

OPTIMAL MONITORING AND BUREAUCRAT ADJUSTMENTS

Wendy Wong*

September 4, 2020

Abstract

Monitoring policies aiming to maximize deterrence of bureaucrat misconduct, under a budget, require accounting for bureaucrats' attempts to evade detection. This paper examines strategic responses of bureaucrats based on their expectations of the likelihood of an audit in India's employment guarantee program. Exploiting random assignment to audits, I find deterrence of misappropriated expenditures is stronger at higher than lower expectations of the likelihood of an audit; and bureaucrats substitute across time and method of misappropriation to evade detection. Using these results and applying a model of Bayesian persuasion, I solve for the optimal design of information communicated on the likelihood of audit. An analysis of counterfactuals shows that a "crackdown" policy, i.e. informing of audits in advance, would have persuaded bureaucrats to forgo misappropriating an additional USD 35m in expenditures (16% of average annual expenditures in the empirical setting) when compared to a policy where the occurrence of audits is unpredictable, i.e. no information is provided.

*University of Chicago, Harris School of Public Policy. Email: wendywong@uchicago.edu. I am grateful to my committee chairs, Canice Prendergast and Konstantin Sonin, and committee members, Chris Blattman and Luis Martinez, for continuous guidance. I am grateful to Ujjwal Pahurkar, Gurjeet Singh, Ganauri Vishwakarma, and the team at Jharkhand's Social Audit Unit for field support. I am grateful for feedback from Scott Ashworth, Dan Black, Oeindrila Dube, Steven Durlauf, Jeffrey Grogger, Damon Jones, Donald Moynihan, Guillaume Pouliot, James Robinson, Raul Sanchez de la Sierra, Shaoda Wang, Austin Wright, and workshop participants at the Harris School's PhD Workshop and Student Working Group, and the University of Chicago's Development Economics Lunch. I am grateful to Eric Dodge, Rohini Pande, Charity Troyer Moore, and current and former colleagues at EPoD India for allowing me access to their resources in India. Finally, I gratefully acknowledge financial support from the University of Chicago's Committee on Southern Asian Studies and the Becker Friedman Institute Development Economics Initiative. Any opinions expressed are those of the author.

1 Introduction

Many current and historical examples of the failure of government to provide public services can be attributed to the misappropriation of public authority and resources by bureaucrats (Wilson, 1991; Rose-Ackerman and Palifka, 2016). The potential distortions caused by this kind of misconduct are exacerbated when populations are reliant on the state for economic development and redistribution (Shleifer and Vishny, 1993; Svensson, 2005; Olken, 2006; Olken and Pande, 2012). Governments adopt monitoring programs to identify, deter, and punish misconduct (Finan et al., 2017). However, less is known about the unintended consequences of monitoring. In particular, bureaucrats’ may strategically adjust their behavior when they believe their actions will go undetected by monitors. Even less is known about what these deviations by bureaucrats imply for designing a policy that maximizes the deterrence of misconduct. This matters because, while monitoring is effective, constant and extensive monitoring is not possible under budget constraints.

If we know how bureaucrats adjust in response to monitoring, then is it better to inform bureaucrats when the auditor is coming or make it hard to predict? It may be tempting to conclude that the best strategy is one that maintains unpredictability. But, the answer depends on the relationship between bureaucrats’ expectations of the likelihood of an audit and their deterred misconduct (Lazear, 2006; Eeckhout et al., 2010). Yet, auditing guidelines routinely advise to withhold information from audit subjects to maintain unpredictability.¹

To advance on this question, this paper examines strategic responses of bureaucrats based on their changing expectations of the likelihood of an audit. Analyzing a monitoring policy of India’s Mahatma Ghandi National Rural Employment Guarantee Scheme (NREGS), the largest public employment program in the world (Sukhtankar, 2017)², I show that the deterrence of bureaucrats’ misconduct is more responsive at higher than lower expectations of the likelihood of an audit. Using

¹See as an example: U.S. auditing standards adopted and approved by the Public Company Accounting Oversight Board and the U.S. Securities and Exchange Commission to maintain unpredictability of when audits may occur.

²Sixty-five percent of the population in India, or about 11.5% of the world population, are eligible for NREGS. NREGS expenditures make up about 0.4% of India’s GDP.

this result, I show that a simple change in the design of information, which perfectly informs bureaucrats when they will be audited or not, persuades them to be less corrupt in aggregate than when leaving them to guess—while holding the budget and rules of the audit fixed.

This improvement to monitoring, which strengthens the implementation of NREGS, has impacts that are far-reaching. The NREGS program insures against income shocks by guaranteeing employment to rural households to work on public projects. Policymakers have been reliant on NREGS to curb unemployment as millions of migrant workers return home due to the ramifications of COVID-19.³ Far from perfect implementation, evidence in data and anecdotes reveal issues along various dimensions of NREGS program performance.⁴ Issues range from beneficiary payment delays to poor workplanning to fabricated employment and material procurement. Existing literature documents the extensive problem of fabricated employment (e.g. Niehaus and Sukhtankar, 2013; Banerjee et al., 2020; Muralidharan et al., 2016).

The empirical strategy of this paper leverages a monitoring policy where audits were staggered across years and randomly selected *without* replacement until all GPs received an audit. The monitoring policy audited gram panchayats (GPs), the smallest implementing unit of NREGS⁵, beginning in the state of Jharkhand in 2016 where the first audits across GPs completed within three years. The staggered implementation of audits combined with the randomization design generated random variation in bureaucrats’ expectation of the likelihood of an audit. I estimate the associated responses of bureaucrats by employing a difference-in-differences specification using detailed administrative data at the GP-month level.

This paper combines the administrative data with audit reports to show that bureaucrats’ responses, driven by incentives from the monitoring policy, can be interpreted as changes in deterrence. The main outcome of interest is program expenditures, which focuses the study of bureaucrat misconduct on financial mis-

³This article reports a surge in demand for work from NREGS in the height of the national lockdown. This article reports an additional allocation of \$5.3 billion USD (Rs 40,000 crore or 58% of the national FY2019-20 expenditures) to NREGS this year.

⁴See Sukhtankar, 2017 for a synthesis of existing research on India’s workfare program.

⁵The GP-level government sits under the block-, then district-, then state-level administrative units. The GP comprises wards. According to the 2011 Census of India, the median population of a GP in Jharkhand is about 6,100.

appropriation. Program expenditures comprise both honest and misappropriated expenditures. But, paired with information from audit reports and other administrative data, we can infer whether incentives from the monitoring policy are driving changes in honest or misappropriated expenditures. Program expenditures are also routinely measured, hard to manipulate ex-post, and capture bureaucrat behavior when they think their misconduct is likely to go undetected. Whereas, using data only from audits provides an incomplete view of the bureaucrats' response. This is a concern when bureaucrats' performance falls outside the scope of the audit and they know their actions will go undetected by auditors.

Using this empirical strategy, I find that *the anticipatory effects of the audit on deterrence are substantial; they are more elastic at higher than lower expectations of the likelihood of an audit*. There is a 15% decline in total expenditures during high expectations of the likelihood of an audit. In contrast, the anticipatory behavior during periods of lower likelihoods of an audit are statistically indistinguishable from zero. This observed anticipatory behavior is consistent with changes in deterrence when verified with audit report outcomes.

Bureaucrats substitute misappropriation in wages for misappropriation in materials. This is observed when bureaucrats anticipate an audit. Material procurement during this period increases while wage expenditures experience a slight decline. Audit reports reflect that issues related to material procurement may be harder to detect, which explains why bureaucrats adjust along this margin. The reports also show that issues related to material procurement increase during this period.

Intertemporal substitution toward material expenditures diminishes the deterrence effect during the audit. Substitution of wage for material misappropriation is also observed during an audit. Bureaucrats have an incentive to respond during the audit because auditors both verify the previous fiscal year's and document issues with ongoing work. Total expenditures decrease by 11% and is then followed by a 5% increase during this period. This is driven by a decline in employment then by an increase in material expenditures. Results to disentangle mechanisms support that these are changes in misappropriated expenditures. These effects cannot be explained by alternative mechanisms such as multi-tasking issues while auditors

are present and they also show real output remains unchanged during this period. Bureaucrats are substituting less misappropriation today (in wages) for more misappropriation tomorrow (in materials).

If audits were randomized *with* replacement, would there have been greater deterrence? If bureaucrats are only responsive when they believe the likelihood of an audit is very high, as the estimated anticipation effects show, then it is better to inform them about their audit in advance. I show this result by modeling an information design problem and deriving the optimal signal. In this model, the principal is concerned with maximizing deterrence among bureaucrats by providing information about the likelihood of an audit. The following example illustrates the main intuitions of the model and provides conditions under which selecting bureaucrats for audit randomly with or without replacement is better.

Example (*To randomize with or without replacement?*). Consider a principal who is deciding between randomly selecting N bureaucrats for audit with or without replacement over two periods. There is greater predictability when auditing without replacement because bureaucrats observe who is audited and waiting-to-be-audited. The principal is interested in deterring misappropriated expenditures by bureaucrats. In every period, the principal only has the capacity to conduct $M < N$ audits or share $p = \frac{M}{N}$ of bureaucrats. In the first period, a likelihood p of audit for all N bureaucrats leads to a deterrence of $U(p)$. In the second period, deterrence is $U(p)$ under randomization with replacement. Under randomization without replacement, the deterrence in the second period is $pU(0) + (1 - p)U\left(\frac{p}{1-p}\right)$. The share p audited in the first period will not be audited in the second period, they know this, and behave accordingly (first term). While those waiting to be audited (share $1 - p$) believe they will be audited with probability $\frac{p}{1-p}$ and adjust their behavior accordingly (second term). This gives us $U(p) \leq pU(0) + (1 - p)U\left(\frac{p}{1-p}\right)$. Jensen's inequality tells us that auditing without replacement yields more deterrence than with replacement if $U(\cdot)$ is convex. Ultimately, what determines the optimal design of information is the shape or elasticity of deterrence with respect to expectations of the likelihood of an audit. The shape of this object is determined by the benefits bureaucrats get from misappropriating expenditures and their ability to do so.

The model and example above show that the shape or the elasticity of bureaucrats' deterrence with respect to their expectations of the likelihood of an audit characterizes the principal's optimal signal. The shape of this object is an empirical question. It is also a sufficient statistic for determining the optimal signal and assessing welfare under alternative policies. I estimate this sufficient statistic using the estimated anticipation effects.

I find that *randomly informing some bureaucrats that they will be audited and others that they will not (i.e. crackdown policy) is the optimal design of information to maximize deterrence in this empirical setting. Moreover, it is better to select bureaucrats for audit randomly without replacement than randomly with replacement (i.e. equal likelihood of audit).* Designing signals which communicate when the auditor is coming will perform better than signals that maintain unpredictability. When randomizing without replacement, there is more information compared to randomizing with replacement because bureaucrats audited in later rounds anticipate their turn for audit. Sensitivity analyses confirm the robustness of these results.

Welfare estimates show a crackdown policy would have deterred an additional USD 19.8 and USD 35 million (or 9-16% of average annual expenditures from FY2016-19 in Jharkhand) in misappropriated expenditures when compared to the actual policy of randomizing without replacement and to randomizing with replacement, respectively. These potential gains are substantial given wide-prevailing audit standards to withhold information from audit subjects to maintain unpredictability. This paper makes a strong case for evaluating the possibility in other settings that changes in the design of information, like implementing a crackdown, may yield significant returns.

This paper's key contributions to the literature are: (1) empirical evidence on strategic responses by bureaucrats to monitoring; and (2) showing how a monitoring policy, which takes into account these strategic responses, can be optimally designed with a budget-constrained policymaker in mind.

This paper complements the literature studying the effectiveness of various rules of monitoring on deterrence. Previous studies demonstrated the importance of knowing with certainty that your audit will occur (Olken, 2007), having a reliable monitor

that can discover and accurately report findings (Banerjee et al., 2008; Duflo et al., 2012; Duflo et al., 2013); having a reliable system for imposing penalties when infractions are found, including an informed electorate (Ferraz and Finan, 2008; Ferraz and Finan, 2011; Afridi and Iversen, 2014; Bobonis et al., 2016); and having a persistent threat from monitoring over time (Avis et al., 2018).

But, less is known about strategic responses to a monitoring policy when audit subjects believe their actions are likely to go undetected by auditors. The challenge is that measures of verified performance beyond audits are limited. The tax literature has made use of third-party data to examine how tax reports change with the threat of audit (Casaburi and Troiano, 2016; Carrillo et al., 2017). Other studies have made use of administrative data, and find audits lead to strategic adjustments in procurement by bureaucrats (Gerardino et al., 2017; Lichand and Fernandes, 2019). This paper leverages administrative and audit data to measure strategic responses in combination with sufficient variation in the principal’s policy parameter of interest: bureaucrats’ expectations of the likelihood of an audit. This combination is what allows us to determine the optimal design of information.

This paper performs an empirical test of a model of information design, which complements the theoretical literature on optimal information design (e.g., Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Kamenica, 2019). The intuitions from the model in this paper draw on prior theoretical work. Lazear (2006) theoretically analyzes the trade off between provision of concentrated (e.g. crackdown) versus dispersed (e.g. randomizing with replacement) incentives through monitoring, and finds that the optimal design depends on the shape of deterrence as a function of expectations of the likelihood of an audit. Eeckhout et al. (2010) develop and empirically test a similar model with data on car-speed monitoring. However, constraints in their empirical setting allow them to only test positive predictions of their model assuming an optimal policy is being implemented.

