

Pre-Analysis Plan

Mechanisms for Strengthening Accountability to the Rural Poor: Evidence from Public Expenditures in Mali

Jessica Gottlieb*
Texas A&M University

Katrina Kosec†
IFPRI

October 15, 2015

Abstract

Can inducing local governments to compete for scarce central government and donor funds—that is, top-down efforts to encourage inter-jurisdictional competition—improve public service delivery? While this question has been examined using a variety of identification strategies to address endogeneity, we know of no experimental evidence to date. In contrast, a growing experimental literature suggests that auditing local governments and sharing their performance with citizens creates bottom-up pressure that improves service delivery. Unfortunately, bottom-up mechanisms are costly in practice, may require civic education campaigns, and may be susceptible to local policymakers manipulating citizens’ understanding of the results. We build on this literature in two ways. First, we examine whether auditing local governments and using a top-down mechanism can effectively induce inter-jurisdictional competition and improve service delivery. Second, we examine whether there is any evidence of a statistically significantly different impact of top-down vs. bottom-up mechanisms. If a top-down mechanism is effective, it would suggest a valuable and relatively cost-effective mechanism to improve service delivery.

*Texas A&M University. jgottlieb@tamu.edu.

†International Food Policy Research Institute. k.kosec@cgiar.org.

1 Motivation and Background

Can inducing local governments to compete for scarce central government and donor funds—that is, top-down efforts to encourage inter-jurisdictional competition—improve public service delivery? Studying the determinants of effective local policymaking is important since a growing number of countries are devolving administrative, fiscal, and political authorities to local governments. The movement toward decentralization is particularly dramatic in the developing world (Oates, 1999; Dillinger, 1994). Decentralization has the potential to create more efficient, responsive, and participatory government. Local governments have better information about optimal public expenditure levels (Hayek, 1945), and citizens can monitor performance by benchmarking policy outcomes against those of neighbors (Besley and Case, 1995; Seabright, 1996). Decentralization may also lead governments to compete to attract citizens and investments (including transfers and donor funds), resulting in more efficient service delivery and poverty reduction. However, in many developing countries, local policymakers are unaccountable due to a lack of transparency and elite collusion (Bardhan, 2002; Bardhan and Mookherjee, 2006; Mansuri and Rao, 2004). As a result, decentralization does not always lead to better policies or economic growth.

Reaping the benefits of decentralized service delivery while avoiding mismanagement of funds poses an important development challenge. There are two primary mechanisms to strengthen accountability. Top-down mechanisms have higher tiers of governments, donors, or NGOs hold local policymakers accountable by restraining, monitoring, and rewarding or sanctioning them. Bottom-up mechanisms help citizens hold local policymakers accountable by giving them more information, voice, and involvement in policymaking.

A burgeoning empirical literature suggests mechanisms of bottom-up accountability that can successfully improve service delivery. A number of studies show that increasing voter access to information increases the responsiveness of policymakers to their preferences (Besley and Burgess, 2002; Strömberg, 2004; Reinikka and Svensson, 2004, 2005). Voters also reward or punish policymakers when provided with information about their

performance, and these effects are especially large when there is a strong media presence (Ferraz and Finan, 2008; Chong et al., 2010; Banerjee et al., 2011). There is less evidence that directly involving citizens in governance improves policymaker accountability. While Díaz-Cayeros et al. (2014) find that participatory democracy improves access to public services, Olken (2007) finds that increasing grassroots participation in monitoring of public officials does not reduce the misappropriation of funds. And Mansuri and Rao (2004) find that community-driven development projects are generally dominated by local elites and do not target the poor.

Less is known about the impacts and effectiveness of mechanisms of top-down accountability. This is an important omission given that top-down accountability might be enforced using the results of government audits frequently used in bottom-up mechanisms, but without the added costs of civic education and the added threat of local policymakers obscuring audit outcomes (Humphreys and Weinstein, 2012). Most related to this research project, Olken (2007) finds that a higher probability of a central government audit of public expenditures significantly reduces corruption. While audits carry the possibility of sanctions by the central government—and thus would appear to represent a top-down accountability mechanism—Olken notes that “this is quite rare; more important, the results of the audit were read publicly to an open village meeting by the auditors and so could result in substantial social sanctions.” In no villages were the results of audits concealed from villagers, and thus all findings may be due to bottom-up accountability.

A lack of accountability among local government officials is a key constraint on economic growth in Mali. Mali is a politically decentralized country with 703 commune governments, each led by an elected commune council with between 11 and 45 representatives (6 on average). Each year, communes receive a significant share of their budget from the central government, but fund mismanagement is a perennial problem. Corruption and inefficiency among Mali’s commune governments have been measured and proven by external audits by the European Commission (2010, 2012, 2014). These audits demonstrate that much

of the funding that is transferred to communes is lost due to mismanagement, including improper procurement practices, overbilling, faulty construction of development projects, and insufficient project maintenance.

Mali's central government is keen to reduce inefficiencies in public expenditures while also permitting local commune governments autonomy in a number of public expenditure decisions. Donors share this concern about inefficiency and fund mismanagement. In particular, our sources suggest that this may have contributed to donor reluctance to give larger or more steady streams of funds. If a lack of perceived policymaker accountability is driving out donor funds, it is likely to reduce economic growth and the quality of Mali's business climate even further. As such, reducing corruption and mismanagement of public spending are critical if Mali is to experience sustainable economic growth in the coming years.

Our research project has several goals, motivated by both the pressing policy issue of how to make public expenditure more effective in a decentralized setting as well as gaps in the existing empirical and theory literatures on this question. First, given limited evidence on the potential of top-down accountability mechanisms, our primary goal is to estimate the impacts of a strong, top-down accountability intervention on public expenditures and corruption in the context of Mali. Specifically, we will test the impact of notifying local leaders that they will be audited and their relative performance on the audit shared with heads of key government agencies and donors—with awards for the best performers and the possibility of relative performance being used in the allocation of future awards. This will provide a robust test of whether one form of increasing intra-jurisdictional competition in a decentralized setting can improve public expenditure outcomes and reduce corruption.

As a secondary contribution, we will examine whether a more costly, bottom-up intervention has a significantly greater impact than does a top-down intervention. Little is known about the relative benefits of bottom-up versus top-down mechanisms in an electoral context. This knowledge gap persists in part because it is difficult to design an intervention that involves both types of mechanisms being implemented in a way that permits acceptable

comparisons. Olken and a recent study by [Sanchez de la Sierra et al. \(2015\)](#) do examine relative impacts of top-down and bottom-up accountability interventions within the same country context. An important difference of our study is the analysis of the behavior of formally elected local officials rather than village committees or elites.

