Updated Pre-analysis Plan for "Can Voters Reward Prudent Managers of Public Funds?"

Kelly Zhang

February 18, 2014

Abstract

This updated version makes two primary updates to the previous pre-analysis plan. First, it follows up on pre-specified modifications. It makes adjustments based upon covariate balance for the time treatment and elaborates on the estimation of heterogenous treatment effects. Second, it corrects some of the previous model specifications. It drops constituency fixed effects when the covariate of interest also varies at the constituency level, and modifies "sublocation" analyses to "location/ward" analyses because the audit data are only available at the "location/ward" level.

1 Overview

Mismanagement of public spending is a salient issue in many developing countries (World Bank Development Report, 2004). In a study across sub-Saharan African countries, up to 80% of public funding did not reach beneficiaries (Reinikka and Svensson, 2004).[1] Generally, waste in public spending can be especially costly to the beneficiaries, who rely on the public sector for basic services such as health, water, education, and infrastructure. Bandiera et al. (2009) distinguish passive waste from active waste—the latter benefiting the decision maker, while the former does not—and find that the type of waste tends to vary with how the public body is set up.

There tends to be a weak correlation between perceived and actual corruption for missing expenditures in public spending (Olken, 2009). The ability of citizens to infer incumbent performance in public spending can be critical in ensuring politician discipline while in office, as informational asymmetries have the potential to hinder electoral accountability. When politicians manage funds and voters are able to hold them accountable for their actions, having information on politician corruption can enable voters to reward well-performing and to sanction poorly-performing politicians (Ferraz and Finan, 2008).Yet, some experimental studies also demonstrate that information campaigns about corruption are not always effective in improving political accountability. Distribution of audits on leakages in politician spending appeared to reduce voter turnout and to alienate voters in a field experiment in Mexico; an information campaign on politician corruption

[1]Such leakage in public spending is typically measured using estimation by subtraction–where government allocation records are compared with local beneficiary records.

scandals appeared to reduce the salience of corruption in vote choice in Brazil (Chong et al., 2011; de Figueiredo et al., 2011).

This project's approach develops the notion that voters may be biased against incumbents in developing democracies. Although there appears to be an incumbency advantage in developed democracies such as the United States, Canada, Finland, Italy, Japan, and the United Kingdom, where voters tend to be biased in favor of the incumbent (Fiorina, 1981; Gelman and King, 1990; Ansolabehere and Snyder, 2003; Ariga, 2010), in developing democracies such as India and Brazil, there appears to be an incumbency disadvantage (Uppal, 2009; Klašnja, 2011; Titiunik, 2011).

When public sector failures (i.e. pot-holed roads, dilapidated medical centers, missing stolen education funds) are evident in everyday life, voters may assume that incumbents are responsible for such failures even when they are not. Politicians in developing democracies often lack credibility in promising public goods, and will often use strategies of vote-buying, clientelism, and identity-based voting for electoral gain (Keefer and Vlaicu, 2008; Kramon, 2011; Wantchekon, 2003, 2009; Pande, 2003). Given high skepticism towards politicians, even incumbents with a strong record of public goods provision may find it difficult to get credit for what they have done for their constituents. When the public sector is perceived to be highly corrupt, it is possible that voters not only lack information about poor performance, but also information about good performance. This project proposes that in contexts where corruption perceptions are highly pessimistic, an information campaign which enables citizens to distinguish the well-performers from poor-performers can improve political accountability.

The analysis here focuses on how citizens respond to politician management of passive waste in public spending, in the context of a developing democracy. It uses data from a field experiment in Kenya, and builds on a growing body of experimental studies, which examine how information campaigns on politician performance can improve political accountability (Banerjee et al., 2011; Chong et al., 2011; Humphreys and Weinstein, 2012). The theory also utilizes insights from the political economy literature on public spending, which posit that citizens tend to reward politicians for responsiveness and for public programs in which they are beneficiaries (Banerjee et al., 2011; Besley and Burgess, 2002; Fiorina and Noll, 1979; Tufte, 1978).