The theoretical model in this paper provides a sufficient statistic that I estimate to determine the optimal policy and analyze welfare. To my knowledge, the sufficient statistics approach has not been applied to studies motivated by the optimal design of information. This approach is related to a literature in public finance that uses

models to develop sufficient statistics to evaluate welfare as functions of reduced-form elasticities and not structural primitives (Chetty, 2009). This approach can be applied to other settings where communication on monitoring can play a role. Examples range from oversight for fraud in campaign financing, tax returns, and the allocation of social insurance to police use-of-force.

The rest of the paper is organized as follows. Section 2 describes the institutional background, information environment of the policy, and data sources. Section 3 develops the model. Sections 4 and 5 present the empirical strategy and reduced-form results. Section 6 presents the assumptions required to estimate the sufficient statistic, determines the optimal policy using the sufficient statistic, conducts robustness checks, and estimates welfare consequences of alternative policies. Section 7 concludes.

2 Background

2.1 The public employment program and the audit agency

Launched in 2006, the Mahatma Ghandi National Rural Employment Guarantee Scheme (NREGS) guarantees 100 days of work per year to rural households. Beneficiaries provide unskilled labor on projects commissioned by the local government. Wages set by the state government are generally below the minimum wage for unskilled work in agriculture and other industries.⁶ Beneficiaries construct or maintain assets that go toward the improvement of rural livelihood. Examples of assets include structures for water conservation and harvesting, homes, latrines, and animal shelters.

Jharkhand is a state in the eastern part of India. The population is close to 33 million with 76% of people living in rural areas. Sixty-one percent of the population relies on agricultural work.⁷ NREGS in Jharkhand has served around 7.7 million

⁶According to minimum wages set by the Ministry of Labour and Employment of the Government of India.

⁷Sources: Jharkhand Economic Survey 2017-18; Department of Agriculture of the Government of Jharkhand. Jharkhand has close to 40% of the mineral reserves found in India. The mining and manufacturing sectors contribute to over 33% of the state's economy.

people and produced over 1 million projects. In 2016, the state government of Jharkhand began randomly selecting GPs implementing the NREGS program for audit.

The Social Audit Unit (hereinafter referred to as the audit agency) is a separate government agency that conducts the audits.⁸ The audit agency is funded independently of NREGS and managed by a steering committee of various stakeholders across the state government and civil society. Competitive compensation for auditors and quality assurance mechanisms suggest it is likely the audit agency conducted audits at-scale with credibility and integrity. More details on the audit agency and their processes are in Appendix A.2.

The goal of the audit agency was to audit all GPs for the first time before selecting GPs for audit for the second time. This goal was consistent with the NREGS national act to ensure regular auditing of all implementing bodies. To do so, the audit agency randomly selected GPs without replacement for audit from 2016-2019 until all GPs were audited. This paper focuses on the effect this monitoring policy had on bureaucrat behavior during these 3 fiscal years (FY) when GPs were receiving their first audit.⁹ During this period, 4,180 GPs were audited and informed that the previous FY's work would be part of the audit.

2.2 The process when auditors arrive

Auditors typically spend a week at the GP to verify administrative reports from the previous FY, document other observed issues, and conduct the public hearings. The average number of auditors per audit is 2.58 and the distribution ranges from 2-9 auditors. Auditors gather information from GP office records; field observations; and household interviews. Their tasks include matching receipts with materials reported to be procured, checking adherence of job advertisements to program guidelines, physically verifying output at project sites (e.g. measuring dimensions of a

⁸"Social audits" incorporate community and beneficiary feedback, hence the name. As noted in Section 2.2, the audit gathers information through household interviews. Public hearings are also held to announce audit findings and adjudicate issues found during the audit in a public forum.

⁹Fiscal years in the Indian government go from the beginning of April to the end of March the following year.

dug pond), and interviewing households to verify past employment and document complaints.

The audit process is not designed to be disruptive of normal program operations. According to audit guidelines, providing paperwork for auditors in advance and participating in a couple days of public hearings are the only tasks required of GP bureaucrats during the audit process. About 64 percent of GPs fail to comply with parts of the audit process, but auditors can still proceed with verification of administrative reports using information gathered from field visits of project sites and household interviews.

Notably, information gathered during the audit process is also reflective of concurrent program performance. The auditors' scope of evaluation is not limited to verifying administrative reports from the previous fiscal year. For example, auditors evaluate quality of record-keeping and proper advertising of available jobs, and they conduct household interviews. During these interviews, households raise ongoing issues even though auditors are investigating employment from the previous year. Problems of recall may make it hard for households to speak only to last year's work. Based on the audit reports, 50-60% of the identified issues reflect ongoing problems.

The threat of punishment to bureaucrats from the audit is credible. Issues identified from the audit are presented and adjudicated in a public hearing with auditors, bureaucrats, and beneficiaries in attendance.¹⁰ Issues are resolved when evidence is provided to show the issue is unfounded or when those culpable agree to take corrective action. Bureaucrats charged with financial misappropriation face paying a penalty commensurate with the amount misappropriated or risk losing their job.¹¹ Unresolved issues are escalated to be adjudicated at a higher-level public

¹⁰Among all audit reports, there were 68,231 documented issues. The mean number of issues per audit is 21. Nineteen percent of issues are related to concerns about competency in implementation; 18% of issues are related to issues obtaining work and payments, not obviously related to misappropriation; 16% of issues are related to misappropriation of wages and allocated employment; 6% of issues are related to misappropriation in material procurement; 11% of issues are related to whether the project was actually constructed with expended funds; and 12% of issues are related to officials refusing to cooperate in some way with the audit process.

¹¹While the audit reports provide information on the identified issues, they currently do not contain information about the resolution or follow-up on the issue. Anecdotally, elected and appointed bureaucrats at the GP have lost their jobs (or were potentially transferred) as a result

hearing. The audit agency reports around \$1.5 million USD (11 crore INR) have been recovered through the audits, which is around 0.8% of Jharkhand's NREGS expenditures in FY2018-19.

2.3 The information environment and bureaucrat incentives

Bureaucrats can anticipate their next audit when selection for audit is predictable. In this setting, selection is based on whether all GPs have had their turn to be audited before selecting GPs for audit again. In response, bureaucrats may adjust opportunistic behavior to influence the outcome of an anticipated audit. These strategic adjustments are possible because past performance is part of the audit evaluation. This section summarizes what bureaucrats knew and what we can infer about their incentives in response to the monitoring policy.

Every year from FY2016-2019, all GPs received an announcement that stated who has been selected for audit, that last year's performance would be audited, and the audit dates. The announcement in Year 1 stated that the audit agency plans to eventually target 50% of GPs for audit every year.¹² It also stated that this fell short of the benchmark in the 2006 NREGS Act Section 17 requiring all GPs be audited twice a year. With this information, it would have been reasonable to expect that the audit agency would complete the round of first audits in all GPs before starting the round of second audits.

One's selection for audit could be anticipated, but was not perfectly predictable. Current audit capacity was observed, future audit capacity was unknown. So, the number of years it would take to complete the round of first audits was not known ex-ante. Even the leadership of the audit agency was uncertain about future audit capacity. In Year 1 (FY 2016-17), 548 GPs were audited; 1,495 GPs in Year 2 (FY 2017-18); and 2,137 GPs in Year 3 (FY 2018-19). As it happened, by Year 3, the remaining GPs were all selected for audit, then the audits would restart in Year 4 (FY2019-20).¹³ It was very likely that those not part of Wave 3 audits would be

of issues uncovered during the audit.

¹²This announcement included a press and video conference with all district officials to disseminate the announcement.

¹³Around 300 GPs were not selected for audit under this policy for various reasons. They are

selected for audit in Year 4.

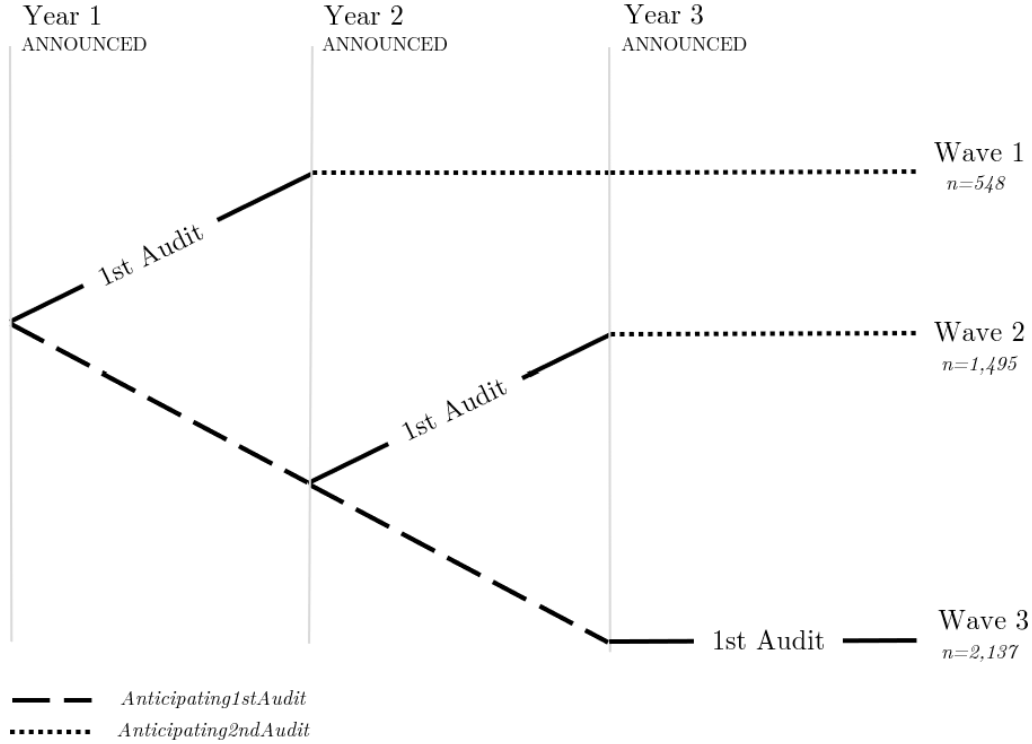


Figure 1: Roll-out of the round of first audits and evolution of beliefs for each wave. Dashed and dotted lines denote horizons of anticipating the first and second audits, respectively. Wave 1 follows the path on the top; Wave 2 follows the path in the middle; and Wave 3 follows the path on the bottom.

With predictability over when one's audit might occur, we can expect bureaucrats to act in anticipation of future audits. Figure 1 illustrates how expectations of a future audit for each wave may evolve over the three years of roll-out. Dashed and dotted lines denote horizons of anticipating the first and second audits, respectively. Wave 1 follows the path on the top; Wave 2 follows the path in the middle; and Wave 3 follows the path on the bottom.

In particular, as we move from Year 1 to Years 2 and 3, bureaucrats who have not had their first audit have increasing expectations that this year's performance will be audited next year. E.g. Wave 3's expectations that current performance will be audited increases over time (i.e. horizon denoted by the dashed lines). While described in detail in Section 4.2 and dropped from analysis. More details on notices that were publicly disseminated can be found in Appendix A.3.

Wave 1 bureaucrats believe the likelihood they will be audited a second time in Year 2 is very low since there are a sufficient number of other GPs waiting for their first audit (i.e. horizon denoted by the dotted line during Year 2). On the other hand, as bureaucrats observe the completion of the round of first audits in Year 3, Waves 1 and 2 bureaucrats have a very high expectation of receiving a second audit in Year 4 (i.e. horizons denoted by the dotted lines during Year 3). About 84% of Waves 1 and 2 GPs compared to 18% of Wave 3 GPs were audited in Year 4.

Additionally, bureaucrats also have an incentive to adjust behavior while their audit is occurring. During the audit, auditors work and sleep at the GP office where NREGS administrative matters takes place. And part of the information gathered during the audit process is reflective of concurrent program performance, as discussed in Section 2.2.

3 A Model of Information Design and Deterrence

This section presents a model of information design used to study the monitoring policy of the public employment program. In the model, the Principal (e.g. audit agency or citizens) communicates information about the monitoring policy that allows the Bureaucrat to learn about the likelihood of an audit. This information affects the Bureaucrat's choice on expenditures to misappropriate. The Bureaucrat's choice affects payoffs to both the Principal and the Bureaucrat. Section 3.2 provides conditions for when the Principal would prefer to send informative signals, i.e. information that helps the Bureaucrat better predict when the audit will occur, and how to solve for the optimal signal. The model provides a sufficient statistic which allows us to evaluate how various signals ultimately affect the Principal's welfare. This sufficient statistic is estimated in this empirical setting in Section 6.

3.1 Setup

There is a Principal who oversees implementation of a government program by N bureaucrats. The Principal is budget-constrained and can only audit M bureaucrats, where $M < N$. Consider the Principal's interaction with a single, arbitrary

Bureaucrat. This model is static and two-player but accommodates settings with N Bureaucrats and T periods under reasonable assumptions (see Appendix B.1 for details). The Bureaucrat privately benefits from misappropriated expenditures, while the Principal is made worse off.

The Principal chooses a communication policy, π , that conveys to the Bureaucrat the likelihood of the audit. Formally, the Principal announces π , a chosen probability distribution over likelihoods of audit. We assume that the Principal commits to the communication policy: once π is announced, the signal received by the Bureaucrat is obtained from the distribution π .