A third aim of our project is to inform a theory-building exercise that better formalizes predictions of how these two accountability mechanisms perform relative to one another, and whether and how this relative performance is conditioned by the electoral context. Namely, we discuss the importance of political competitiveness, electoral volatility, and voter information to the ability of either type of mechanism to yield dividends. Our theoretical discussion makes the novel prediction that the level of political competitiveness should condition the relative effectiveness of top-down and bottom-up accountability mechanisms. Because the political competitiveness and electoral volatility are difficult to measure and measurement is particular to country context, our data collection will include ways of validating measures for use in future research.

2 Theory

In this section, we informally describe a simple theory that motivates two predictions which we will test with our experimental design. The theory, however, remains incomplete with respect to extensions we find interesting and important. In particular, how accountability relationships are conditioned by intervening factors such as information and competition is underdeveloped, and a dynamic extension has yet to be considered. We intend to inform future assumptions and modeling choices with data we collect in the course of the field experiment. As such, this informal theory will eventually be formalized with the aforementioned extensions, with the aim of testing additional predictions in future research.

As in standard moral hazard models of voter control ([Barro, 1973](#); [Ferejohn, 1986](#)), an elected (local) politician is assumed to be motivated both by being in office and by any

personal benefits he can obtain through the misappropriation of public funds for private purposes (corruption). He is in office in a first period, and faces an election in the next period which will determine whether or not he remains in office. If he loses in the second period, he can continue to stand for election in future rounds. The higher the incumbent's budget, the more he is able to misappropriate, which increases the potential payoff from holding office. Misappropriation leads to fewer and less productive public investments, and is therefore not in the interest of voters. The politician may attempt to influence voters through direct transfers to them, though these must come out of his own misappropriation. A politician chooses his type to maximize his expected utility. The higher it is, the less likely he is to misappropriate.

2.1 Conditioning parameters

The electoral context in which the incumbent faces re-election in the second period is defined by three parameters: political competitiveness, voter information, and electoral volatility. [Ferejohn \(1986\)](#) argues that politician accountability is increasing in the number of parties. The logic is that with fewer parties, an incumbent can misappropriate a lot in the first period, lose in the second period, but have a higher probability of getting re-elected in later periods. Because having more parties reduces the probability of holding office (for the out of office party), the loss of office is more consequential. Valuing holding present office more, the incumbent will be more sensitive to accountability pressures from voters in order to get reelected. While we do not dispute this logic, there is a competing factor we observe in some contexts – electoral volatility – that may interfere with the original prediction.

Defined as the share of voters that switch parties across elections, electoral volatility may make reelection in the second period so unlikely that the incumbent is insensitive to any accountability pressures. Such volatility can arise from voters engaging in party-switching among a stable set of parties, the entry and exit of parties, or both. Whereas party competition, as defined above, conditions the likelihood of an out-of-office party getting reelected,

electoral volatility conditions the likelihood of an incumbent ever getting reelected – independent of the incumbent’s choice of type or level of misappropriation. And, whereas incumbents can form expectations about how different party configurations will impact their chances of winning back office in subsequent rounds, electoral volatility increases the uncertainty with which incumbents can predict how their actions today will impact prospects of officeholding in the next period. Put another way party competition affects probabilities of future officeholding while electoral volatility affects the uncertainty over that future probability.

Electoral volatility is particularly high in Mali – in fact, Mali is ranked last in all of Africa by [Riedl \(2014\)](#) in her index of party system institutionalization. If we think of local government elections as each having a unique electoral volatility measure, then a locality with a large number of competing parties could be an indicator of either high levels of political competitiveness, high levels of electoral volatility, or both. Given that these two conditioning variables have predicted effects on accountability that go in opposite directions, it will be important to find distinct measures of each that do not confound the two constructs and yield biased conclusions. One aim of our research design is thus to investigate appropriate measures for these constructs in our particular context. Another, will be to better understand the drivers of electoral volatility. Anecdotal evidence suggests volatility is driven by politicians switching parties due to personal conflicts, independents creating new parties opportunistically, or external shocks to party income that make some parties temporarily better or worse at buying votes. Another aim of the research design will be to better understand the determinants of electoral volatility in our case.

A straightforward prediction from the existing literature with respect to voter information is that an electoral environment that provides voters greater access to information about politician performance will yield more accountable politicians. The incumbent knows that voters are receiving a better signal of performance and so are less likely to misappropriate to ensure future reelection relative to a context in which voters receive a weak signal of performance. While the parameters of voter information and political competitiveness are

likely endogenous to each other, and to the level of public service provision, we will treat them as initial conditions or parameters of the model when evaluating short-term impacts, but will later consider how they may change endogenously in a dynamic extension.

2.2 Reducing misappropriation

There are two distinct mechanisms that could be used to reduce misappropriation, or incentivize politicians to choose a higher type. First, a top-down accountability intervention audits a politician and reveals his misappropriation level to *the central government and donors*. The central government and donors use this information as follows: if the politician is of a lower type than the average politician in the country, then they will reduce the amount of funding that they give to his local government in the next period, and re-direct that funding to better-performing local governments. As a result, this intervention can affect the payoff the politician receives from holding office.

Second, a bottom-up accountability intervention audits a politician and reveals his type to *voters*. Voters can use this information as follows: if the politician's type is below either the expected type of competitor(s) (if they exist) or a certain threshold of performance voters choose to maximize politician performance in equilibrium,¹ then they will vote the incumbent politician out of office in the next period. If it matches that of competitor(s), then the voter will randomly select an individual from among the incumbent politician and the full set of competitors in the next period. If it is higher than that of competitors, or if there are no competitors, then they will vote for the incumbent politician in the next period. As a result, this intervention can affect the politician's probability of holding office in the next period.

¹Fearon (1999) describes the differences in sanctioning and selection strategies voters may use to discipline the incumbent. We remain agnostic about mechanisms for now.

2.3 Theoretical expectations

We have several theoretical expectations we will later formalize. For now, we state and then informally talk through the intuition of each, in turn:

Proposition 1. Both a top-down intervention to encourage inter-jurisdictional competition and a bottom-up intervention to inform voters about leader performance will improve public service delivery outcomes among low politician types.

Both of these mechanisms provide more information to an agent responsible for controlling the level of some aspect of the politician's utility function: the ability to stay in office by being reelected (voters have control over this) and the ability to obtain a large budget from which to misappropriate in the future (the central government and donors have control over this). Each of these agents can choose to ignore the new information that they received from a top-down or a bottom-up accountability intervention, so they cannot be made worse off. However, they may strictly better off if they can use this information to induce the politician to act in their best interest.