2 Sample and Data

The baseline survey documents respondents' baseline beliefs about how much leakage there is in politician spending for the Constituency Development Fund, with the following question:

- *In the past 5 years, the National Taxpayer's Association has examined Constituency Development Fund (CDF) spending for MPs across Kenya. In their assessments, they compare the government records of CDF allocations with the project spending records of each constituency. Suppose that your constituency was allocated 50,000,000 Ksh for the CDF. On average, how much do you think actually reached the assessed CDF projects in your constituency, in one year?*

Baseline data collection was embedded in a nationwide public opinion poll of approximately 6000 individuals, carried out in late February 2013 by Ipsos Synovate. The survey

company used a randomized multi-stage stratified sampling design using probability proportional to size, with face-to-face interviews nationwide. This public opinion survey had a battery of over 100 questions, the majority of which were on political attitudes towards the upcoming elections.

Due to budget constraints, only 2500 individuals were interviewed from this sample for the endline.[2] The treatment SMS messages were delivered during the first week of March 2013. The endline surveys took place during the second week of March 2013, via a phone interview.

3 Experimental Design

The information intervention was embedded in a series of factual SMS messages. Treatment was randomly assigned at the individual level, and delivered during the 2013 Kenyan national elections. Treatment individuals received information on how much their constituency received in Constituency Development Funds (CDF), and how much actually reached projects for a given fiscal year.[3] Control individuals received no information. For the entire SPEC sample (n=5000), treatment was stratified by urban/rural, gender, wage income, above median education, overestimating leakage relative to the audit, above constituency median beliefs about leakage in public spending, and if the MP was prudent manager in having 0% of public funds unaccounted for in the audit. The follow-up interviews were carried out over 5 days; the random assignment to day of interview was stratified by treatment.

The factual SMS messages had the source of information listed at the end of each message. The entire sample received the initial message, which stated that citizens could vote for 6 elected positions (Senator, Ward Representative, Governor, Women's representative, President, and Member of Parliament/National Assembly). This was to introduce respondents to the SMS delivery of factual information, such that initial mobilization via SMS would be equal across both treatment and control groups. Thus, both the treatment and control groups had the same initial "civic education" message on Day 1. The treatment group then received a series of SMS messages as follows[4]:

1. **Day 1 (Full Sample)**
   Did you know that you can vote for 6 elected positions? President, Senator, MP, Women Rep, Governor, Ward Rep. SOURCE: IEBC, Handbook on Elective Positions

2. **Day 2 (Treatment Only)**

[2]Of the original 6000, about 1000 individuals without phones or valid phone numbers were excluded from the study, because the intervention took place through an SMS platform. While this limits the generalizability of the study sample to those with mobile phones, we believe that this is reasonable population to study, since mobile phone penetration is at around 78% (Communications Commission of Kenya 2013).

[3]The information was drawn from the National Taxpayer's Association, a reputable and well-known local civil society organization, funded by international aid agencies such as Department for International Development (DfID) and Swedish International Development Cooperation Agency (SIDA).

[4]Note: Not all treatment individuals received messages 4 and 5. Due to technical difficulties the list of treatment phone numbers to SMS was randomly truncated by the telecommunications company, whose servers were overloaded by the electronic transmission of voting results from around the country.

- **Audited Constituencies**
  Did you know [constituency]'s CDF got XXXXXXXXKSH in XXXX/XX, with XXXXXXXXKSH reaching project budgets? SOURCE: National Taxpayers Association CDF Report Card [by constituency]
- **Non-Audited Constituencies**
  Did you know parental involvement is key for KCPE performance? [COUNTY-NAME] ranks XX, scoring Y/10. SOURCE: National Taxpayers Association School Report Card [by county, SRC-10. Parental Responsibility]

3. **Day 3 (Treatment Only)**
   Did you know that a school is supposed to receive 1,020 Ksh per pupil, from the Free Primary Education capitation grant? SOURCE: Ministry of Education