If the Principal provides no information, each bureaucrat is expected to be monitored with equal likelihood, given the Principal's capacity constraint. That is, the Bureaucrat has a prior about the the likelihood of an audited, $q_0 = \frac{M}{N}$. Upon receiving a signal, the Bureaucrat forms the posterior belief about the likelihood of an audit, q , according to Bayes' rule. Using updated information, the Bureaucrat chooses an action $a \in [0, 1]$ on share of expenditures to misappropriate.

The Bureaucrat's expected utility, $V(a, q)$, is a function of the benefits from the misappropriated expenditures (a) and the expected punishment from audit that happens with probability q . Assume that $V(a, q)$ satisfies the *single-crossing property*, i.e., for all $\bar{q} > \underline{q}$, $\bar{a} > \underline{a}$, if $V(\bar{a}, \underline{q}) - V(\underline{a}, \underline{q}) \leq 0$ then $V(\bar{a}, \bar{q}) - V(\underline{a}, \bar{q}) \leq 0$. This assumption means that if a Bureaucrat who expects the audit with probability \underline{q} prefers \underline{a} (where $\underline{a} < \bar{a}$), he would make the same choice if the probability of audit was \bar{q} (where $\bar{q} > \underline{q}$). For example, $V(a, q) = v(a) - qc(a)$ where $v(a)$ is the net benefit from misappropriating a and $c(a)$ is the punishment for choosing a . We abstract away from heterogeneity in the Bureaucrat's propensity to misappropriate because assignment to a signal from π is random; we can think of the Bureaucrat's response, $a(q)$, given belief q , as holding all other factors affecting $a(\cdot)$ equal.

The Principal's utility is $u(a(q)) = -ka(q) + b$, where $k \in \mathbb{R}^+ \setminus \{0\}$ and $b \in \mathbb{R}$ are constants. Naturally, the Principal's utility is decreasing in $a(\cdot)$, i.e. $\frac{\partial u}{\partial a} = -k < 0$. The Principal's utility is a particular affine transformation of the Bureaucrat's choice of misappropriated expenditures, i.e. $u(\cdot)$ is negatively proportional to $a(q)$. Note that elasticity of $a(q)$, the Bureaucrat's choice of misappropriated expenditures, is

the negative of the elasticity of $u(a(q))$, the Principal's utility. We can interpret $u(a(q))$ as the Bureaucrat's deterrence scaled and shifted by constants k and b , respectively. The Principal's utility is net of the costs for conducting M audits. The Principal's expected utility, $\mathbb{E}_{q \sim \pi(\cdot)}[u(a(q))]$, sums over the likelihood of posterior beliefs q given the set of signals that can be drawn from π .

Example (*Signals*). For example, the Principal may consider a π that concentrates ("a crackdown") or disperses incentives (uninformative signal where everyone knows the prior). To implement a crackdown, the Principal can choose a signal space where the Bureaucrat could draw signal $s \in \{H, L\}$: being audited with high (H) or low intensity (L). The Principal can design π such that:

$$\begin{aligned}\pi(H|Audit) &= 1 & \pi(H|NotAudit) &= 0 \\ \pi(L|Audit) &= 0 & \pi(L|NotAudit) &= 1\end{aligned}$$

Then, by Bayes' Rule, the Bureaucrat that drew signal H would have posterior beliefs of the likelihood of an audit of $q_H = 1$ and one who drew signal L would have posterior belief $q_L = 0$.

The timing of this game is as follows:

- (1) Principal chooses and commits to communication policy, π , about the likelihood of an audit.
- (2) Nature draws the signal for Bureaucrat consistent with π .
- (3) Bureaucrat observes the signal and forms posterior beliefs q about the likelihood of the audit.
- (4) Bureaucrat chooses, $a \in A$, the share of resources to misappropriate.
- (5) Monitoring takes or does not take place. All payoffs are realized.

The equilibrium concept for this model is Bayes-Nash perfect equilibrium.

3.2 Analysis

The optimal policy depends both on the elasticity of the Bureaucrats' response to likelihood of monitoring and the Principal's capacity to conduct audits. This is consistent with insights from Lazear (2006) and Eeckhout et al. (2010).

The Bureaucrat chooses an action that maximizes expected utility:

$$a^*(q) = \arg \max_{a \in A} V(a, q)$$

Since $V(a, q)$ satisfies the single-crossing property, then we know that $a^*(q)$ is weak decreasing in q (Milgrom and Shannon, 1994). With this and the fact that $u'_a < 0$, then $u(a^*(q))$ (the Principal's value function given $a^*(q)$) is weakly increasing in q . For notational convenience, let $U(q) = u(a^*(q))$. Taking the receiver's optimal response into account, the Principal will choose the best communication policy, π^* , in order to maximize:

$$\max_{\pi \in \Pi} \mathbb{E}_{q \sim \pi(\cdot)} U(q)$$

Define $co(U(q))$ to be the convex hull of the function $U(q)$, i.e. the set of all convex combinations of the values of $U(q)$. $\{u|(q_0, u) \in co(U(q))\}$ is the set of attainable payoffs for the Principal, where we are restricted to $(q_0, u) \in co(U(q))$ by Bayes' Rule.¹⁴ These payoffs are achievable by the Principal's chosen policy π , which can induce payoffs of convex combinations of different values of the function $U(q)$. For example, the Principal can choose π that divides Bureaucrats into two groups where some have posterior beliefs of audit with q' and others with q'' , as long as the mean of posterior beliefs is equal to q_0 . The largest payoff possible to the Principal for some given prior q_0 is $\hat{U}(q_0) = \max\{u|(q_0, u) \in co(U(q))\}$.

Proposition 1 (Persuasion works). *The Principal benefits from sending an informative signal (concentrating incentives) if and only if $U(q_0) < \hat{U}(q_0)$. Moreover, if the function is convex at $U(q_0)$, then $U(q_0) < \hat{U}(q_0)$ and an informative signal is better.*

¹⁴Bayes' Rule tells us that the mean of the posterior beliefs of the likelihood of an audit is equal to q_0 . Using the Example on signals above: $q_0 = Pr(Audit) = \pi(Audit|H)Pr(H) + \pi(Audit|L)Pr(L) = \pi(H|Audit)Pr(Audit) + \pi(L|Audit)Pr(Audit)$. The second equality follows from the law of iterated expectations and the third equality follows from Bayes' Rule.

All proofs are in Appendix B.3. The condition for sending an informative signal in Proposition 1 is derived in Kamenica and Gentzkow (2011) and echoes the findings in Lazear (2006) and Eeckhout et al. (2010). This finding implies the following result that we need. This results and the discussion to follow help to understand the intuition behind Proposition 1.

Proposition 2. *The Principal prefers concentrated over dispersed incentives if $U(q)$ is convex. The converse is true if $U(q)$ is concave.*

To construct the optimal signal, we can reframe the Principal's problem by searching over the set of distributions of posterior beliefs to maximize the Principal's expected utility. With the posterior distribution of interest, we can backout a signal that induced it. A geometric interpretation of this solution means that a Principal prefers to send an informative signal if there is a convexity in $U(q)$ evaluated at q_0 . The shape of $U(q)$ determines the optimal signal. Moreover, the signal that induces the payoff $\hat{U}(q_0)$ is the optimal signal.

Figure 2 provides an illustration of when concentrated versus dispersed incentives are better. The shaded area represents $co(U(q))$; \hat{U}_π denotes the payoff generated by a signal that concentrates incentives at q_H and q_L ; and the bold dotted line represents $\{u|(q_0, u) \in co(U(q))\}$, the payoffs attainable by the Principal. The figure on the left shows that $\hat{U}(q_0) = \hat{U}_\pi$, i.e. concentrating incentives by telling some they will be selected for audit with a very high likelihood (signal H) and others they will not be audited (signal L), is greater than $U(q_0)$, i.e. dispersed incentives when everyone believes they will be audited with probability q_0 . To achieve utility $\hat{U}(q_0)$, our signal must place weight on posterior beliefs q_H and q_L such that $q_0 = Pr(H)q_H + (1 - Pr(H))q_L$. Solving for $Pr(H)$, we can then use Bayes' Rule to deduce the optimal signal by solving for $\pi(H|Audit)$ and $\pi(L|Audit)$. The figure on the right shows the converse where a signal where everyone believes q_0 is the likelihood of an audit is optimal.

$U(q)$ is increasing in q , but the shape of $U(q)$ is an empirical question. It is a sufficient statistic for determining the optimal signal and marginal welfare changes, and does not require specifying additional primitives underlying the Principal's choice problem. Section 6 discusses the approach for estimating $U(q)$ using the reduced

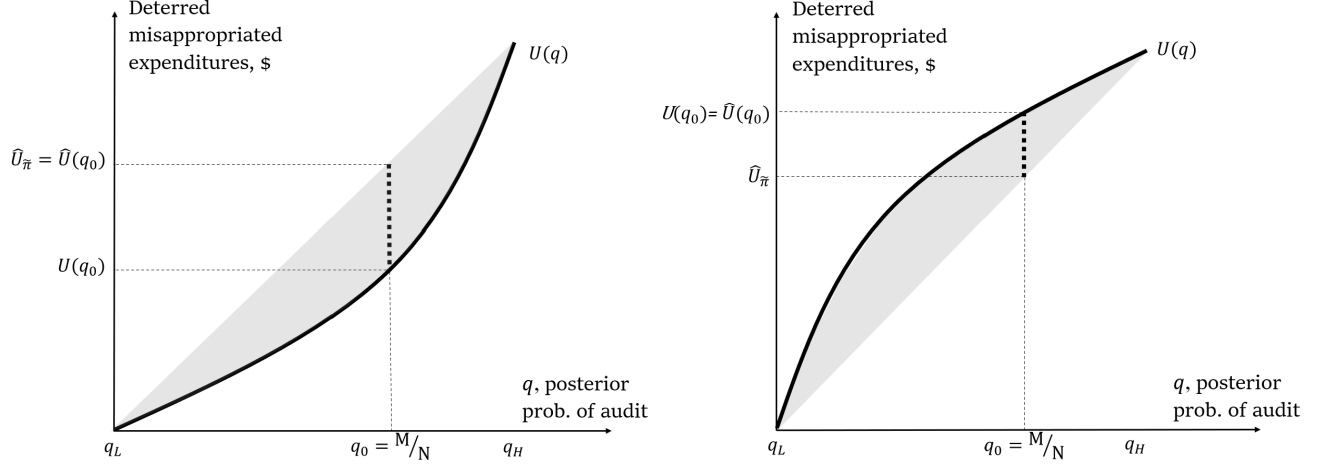


Figure 2: When does the Principal prefer concentrated versus dispersed incentives? These figures represent scenarios of the Principal's utility as a function of the Bureaucrat's posterior beliefs. Concentrated incentives through communicating more information on the likelihood of monitoring is preferred in the figure on the left. Dispersed incentives where the Principal communicates no additional information are preferred in the figure on the right.

form estimates from Section 5.

4 Empirical Strategy

This section discusses the identification strategy and econometric specifications; the challenges of measuring corrupt behavior in response to the monitoring policy; and the approach for using changes in administrative measures of bureaucrat performance as a proxy for changes in corrupt activity in response to the monitoring policy.

4.1 Estimating equation and identification assumptions

We are interested in estimating how bureaucrats respond during time t with varying expectations of the likelihood that performance in time t will be audited. I use a differences-in-differences model to estimate bureaucrat responses as expectations change during various stages of the monitoring policy. The following specification

estimates the intent-to-treat from learning one has been selected for their first audit:

$$y_{it} = \alpha_i + \alpha_{dt} + \textit{Anticipating1stAudit}'_{it}\beta + \delta_1 \textit{Post1stAnnounce}_{it} + \epsilon_{it} \quad (1)$$

where α_i denotes fixed effects for GP i ; α_{dt} denotes fixed effects for district d in month t ; ϵ_{it} is the residual error term.¹⁵ Standard errors are clustered by block—the unit of stratification and one administrative level above the GP. y_{it} denotes the GP performance outcome. The main outcome is total expenditures, measured as the sum of wage and material expenditures. The excluded component is administrative expenditures because panel data are currently not available for this measure. This is of limited concern because administrative expenditures are a negligible share of total expenditures (0.4% on average). Each component of bureaucrat behavior captured by the remaining variables and their identification will be discussed in turn.

First, consider $\textit{Post1stAnnounce}_{it}$, a dummy variable capturing the period after each GP learns from announcement of their selection for the first audit. This variable captures the intent-to-treat of informing bureaucrats of their audit. The effect of $\textit{Post1stAnnounce}_{it}$ on GP performance y_{it} would be identified under the assumption of parallel trends between those announced for their first audit and not yet announced.

However, being able to anticipate one’s first audit is a potential threat to the parallel trends assumption needed to identify $\textit{Post1stAnnounce}_{it}$. With the random assignment to audit in Year 1, we should expect parallel trends between Wave 1 and non-Wave 1 groups prior to announcement.¹⁶ But, we should not expect par-

¹⁵The GP fixed effects also help account for stratification in the randomized roll-out at the block administrative level. In the 2016-17 audit, the randomization was stratified by block (one administrative level higher than GP). In the 2017-18 audit, the randomization was stratified by block with an additional rule and selection was among the GPs not incorporated in the 2016-17 audit. The additional rule was that all GPs within a block would be selected for audit if there were 10 or fewer GPs remaining to be audited within that block (27% of the blocks in Jharkhand have ≤ 10 GPs; 42% of blocks had ≤ 10 GPs left to be audited by 2017-18). Using the announcement data to check, an average of 98% of GPs within these blocks were audited. For the 2018-19 audit, the remainder of unaudited GPs were selected for audit. In these waves of the audit, we would expect to have independence in observed and unobserved variables between the treatment and control groups conditional on block fixed effects, which also controls for the number of GPs in a given block during this time.