Proposition 2. The effectiveness of a bottom-up accountability mechanism at reducing misappropriation should be

1. increasing in levels of political competitiveness – a viable outside option makes the threat of sanctions from voters bind and the less likely an out-of-office politician is to get elected in the future makes the value of present office greater;
2. decreasing in electoral volatility – too volatile an electoral system reduces the attractiveness of future political office beyond the payoffs of misappropriating everything today; and
3. decreasing in prior levels of voter information – more knowledgeable voters will learn less from new information and thus recalibrate their evaluation of the incumbent less relatively to more poorly-informed voters.

We expect incumbent politicians to behave optimally and maximize their own utility rather than a social welfare function. If we take the simple case in which an incumbent politician makes a discrete choice between being a low type or a high type, we can predict likely behavior. An incumbent politician subjected to a bottom-up intervention will choose to be a high type if and only if being a low type will not get him reelected, being a high-type will, and reelection offsets the losses from reduced misappropriated income (rents) due to being a high type. This is unlikely to be the case in two settings. First, it is unlikely to be the case if there is little political competition or no viable opposition party; in that case, the incumbent will not be voted out of power by choosing to be a low type, so being a low type is a dominant strategy, and a bottom-up intervention ineffectual. Second, it is unlikely to be the case if there is extreme electoral volatility; then, the incumbent is less certain that choosing a high type will translate into high reelection prospects in the second period.

In the more realistic case in which an incumbent politician being subjected to a bottom-up intervention chooses his type from a continuum, we expect the following. A politician will either (a) misappropriate everything and not be reelected in the second period, (b) misappropriate everything and still be reelected in the second period (will only occur if there is no competitor), or (c) misappropriate just enough to make voters indifferent between leaving him in office and electing a competitor, and will accordingly remain in office. Both (a) and (b) will yield the highest levels of misappropriation, while (c) will yield relatively lower misappropriation. We predict that (a) is most likely when there is high electoral volatility, (b) is most likely when there is low political competitiveness, and (c) is most likely when there is high political competitiveness and low electoral volatility. This again supports Proposition 2.

The threat of exposure inherent in a bottom-up intervention also depends on the extent to which voters value public service delivery. While we often assume they do, voters may also value targeted transfers that are funded, in part, by politician corruption. Thus, a bottom-up mechanism will only shift a politician's behavior to the extent he believes voters

will change their electoral behavior in light of the news. We will accordingly incorporate side payments into our formal model.

Now, we consider the actions of an incumbent politician subjected to a top-down accountability intervention. The politician knows that his misappropriation in each period carries a threat of having a smaller pie from which to extract rents in future periods. We consider a case in which the cycle of the audit and subsequent transfer allocation is shorter than the electoral mandate. This follows the case of Mali in which resources are transferred to local governments annually, but local politicians are only elected every five years. To maximize rents over the five-year period, the corrupt politician will only reduce corruption in each year to the point that the net present value of cumulative rents extracted in the five years is greater than that of extracting everything in the first period and receiving nothing in latter periods. There should thus exist a formula for performance-based transfers that reduces (but does not eliminate) politician corruption, which is consistent with Proposition 1. Electoral considerations in the last period can have effects in both directions. Extracting rents for the purposes of vote-buying can induce higher rates of corruption in spite of the incentives generated by inter-jurisdictional competition. However, if voters value public service delivery over targeted transfers, the politician may prefer to reduce his own rent extraction.

Because we do not have strong expectations about what happens in the last period of the incumbent's mandate, and because the conditionality of the incumbent's behavior on electoral context in the last period is similar to that in the bottom-up case, we turn our attention now to the incumbent's behavior in the first four year's of the mandate. During this part of the mandate, the incentives to reform that inter-jurisdictional competition places on the corrupt incumbent should be unconditional on electoral context. Combining this with insight from Proposition 2, we now propose:

Proposition 3. In an environment of low political competitiveness, a top-down mechanism will be more effective at reducing misappropriation than a bottom-up mechanism.

In the aggregate, however, we do not have clear theoretical expectations about whether

bottom-up or top-down mechanisms will perform better in the short-term. We accordingly take this as an important empirical question, especially since bottom-up mechanisms tend to be more costly than top-down ones. If the effects are indistinguishable, the latter should dominate the former for policymakers. We will thus interrogate the following proposition:

Proposition 4. A bottom-up mechanism of accountability will not have a greater average impact on public service delivery than will a top-down mechanism.

2.4 Developing and testing theoretical expectations

The principal output of this project will thus be an experimental study testing one observable implication of Proposition 1 – the impact of a top-down mechanism of accountability on local public expenditures, with some additional analysis of whether or not these impacts are appreciably larger (or smaller) than those of a bottom-up mechanism as a test of implications of Proposition 4. The hypotheses directly tested by the experiment are described in Section 6.

While the research design is not set up to robustly test the implications of Propositions 2 and 3, underlying variation in political competitiveness (and possibly volatility and voter information) across communes will allow us to garner some qualitative insights into their plausibility. One additional project output will be a qualitative study of where—that is, under which levels of political competitiveness—top-down and bottom-up accountability mechanisms are most likely to be successful. Another will be understanding the drivers of electoral volatility and its impacts on politician time horizons, as well as finding reasonable proxies that can be used to measure it. We will use this evidence as input into developing a formal model of politician behavior under these two very different mechanisms of accountability along the lines of what we describe informally above. Section 5.4 describes the qualitative study and how it will interrogate our model’s assumptions and further assess the plausibility of our hypotheses.

3 Malian Context

Mali’s decentralization of political authority to 703 commune governments allows us to carry out a governance experiment across similar cross-sectional units. Below, we describe how local governments are elected and how they raise resources to spend on public service delivery.

3.1 Electoral Context

Since 1999, local government councilors have been elected with five-year mandates. In each election, parties submit closed candidate lists and seats are accorded to parties based on proportional representation. The number of seats on a commune council ranges from eleven to forty-five, based on population, and in every commune an average of six parties compete for those seats. For the first three of these elections, the mayor was elected indirectly from among the commune councilors. The 2014 elections, which have been postponed twice due to conflict in the North of the country (most recently in October 2015), will institute direct mayoral elections for the first time.