4. **Day 4 (Treatment Only)**
   Did you know that Nigeria won the 2013 Africa Cup of Nations? SOURCE: Daily Nation, 14 February 2013

5. **Day 5 (Treatment Only)**
   Did you know that Kenya won 11 medals in the 2012 London Olympics? 2 gold, 4 silver, and 5 bronze medals. SOURCE: BBC SPORT

The SMS messages were delivered during the first week of March 2013. Day 1 and Day 2 were delivered just prior to the national elections on March 4th. The remaining messages (Days 3-5) were delivered in the days immediately following the March 4th elections. The endline surveys were carried out in the week following the information campaign, during the second week of March 2013.

4 Specifications

In this section, the specifications are for individual i in audited constituency j. To facilitate interpretation and comparison across estimates, OLS (with sampling weights) is used to estimate the average treatment effect, for those assigned to treatment T_{ij}. After a preliminary analysis of the control data, an amendment may be filed to drop outcome variables that have little or no variation.[5]

4.1 Model Specification

For the primary specification, the average treatment effect is estimated using OLS:

$$Y_{ij} = \alpha + \beta_1 T_{ij} + \eta_j + \varepsilon_{ij} \tag{1}$$

For heterogenous treatment effects using OLS:

$$Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_i + \beta_3 T_{ij} \times B_i + \eta_j + \varepsilon_{ij} \tag{2}$$

[5]This provides an intent-to-treat estimate. Using SMS receipt records, it is possible to identify the individuals who received the messages, to estimate the impact of treatment on the treated.

Where Y_{ij} is the outcome of interest, T_{ij} is the treatment indicator, and B_i is the baseline characteristic of interest, with η_j denoting constituency fixed effects. Here, B_i can also be B_j at the constituency level and B_l at the location/ward level. Constituency fixed effects η_j will be dropped from the specification when the covariate of interest B_j also varies at the constituency level.

Testing the hypotheses:

- $\beta_1 = 0$ in equation 1.
- $\beta_3 = 0$ for each covariate B_i in equation 2.

4.2 Primary Outcomes

Information uptake

Information uptake is measured in absolute deviations from the audit estimate, where respondent guess R_{ij} is subtracted from the audit estimate from treatment T_j:

$$Y_{ij} = |R_{ij} - T_j|$$

- Y_{ij} = How much money was given to respondent's constituency for the CDF fund (absolute deviation)
- Y_{ij} = Percentage of money that actually reached CDF projects (absolute deviation)

Information uptake can also be measured as a binary outcome:

- Y_{ij} = Willingness to provide a guess on how much money actually reached projects (binary indicator)
- Y_{ij} = Ability to name incumbent MP from 2007/8 to 2012/3 (binary indicator for correct name)

Political support

- Y_{ij} = Re-elected incumbent MP for elected office (binary indicator for re-elected)
- Y_{ij} = Willingness to extend interview to state support for incumbent MP (time in seconds)
- Y_{ij} = Willingness to vote (binary indicator: voted self-report & ability to name elected positions & willingness to name parties that they supported)

Beliefs about incumbent politician quality (CDF management)

- Y_{ij} = Listing CDF management as a good aspect of their former MP (binary indicator for good aspect)
- Y_{ij} = Listing CDF management as a bad aspect of their former MP (binary indicator for bad aspect)
- Y_{ij} = Score of former MP in public fund management (x/10)
- Y_{ij} = Score of former MP in public good provision (x/10)

4.3 Secondary Outcomes

Expectations of MP

- Y_{ij} = Would ask MP about CDF (binary indicator)
- Y_{ij} = Would ask MP for cash (binary indicator)

Beliefs about incumbent politician quality (non-CDF management)

- Y_{ij} = Lists more good things about former MP (count of items listed)
- Y_{ij} = Lists more bad things about former MP (count of items listed)

Beliefs about top contender quality

- Y_{ij} = Score of top contender in public good provision (x/10)