¹⁶To illustrate, under the standard difference-in-differences estimator with two periods in the

allel trends for subsequent waves because random assignment without replacement under the information environment can lead to anticipation. See discussion of this anticipatory behavior in Section 2.3.

We can control for this horizon of anticipation with $Anticipating1stAudit_{it}$ in order to identify $Post1stAnnounce_{it}$. $Anticipating1stAudit_{it}$ is a vector of two dummy variables which capture anticipation for those waiting for their first audit separately as they observe the announcements in Years 2 and 3. Importantly, the estimates of $Anticipating1stAudit_{it}$ are also parameters of interest because they capture strategic responses by bureaucrats as expectations of the likelihood of an audit vary. Given random assignment without replacement to audit each year, bureaucrat expectations of when their first audit will occur are also randomly assigned, allowing us to identify $Anticipating1stAudit_{it}$.

The following is the main specification, which disaggregates behavior during $Post1stAnnounce_{it}$ and estimates the profile of bureaucrat responses as expectations of the likelihood of an audit change during the roll-out.

$$y_{it} = \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \delta_1 Post1stAnnounceBefore1stAudit_{it} + \delta_2 1stAudit_{it} + Anticipating2ndAudit'_{it}\gamma + \epsilon_{it} \quad (2)$$

$Anticipating1stAudit_{it}$ is as described above. The remaining variables are mutually exclusive groups and are defined such that $Post1stAnnounce_{it} = Post1stAnnounceBefore1stAudit_{it} + 1stAudit_{it} + Anticipating2ndAudit_{it}$. The variable $1stAnnounceBefore1stAudit_{it}$ captures the period once one learns they will be

potential outcomes framework, we have that

$$\hat{\delta}^{DiD} = (\bar{Y}_{1,D=1,t_0} - \bar{Y}_{0,D=1,t_0}) + [(\bar{Y}_{0,D=1,t_0} - \bar{Y}_{0,D=1,t_0-1}) - (\bar{Y}_{0,D=0,t_0} - \bar{Y}_{0,D=0,t_0-1})]$$

where the second bracketed term is the difference between the counterfactual trend of the treatment group and the trend of the control group; D denotes treatment status; t_0 is the period of treatment; and (Y_0, Y_1) denote potential outcomes of receiving treatment or not. From randomization of the audit, the following must hold:

$$(Y_{0t}, Y_{1t}) \perp\!\!\!\perp D$$

which implies that $[Y_{0,D=0,t_0}] = [Y_{0,D=1,t_0}]$ and $[Y_{0,D=0,t_0-1}] = [Y_{0,D=1,t_0-1}]$, and is sufficient for $\hat{\delta}^{DiD}$ to be an unbiased estimate of the treatment effect.

receiving their first audit but before the first audit occurs. $1stAudit_{it}$ is a vector of two dummy variables capturing the periods during the month of and months following the first audit. $1stAudit_{it}$ estimates the effect of an active audit on bureaucrat behavior as well as any persistent effects of experiencing an audit. While GPs were randomly assigned for audit each year, the order in which audits occur within each year is non-random. According to the audit agency, the schedule was designed to limit travel time in between audits. So, teams of auditors were not traveling from one end of the state to the other. The district-specific time fixed effects accounts for this timing of audits.

Finally, consider $Anticipating2ndAudit_{it}$, a vector of two dummy variables which capture the anticipatory horizons in Years 2 and 3 for those anticipating their second audit. $Anticipating2ndAudit_{it}$ is mutually exclusive from the behavior captured by $1stAudit_{it}$. The estimates of $Anticipating2ndAudit_{it}$ are parameters of interest because they capture strategic responses by bureaucrats anticipating second audits as the round of first audits progress. $Anticipating2ndAudit_{it}$ is identified if it is not confounded with persistent effects of the audit. To do so, we can estimate an event study and test whether behavior after the audit (controlling for variation from $Anticipating2ndAudit_{it}$) is statistically different from behavior before the audit (result reported in Section 5.3).

Of note, all regression will treat ‘ $Anticipating1stAudit$ - Year 1’ as the reference group. Periods of anticipation are defined by when annual announcements for audit are released, and this affects all GPs simultaneously. This implies that the sum of $Post1stAnnounce$ and $Anticipating1stAudit$ variables are collinear with a linear combination of the time fixed effects. So, using ‘ $Anticipating1stAudit$ - Year 1’ as the reference group allows for a solution to our coefficient estimates.

4.2 Data sources and sample restrictions

2011 Village Census of India The 2011 Village Census of India provides data on GP demographics, local economy, household and village amenities, and natural resources. Census data are at the ward level, one administrative unit below the GP. Ward-level data aggregate to provide GP-level data. These data help us check the

integrity of the random assignment in tests of balance on observable characteristics.

Annual audit announcements The audit agency provided documentation on the announcements from 2016-2020. Announcements detail GPs selected for audit and audit dates. Together with the announcement dates, this information helps us capture the effect during periods of anticipation discussed in Section 2.3; and the effect of learning of an upcoming audit and experiencing an audit.

NREGS administrative data NREGS management information system (MIS) provides the data for all bureaucrat performance outcomes. MIS is a national government data portal that tracks detailed information on program implementation in each GP.¹⁷

Outcomes on employment include wage expenditures; person-days of employment; and days of delayed payment across all households. Data on projects include details on material procurement, and expenditures on labor and materials. Data on expenditures correspond to program outlays and are an upper bound of actual employment and materials provided and paid for through the program. Anecdotal evidence from interviews with government officials suggest that once expenditures have been paid, they cannot be manipulated by bureaucrats. Panel datasets constructed with these outcomes are by GP-month from April 2014 to March 2019. The MIS job card register provides historical employment data at the household-level. This is helpful for analyzing outcomes by whether the household belongs to a marginalized social group. The household data are aggregated to construct a GP-month-household social group dataset.

Audit characteristics and outcomes from audit reports include share of portfolio audited, number of auditors, documented issues, and attendance at public hearings. The audit reports are from MIS and used to construct a GP-level dataset. This is helpful to compare audit performance across waves with differing anticipatory behavior.¹⁸

Sample restrictions By the end of Year 3, audits were conducted in 4,180 GPs

¹⁷Access MGNREGA MIS here and an MGNREGA Village View Dashboard MGNREGA MIS Dashboard here.

¹⁸Currently, only a subset of audit reports are available for analysis, and are only from Waves 2 and 3 of the audit.

or 93% of all GPs in Jharkhand. Around 300 GPs were not selected in the audit calendar for the following reasons: 1) 220 GPs had special audits conducted at the request of upper-level government officials¹⁹; 2) 49 GPs were audited during the pilot; and 3) an even smaller number of remaining GPs did not have any NREGS expenditures or were undergoing an administrative boundary change. Furthermore, 42 GPs were selected for audit twice over the audit roll-out. A subset of these 42 GPs were audited twice because they were also selected for special audit; for the remainder, they seem to have been selected for audit twice by mistake based on conversations with technical specialists at the audit agency. The sample for analysis omits observations from GPs that meet the following criteria: (1) were ever selected for a special audit; (2) were audited during the pilot; or (3) were audited more than once. This leaves 4,052 GPs in the sample for analysis in the unbalanced panel data, and 3,897 GPs in the sample for analysis in the balanced panel data. All analyses use the balanced panel data.

4.3 Tests for violations of identifying assumptions

Using census and administrative data, statistical tests show balance on observable characteristics across waves (Table 1). Except there are statistical differences between Waves 2 and 3 Scheduled Tribes population, share of person-days allocated to females in FY 2015-16, and share of person-days scheduled caste and days of delayed payment in FY 2014-15. The number and extent of these differences are consistent with arising by chance. Overall, the differences are small (4% difference in Scheduled Tribe population and a 3% and 6% difference in share of person-days scheduled caste and female, respectively), and differences in variation in demographic parameters that tend to be stable over time can be accounted for with GP fixed effects in our main specification.

¹⁹Requests for audits can be submitted to the audit agency by higher-level government officials; they are referred to as special audits. A majority of the special audits that took place during the study period were initiated when Chief Secretary of the Government of Jharkhand requested special audits in two districts in 2017-18 upon observation of suspicious behavior during a statewide progress report meeting. These special audits were also publicly announced. While GPs receiving special audits are not included in the sample for analysis, I account for the information learned by bureaucrats from the announcement on special audits.

Table 1: Tests of balance in observables across waves

	<i>Wave1</i>	<i>Waves2 – 1</i>	<i>p-value</i>	<i>Waves3 – 1</i>	<i>p-value</i>	<i>Waves2 – 3</i>	<i>p-value</i>	Observations
<i>Panel A: 2011 Village Census</i>								
Number of households	1151.171	3.357	0.808	4.912	0.720	-1.555	0.907	3,806
Total population	6085.480	9.501	0.898	35.530	0.620	-26.029	0.712	3,806
Scheduled castes population	756.449	-19.788	0.423	-16.176	0.454	-3.612	0.863	3,806
Scheduled tribes population	1994.775	87.028	0.112	2.013	0.969	85.016	0.029**	3,806
Literate population	3036.874	17.488	0.667	21.162	0.622	-3.674	0.924	3,806
Total working population	2660.462	-22.182	0.515	-10.193	0.751	-11.989	0.696	3,806
Main working population	1210.658	-19.073	0.514	4.701	0.846	-23.774	0.303	3,806
Main working population, cultivation	515.584	-11.378	0.463	-4.440	0.782	-6.938	0.557	3,806
Main working population, agriculture	297.796	0.104	0.993	7.806	0.487	-7.701	0.439	3,806
Main working population, household industries	49.216	-5.772	0.260	-0.900	0.852	-4.872	0.259	3,806
Marginal working population	1449.804	-3.110	0.911	-14.895	0.536	11.785	0.645	3,806
Marginal working population, cultivation	430.808	14.031	0.372	17.099	0.272	-3.069	0.828	3,806
Marginal working population, agriculture	788.113	-12.689	0.538	-23.627	0.225	10.938	0.522	3,806
Marginal working population, household industries	47.520	-3.562	0.286	-2.536	0.416	-1.026	0.688	3,806
Total geographical area (sq. km.)	1893.787	-8.928	0.864	-24.048	0.629	15.120	0.721	3,817
Forest area (hectares)	500.300	-22.264	0.512	-24.568	0.433	2.304	0.930	3,817
Barren, uncultivable land area (hectares)	84.802	-0.840	0.875	-4.633	0.397	3.793	0.403	3,817
Permanent pastures/grazing land area (hectares)	30.474	3.260	0.286	2.431	0.281	0.829	0.732	3,817
Total unirrigated land area (hectares)	584.179	10.940	0.674	14.793	0.584	-3.853	0.891	3,817
Wells and tubewells area (hectares)	32.112	-3.335	0.122	5.290	0.339	-8.626	0.143	3,817
Tanks and lakes area (hectares)	33.122	4.918	0.230	12.658	0.102	-7.740	0.278	3,817
<i>p-value of F-test of joint orthogonality</i>			0.77		0.73		0.97	
<i>Panel B: MGNREGA MIS 2014-15</i>								
No. HHs with registered demand for employment	316.658	-10.501	0.231	-6.938	0.406	-3.563	0.623	2,980
Approved labor budget (lakhs)	18287.034	-33.614	0.940	-33.781	0.932	0.167	1.000	2,980
No. HHs provided employment	288.075	-10.473	0.206	-8.399	0.265	-2.074	0.761	2,980
Person-days of work generated	316.556	-10.409	0.235	-6.880	0.409	-3.529	0.627	2,980
Share of person-days, scheduled caste	13.175	-0.622	0.314	0.158	0.759	-0.779	0.052*	2,980
Share of person-days, scheduled tribe	37.733	0.258	0.795	0.599	0.568	-0.341	0.685	2,980
Share of person-days, female	32.413	0.604	0.197	0.471	0.370	0.132	0.765	2,980
No. HHs with 100 days completed	20.328	-0.158	0.948	-1.180	0.528	1.023	0.615	2,980
Days of delayed payment	28033.808	-30.562	0.989	-3622.287	0.175	3591.726	0.094*	2,980
Amount of delayed payment	13600.405	157.695	0.873	-1381.928	0.230	1539.623	0.127	2,980
Work completion rate	65.568	-0.533	0.385	-0.865	0.131	0.332	0.488	2,980
Total expenditures (lakhs)	21.920	-0.484	0.617	-1.098	0.174	0.614	0.459	2,980
Share of expenditures, wages	75.240	-0.412	0.613	-0.163	0.847	-0.250	0.690	2,980
Share of expenditures, admin	0.010	0.007	0.520	0.006	0.452	0.001	0.912	2,980
<i>p-value of F-test of joint orthogonality</i>			0.55		0.13		0.53	
<i>Panel C: MGNREGA MIS 2015-16</i>								
No. HHs with registered demand for employment	323.945	-9.509	0.205	-11.023	0.115	1.514	0.818	3,704
Approved labor budget (lakhs)	20030.273	145.763	0.735	-188.662	0.705	334.424	0.450	3,704
No. HHs provided employment	286.395	-8.142	0.239	-10.799	0.102	2.657	0.670	3,704
Person-days of work generated	323.772	-9.538	0.203	-11.118	0.112	1.580	0.810	3,704
Share of person-days, scheduled caste	12.813	-0.606	0.385	0.107	0.856	-0.713	0.138	3,704
Share of person-days, scheduled tribe	32.717	0.609	0.508	0.046	0.962	0.562	0.462	3,704
Share of person-days, female	31.405	0.540	0.202	-0.291	0.522	0.831	0.028**	3,704
No. HHs with 100 days completed	43.652	1.444	0.591	-1.285	0.594	2.729	0.293	3,704
Days of delayed payment	31239.766	1393.340	0.653	928.067	0.658	465.273	0.890	3,704
Amount of delayed payment	15863.149	565.533	0.708	582.585	0.583	-17.052	0.992	3,704
Work completion rate	64.943	-0.366	0.510	-0.575	0.294	0.209	0.622	3,704
Total expenditures (lakhs)	30.038	-0.027	0.977	-1.108	0.229	1.081	0.196	3,704
Share of expenditures, wages	71.459	-0.884	0.124	-0.639	0.212	-0.245	0.658	3,704
Share of expenditures, admin	0.001	0.001	0.623	-0.001	0.248	0.001	0.223	3,704
<i>p-value of F-test of joint orthogonality</i>			0.28		0.79		0.59	
<i>p-value of F-test of joint orthogonality on all covariates</i>			0.33		0.71		0.98	
<i>p-value of Likelihood ratio test on multinomial logit with and without all covariates</i>	0.96							

Note: All regressions include block administrative level fixed effects to account for randomization design. Standard errors are clustered by block to account for correlation within block.