There are a few regularities in Mali’s local elections that are worth noting. First, as evidence of the importance of coalition politics, when the largest party has a plurality of seats rather than a majority, the mayor is more likely to come from one of the smaller parties. Second, as possible evidence of electoral volatility, incumbent mayors won in only 29% of communes in 2009. In this context of high volatility, [Gottlieb \(2015\)](#) finds evidence that commune councilors from different parties sometimes have incentives to collude to jointly reap the spoils of office rather than compete and drive down the size of the pie, leading to fewer and less productive public investments. Such behavior is especially pervasive where there is a particular type of low political competitiveness. In a commune in which all active political parties are represented on the commune council (one-third of cases in 2004), there is no “out party” that—armed with information that is superior to that of citizens about the degree of corruption—has both the knowledge and incentives to expose corrupt commune

council members. While [Gottlieb \(2015\)](#) uses a regression discontinuity design to show that this form of low political competitiveness leads to lower provision of local public goods, she also finds that more standard measures of competitiveness such as the effective number of parties are also correlated with higher public goods provision, consistent with the implications of Proposition 2.

The change in electoral rules with respect to the direct election of the mayor may reduce incentives for collusion and increase the likelihood that minority parties on the council will serve as a real opposition. Thus, the best measure of political competitiveness in the context of the 2016 local elections is an open empirical question. For future commune councils, there may be different dimensions of competitiveness that matter more to the time horizon of elected officials and thus the value of holding office; e.g., the likelihood they will maintain seats on the commune council and the likelihood they will maintain or obtain the mayoral seat.

3.2 Intra-governmental Transfers

Public policy in each commune is made by a commune council. Councils receive about \$30,000 per annum for use on community development projects from the Agence Nationale d'Investissement des Collectivités Territoriales (ANICT), which represents a large share of their budget. Communes include an average of 18 different villages, and these project funds typically allow them to select one village in which to build a community development project per annum. Our project will consider how to make commune leaders more accountable in their expenditure of these development project funds.

ANICT project funds are currently allocated on the basis of a formula that takes into account population size, remoteness, and level of development. Fund allocation amounts are announced annually in May/ June. By the Fall, commune leaders submit proposals for projects to be funded with these monies. If approved, funds are typically transferred to the commune by the end of the year.

3.3 Partners

Our partners will include both the ANICT and the Direction Générale des Collectivités Territoriales (DGCT), which is part of the national decentralization ministry and makes and implements policy pertaining to decentralization. We will work with the head of ANICT's administrative council and DGCT's deputy director for local development. Both recognize existing inefficiencies and are committed to furthering initiatives to improve local governance. We will also partner with Innovations for Poverty Action (IPA) for the implementation of the audit and other data collection. Finally, we will continue to be in regular contact with several important donors that we have already approached and who are excited about the policy lessons that may be gleaned from our research. These include the European Union, UNCDF, and the bilateral donors of Belgium and Luxembourg.

The ANICT is keenly interested in adding performance measures to its existing formula for spatially allocating funds. Of his own initiative, the councilor to the Director General of the ANICT, our main contact within the organization, proposed the idea to the governing council of the ANICT in June 2015. In particular, he wants to add to the existing formula used to allocate funds: i) an indicator for the quality of the procurement process for the preceding year's development project, and ii) an indicator for the extent to which citizens were involved in the decision-making process. These suggestions are more theoretical than practical at this stage because the government does not presently have the financial or human capacity to measure such indicators for every commune government. One of the contributions of our project would be to test cost-effective and feasible mechanisms to collect the kinds of data necessary to move the ANICT to a partially performance-based funding mechanism.

In addition to the government's interest in performance-based financing, donors, who together provide about 90 percent of the ANICT budget, are also interested in our proposed research project. Our contacts in the European Union Commission, the Belgian and Luxembourg bilateral donor agencies, and the United Nations Capital Development Fund (UNCDF) all expressed support for this idea. The Belgian bilateral was the most enthusiastic given

previous experience and success with performance-based financing in other countries. They offered as a potential policy change, contingent on the results of our experiment, the possibility of conditioning their 2017 contribution to the ANICT on the adoption of performance measures in the ANICT’s formula for allocating funds.

After the European Union’s discouraging audit of ANICT expenditures across a representative sample of communes (2010, 2012, 2014), they have been working with the DGCT to devise regular supervision and control mechanisms for commune governments. The recommendation of one of their consultants was to more intimately involve sub-prefects in the control function of elected local governments. By law, sub-prefects are the appropriate deconcentrated state authority that oversee the activities of commune governments. However, the sub-prefects currently have little capacity and motivation to do their job well. As such, the DGCT is also interested in learning from our project about feasible ways to sustainably implement oversight and monitoring of commune governments.

4 Research Design

4.1 Description of Treatments

We will carry out a field experiment with three treatment groups:

1. Control: no intervention
2. Treatment 1 (T1): Top-down accountability treatment (i.e. inter-jurisdictional competition treatment)
3. Treatment 2 (T2): Bottom-up accountability treatment (i.e. voter control treatment)

All treatment arms will receive an audit, though only T1 and T2 will be told in advance that an audit will occur. The benefit of a pure control is estimating the impact of treatment relative to the status quo. [Sanchez de la Sierra et al. \(2015\)](#) find that when communities are forewarned of an audit by explicitly told the audit information will not be shared, con-

stituents still have strong priors that the audit information will be shared, thus decreasing the likelihood of being able to detect an effect of treatment.

The treatments are informational in nature. The variation we exploit is the differential effect on commune performance of being warned about whether and how audit results will be disseminated. Our primary outcomes will be measured using audit results, though we will additionally employ as outcomes the responses to several household survey questions. While each treatment group's audits will be handled differently – disseminated (or not) to different audiences – this only occurs after the effects of treatment are measured.

In T1, commune leaders will be notified that they have been selected to be audited and that audit results will be shared with the central government and major donors. They will be informed that commune performance is assessed relative to other communes – and receive a list of the communes in their treatment group against which performance will be compared. They will be read a script by project staff and be provided a letter stating:

- (a) The approximate date and content of the audit (about one year from receipt of the funds)
- (b) That audit results will subsequently be presented to a high-level meeting of heads of key government agencies (ANICT, DGCT) and donors (to be listed by name)
- (c) That a commune leader from each of the top 5 performing communes (about the top 3 percent) will be invited and hosted for a visit to Bamako, be formally recognized for their superior performance, and have an opportunity to meet with donors and the federal government about their needs; and
- (d) That the federal government and donors may use this relative audit performance information in the future to determine allocations (commune leaders will be informed of any stated, specific donor interests in doing so in the future).

This will allow us to test whether commune governments can be induced to compete with one another for recognition and potential future increases in funds due to good performance.

T2 involves a more traditional, bottom-up accountability treatment. In these communes, we will publicize the results of the audit to commune residents in town hall meetings and by posting them in a public place, notifying residents of how their leaders compare to other

leaders in their treatment group. Through a similar script and letter to the one described above, commune leaders will be notified of the date and content of the audit, but this time how audit results will be announced at a town hall meeting and posted in a public space.