Beliefs about general politician quality

- Y_{ij} = How many politicians would steal public funds (out of 10)
- Y_{ij} = How many politicians are competent (out of 10)
- Y_{ij} = How many politicians are honest (out of 10)

4.4 Treatment

- T_{ij} = Treatment
  (Assignment to treatment, to receive SMS messages 2 to 5)

4.5 Covariates of Interest

Individual level

- B_i = Respondent was especially cynical about corruption
  (binary indicator for prediction above consituency median)
- B_i = Respondent overestimated corruption
  (binary indicator for predicting more funds stolen than audit estimate)
- B_i = Education
  (binary indicator for above median)
- B_i = Urban resident
  (binary indicator for urban)
- B_i = Earns wage income
  (binary indicator for wage income)

Constituency level

- B_j = Incumbent MP was prudent manager
  (binary indicator if MP had 0% funds unaccounted for)
- B_j = Incumbent MP had funds classified as badly used or wasted
  (binary indicator if MP had funds badly used or wasted)
- B_j = Incumbent MP ran for re-election
  (binary indicator if re-ran for election in 2013)
- B_j = Incumbent MP has held elected office previous to this term
  (binary indicator if MP has held an elected office prior to being elected as MP)

Location/Ward level

- B_l = Location/ward has CDF projects with leakage
  (binary indicator for if location/ward k has any funds unaccounted for)
- B_l = Location/ward has badly managed or wasted funds for CDF projects
  (binary indicator for if location/ward k has any funds badly used or wasted)

4.5.1 Additional Specifications

Controls and strata fixed effects can be added to enhance precision of the estimates. Any variables that are found to be imbalanced at the baseline will be included as controls as well. Controls are designed by the vector X_{ij} and include the following:

- Language of interview (binary indicator for English)
- Gender (binary indicator for male)
- Age (binary indicator for above 45 years of age)
- Year of CDF audit data (binary indicator for year of constituency audit)

For additional specifications, OLS is used to estimate the average treatment effect with controls and strata fixed effects:

$$Y_{ijk} = \alpha + \beta_1 T_{ij} + X'_{ij}\gamma + \eta_j + \theta_k + \varepsilon_{ijk} \tag{3}$$

Heterogenous treatment effects with controls and strata fixed effects,

$$Y_{ijk} = \alpha + \beta_1 T_{ij} + \beta_2 B_{ij} + \beta_3 T_{ij} \times B_{ij} + X'_{ij}\gamma + \eta_j + \theta_k + \varepsilon_{ijk} \tag{4}$$

Where Y_{ij} is the outcome of interest, T_{ij} is the treatment indicator, and B_{ij} is the baseline characteristic for individual i in audited constituency j, with η_j denoting constituency fixed effects and θ_k denoting strata fixed effects. Constituency fixed effects η_j will be dropped from the specification when the covariate of interest B_j also varies at the constituency level.

Testing the hypotheses:

- $\beta_1 = 0$ for equation 3.
- $\beta_3 = 0$ for equation 4.

4.6 Triple Interaction Terms

In some scenarios, triple interaction terms will be illustrative for particular subgroups, using data from the audits. The primary specifications are presented below, but additional specifications can include controls and strata fixed effects.

- B_{1j} = Incumbent MP ran for re-election
- B_{2j} = Incumbent MP was prudent manager
- B_{3j} = Incumbent MP had badly used or wasted funds at the *constituency level*
- B_{1l} = Incumbent MP had badly used or wasted funds at the *location/ward level*
- B_{2l} = Location/ward has CDF projects with leakage

Incumbent MP ran for re-election

Of particular interest is if individuals living in constituencies where MPs ran for re-election are more responsive to the incumbent being a prudent manager, having funds classified as badly used or wasted at the constituency or location/ward level, or having funds unaccounted for at the location/ward level.

$$\begin{aligned} Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_{1j} + \beta_3 B_{2j} + \beta_4 (T_{ij} \times B_{1j}) + \beta_5 (T_{ij} \times B_{2j}) \\ + \beta_6 (B_{1j} \times B_{2j}) + \beta_7 (T_{ij} \times B_{1j} \times B_{2j}) + \varepsilon_{ij} \end{aligned} \tag{5}$$