To test for parallel trends, Figure 3 shows an event study of total expenditures with lags and leads around the month of announcement and including only ‘*Anticipating2ndAudit* - Year 2’ as a control for anticipatory behavior.²⁰ There is no statistically distinguishable trend during the months before the announcement (p -value = 0.45). This lends credibility to our difference-in-differences approach. During the months following the announcement, expenditures decline which will be explored in Section 5. Additionally, there is no evidence of pre-trends for wage and material expenditures (p -value = 0.19 and 0.99, respectively; see Appendix Figure C.3).

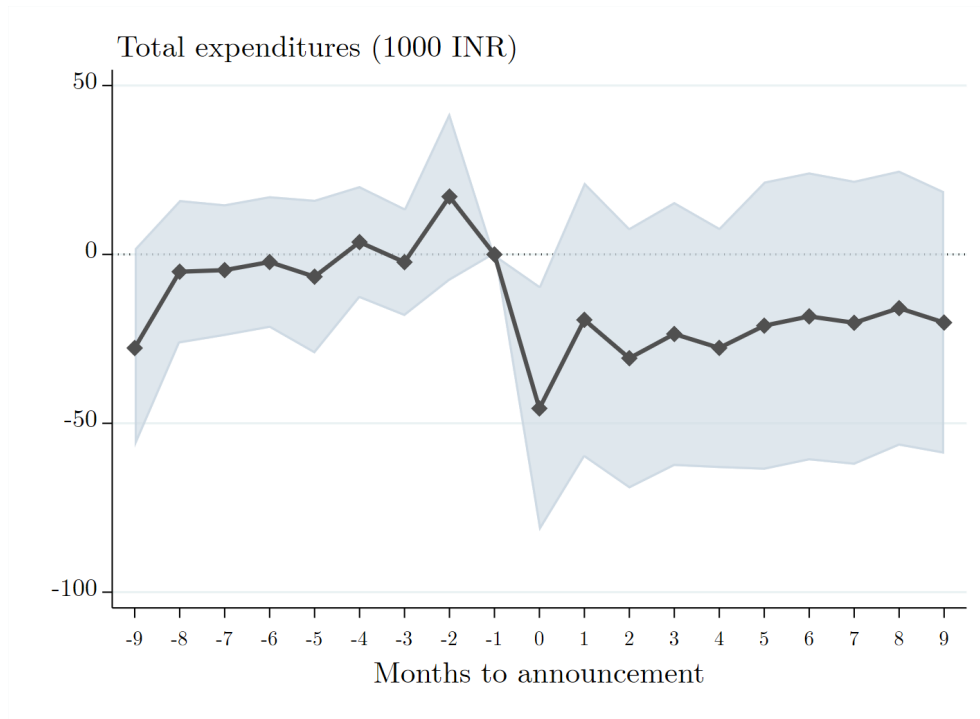


Figure 3: Announcement event study for total expenditures (1000 INR). The omitted category in this regression is the one month lead before the announcement. The regression includes GP and district month-year fixed effects. Standard errors are clustered by block.

For robustness, I estimate the same event study excluding variables capturing potential anticipatory behavior of the first audit and also cannot reject the null

²⁰Recall that ‘*Anticipating2ndAudit* - Year 1’ is the reference group. See discussion in Section 4.1.

(Appendix C.1). Due to potential anticipatory behavior in subsequent waves, we should not expect parallel trends between GPs in Wave 2 and Wave 3 without controlling for their horizons of anticipation. Analyzing the event studies by wave, there is evidence of pre-trends leading up to the Year 2 announcement. This supports our primary difference-in-differences specification in Equation 1 which accounts for horizons of anticipating the first audit. These findings also validate our interest in estimating anticipatory behavior.

4.4 Inferring deterrence from administrative data

The main analysis of this paper estimates the effect of the monitoring policy on total expenditures. Not misappropriated expenditures, a measure we cannot observe. I argue that changes in expenditures driven by the monitoring policy are a reasonable proxy for *changes* in misappropriated expenditures.

Bureaucrats misappropriate NREGS public funds by over-reporting employment (Niehaus and Sukhtankar, 2013; Banerjee et al., 2020; Muralidharan et al., 2016). Audit reports suggest expenditures can be misappropriated through material procurement. Example issues from audit reports include fake receipts, higher-than-expected prices for low-price goods, or procured materials are missing from the worksite. Interviews and media outlets provide anecdotal evidence that corrupt bureaucrats will expense overreported labor for manual-labor-intensive projects and use machines to complete the public projects instead.²¹ Auditors have also documented the use of machines to complete otherwise non-existent but billed projects.

An independent verification of employment and material procurement over time would be the ideal measure of the effect of the monitoring policy. It would allow us to compare reported versus actual outcomes, and measure overreporting. But, this information is ex-post unavailable, as are third-party sources of verification over time. Even with the necessary resources, routine (e.g. monthly) verification surveys is subject to measurement error. For instance, if bureaucrats perceive the survey to be like an audit, bureaucrats may respond accordingly. This would defeat the

²¹E.g. see news article [here](#) describing how corruption prevents beneficiaries accessing benefits of the program.

attempt to measure misappropriated expenses under varying expectations of the likelihood of an audit. In addition, Appendix C.3 emphasizes the importance of using monthly data to study the response of bureaucrats to the monitoring policy over time. It shows that measures of annual performance are too coarse to draw conclusions about bureaucrats’ strategic adjustments to the monitoring policy from year to year. To my knowledge, the administrative data are the only data that provides measures of GP by month performance for the duration of the study period.

Other studies construct proxies for over-reported employment by estimating the share of fake households under NREGS (Niehaus and Sukhtankar, 2013; Banerjee et al., 2020; Muralidharan et al., 2016). Banerjee et al. (2020) do so at scale by comparing administrative records of working households with households documented in the 2012 Socioeconomic Caste Census (SECC).²² A similar exercise in our empirical setting would not work for two reasons. First, during the beginning of our study period, almost all (99.95%) wage payments were made directly to bank accounts. This follows a program shift from paying wages in cash to direct deposits in bank accounts. With challenges to opening fake bank accounts, this suggests that payment of overreported wages is less likely due to fake workers and more likely due to collusion with or coercion of beneficiaries to overreport employment. Second, population changes since 2012 would make the comparison less reliable. For these reasons, estimating the share of fake households with the SECC is not suitable here.

To address these challenges, I study changes in monthly expenditures in response to the monitoring policy. Random assignment of audits isolates the effect of incentives created by the policy on expenditures. To be clear, let c_p denote the true share of total expenditures that is misappropriated under a given audit policy p . Suppose $p \in \{0, 1\}$ where 0 denotes no audit policy and 1 denotes some audit policy. Our goal is to make inferences about $c_1 - c_0$ using administrative reports on expenditures. Let γ_p represent expenditures from the administrative data under policy p . γ_p aggregates honest and misappropriated expenditures under the program.

We need to rule out whether changes in p , holding all else equal, also changes

²²Imbert and Papp (2011) use another approach to compare administrative reports of NREGS employment with household survey data from the NSS on reported employment on public works, but this particular survey has not been released in recent years.

behavior in honest expenditures. If we can rule out other mechanisms, then using $\gamma_1 - \gamma_0$ is a viable strategy to make inferences about changes in misappropriated expenditures, $c_1 - c_0$, due to the policy change. I test hypotheses on whether estimated changes in expenditures in response to the audit policy are consistent with financial misappropriation or not. This allows us to refine our interpretation of the mechanisms driving the estimated effect of the audit policy.

Finally, the measures of program output for analysis are total, wage, and material expenditures. This restricts our study of corrupt behavior to financial misappropriation. I do not have measures for other activity where the bureaucrat unlawfully leverages their position for personal gain. For example, approving public projects to construct private assets for a household as part of a collusive agreement. Furthermore, since data are at the GP-level, I can only study the collective actions of GP-level bureaucrats.²³

5 The Impact of Changing Expectations

This section shows that the effect of the audit policy on deterrence of misappropriated expenditures is elastic to varying expectations of the likelihood of an audit. All results examine the effect of the policy on total expenditures alongside the effect on wage and material expenditures for insights into the response of bureaucrats. I provide evidence we can interpret these effects as changes in misappropriated expenditures. Anticipatory effects are strongest in deterring misappropriated expenditures when one is almost certain of an audit, while the response under less certainty is smaller. On the other hand, deterrence does not come without substitution across time and method of misappropriated expenditures.

5.1 Bureaucrats' response to changing likelihood of audit

Table 2 presents difference-in-differences estimates: Panel A shows Equation 1 and Panel B shows Equation 2 disaggregating behavior during *Post1stAnnounce* into mutually exclusive groups. The outcomes for both panels are total, wage, and

²³See Section A.1 for description of bureaucrats at the GP level.

material expenditures. As discussed in Section 4.1, the reference group for regressions in both panels is the horizon of anticipating the first audit in Year 1 (*Anticipating1stAudit* - Year 1).

There is a decline in expenditures after learning of selection for audit. Table 2 Panel A Column 1 shows a 7% decline (-17.9/269.5, p -value=0.054) relative to baseline total expenditures once a GP learns about selection for their first audit (*Post1stAnnounce*). This effect is driven by a drop in wage expenditures (Column 2).²⁴

Expenditures are not substantially affected when GPs anticipate with low to moderate likelihood that they will be audited. That is, expenditures while Wave 3 GPs anticipate their first audit in Year 2 (*Anticipating1stAudit* - Year 2) increase, but the difference is not significant. Notably, this estimate is driven by a drop in wage expenditures (Panel A Column 2) and an increase in material expenditures (Panel A Column 3) on average, though estimates are not statistically significant. However, if we turn to Panel B, estimates for *Anticipating1stAudit* - Year 2 are qualitatively consistent with Panel A, except the increase in material expenditures is significant and 24% (p -value=0.02) of baseline material expenditures. Total expenditures increase on average by 5% (p -value=0.15) during this period, but the difference is again not statistically significant. Similarly, consider behavior the year after Wave 1 receives their audit (*Anticipating2ndAudit* - Year 2), when it is unlikely they will be a second time as first audits continue to roll out. There is a 1.7% (p -value=0.73) decline in total expenditures coupled with a decline in wage and increase in material expenditures. This suggests bureaucrats resume business as usual when they can reasonably expect to not be audited next year.

Expenditures decline substantially when GPs anticipate with high likelihood that they will be audited. As the roll-out completes in Year 3 and the likelihood of a second audit in Year 4 is extremely likely (*Anticipating2ndAudit* - Year 3'), there is a 15% (p -value=0.003) drop in total expenditures. This is driven by a 15% (p -value<0.0001) decline in wage expenditures (or 14% decline in person-days of employment) and a 13% (p -value=0.24) statistically insignificant decline in ma-

²⁴Recall that identification of *Post1stAnnounce* comes from accounting for behavior captured by *Anticipating1stAudit* - Year 1.