4.2 Experimental Design

Due to the potential for a negative interaction effect between the two treatment arms, or a ceiling effect in which a combination of treatments would not much surpass the effect of any single treatment, a factorial design is not used. Instead, each commune is randomly assigned to only one treatment condition.

Myriad studies have examined the effect of bottom-up mechanisms on government accountability with mixed, but mostly positive results. We thus do not aim to directly test this relationship in our context. Instead, we prioritize evaluating the average impact of top-down mechanisms, which are less often studied. To detect a reasonable average treatment effect of T1 relative to control, our power calculations in Section 4.4 suggest we thus need 168 communes in each of the T1 and control groups.

We do not anticipate having a sufficiently large sample of communes (due to security concerns) to additionally test the causal impact of T2 relative to control using commune-level audit outcomes. Instead, we will use a cluster-randomized design with individual-level outcomes (collected as part of a household questionnaire) to detect whether this more costly, bottom-up intervention has a significantly greater impact on survey outcomes than does a top-down treatment. Our power calculations require 79 communes to be assigned to T2, yielding a cumulative sample size of 415 communes.

The total number of communes in the four southern Malian regions in which it is currently feasible to intervene given current security constraints is 502. So that our experimental sample is at least representative of these southern regions, each commune in these four regions will have about a 33% probability of being assigned to control, a 33% probability of being assigned to T1, and a 16% probability of being assigned to T2.

Given the importance of political competitiveness to our conditional hypotheses, we will stratify treatment assignment on whether or not the commune council has an “out party” in the 2016 local elections, e.g. whether a party ran but did not win seats on the council. And, given the important differences in the stakes of commune government and the information resources across rural and urban settings, we will also stratify treatment assignment on whether the commune is urban or rural. To account for potential differences across the four southern regions, elements of each block will come from the same region.

4.3 Spillovers

Though the result of treatment spilling over into control communes would bias us against finding an effect of treatment – and thus is of less concern for making robust causal inference – we discuss several reasons why we expect spillovers to be relatively limited. First, at the time each treated unit is informed of the treatment, they will be given a list of everyone in their treatment arm. This will make them aware that not all communes will receive the same treatment, and will likely head off discussions with other communes leaders not in their treatment arm about whether or why they are not receiving the same treatment. Control communes, however, will not be advised about their status as a control commune.

Second, and most importantly, as the audit comprises our primary outcome measure – and the secondary measurement instrument, a household survey, is carried out at the same time as the audits – our outcomes will be measured before any commune leader is able to share audit results with each other or their citizens (something that might occur if they like the audit results). As a result, an important element of T2—knowledge that citizens will be made aware of the audit results—should not occur in control and T1 communes. Indeed, in the case of control communes, commune leaders will not even know that audits are to occur, so therefore should not change their behavior due to being in our sample. While we cannot rule out that leaders of T1 communes will form the plan to share their audit results if they are good, we will at least be surveying households before they actually do so.

Finally, given that the vast majority of our sample communes will be rural and remote (667 of Mali’s 703 total communes are rural, and they have an average population of only 14,000), and given low internal mobility in Mali, we do not anticipate that contacts between neighboring communes will be frequent. It is true that commune leaders in a given locality have more opportunities to exchange with each other than does the average citizen. They may encounter each other at district or regional meetings, or be connected through political party or other associational networks. These political elites are often better off economically and thus more mobile and well-connected. One worry is that commune leaders within a given treatment group could collude to remain at a low, or status quo level of performance. However, the sampling strategy ensures that members of each treatment group will be evenly distributed across all of Mali’s southern region, making it nearly impossible for such coordination to succeed. Further, in the treatment of greatest interest to us – the top-down, T1 treatment – collective action difficulties and other coordination problems should break any collusive agreements.

4.4 Power Analysis

Commune-level analysis

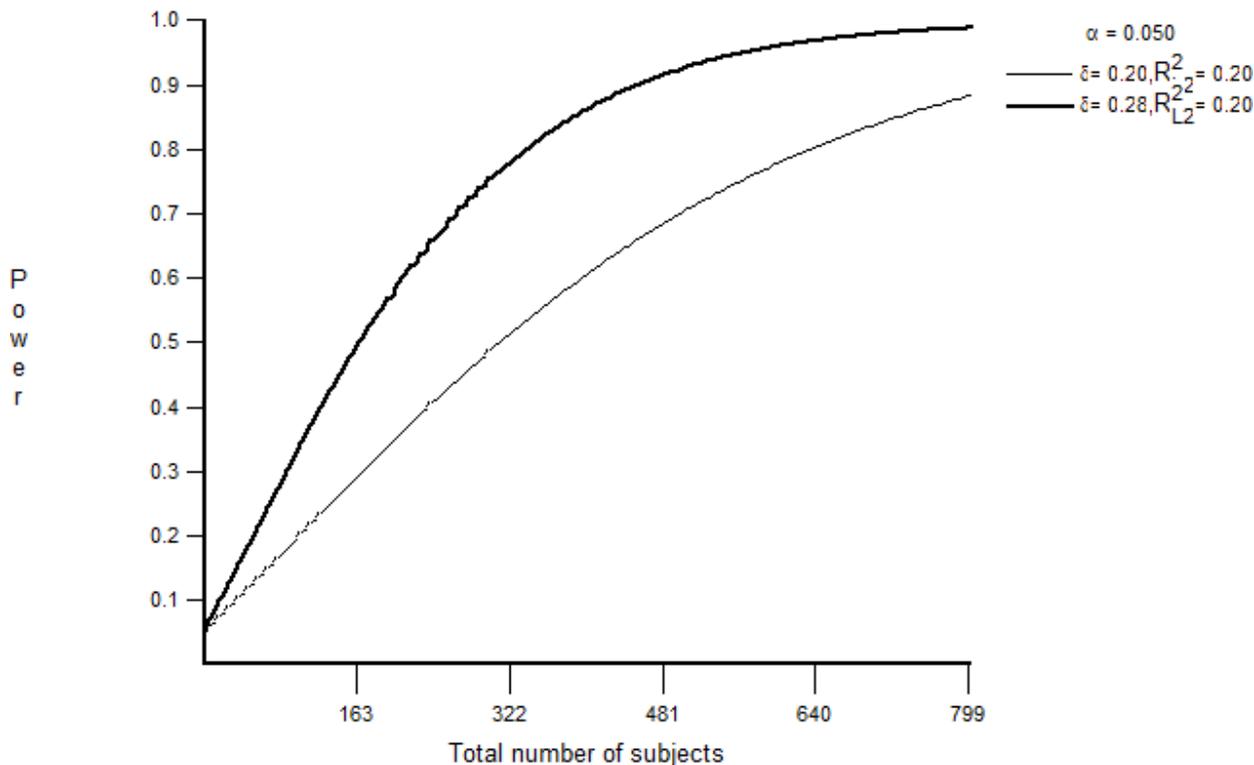
Building upon two prominent studies in the literature that evaluate the effect of audits on local government behavior, we use a minimum detectable effect size of 0.275 standard deviations (the midpoint of the two from the literature). [Olken \(2007\)](#), found that treatment reduced missing expenditures in local budgets by 0.25 standard deviations. [Ferraz and Finan \(2011\)](#) found effect sizes of up to 0.3 standard deviations for their audit outcomes.