Testing the hypotheses for equation 5:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_4 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election.
- $\beta_5 = 0$ to see if treatment differs for constituencies where the incumbent was a prudent manager.
- $\beta_7 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election and where the incumbent was a prudent manager.

$$\begin{aligned} Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_{1j} + \beta_3 B_{3j} + \beta_4 (T_{ij} \times B_{1j}) + \beta_5 (T_{ij} \times B_{3j}) \\ + \beta_6 (B_{1j} \times B_{3j}) + \beta_7 (T_{ij} \times B_{1j} \times B_{3j}) + \varepsilon_{ij} \end{aligned} \tag{6}$$

Testing the hypotheses for equation 6:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_4 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election.
- $\beta_5 = 0$ to see if treatment differs for constituencies where the incumbent had funds badly used or mismanaged at the *constituency level*.
- $\beta_7 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election and had funds badly used or mismanaged at the *constituency level*.

$$\begin{aligned} Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_{1j} + \beta_3 B_{1l} + \beta_4 (T_{ij} \times B_{1j}) + \beta_5 (T_{ij} \times B_{1l}) \\ + \beta_6 (B_{1j} \times B_{1l}) + \beta_7 (T_{ij} \times B_{1j} \times B_{1l}) + \varepsilon_{ij} \end{aligned} \tag{7}$$

Testing the hypotheses for equation 7:

- $\beta_1 = 0$ to see if treatment differs from control.

- $\beta_4 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election.
- $\beta_5 = 0$ to see if treatment differs for constituencies where the incumbent had funds badly used or mismanaged at the *location/ward level*.
- $\beta_7 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election and had funds badly used or mismanaged at the *location/ward level*.

$$\begin{aligned} Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_{1j} + \beta_3 B_{2l} + \beta_4 (T_{ij} \times B_{1j}) + \beta_5 (T_{ij} \times B_{2l}) \\ + \beta_6 (B_{1j} \times B_{2l}) + \beta_7 (T_{ij} \times B_{1j} \times B_{2l}) + \varepsilon_{ij} \end{aligned} \tag{8}$$

Testing the hypotheses for equation 8:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_4 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election.
- $\beta_5 = 0$ to see if treatment differs for constituencies where the incumbent had funds unaccounted for at the *location/ward level*.
- $\beta_7 = 0$ to see if treatment differs for constituencies where the incumbent ran for re-election and had funds unaccounted for at the *location/ward level*.

Incumbent MP was a prudent manager

Also of interest is whether or not voters within constituencies or location/wards, where funds are badly used or mismanaged, respond differently to the information treatment. Recall that prudent manager is defined as having 0% unaccounted for funds; however, it is possible that the accounted for funds were still classified by technical assessments as being mismanaged or wasted.

$$\begin{aligned} Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_{2j} + \beta_3 B_{3j} + \beta_4 (T_{ij} \times B_{2j}) + \beta_5 (T_{ij} \times B_{3j}) \\ + \beta_6 (B_{2j} \times B_{3j}) + \beta_7 (T_{ij} \times B_{2j} \times B_{3j}) + \varepsilon_{ij} \end{aligned} \tag{9}$$

Testing the hypotheses for equation 9:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_4 = 0$ to see if treatment differs for constituencies with prudent managers.
- $\beta_5 = 0$ to see if treatment differs for constituencies where the incumbent had funds badly used or mismanaged at the *constituency level*.
- $\beta_7 = 0$ to see if treatment differs for those who live in constituencies with prudent managers (funds fully accounted for) but still funds classified as badly used or mismanaged *constituency level*.

$$Y_{ij} = \alpha + \beta_1 T_{ij} + \beta_2 B_{2j} + \beta_3 B_{1l} + \beta_4 (T_{ij} \times B_{2j}) + \beta_5 (T_{ij} \times B_{1l}) + \beta_6 (B_{2j} \times B_{1l}) + \beta_7 (T_{ij} \times B_{2j} \times B_{1l}) + \varepsilon_{ij} \quad (10)$$