Panel A: Difference-in-differences			
	<i>Expenditures (1,000 INR):</i>		
	(1)	(2)	(3)
	Total	Wages	Materials
<i>Anticipating1stAudit</i> - Year 2	3.13 (10.63)	-9.07 (5.62)	12.20 (8.19)
<i>Post1stAnnounce</i>	-17.91* (9.25)	-15.43*** (4.90)	-2.48 (7.15)
Observations	233,760	233,760	233,760
Baseline mean	269.5	187.1	82.39
Adj. R-squared	0.40	0.47	0.19

Panel B: Difference-in-differences with <i>Post1stAnnounce</i> disaggregated			
	<i>Expenditures (1,000 INR):</i>		
	(1)	(2)	(3)
	Total	Wages	Materials
<i>Anticipating1stAudit</i> - Year 2	13.87 (11.33)	-5.98 (6.05)	19.84** (8.58)
<i>Post1stAnnounce</i> , disaggregated:			
Before <i>1stAudit</i>	-18.83** (9.46)	-10.21** (5.04)	-8.62 (7.14)
Month of <i>1stAudit</i>	-42.07*** (11.68)	-35.33*** (5.99)	-6.74 (9.30)
After <i>1stAudit</i>	-8.42 (11.26)	-17.54*** (5.62)	9.12 (8.70)
<i>Anticipating2ndAudit</i> - Year 2	-4.78 (13.83)	-14.54* (7.67)	9.76 (9.07)
<i>Anticipating2ndAudit</i> - Year 3	-39.38*** (13.30)	-28.06*** (7.28)	-11.32 (9.02)
Observations	233,760	233,760	233,760
Baseline mean	269.5	187.1	82.39
Adj. R-squared	0.40	0.47	0.19

Table 2: Effect of stages of the monitoring policy on Bureaucrats' response in program expenditures. This table estimates the main differences-in-differences specification for three outcome variables: total (wages + materials), wage, and material expenditures. The regressions in Panel A estimate Equation 1 and the regressions in Panel B estimate Equation 2 which includes variables disaggregating behavior during *Post1stAnnounce* into mutually exclusive groups. All regressions include district-month-year and GP fixed effects. Standard errors are clustered by block. The omitted category is the horizon of anticipating one's first audit during Year 1 (*Anticipating1stAudit* - Year 1). The baseline is the mean from the beginning of the panel (two years prior to first audits) up to and including the period captured by *Anticipating1stAudit* - Year 1. This longer period is included in the baseline to average out seasonal variation in expenditures. *** p<0.01, ** p<0.05, * p<0.1

terial expenditures relative to the reference group.²⁵ Similarly, when the audit is actually occurring and concurrent performance is subject to monitoring ('Month of *1stAudit*'), there is a 16% (p -value<0.0001) decline in total expenditures largely driven by a decline in wage expenditures. These results are robust to tests for spillover effects during the month your neighbors receive an audit (Appendix D.2). The estimates for '*Anticipating2ndAudit* - Year 3' and 'Month of *1stAudit*' are both statistically different from anticipatory behavior captured by '*Anticipating1stAudit* - Year 2' and '*Anticipating2ndAudit* - Year 2'. This suggests the effect from the policy on deterring misappropriated expenditures is more responsive when there is greater certainty about the likelihood of an audit.

Of note, the effects of anticipating the second audit on total expenditures are identified when there are no persistent effects from the audit (discussed in Section 4.1). The effects on total and material expenditures during the months after the audit ('After *1stAudit*') are insignificant and small relative to the reference group. On the other hand, there is a significant decline in wage expenditures during this period, but this effect is being driven by the decline in wage expenditures experienced 1 month after the audit. After parsing out this effect, there is no effect 2+ months after the audit on wage expenditures that is statistically distinguishable from 0 (p -value = 0.25). What happens the months following the audit is explored further in Section 5.3. These results suggests the estimates for *Anticipating2ndAudit* on total expenditures are not confounded with persistent effects from the audit.

Altogether, the results from Table 2 Panel B show that bureaucrats response in program expenditures are more responsive when expectations of the likelihood of an audit are high. Can we interpret these estimates as changes in misappropriated expenditures? The following set of results show that decreases in expenditures can be interpreted as deterrence and increases in expenditures can be interpreted as displaced corrupt activity.

²⁵We also estimate Equation 2 where post-audit anticipatory behavior in Year 3 ('*Anticipating2ndAudit* - Year 3') is disaggregated by Wave 1 and Wave 2's responses. I find that the average decline in total expenditures is 10% and 15% for Waves 1 and 2, respectively, and their responses are not statistically distinguishable from one another.

	Share of issues detected by auditors, by issue type:					
	Employment misappropriation		Material misappropriation		Project non-existent	
	(1)	(2)	(3)	(4)	(5)	(6)
Wave 3 - Wave 2	-0.037*** (0.011)	-0.022 (0.018)	0.022*** (0.006)	0.020* (0.012)	-0.003 (0.010)	-0.023 (0.022)
Audit manager experience		-0.0001 (0.002)		-0.001 (0.001)		0.002 (0.002)
Wave 2 Mean	0.17	0.17	0.04	0.04	0.09	0.09
Audit manager FE		X		X		X
Observations	2,349	1,761	2,349	1,761	2,349	1,761
Adjusted R ²	0.27	0.36	0.17	0.22	0.24	0.29

Table 3: Audit performance across waves is consistent with differences in behavior while anticipating the first audit. Unit of observation is GP. All regressions include the following control variables include: number of employed households and works to verify; share of employed households and works verified by auditors; and number of auditors. Audit manager experience is measured by number of audits conducted to date. Standard errors are clustered by block. *p<0.1; **p<0.05; ***p<0.01.

5.2 Deterrence and substitution in anticipation

Anticipatory behavior captured by ‘*Anticipating1stAudit* - Year 2’ in Table 2 Panel B can be interpreted as changes in misappropriated expenditures. Using data from audit reports, Table 3 tests for differences in audit performance between Wave 2 and Wave 3 using number of issues on wage and material expenditures documented by auditors. Recall that Wave 3’s anticipatory behavior is estimated with *Anticipating1stAudit* - Year 2 relative to the reference group (*Anticipating1stAudit* - Year 1) which captures Wave 2’s behavior that was audited in Year 2.

The number of issues related to wage misappropriation found across both groups is statistically indistinguishable where Wave 3 on average has fewer issues of wage misappropriation. This is consistent with our estimate from Table 2 Panel B of anticipatory behavior in Wave 3 (*Anticipating1stAudit* - Year 2) of less wage expenditures on average. Additionally, Wave 3 is 50% more likely to be cited for issues related to material misappropriation (as reported in Table 3: 0.06 share of total is-

sues due to material misappropriation for Wave 3 versus 0.04 for Wave 2, $p\text{-val} = 0.08$). This is consistent with the observation in Table 2 Panel B that Wave 3 spent 25% ($p\text{-val} = 0.02$) more on materials in Year 2 in anticipation.

Should a 25% increase in material expenditures (estimated by ‘*Anticipating1stAudit* - Year 2’ in Table 2 Panel B) convert to a larger increase in detection of issues related to material expenditures? This could depend on many things, including auditors’ ability to detect misappropriated material expenses. With material misappropriation issues making up a 4-6% share of total issues, fake receipts may be harder to detect or materials are a small part of misappropriated expenditures. Anecdotal, in anticipation of audits, bureaucrats have been known to use machines to construct projects that do not physically exist, but where money has been spent. Table 3 shows that Wave 3 has on average 22% fewer issues related to whether the project exists compared to Wave 2, although the difference is not statistically significant. Nevertheless, this could explain part of the surge in material expenditures contributing to an increase in misappropriated expenditures.

These results lend credibility to our interpretation that the estimated differences in bureaucrat behavior during periods of anticipation are interpretable as changes in misappropriated expenditures. Moreover, an average decrease in wage expenditures and employment misappropriation issues coupled with an average increase in material expenditures and material misappropriation issues shows that bureaucrats substitute across methods of misappropriating expenditures for personal gain. The shift from misappropriating wages to materials is specific to this context and likely depends on relative rates of detection by auditors. Audit reports show that issues related to employment misappropriation are more than twice as likely to be documented compared to material misappropriation (13.9% vs 6%, respectively).

5.3 Deterrence and substitution during the audit

This section shows how the decline in expenditures driven by a decline in employment during the months around the audit, estimated by ‘0 – 1 Months After *1stAudit*’ in Table 2 Panel B, can be interpreted as changes in misappropriated expenditures. Bureaucrats may be deterred from misappropriating expenditures

because concurrent performance is being evaluated by auditors (see discussion in Section 2.2). However, the actual act of experiencing an audit introduces potential confounders to interpreting bureaucrat behavior as changes in misappropriated expenditures. Two alternative mechanisms potentially confound our interpretation and are tested: a disruptive audit leads to difficulties completing usual tasks or multi-tasking issues (Holmstrom and Milgrom, 1991); and the audit helps bureaucrats learn and improve productivity (Arrow, 1962; Syverson, 2011).²⁶

There are two aspects to the multi-tasking issue while auditors are present: (1) a disruptive audit keeps bureaucrats from getting honest work done; and (2) a disruptive audit keeps bureaucrats from misappropriating expenditures as they try to minimize detection while auditors are present.²⁷ While the audit process is designed to prevent the audit from being disruptive to honest work (see discussion in Section 2.2), the former is still a potential concern as a confounder. The latter is not because it is an aspect of the bureaucrat’s multi-tasking problem that contributes to changes in misappropriated expenditures.

To disentangle among potential mechanisms, I estimate event studies to examine responses around the time of audit:

$$y_{it} = \alpha_i + \alpha_{dt} + \textit{Anticipating1stAudit}'_{it}\beta + \sum_{k \in \tau} \delta^k \textit{1stAudit}^k_{it} + \epsilon_{it} \quad (3)$$

where $\textit{1stAudit}^k_{it}$ is an indicator taking the value 1 if i is k months from audit at time t . As discussed in Section 4.1, the set of fixed effects account for potentially

²⁶Another possible explanation is that the decline in employment is demand (beneficiary) driven and may have been caused by the audit. If this were the case, it would not reconcile with the fact that employment began to decline a month before auditors arrived where it seems unlikely for beneficiaries to adjust behavior in advance of the audit. If they did anticipate the occurrence of the audit, then we might also expect, but do not observe, behavioral adjustments once the announcements were made and prior to the audit. Furthermore, there is ample documentation of citizen complaints in the audit reports which make it unlikely that the audit deterred households from seeking employment or the benefits they are entitled, or reduced their need to be employed. The audit is likely to affect the behavior of bureaucrats over the behavior of citizens, as punishment from audits sought to punish deviations in implementation and improve access to resources for intended beneficiaries.

Another potential, but unlikely, explanation is that the bureaucrat increased their effort in the presence of auditors. We should expect to see employment increase because more honest employment is reflective of better performance.

²⁷Field interviews with beneficiaries confirm that this is potentially a widespread problem.

endogenous timing of audits due to scheduling constraints of the audit agency.

Figure 4 presents the leads and lags of the regression specified in Equation 3 for total, wage, and material expenditures. Relative to 10 and more months prior to audit (the omitted category), total expenditures decrease by 9-12% (p -value $\in [0.003, 0.013]$) during the month of the audit, proceed to increase by 3-6% (p -value $\in [0.16, 0.46]$) 2-4 months after the audit, and revert to pre-audit levels afterward. This effect is driven by a 17% decline (p -value < 0.0001) in wage expenditures during the month the audit and an 24-30% increase (p -value $\in [0.003, 0.013]$) in material expenditures 2-4 months after the audit.

We can rule out improvements in productivity by learning from the audit as a confounder. Anything learned should persist, but the observed changes are specific to the months around the time of audit. F -tests cannot rule out the hypothesis of equal trends for all outcomes during the periods before and after the months where bureaucrats respond to the occurrence of the audit (p -value $\in [0.14, 0.95]$).

There are several reasons why multi-tasking issues do not explain the decline in expenditures during the month of the audit. First, while the presence of auditors affects employment and material procurement, it does not affect efficiency in making wage payments. If multi-tasking explains the decline in employment, then it may also affect other tasks. Delays are a low-cost task for bureaucrats to shirk while auditors are present. They are a lagged indicator of performance and current measures of delay are not detectable by auditors. Studying delays in making wage payments also allows us to disentangle GP bureaucrat effort from incentives to be corrupt.²⁸ Delayed payments are often cited to reflect a lack of a core competency of the government in targeting resources²⁹, and not a known mechanism through which bureaucrats misappropriate finances.

There is no statistically significant change in average days of delayed payment per household during the month around the time of audit; the average increase

²⁸Days of delayed payment is the number of extraneous days it takes to process a payroll (Banerjee et al., 2020; Narayanan et al., 2019). Once the payroll is processed, the central government deposits wages to beneficiary bank accounts. Appendix D.4 explores delays in greater detail, and the conclusion that audits do not affect delays remains unchanged.