We also include covariate adjustments in our power calculation to reduce variance in the dependent variable, also conserving power. Our primary covariates will include a commune wealth index, remoteness, previous experience with governance-focused NGOs, and fixed effects for the 8 blocks created by crossing region (4 total) and political competitiveness (2 levels). Given previous studies in Mali ([Gottlieb, 2015](#)), these covariates can explain up to 30

percent of variation in local public expenditures. We use a more conservative 0.2 covariate adjustment.

Figure 1 shows that for a power of 0.8 and a statistical significance level of 0.05, we would need 336 communes total, or 168 communes in each of two treatment arms.² For three treatment arms, we would need a minimum of 504 communes, or more than 700 accounting for multiple comparison corrections.

Figure 1: Power analysis at the commune level



The total number of communes in the 4 southern Malian communes in which it is currently feasible to intervene given current security constraints is 502. Thus, we do not have sufficient power to conduct a rigorous test with three arms. Therefore, we will focus our main commune-level test on the more novel intervention – the effect of increasing intra-jurisdictional competition on local public expenditures.³ For this, we need a minimum of

²Power calculations are conducted using Optimal Design http://sitemaker.umich.edu/group-based/optimal_design_software.

³There are existing studies in the literature that already test the impact of an informational treatment that stimulates bottom-up accountability.

336 communes, or 168 in control and 168 in T1.

Individual-level analysis

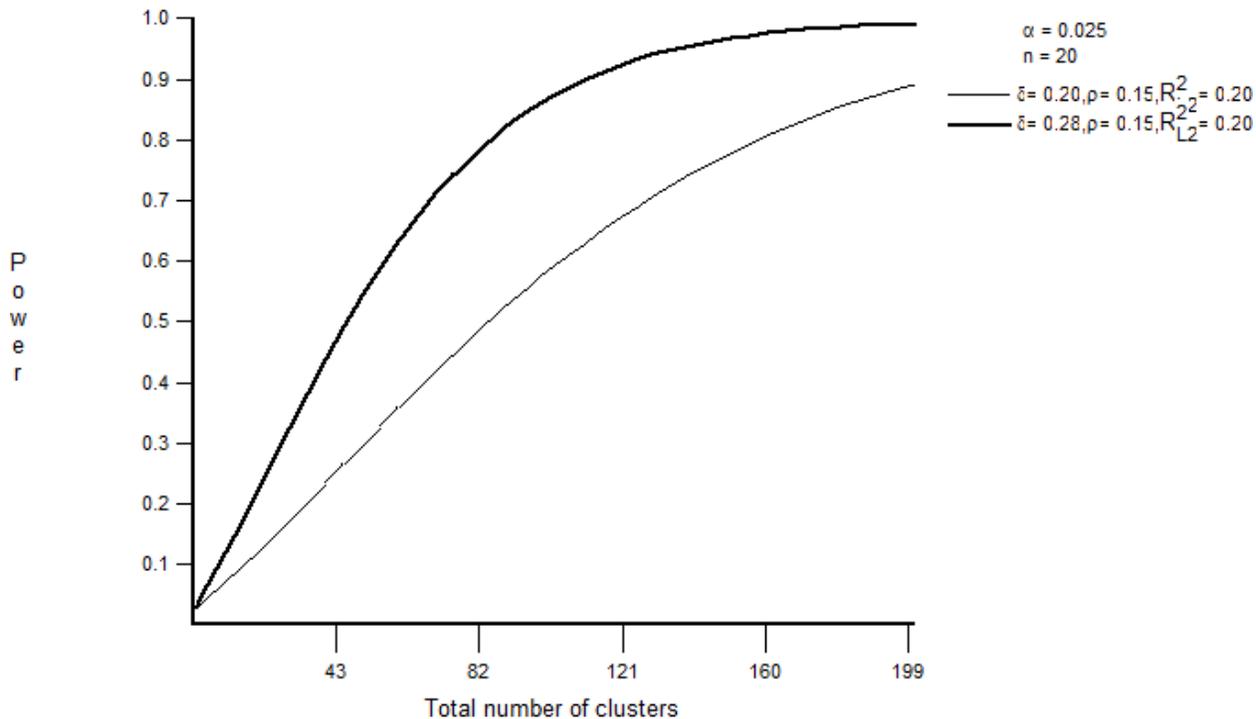
Our second priority is to compare the effect of a top-down accountability treatment to that of a bottom-up accountability treatment. We do not expect that the difference in impacts between the treatments will be large. A standard in the empirical political science literature is that, for a mean difference to be substantively meaningful, it should be at least 0.2 standard deviations. Thus, we examine our power requirements if we were to be able to detect a difference as small as 0.2 standard deviations at the commune level. Figure 1 shows we would need 630 communes total, or 315 in each of two arms, to make this sole comparison.

Given the infeasibility of rigorously testing the above hypothesis – that there is no substantial difference between T1 and T2 – at the commune level, we will use the power-saving technique of a cluster-randomized design, instead using outcomes at the individual level. Such outcomes include perceptions of government performance, information about public investments, participation in the decision-making process, etc. While survey-based outcomes are not as externally valid as a behavioral outcome, we find it sufficiently interesting to be able to compare T1 and T2 using survey measures.

Figure 2 shows the sample size requirements of a cluster-randomized design in which 20 individuals are surveyed in each cluster or commune. As before, we set our standard for power at 0.8 and we implement a covariate adjustment of 0.2. Because we will also compare the effect of each arm on control, we employ Bonferroni multiple comparison corrections and divide our alpha by 2. Finally, we assume a reasonably conservative intra-cluster correlation of 0.15. Given these parameters, to detect a minimum of a 0.2 standard deviation difference between T1 and T2, we need 79 communes in each of our three arms.

Because our commune-level analysis already requires 168 communes in each of Control and T1, we will randomly assign 79 from each arm to receive individual-level surveys. We will also add an additional 79 T2 communes to the 336 T1 and C, bringing our total sample

Figure 2: Power analysis for a cluster-randomized design: outcomes measured at individual level but treatment at the commune level



size for the interventions to 415 – a feasible size given security constraints. Because all T2 communes require an audit for the intervention to be credible, we will conduct the audit in all of the 415 communes in the sample. We can thus compare results from the 79 T2 audits to the 168 in T1 and C, but would only have power of about 0.6.