Testing the hypotheses for equation 10:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_4 = 0$ to see if treatment differs for those who live in constituencies with prudent managers.
- $\beta_5 = 0$ to see if treatment differs for those who live in constituencies where funds were badly managed or misused in CDF projects at the *location/ward level.*
- $\beta_7 = 0$ to see if treatment differs for those who live in constituencies with prudent managers and where funds were badly managed or misused in CDF projects at the *location/ward level.*

5 Extensions

5.1 Capturing information dissipation

Here, we are interested in the extent to which time diminishes the impact of treatment (T_{ij}). The specification below is for the previously listed primary outcomes of information uptake, political beliefs, and political support listed in section 4.2. Additionally, a more list of comprehensive information uptake outcomes are specified below, just to capture the extent to which information may generally dissipate after an information intervention. For the outcomes listed for comprehensive information uptake, the full sample (audited and non-audited constituencies) will be used to estimate the average treatment effect.

Technical errors on the part of the mobile phone provider randomly truncated SMS delivery lists for SMS messages 4 and 5. Thus, the primary specification differs here in the treatment indicator; the primary specification here will also create a new treatment indicator for "treatment" for SMS messages 4 and 5: Only those to whom the SMS message was delivered will be coded as being "treated" with the information for SMS messages 4 and 5. This new treatment indicator provides a better estimate for the true impact of information, because it excludes those individuals who never received the messages due to the technical error. The analysis will still be done for the original treatment indicator; however, inclusion of the excluded individuals is likely to understate the impact of the SMS treatment.

Comprehensive Information Uptake

In addition to the information uptake outcomes listed under Primary Outcomes (section 4.2), we also utilize the variables listed below for information uptake. Absolute deviations are again measured by subtracting the respondent guess R_{ij} from the treatment information T_j:

$$Y_{ij} = |R_{ij} - T_j|$$

- Y_{ij} = How much money a school receives for each pupil for Free Primary Education (absolute deviation)
- Y_{ij} = Knows Nigeria won the 2013 Africa Cup of Nations (binary indicator)
- Y_{ij} = How many medals Kenya won in the 2012 London Olympics (absolute deviation)

Interview Days

This can be measured by looking at the categories day-by-day:

- D_1= Randomly assigned to be interviewed 1 day after Day 1
- D_2= Randomly assigned to be interviewed 2 days after Day 1
- D_3= Randomly assigned to be interviewed 3 days after Day 1

Specification

For the random assignment of each individual to the subsequent day of interview D_k, after Day 1, OLS is used to estimate:

$$Y_{ij} = \alpha_j + \beta_1 T_{ij} + \sum_{n=1}^{3} \beta_{1+n} D_{in} + \beta_{5+k} T_{ij} \times \sum_{k=1}^{3} D_{in} + \varepsilon_{ij} \tag{11}$$

Testing the hypotheses:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_6 = 0$ to see if treatment differs for being interviewed 1 day after Day 1.
- $\beta_7 = 0$ to see if treatment differs for being interviewed 2 days after Day 1.
- $\beta_8 = 0$ to see if treatment differs for being interviewed 3 days after Day 1.

In the endline survey, the individuals were randomly assigned to an interview Day 1, Day 2, Day 3, Day 4, and Day 5. If they could not be contacted on the assigned day, they were dropped from the list, with a minimum daily yield of about 75% of individuals from the original call list.

There appears to be imbalance in covariates between individuals interviewed on Day 5 and individuals interviewed during the first four days. Although treatment assignment was randomized, the survey company experienced a service outage and was unable to reach its quota on Day 5, instead including respondents from subsequent days. Interview descriptives (no. interviews per interviewer, interview duration, etc.) appear to be irregular. Due to this irregularity in data collection, we exclude Day 5 from the main analysis and focus on Days 1-4. The full analysis including Day 5 will also be reported, but will not be the primary analysis since Day 5's sample may not be comparable to the sample from Days 1-4.