²⁹Aggarwal, A. (2017). *Ten Ways MGNREGA Workers Do Not Get Paid*. Economic&Political Weekly

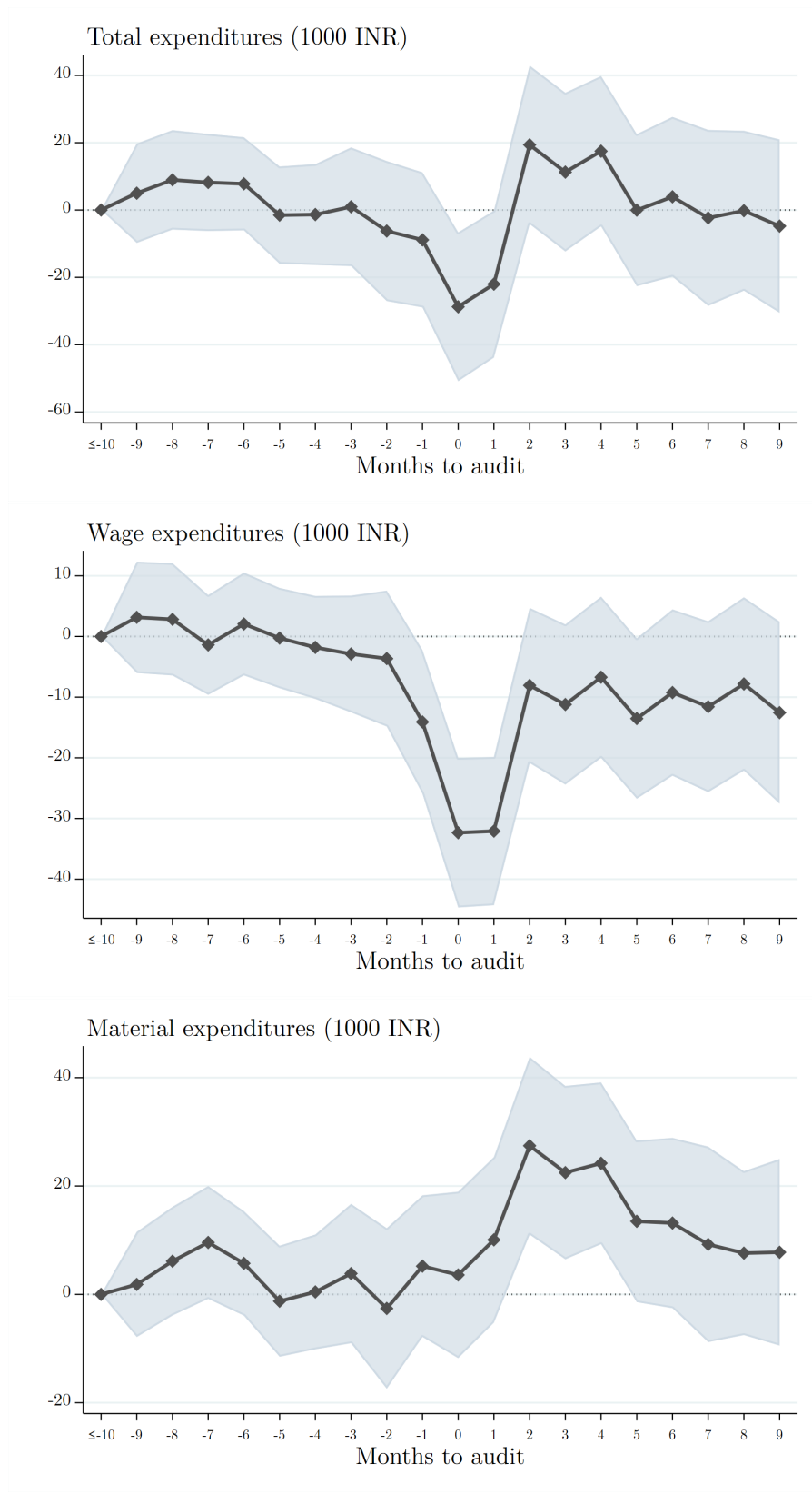


Figure 4: Changes in expenditures around the time of audit (Equation 3). The omitted category is 10 or more months before the audit. The raw mean of the omitted category is 272, 187, and 85 (1,000 INR) for total, wage, and material expenditures, respectively. The regressions include GP and district month-year fixed effects. Standard errors are clustered by block.

during the month of audit is less than one day of delayed payment per household (Appendix Figure D.5). This provides a piece of evidence that multi-tasking issues during the audit are not driving the measured decline in employment and increase in material procurement.

Second, bureaucrats' responses in the months around the audit are unlikely to have impacted program output, providing further evidence that multi-tasking is not a concern. If we were worried that the decline in labor input during the month of the audit was a decline in real employment and thus would lead to a decline in real output, then we would also expect a corresponding decline in material input.³⁰ During the period prior to the first wave of audits, a 1% increase in material expenditures is associated with a 0.06% increase in person-days of work provided (t -stat = 8.12). If material and labor inputs are positively correlated, then real output will tend to depend on both inputs.

Figure 4 shows no decline in materials procured around the time of audit corresponding to the observed decline in employment. If procurement of materials were not in sync with employment, we would expect a corresponding decline in materials procured some months before or after the decline in employment (if this were a decline in real employment). However, an increase in procured materials is the only observed response occurring the month after the decline in employment. This suggests that real output did not decline during the month of audit and multi-tasking while auditors were present was not an issue.

The findings are consistent when examining employment by whether the project worked on required materials or only required labor. Figure 5 shows that the decline in employment is observed across projects of both types. Moreover, the observed increase in materials 2-4 months after the audit (third figure of Figure 4) does not correspond to an increase in employment on projects requiring materials as shown in Figure 5 (a). Likewise, this suggests that the increase in material expenditures 2-4 months after the audit did not result in an increase in real output.

³⁰The different projects carried out under NREGS have guidelines on the ratio of material to labor expenditures. This ratio is used as a performance indicator for bureaucrats to ensure that unskilled labor is used to execute projects, which is the intention of the program. It is not uncommon for bureaucrats to fake the works funded on the payroll by paying wages for fabricated work and having machines complete the work.

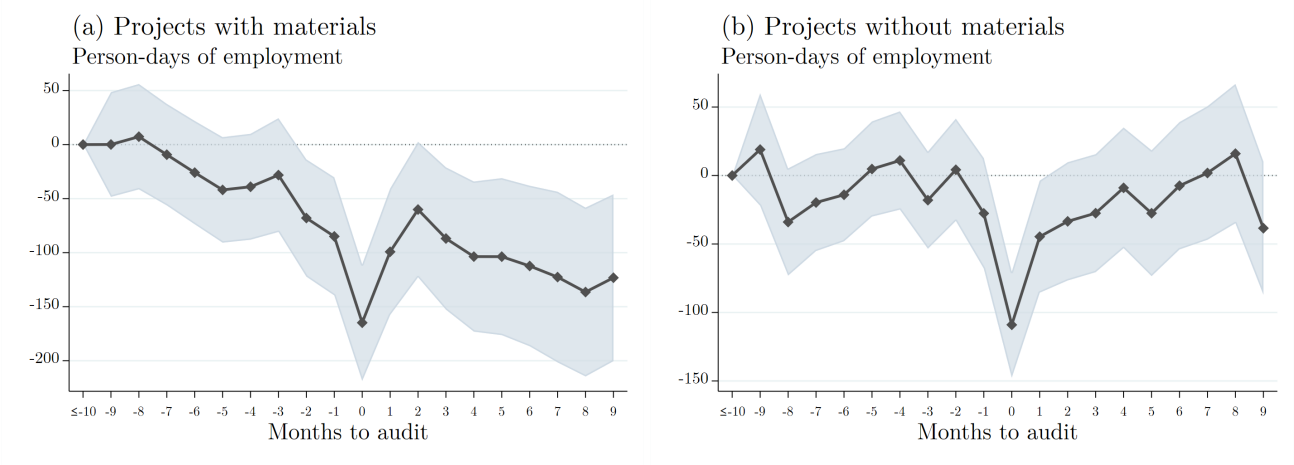


Figure 5: Person-days of employment around the time of audit, by whether the project requires materials. The omitted category is 10 or more months before the audit was conducted. The raw mean of the omitted category is 855 and 310 work-days for projects requiring and not requiring materials, respectively. The regression includes GP and district month-year fixed effects as specified under Equation 3. Standard errors are clustered by block.

Taken together, these results show that the changes in expenditures in the months around the audit left program output unaffected and can be interpreted as changes in misappropriated expenditures. Furthermore, the decline then increase in total expenditures around the time of audit suggests substitution of misappropriation by bureaucrats across time and across method of misappropriation (from wage to material misappropriation). The intertemporal substitution of misappropriated expenditures crowds out the estimated deterrence observed during the month of the audit. These adjustments are consistent with the observed behavior of bureaucrats anticipating their first audit (Section 5.2).

Lastly, results show that changes in misappropriated wages largely affect beneficiary households of upper-caste. The caste system in India has historically driven some social groups to be socially and economically marginalized. Government programs in India take caste into consideration when allocating public resources to address these disparities (Jha and Ambedkar, 2019). Households that are part of historically marginalized groups are categorized into Scheduled Caste, i.e. Dalits, or Scheduled Tribe, i.e. non-Hindu local tribal groups. All other households are labeled as 'Other', which comprises of Upper Caste Hindus or Other Backward Castes.

Other Backward Castes can include other socially marginalized groups, including populations of non-Hindu religions like Muslims. But, they were not historically oppressed by institutions like Dalits have been.

Using household-level NREGS administrative data, an event study of employment by household social group shows that the employment pattern is sensitive to the beneficiary's social group. Figure 6 below shows the audit event study for wage expenditures by household socialgroup: categorized in administrative data as Scheduled Caste, Scheduled Tribe, or Other. Households that tend to be of higher caste or labeled as 'Other' experience a statistically significant 33% decline in employment during the month of audit, which recovers immediately afterward. Scheduled Tribe households experience a statistically significant 16% decline in employment, which gradually recovers by 4 months post-audit. Scheduled Caste households experience a 21% average decline in employment, but this is not statistically distinguishable from 0.

First, these results provide additional evidence against multi-tasking issues as a potential confounding mechanism. The cost of time in processing applications and allocating employment do not differ administratively across beneficiary social group; this points to bureaucrats practicing discriminatory behavior in allocation of employment around the time of audit, unrelated to difficulties in accomplishing tasks. The sharp decline in employment for upper-caste households (who are not expected beneficiaries of NREGS) suggest wages are misappropriated with or using the names of upper-caste households. This exacerbates difficulties in accessing resources for the poor and marginalized groups.

Interpreting the measured decline in employment (Figure 4 as a decline in misappropriated employment, the audit averted false payments equivalent to 18% of the level of employment in the months before the audit. This decline is consistent in magnitude with what Imbert and Papp (2011) measure in Khera (2011). They estimate overreported employment in person-days spent on public works using household survey data to be between 20-29% of the reported person-days of NREGS employment across the states in their sample.

The results from this section suggest that bureaucrat adjustments through de-

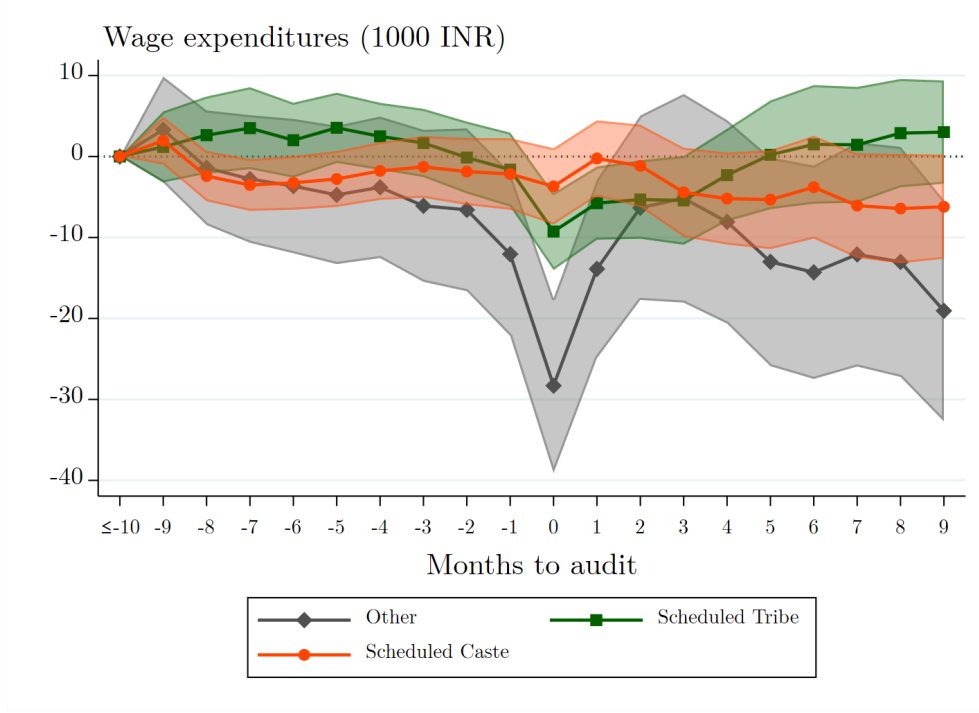


Figure 6: Audit event study for wage expenditures by social group. The omitted category is 10 or more months before the audit was conducted. The mean of the omitted category is 86, 58, and 18 in wage expenditures (1000 INR) for Other, Scheduled Tribe, and Scheduled Caste households, respectively. The regression includes GP-social group, district-month-year, and social group-month-year fixed effects. Standard errors are clustered by block.

creases in misappropriated expenditures around the time of audit are short-lived at best. But, results also suggest lost rents are recovered later as bureaucrats substitute their behavior across time and method of misappropriation.

6 Information Design and Counterfactuals

This section provides conditions under which the estimates of deterrence and assumptions on Bureaucrats' beliefs can be used to estimate the sufficient statistic from the model in Section 3 and determine the optimal monitoring policy. Results show that designing monitoring policies which concentrate incentives at being audited with certainty or not will deter more corrupt behavior than policies that maintain uncertainty of the likelihood of an audit (Section 6.2). Section 6.3 analyzes

the sensitivity of the conclusions on the optimal monitoring policy by relaxing assumptions on both beliefs and estimates of deterrence. Section 6.4 estimates welfare consequences to the Principal under alternative monitoring policies that change the design of information.

6.1 Assumptions for estimating the sufficient statistic

The main results from the model in Section 3 tell us that we care about the shape of bureaucrats' deterrence given their expectations of the likelihood of an audit. This is a sufficient statistic for evaluating the Principal's welfare and determining the optimal design of information. Because of bureaucrats' changing expectations generated by being selected for audit without replacement, the sufficient statistic can be estimated in this empirical setting using the estimates of deterrence from Section 5.1. We can do so without needing to specify additional underlying parameters that determine preferences and constraints.

There are a series of assumptions required to estimate the sufficient statistic in this empirical setting. Each will be discussed in turn.