5 Measurement Instruments

5.1 Audits

We will carry out audits of the ANICT-funded projects in each of our 415 sample communes, approximately nine months to one year after funds are received. Shortly after the 2016 fund allocation amounts are announced (usually around May/June), treatment commune leaders will be told that audits will take place and when; no information will be given to control commune leaders. Data collection for the audits will be carried out jointly by qualified

Malians recruited and trained by IPA and sub-prefects who will also receive training and resources.

Our outcomes of interest measure the extent to which ANICT funds are spent in accordance with federal regulations. Following [Ferraz and Finan \(2011\)](#), we will consider three main categories of violations of federal regulations: fraud in procurement, over-invoicing of goods and services, and diversion of public funds for private gain. Because the first two may occur before funds are actually transferred, it is important that our intervention takes place right after the announcement of available funds, not after the actual funds transfer.

The audit will identify the percent of project expenditures in each of these three non-mutually exclusive categories. In an external audit of ANICT projects ([European Commission, 2010](#)), the first two categories were pervasive problems for local public expenditure; in a sample of 132 development projects awarded during 2001-2009, 90 percent had irregularities in procurement and a third involved overbilling.

We will also collect three indicators of efficiency: the percentage of the grant spent, timeliness of project construction, and the extent to which the project contract is carried out. The findings of the audit will be presented in relative terms using scores from a corruption index and an efficiency index. The comparison group will always be the other communes in the same treatment condition, which are communicated explicitly to the group. In addition to the composite scores on each index which will be communicated to the appropriate audience in each treatment, information about the component scores for each index will also be provided.

Specifically, data collected will include physical copies of the request for bids, the actual bids received, the commune's scoring of the competitive bids, the project contract with the winning firm, invoices from the winning firm, and digital photographs (with GPS coordinates) of physical infrastructure to record progress on construction and amounts and quality of materials used. A team of engineers will then use these data to conduct the audit, also using the results of a price survey.

5.2 Household surveys

To complement the commune-level outcomes from the audits, in 79 communes of each type (Control, T1, and T2) we will also collect data at the individual-level in the village in which the ANICT-funded infrastructure project was built. The survey will take place at the same time as the audit, such that audit results cannot sway the opinions of respondents. In each village, we will survey the head in each of 20 randomly-selected households to learn about: 1) their perceptions of the commune government, 2) the transparency with which the ANICT project was conducted, 3) their level of information about the ANICT project, and 4) their level of satisfaction with all aspects of the project, from project selection to completion and delivery.

In addition to providing greater power than outcomes collected at the commune-level, these individual-level surveys will also allow for a manipulation check to determine whether our information treatments were properly administered. We will ask respondents if they were aware that an audit would take place, whether and how results of that audit would be disseminated, and to whom the performance of their commune government would be compared. Finally, if voters care more about one of the two indices – corruption vs. efficiency – we should observe that the voter-preferred index is more highly correlated with the treatment effect than the less-preferred index. As such, we will measure voter priorities with respect to these two dimensions.

5.3 Leader interviews

Prior to providing the experimental information to treated groups, we will conduct a survey of the mayor of each commune as well as an elected member of the second largest party. To learn about the types of political competitiveness that affect leader time horizons and the value of holding office, we will ask each respondent the perceived likelihood of their party winning seats on the council in the next election and the likelihood of their party holding the mayor's seat. We will also ask about perceived political competition in the commune,

evidence of performance-based campaign messages, perceptions of what voters prioritize when electing councilors, and the extent to which parties cooperate or come into conflict on the council in between elections. To gain insight into variation in electoral volatility, we will ask leaders about their perceived likelihood of winning reelection, and then, most importantly, about their confidence level for that guess. We will similarly ask the extent to which they were surprised by the electoral outcomes of the recent election.

5.4 Qualitative data collection

We will additionally carry out a qualitative study to both inform our understanding of the mechanisms underlying our experimental results and to examine institutional and political factors that condition the effectiveness of a top-down or bottom-up accountability intervention. Our qualitative study will use the same insights and discontinuity of [Gottlieb \(2015\)](#) to select a sub-sample of our sample communes that narrowly have vs. narrowly lack a viable “out party,” based on data from the 2016 local elections. In each of the two groups, we will have communes that received a top-down accountability treatment, communes that received a bottom-up accountability treatment, and control communes—likely two of each, providing a total sample of 12 communes. We will carry out both in-depth interviews and focus groups in each selected commune.

In-depth interviews will ask commune leaders, individuals engaged in contracting public works funded by ANICT, and individuals engaged in building those public works about their experiences with the most recent ANICT-funded project. As a manipulation check commune leaders will be asked the extent to which they perceived their performance this year to affect funding in future years. Commune leaders will also be asked to describe the recent local elections, including about the level of competitiveness, the extent of turnover, and details about the campaigns. Focus groups will be held both with a group of citizens who frequently use those public works (intended beneficiaries) and with a separate group of citizens that does not or rarely uses them (non-beneficiaries). These citizens will be asked about the full

process, from the the selection of a development project to the end of construction.

6 Testing Empirical Implications

6.1 Main Hypotheses

The following two hypotheses will be robustly tested by our experimental design:

Hypothesis 1. Inducing local governments to compete for scarce central government and donor funds—that is, top-down efforts to encourage inter-jurisdictional competition—will improve audit scores among T1 communes relative to control.

By construction, T2 may lead to the public being more informed about the ANICT project than in T1. Anticipating the public exposure, the commune council may choose to be more transparent in their activities prior to the actual announcement of audit results. This may also lead to higher levels of perceived government performance in T2 communes relative to T1 communes. However, satisfaction with the actual project itself – both the quality of the infrastructure and the process by which it was constructed – should not be tainted by these potential council activities and be the best proxy for quality of actual public service delivery.

Hypothesis 2. Stimulating bottom-up pressure by publicizing the results of the audit to commune residents in town hall meetings and by posting them in a public place will not have a greater impact on public satisfaction with the 2016 ANICT project than will a top-down mechanism.

6.2 Conditional Hypotheses

We will also examine the following conditional hypotheses, but will not necessarily have the power to generate robust inferences.

Hypothesis 3. The effectiveness of a bottom-up mechanism on reducing misappropriation will be increasing in the number of political parties that ran in the most recent local election.

Hypothesis 4. Among communes where there is no out party, or viable opposition, T1 will have a greater impact than T2 on public service delivery than where there are greater levels of political competitiveness.