- D_0= Randomly assigned to be interviewed on Days 3, 4.

For the random assignment of each individual to being interviewed later on D_0, OLS is used to estimate:

$$Y_{ij} = \alpha_j + \beta_1 T_{ij} + \beta_2 D_0 + \beta_3 T_{ij} \times D_0 + \varepsilon_{ij} \tag{12}$$

Testing the hypotheses:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_3 = 0$ to see if treatment differs for being interviewed during the first two days.

5.2 Do non-audited constituencies differ?

While the National Taxpayer's Association audited over 130 constituencies across Kenya, not all constituencies were audited, due to the intensive resources required to carry out these audits. Although their decision to audit constituencies was geographically determined (starting from the provincial capital, and moving outwards), one concern is that the NTA only audited constituencies where citizens were deemed to be the most responsive to information on public spending. While this would not threaten the internal validity of the experiment, it has potential implications for the generalizability of this study.

The outcome of interest here is:

- Y_{ij} = How much money is allocated per student for Free Primary Education (absolute deviation from correct answer)

Where information uptake is measured in absolute deviations from the audit estimate, where respondent guess R_{ij} is subtracted from the audit estimate from treatment T_j:

$$Y_{ij} = |R_{ij} - T_j|$$

The sample is expanded to include non-audited constituencies here, with treatment (T_{ij}) randomly assigned at the individual level. Thus in both audited and non-audited constituencies, there are treated and untreated individuals. To examine if treated individuals living in non-audited constituencies (N_j) exhibit similar patterns in recalling information about public spending to audited constituencies, we can use the following specification:

$$Y_{ij} = \alpha_j + \beta_1 T_{ij} + \beta_2 N_j + \beta_3 T_{ij} \times N_j + \varepsilon_{ij} \tag{13}$$

Testing the hypotheses that:

- $\beta_1 = 0$ to see if treatment differs from control.
- $\beta_1 + \beta_3 = 0$ to see if the treatment effect differs from control, for non-audited constituencies.

5.3 Exploratory Analyses

5.3.1 Local Average Treatment Effect

Following Imbens (2013), we can use treatment (T_i) as an instrument in two-stage least squares (TSLS) to explore the causal effect of beliefs about politician performance in public fund management (W_i on political support Y_i). Here, Y_i is (1) if the respondent re-elected the incumbent for elected office and (2) respondent willingness to extend interview to state support for incumbent MP, T_i is the treatment indicator, and W_i is the respondent i's score of former MP in public fund management. The following set of equations can be estimated by least squares:

$$W_i = \pi_0 + \pi_1 \cdot T_i + \nu_i$$

$$Y_i = \gamma_0 + \gamma_1 \cdot T_i + \eta_i$$

with the estimate of interest $\hat{\beta}_1^{IV}$ through Indirect Least Squares (ILS) as the ratio:

$$\hat{\beta}_1^{IV} = \hat{\beta}_1^{TSLS} = \hat{\beta}_1^{ILS} = \frac{\hat{\gamma}_1}{\hat{\pi}_1}$$

.

5.3.2 Latent Variable Model

An alternative approach suggested by Imbens (2013) and Cameron and Trivedi (2009) is to estimate an index function model for individual compliance type given a set of covariates X_i. The model:

$$Y_i = \beta_0 + \beta_1 \cdot W_i + \beta_2' X_i + \varepsilon_i$$

can be decomposed into two parts, a model for compliance type given covariates X_i and a model for potential outcomes given covariates for each compliance type. Here, Y_i is (1) if the respondent re-elected the incumbent for elected office and (2) respondent willingness to extend interview to state support for incumbent MP, X_i denotes the following baseline characteristics: respondent was especially cynical about corruption (above median), education above median, and earns wage income, and $T_i = t$. Assuming (η_i, ε_i) are jointly normally distributed:

$$W_i(t) = 1\{\pi_0 + \pi_1 \cdot t + \pi_2' X_i + \eta_i \geq 0\}$$

$$Y_i = \beta_0 + \beta_1 \cdot w + \beta_2' X_i + \varepsilon_i$$

This model imposes strong restrictions on the relationship between type and outcomes for unit i; compliance type will be dependent upon the relationship between $\pi_0, \pi_i, \pi_2' X_i$.

