001 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$\kappa_H$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.645 (0.013)</td>
<td>0.567 (0.008)</td>
</tr>
<tr>
<td></td>
<td>$\lambda$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.340 (0.008)</td>
<td>0.396 (0.007)</td>
</tr>
<tr>
<td>factor loadings</td>
<td>$\mu_1^1$</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>$\mu_2^1$</td>
<td>1.528 (0.046)</td>
</tr>
<tr>
<td></td>
<td>$\mu_3^1$</td>
<td>1.431 (0.044)</td>
</tr>
<tr>
<td></td>
<td>$\mu_4^1$</td>
<td>-1.909 (0.040)</td>
</tr>
<tr>
<td>standard deviations</td>
<td>$\sigma_1$</td>
<td>0.931 (0.005)</td>
</tr>
<tr>
<td>of measurement errors</td>
<td>$\sigma_2$</td>
<td>0.829 (0.015)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_3$</td>
<td>0.853 (0.006)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_4$</td>
<td>—</td>
</tr>
</tbody>
</table>

Standard errors are given in parenthesis.
The equilibria that are implied by our estimates are typically pure-strategy equilibria.

Nevertheless, it is essential to allow for mixed strategy equilibria as you search of the parameter space.

Both types exert effort in equilibrium. The low type only get reelected in the absence of an election shock.
# Good versus Bad Politicians

<table>
<thead>
<tr>
<th>Model</th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model outcome</td>
<td>0.70</td>
<td>0.67</td>
</tr>
<tr>
<td>Employment rate</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>House price per sq ft (in $)</td>
<td>47</td>
<td>43</td>
</tr>
<tr>
<td>Expenditures per capita (in $)</td>
<td>123</td>
<td>55</td>
</tr>
<tr>
<td>Crime rate</td>
<td>-234</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model outcome</td>
<td>0.64</td>
<td>0.57</td>
</tr>
<tr>
<td>Employment rate</td>
<td>0.48</td>
<td>0.42</td>
</tr>
<tr>
<td>House price per sq ft (in $)</td>
<td>43</td>
<td>37</td>
</tr>
<tr>
<td>Expenditures per capita (in $)</td>
<td>113</td>
<td>47</td>
</tr>
<tr>
<td>Crime rate</td>
<td>-199</td>
<td></td>
</tr>
</tbody>
</table>
Policy Responsiveness due to Effort

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

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

- 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 incumbents in first term with the average quality of incumbents that are reelected and serve a second term, holding effort fixed a zero.
## Empirical Results: Policy Responsiveness

<table>
<thead>
<tr>
<th>Model</th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effort</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model outcome</td>
<td>0.224</td>
<td>0.154</td>
</tr>
<tr>
<td>Employment rate</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td>House price per sq ft (in $)</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Expenditures per capita (in $)</td>
<td>39</td>
<td>13</td>
</tr>
<tr>
<td>Crime rate</td>
<td></td>
<td>-54</td>
</tr>
<tr>
<td><strong>Selection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model outcome</td>
<td>0.056</td>
<td>0.059</td>
</tr>
<tr>
<td>Employment rate</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>House price per sq ft (in $)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Expenditures per capita (in $)</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Crime rate</td>
<td></td>
<td>-21</td>
</tr>
</tbody>
</table>
Entry of New Politician Types

- Next, we allow for entry of a new type of politician.
- We replace half of the lower type using a third type whose bliss point is slightly below that of the lower type.
- Overall, the quality of the low types decreases.
- Voters respond by increasing the election threshold for low types and decreasing the election threshold for the high type.
- In this case effort of the lower types will increase and effort of the high type will decrease.
- Whether total effort will increase or decrease depends on the magnitude of these effects.
Effort, Selection, and Average Quality

![Graph showing the relationship between effort, selection, and average quality of pool. The graph indicates a trend where effort decreases as selection and average quality of pool increase.]
Conclusions

- We find that some evidence that there are statistically significant and economically important differences among mayoral types.
- The effort effect is three to four times as large as the selection effect.
- A surprisingly large fraction of low-effort types can survive in equilibrium.
- An increase in the quality of candidate pool leads to an increase in welfare, but may also lead to lower total effort in equilibrium.
## Goodness of Fit

<table>
<thead>
<tr>
<th></th>
<th>Employment</th>
<th></th>
<th>HPI</th>
<th></th>
<th>Spending</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>data</td>
<td>model</td>
<td>data</td>
<td>model</td>
<td>data</td>
<td>model</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.000</td>
<td>-0.017</td>
<td>-0.000</td>
<td>-0.010</td>
<td>-0.000</td>
<td>-0.012</td>
</tr>
<tr>
<td>(1) lost</td>
<td>-0.096</td>
<td>-0.122</td>
<td>-0.336</td>
<td>-0.185</td>
<td>-0.160</td>
<td>-0.173</td>
</tr>
<tr>
<td>(2) re - 1st</td>
<td>0.196</td>
<td>0.112</td>
<td>0.227</td>
<td>0.198</td>
<td>0.217</td>
<td>0.182</td>
</tr>
<tr>
<td>(3) re - 2nd</td>
<td>-0.154</td>
<td>-0.103</td>
<td>-0.086</td>
<td>-0.146</td>
<td>-0.145</td>
<td>-0.138</td>
</tr>
<tr>
<td>(2)-(1)</td>
<td>0.292</td>
<td>0.235</td>
<td>0.564</td>
<td>0.384</td>
<td>0.378</td>
<td>0.356</td>
</tr>
<tr>
<td>(2)-(3)</td>
<td>0.350</td>
<td>0.216</td>
<td>0.314</td>
<td>0.345</td>
<td>0.363</td>
<td>0.321</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.997</td>
<td>1.004</td>
<td>0.995</td>
<td>1.000</td>
<td>0.997</td>
<td>1.001</td>
</tr>
<tr>
<td>(1) lost</td>
<td>1.112</td>
<td>0.966</td>
<td>0.822</td>
<td>0.918</td>
<td>1.095</td>
<td>0.929</td>
</tr>
<tr>
<td>(2) re - 1st</td>
<td>0.974</td>
<td>1.009</td>
<td>0.992</td>
<td>1.008</td>
<td>0.951</td>
<td>1.009</td>
</tr>
<tr>
<td>(3) re - 2nd</td>
<td>0.931</td>
<td>0.999</td>
<td>1.011</td>
<td>0.986</td>
<td>0.953</td>
<td>0.990</td>
</tr>
<tr>
<td>Re-election</td>
<td>0.709</td>
<td>0.702</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>