6.3 Econometric Specification

We will estimate the following empirical specification:

$$a_i = \beta_0 + \beta_1 T1_i + \beta_2 T2_i + \gamma \mathbf{X}_i + \alpha_j + \epsilon_i \quad (1)$$

where i indexes communes. We denote by a_i an outcome of interest from our commune-level audits, by $T1_i$ a dummy for assignment of commune i to T1, by $T2_i$ a dummy for assignment of commune i to T2, and by α_j block fixed effects. We denote by \mathbf{X}_i a vector of control variables for commune i , but will include controls only as a robustness check, not as part of the baseline specification. When we consider individual-level outcomes, i will instead index individuals, and a_i will be an outcome from our household survey data; we will also cluster standard errors at the level of treatment assignment – the commune. β_1 will indicate the effect of T1 relative to control while β_2 will indicate the effect of T2 relative to control.

7 Ethical Considerations

We will apply to the IRBs of both Texas A&M University and the International Food Policy Research Institute as soon as the project is fully funded and before further talks with the government or drafting of survey instruments. We will provide both IRBs with complete survey instruments once available.

A possible perverse consequence of the proposed interventions is that corruption is chan-

neled away from misspending and into bribes of higher level government officials to alter the results of the audit or the calculation of fund allocation. As one measure to prevent this, we will institute an online repository for audit results, including photographs, that will reduce the ability of audit results to be manipulated. We will also increase the transparency of the formula used to allocate funds, and continue to support the existing practice of publishing inter-governmental transfer amounts annually. Since these mechanisms will not necessarily prevent all fraud, we will implement random checks of the audit and future resource allocation to allow us to identify instances of any misbehavior.

8 Timeline

This project will be completed during March 2016 – April 2018, according to the following broad timeline:

- Local and regional elections: early 2016? (postponed from October 2015 due to conflict)
- Meeting with donors and government partners to agree on treatment language and audit instrument: June 2016
- Announcement of community development funds for 2016: June 2016
- Information treatment conducted (informing communes of future audit + treatment): September 2016
- Politician interviews conducted concurrently with information treatment: September 2016
- Mid-project audit of 2016 funds – March 2017
- Final audit of 2016 funds and household survey: September – November 2017
- Qualitative data collection: November – December 2017
- High-level meeting to publicize audit results – December 2017
- Data cleaning and analysis: January – April 2018

References

- Banerjee, A. V., S. Kumar, R. Pande, and F. Su (2011). Do informed voters make better choices? Experimental evidence from urban india.
- Bardhan, P. (2002). Decentralization of governance and development. *Journal of Economic Perspectives*, 185–205.
- Bardhan, P. and D. Mookherjee (2006). Decentralisation and accountability in infrastructure delivery in developing countries. *The Economic Journal* 116(508), 101–127.
- Barro, R. (1973). The control of politicians: An economic model. *Public Choice* 14(1), 19–42.
- Besley, T. and R. Burgess (2002). The Political Economy of Government Responsiveness: Theory and Evidence from India. *The Quarterly Journal of Economics* 117(4), 1415–1451.
- Besley, T. and A. Case (1995). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *The American Economic Review* 85(1), 25–45.
- Chong, A., A. L. De La O, D. Karlan, and L. Wantchekron (2010). Information dissemination and local governments’s electoral returns, evidence from a field experiment in mexico. *Unpublished paper, Yale University, New Haven, CT.*
- Díaz-Cayeros, A., B. Magaloni, and A. Ruiz-Euler (2014). Traditional governance, citizen engagement, and local public goods: evidence from mexico. *World Development* 53, 80–93.
- Dillinger, W. (1994). *Decentralization and its implications for urban service delivery*. World Bank Publications.
- European Commission (2010). 1^{er} rapport annuel de la Mission de Contrôle externe des investissements de Collectivités Territoriales au Mali. Ministère de l’Administration Territoriale et des Collectivités Territoriale. Bamako, Mali.
- European Commission (2012). Rapport annuel 2012 de la Mission de Contrôle externe des investissements de Collectivités Territoriales au Mali. Ministère de l’Administration Territoriale et des Collectivités Territoriale. Bamako, Mali.
- European Commission (2014). Extrait du 1^{er} rapport semestriel de 2014 de la mission

- de Contrôle externe des Collectivités Territoriales financés par l'ANICT. Ministère de l'Administration Territoriale et des Collectivités Territoriale. Bamako, Mali.
- Fearon, J. (1999). Electoral accountability and the control of politicians: Selecting good types versus sanctioning poor performance. In A. Przeworski, S. C. Stokes, and B. Manin (Eds.), *Democracy, accountability, and representation*, pp. 55–97. Cambridge: Cambridge University Press.
- Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice* 50(1-3), 5–25.
- Ferraz, C. and F. Finan (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics*, 703–1451.
- Ferraz, C. and F. Finan (2011). Electoral accountability and corruption:evidence from the audits of local governments. *American Economic Review* 101, 1274–1311.
- Gottlieb, J. (2015). The logic of party collusion in a democracy: Evidence from mali. *World Politics* 67(01), 1–36.
- Hayek, F. (1945). The use of knowledge in society. *American Economic Review* 35(4), 519–530.
- Humphreys, M. and J. Weinstein (2012). Policing politicians: Citizen empowerment and political accountability in uganda preliminary analysis. *Unpublished paper, Columbia University and Stanford University.*
- Mansuri, G. and V. Rao (2004). Community-based and-driven development: A critical review. *The World Bank Research Observer* 19(1), 1–39.
- Oates, W. E. (1999). An essay on fiscal federalism. *Journal of economic literature*, 1120–1149.
- Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in indonesia. *Journal of Political Economy* 115, 200–249.
- Reinikka, R. and J. Svensson (2004). Local capture: evidence from a central government transfer program in uganda. *The Quarterly Journal of Economics*, 679–745.
- Reinikka, R. and J. Svensson (2005). Fighting corruption to improve schooling: Evidence

- from a newspaper campaign in uganda. *Journal of the European Economic Association* 3(2-3), 259–267.
- Riedl, R. B. (2014). *Authoritarian Origins of Democratic Party Systems in Africa*. Cambridge University Press.
- Sanchez de la Sierra, R., P. van der Windt, and M. Humphreys (2015). Who’s watching? effects of monitoring on strategies for corruption: A field experiment in the congo. *Unpublished paper..*
- Seabright, P. (1996). Accountability and decentralisation in government: An incomplete contracts model. *European Economic Review* 40(1), 61–89.
- Strömberg, D. (2004). Radio’s impact on public spending. *The Quarterly Journal of Economics* 119, 189–221.