References

Ansolabehere, S. and J. M. Snyder (2003). The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942–2000. *Election Law Journal 1*(3), 315–338.

Ariga, K. (2010). *Entrenched Incumbents, Irresponsible Parties? Comparative Analysis of Incumbency Advantage Across Different Electoral Systems*. Ph. D. thesis, University of Michigan.

Bandiera, O., A. Prat, and T. Valletti (2009). Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review 99*(4), 1278–1308.

Banerjee, A., S. Kumar, R. Pande, and F. Su (2011). Do Informed Voters Make Better Choices? Experimental Evidence from Urban India. Poverty Action Lab Working Paper.

Besley, T. and R. Burgess (2002). The Political Economy of Government Responsiveness: Theory and Evidence from India. *Quarterly Journal of Economics 117*(4), 1415–1451.

Cameron, A. C. and P. K. Trivedi (2009). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chong, A., A. L. De La O, D. Karlan, and L. Wantchekon (2011). Looking Beyond the Incumbent: The Effects of Exposing Corruption on Electoral Outcomes. *NBER Working Paper 17679*.

de Figueiredo, M. F. P., F. D. Hidalgo, and Y. Kasahara (2011). When Do Voters Punish Corrupt Politicians? Experimental Evidence from Brazil. *Unpublished Manuscript, University of California Berkeley*.

Ferraz, C. and F. Finan (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *Quarterly Journal of Economics 123*(2), 703–745.

Fiorina, M. (1981). *Retrospective Voting in American National Elections*. Yale University Press.

Fiorina, M. P. and R. G. Noll (1979). Majority rule models and legislative elections. *The Journal of Politics 41*(4), 1081–1104. ArticleType: research-article / Full publication date: Nov., 1979 / Copyright © 1979 Southern Political Science Association.

Gelman, A. and G. King (1990). Estimating incumbency advantage without bias. *American Journal of Political Science 34*(4), 1142–1164.

Humphreys, M. and J. M. Weinstein (2012). Policing politicians: Citizen empowerment and political accountability in uganda. *Mimeo*.

Imbens, G. (2013). Methods for Estimating Treatment Effects IV: Instrumental Variables and Local Average Treatment Effects.

Keefer, P. and R. Vlaicu (2008). Democracy, credibility, and clientelism. *Journal of Law, Economics, and Organization 24*(2), 371–406.

Klašnja, M. (2011). Increasing Rents and Incumbency Disadvantage. *Mimeo*.

Kramon, E. (December 2011). Why Do Politicians Buy Votes When The Ballot Is Secret? Theory And Experimental Evidence From Kenya. *Mimeo*.

Olken, B. A. (2009). Corruption Perceptions vs. Corruption Reality. *Journal of Public Economics 93*(7), 950–964. doi: 10.1016/j.jpubeco.2009.03.001.

Pande, R. (2003, November). Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India. *American Economic Review 93*(4), 1132–1151.

Reinikka, R. and J. Svensson (2004). Local capture: Evidence from a central government transfer program in uganda. *The Quarterly Journal of Economics 119*(2), 679–705.

Titiunik, R. (2011). Incumbency Advantage in Brazil: Evidence from Municipal Mayor Elections. *Mimeo*.

Tufte, E. (1978). *Political Control of the Economy*. Princeton University Press.

Uppal, Y. (2009). The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures. *Public Choice 138*, 9–27.

Wantchekon, L. (2003). Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. *World Politics 55*(3), 399–422.

Wantchekon, L. (2009). Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence From Benin. Unpublished Manuscript, New York University.

World Bank Development Report (2004). Making Services Work for Poor People. Technical report, The World Bank.