Studying the Electoral Effect of Incomplete Turnout Using Aggregate Data

Peter Selb, University of Konstanz
Marco Steenbergen, University of Zurich

EPOP Conference, September 7-9, 2012
Introduction

Motivation
- Many forms of political participation are socially selective, while SES is often tied to specific political preferences
- "Democracy’s unresolved dilemma" (Lijphart, 1997)
- Severe identification problems in previous research

Electoral effect of incomplete turnout

\[ Y^* = \pi_1 T + \pi_0 (1 - T) \]
\[ \pi_1 - Y^* = (1 - T)(\pi_1 - \pi_0) \]
Previous approaches

Individual level

- Use of survey data to impute nonvoter preferences based on choice models fitted to voters

- Different methods: propensity-score reweighting (Brunell/DiNardo, 2004), Blinder-Oaxaca decomposition (Citrin et al., 2003), MNL (Martinez/Gill, 2005), MI (Bernhagen/Marsh, 2007; 2010)

- Critical MAR-assumption; selectivity and misclassification problems in surveys

- No hope of progress short of collecting new data?
Previous approaches

Aggregate level

- Use of election statistics to assess the relationship between election results and turnout rates across electoral units (districts, countries)


- Methodological advances address potential simultaneity issue at individual level (Anzia, 2012; Berry/Gersen, 2011; Fowler, 2012; Hansford/Gomez, 2010)

- Otherwise, no particular worries about the ecological inference involved (Grofman et al., 1999)
Identification problem & strategy

Classical EI-problem

<table>
<thead>
<tr>
<th>$y = 1$</th>
<th>$x = 1$</th>
<th>$x = 0$</th>
<th>$y = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 = E(y = 1</td>
<td>x = 1)$</td>
<td>$\beta_0 = E(y = 1</td>
<td>x = 0)$</td>
</tr>
<tr>
<td>$y = 0$</td>
<td>$1 - \beta_1$</td>
<td>$1 - \beta_0$</td>
<td>$1 - X$</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>$X$</td>
<td>$1 - X$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

Goodman’s (1953) regression

\[
Y_j = \beta_0(1 - X_j) + \beta_1 X_j \\
E(Y_j) = \beta_0 + (\beta_1 - \beta_0)X_j + e_j \\
corr(X_j, e_j) = 0, \ corr(e_j, e_k) = 0
\]
Identification problem & strategy

Problem of inference on nonvoter behavior

<table>
<thead>
<tr>
<th>$y^* = 1$</th>
<th>$t = 1$</th>
<th>$t = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1 = E(y^* = 1</td>
<td>t = 1)$</td>
<td>$\pi_0 = E(y^* = 1</td>
</tr>
<tr>
<td>$y^* = 0$</td>
<td>$1 - \pi_1$</td>
<td>$1 - \pi_0$</td>
</tr>
<tr>
<td>$T$</td>
<td>$1 - T$</td>
<td>1</td>
</tr>
</tbody>
</table>

Modelling strategy

- Regression of $Y$ (not $Y^*$) on $T$ does not identify $\pi_0$
- Potential solution: re-specification in terms of a sequence of classical EI problems using an auxiliary variable
- Double regression (Kousser, 1973; Grofman et al., 1985); double EI (King, 1997); EI in $R \times C$ tables (Rosen et al., 2001)
Identification problem & strategy

**EI problem in an $R \times C$ table**

<table>
<thead>
<tr>
<th></th>
<th>$x = 1$</th>
<th>$x = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t = 1$</td>
<td>$\gamma_1 = E(t = 1</td>
<td>x = 1)$</td>
</tr>
<tr>
<td>$t = 0$</td>
<td>$1 - \gamma_1$</td>
<td>$1 - \gamma_0$</td>
</tr>
<tr>
<td></td>
<td>$X$</td>
<td>$1 - X$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$x = 1$</th>
<th>$x = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y = 1$</td>
<td>$\lambda_1 = E(y = 1</td>
<td>x = 1)$</td>
</tr>
<tr>
<td>$y = 0$</td>
<td>$1 - \lambda_1$</td>
<td>$1 - \lambda_0$</td>
</tr>
<tr>
<td></td>
<td>$X$</td>
<td>$1 - X$</td>
</tr>
</tbody>
</table>

$T = 1 - T$

$Y = 1 - Y$
Identification problem & strategy

Double regression

\[ E(T_j) = \gamma_0 (1 - X_j) + \gamma_1 X_j + u_j \]
\[ E(Y_j) = \lambda_0 (1 - X_j) + \lambda_1 X_j + v_j \]

\[ \text{corr}(X_j, u_j) = \text{corr}(X_j, v_j) = \text{corr}(u_j, u_j) = \text{corr}(v_j, v_j) = 0, \]
\[ \text{corr}(u_j, v_j) = 0 \]

\[ \beta_0 = E(y = 1|x = 0, t = 1) = \frac{\lambda_0}{\gamma_0} \]
\[ \beta_1 = E(y = 1|x = 1, t = 1) = \frac{\lambda_1}{\gamma_1} \]
Identification problem & strategy

Projection of election results under full turnout

\[ \hat{Y}^*_j = \hat{\beta}_0 (1 - X_j) + \hat{\beta}_1 X_j \]

\[ \hat{Y}^* = \frac{\sum_{j=1}^{J} n_j (\hat{\beta}_0 (1 - X_j) + \hat{\beta}_1 X_j)}{\sum_{j=1}^{J} n_j} \]

Participatory bias

\[ Y - \hat{Y}^* \]
Empirical application

Swiss direct democracy as an empirical case

- Research has focused on elections rather than referenda and initiatives (R&I)
- R&I are on their rise around the globe
- Turnout is often low in R&I
- Turnout-policy link is much more immediate in R&I
- Municipal-level data for all 257 national proposals, 1981-2011
- Auxiliary variable: population share of highly educated
Empirical application

Diagnostics

- Check for out-of-bound predictions for $\beta_0$, $\beta_1$
- Plot residuals versus $X_j$ for indications of aggregation effects
- Use tomography plots (King, 1997) to see (a) how informative the data are about $\beta_0$, $\beta_1$, and (b) how plausible it is to assume constant $\beta_0$, $\beta_1$
- 74 of 257 cases appear relatively unconspicious
- These checks are very preliminary and cannot prove the model’s appropriateness!
Empirical application

Figure: Actual versus projected approval rates under full turnout in 74 proposals.
Conclusion

Summary

- Interest is in relationship between individual participation propensities and preferences
- Survey data is problematic due to selectivity and misclassification
- Previous aggregate approaches are ill-defined
- One solution is to re-specify in terms of a sequence of classical E1 problems using an auxiliary variable, use double regression
- Empirical application suggests sometimes tremendous turnout effects, particularly in initiatives (unlike referendums) on environmental and social security issues
- Results are very preliminary
Conclusion

Next steps

- Auxiliary information pertains to whole local populations, not eligibles (“triple regression”?)
- Use of more critical auxiliary information, infer income distribution from tax statistics
- Inclusion of covariates to account for aggregation bias
- Stratification by region
- Use, develop of more advances methodological approaches
- Contrast with, use of results from survey data