Comparing bureaucrat responses by varying expectations of the likelihood of an audit may be confounded with changing perceptions of the audit each year. Interpreting the differences in anticipatory and direct responses to be a result of differing expectations requires assuming: bureaucrat perceptions of audit quality remains unchanged across years. This assumption means that the estimated parameters for anticipatory and direct effects of the audit on total expenditures are being driven by varying expectations of the likelihood of an audit. And they are not confounded with perceived changes in quality or credibility of the audit each year.³¹

³¹We can think of the perceived cost of corruption from the audit is a product of the likelihood of incurring a penalty and the penalty itself:

$$\text{Cost to corruption via audit} = Pr(\text{Penalty incurred}) \times \text{Penalty}$$

The probability of incurring a penalty is a function of expectations of whether one will be audited and expectations of whether an audit is credible.

$$Pr(\text{Penalty incurred}) = f(\text{whether Audited, whether audit Credible}, \varepsilon)$$

Empirical tests show it is reasonable to make this assumption (See Appendix D.3). In particular, there are no significant differences in measures of audit quality or bureaucrat response to the stages of the monitoring policy across years. For robustness, this assumption is relaxed in a sensitivity analysis (Section 6.3) when determining the optimal policy in Section 6.2. Furthermore, the test for spillover effects in Appendix D.2 shows anticipatory and direct effects of the audit on total expenditures are unaffected after accounting for concentration of audits within one’s block. This provides additional evidence that the anticipatory effects are driven by changes in expectations of the likelihood of an audit rather than perceptions about the audit to the extent they are informed by peer experiences.

To use the deterrence estimates from Table 2 to estimate the sufficient statistic, it requires assuming: the only payoff-relevant parameter for the bureaucrat for deciding today’s action is their current expectation over the likelihood today’s performance will be monitored. This has two implications. First, the Bureaucrat’s response only depends on their current beliefs and not the policy that generated these current beliefs. That is, conditional on expectations of the likelihood of an audit, the Bureaucrat’s response is policy-invariant. For example, if a Bureaucrat’s beliefs are that they may be audited with probability $\frac{1}{2}$, then their best response would be invariant to whether the underlying audit policy inducing those beliefs was one where units were selected for audit randomly with or without replacement. Second, the Bureaucrat’s response only depends on their current beliefs and not the history of actions or beliefs leading up to the current period. That is, the Bureaucrat’s best response has a Markov property.

Furthermore, the random assignment of expectations of the likelihood of an audit in our empirical setting parallels the primitives of the model underlying the sufficient statistic. In particular, the signal structure chosen by the Principal randomly assigns signals to each receiver, so the difference in Bureaucrat response can only be explained by differences in posterior beliefs (holding all else equal).

Finally, assumptions need to be made on bureaucrats’ beliefs. The deterrence estimates for each anticipatory group from Table 2 Panel B are based on varying expectations of likelihood of an audit driven by stages of the monitoring policy.

To approximate Bureaucrat beliefs on the likelihood of an audit, I draw on the information Bureaucrats’ received through audit agency announcements to infer next year’s audit capacity and the remaining number of GPs to be selected for audit.

Table 4 presents the range of beliefs for each anticipatory group under the various assumptions. Assume a range in beliefs over next year’s audit capacity, allows us to be more agnostic about Bureaucrats’ expectations. Let $K_{\tau+1}$ denote next year’s audit capacity where time τ is the current year. Assume that $K_{\tau+1}$ can be equal to any of the following: (i) $K_{\tau-1}$, last year’s audit capacity; (ii) $\frac{1}{2}(K_{\tau-1} + K_{\tau})$, the average of last year’s and this year’s audit capacity; (iii) K_{τ} , this year’s audit capacity; (iv) $\text{Trend}_{\tau} \times K_{\tau}$, this year’s audit capacity multiplied by recent growth in capacity; and (v) $K_{\tau+1}$, beliefs about next year’s audit capacity are ex-post consistent.

	Probability today’s work will be audited, q				
	<i>Assumptions on next year’s audit capacity, $K_{\tau+1} =$</i>				
	$K_{\tau-1}$	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	K_{τ}	$\text{Trend}_{\tau} \times K_{\tau}$	$K_{\tau+1}$
<i>Anticipating1stAudit</i> - Year 1	0.00	0.07	0.14	0.14	0.39
<i>Anticipating1stAudit</i> - Year 2	0.24	0.44	0.65	1.00	0.92
Month of <i>1stAudit</i>	1.00	1.00	1.00	1.00	1.00
<i>Anticipating2ndAudit</i> - Year 2	0.00	0.00	0.00	0.00	0.04
<i>Anticipating2ndAudit</i> - Year 3	0.67	0.82	0.96	1.00	1.00

Table 4: Assumptions on Bureaucrats’ beliefs on likelihood today’s work will be audited.

Probabilities for *Anticipating1stAudit* - Year 1, *Anticipating1stAudit* - Year 2, *Anticipating2ndAudit* - Year 3 are calculated as follows: The denominator is the remaining number of GPs yet to be audited after learning who is selected for audit that year. The numerator is based on the assumption of future audit capacity ($K_{\tau+1}$) using information on past audit capacity. E.g. when $K_{\tau+1} = K_{\tau}$, expectations for *Anticipating1stAudit* - Year 1 are $\frac{548}{3807} = 0.14$, where 548 is the observed number of

audits conducted in Year 1 and 3807 is the remaining number of GPs to be audited after Year 1 audits.³²

Recent growth used for $Trend_\tau \times K_\tau$ is the grow rate from last year's to this year's audit capacity. These assumptions only draw variation in beliefs for the '*Anticipating1stAudit* - Year 1', '*Anticipating1stAudit* - Year 1', and '*Anticipating2ndAudit* - Year 3' groups. The two group's whose beliefs do not vary are '0 – 1 Months After 1stAudit' and '*Anticipating2ndAudit* - Year 2'. As discussed in Section 2.3, during the month of observation, concurrent Bureaucrat performance is susceptible to detection by auditors. So, while an audit is happening, assume bureaucrats believe today's performance will be audited with probability 1. While for Wave 1 GPs in the year following their audits (Year 2), they believe with probability 0 that today's performance will be audited because the audit roll-out is not near completion.

We will estimate the sufficient statistic under these various assumptions on beliefs from Table 4 and their associated estimates of deterrence from Table 2 Panel B. Put simply, the sufficient statistic is a graph of the beliefs ("x" values) against their associated estimates of deterrence ("y" values).

6.2 The optimal design of information

Figure 7 plots the deterrence estimates for each beliefs assumption. Using the sufficient statistic from the model in Section 6, we can interpret the graph as a plot of the Principal's expected utility as a function of the Bureaucrats' beliefs about being audited. The deterrence estimates from Table 2 Panel B are relative to the reference group, '*Anticipating1stAudit* - Year 1', and reported as a share of the baseline mean. We can interpret this transformation as the relative decline in misappropriated expenditures per month as a share of the reference group level of total expenditures. Higher levels of a decline in expenditures means more deterrence and

³²The special audit in FY2017-18 audited an additional 175 GPs for the first time. Likewise, 21 GPs from Wave 1 were selected for their second audit in Year 3, reportedly by accident. Under assumption $K_{\tau+1} = K_{\tau+1}$, Wave 1 would have anticipated this, hence their beliefs of 0.04. The special and duplicate audit GPs are accounted for when constructing expectations the denominator, but not included in the regressions as described in sample restrictions in Section 2.

higher expected utility for the Principal.

As illustrated in the model of persuasion in Section 3, to determine the optimal policy, we need to ask whether a convex combination between any two points on the curve could achieve a higher expected utility for the Principal. That is, rather than communicating some fixed probability of audit for all Bureaucrats (better when concavities present), could the Principal do better by randomly allocating some to a higher probability group and others to a lower probability group (better when convexities present)?

If we considered the average deterrence response (black curves in Figure 7), then under all belief assumptions the curves are globally convex. In other words, the Principal’s optimal monitoring policy is to communicate and implement a crackdown for a random subset of Bureaucrats and not audit the remaining Bureaucrats. The average elasticity under each assumption ranges from 0.07-0.14, where the elasticity for the increasing part of the graphs with more than 3 points of support range from 0.64-5.7. With a constrained budget, this implies that the optimal signal must place weight on communicating the event that some will be monitored with certainty.

We have at maximum five points of support to interpolate the shape of the sufficient statistic using a piece-wise linear function. A limitation of this approach is that this interpolation may smooth out any local concavities that make dispersed incentives optimal. If we are correct about our theoretical assumption that bureaucrat deterrence is monotone in expectations of the likelihood of an audit, then this is of limited concern.

6.3 Sensitivity analysis and robustness of conclusions

This section evaluates the sensitivity of our conclusions about the optimal monitoring policy to assumptions on: bureaucrat beliefs and constant perceptions of audit quality in Section 6.1. How wrong would our assumptions have to be in order for us to conclude information designed to disperse incentives was better? Technically, we want to know under what belief and deterrence parameters would the sufficient statistic be a line (i.e. Principal indifferent) or locally/globally concave (i.e. Principal prefers dispersed incentives)? I evaluate the sufficient statistic under deviations

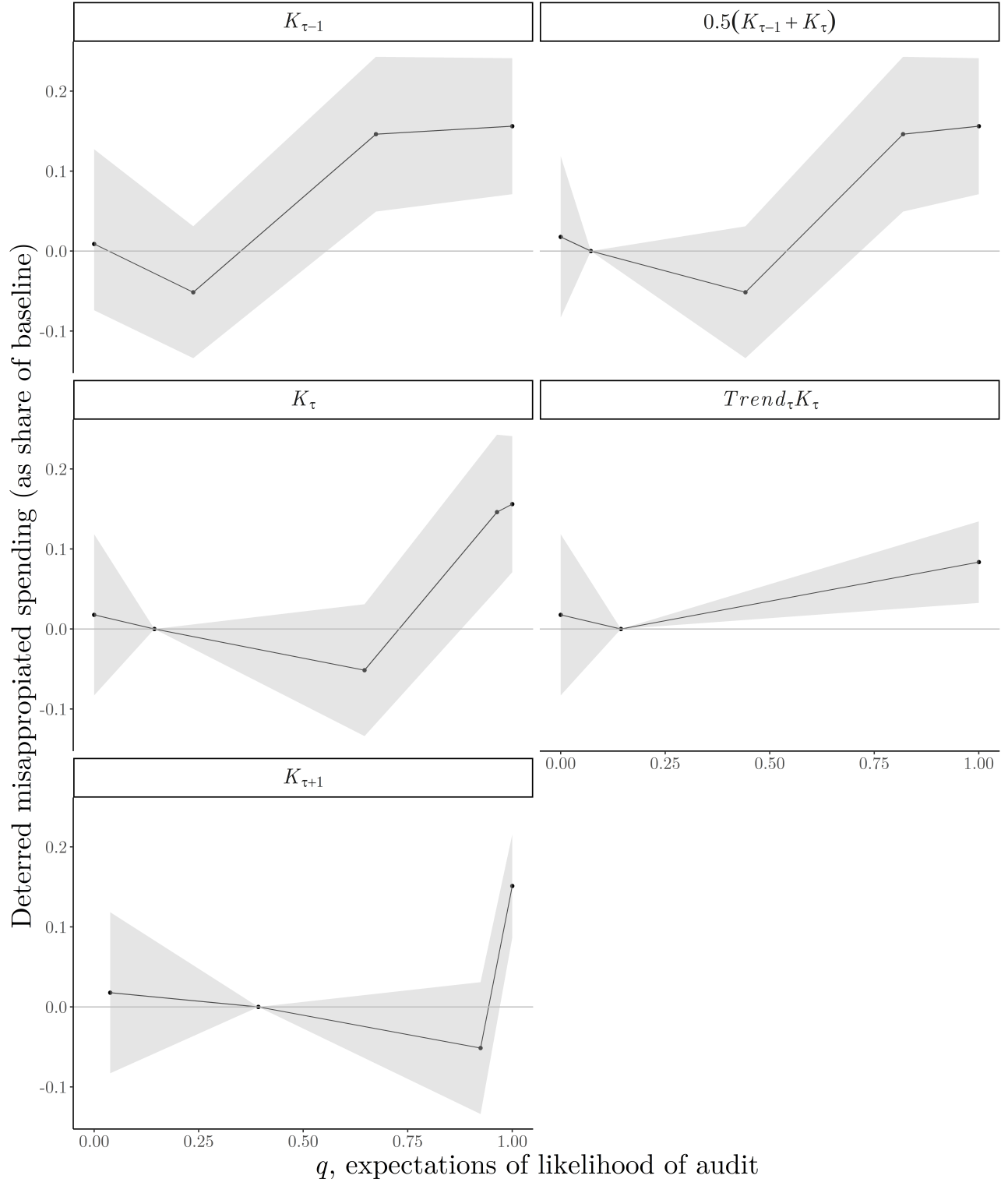


Figure 7: Principal's expected utility as a function of Bureaucrats' beliefs, under different beliefs assumptions. Black dotted line plots the mean of regression estimates with grey shaded area representing the 95% confidence interval for each parameter estimate. Under assumptions $K_{\tau+1} = \{K_{\tau-1}, Trend_{\tau}K_{\tau}, K_{\tau+1}\}$, some anticipatory groups are assumed to have the same beliefs. So, the mean of those groups' deterrence estimates and associated standard errors are plotted.

from belief and deterrence parameters, each in turn. These deviations are designed to challenge our conclusion in Section 6 that concentrated incentives are optimal. The likelihood of these deviations determine the robustness of our conclusion.

The results provide greater certainty that the sufficient statistic is a convex curve as shown in Figure 7 and that the optimal design of information is one which concentrates incentives and implements a crackdown. Simulation results are reported below; details on the simulation procedures and results are in Appendix E.1.

First, I assess deviations from the mean of deterrence estimates and then relax the assumption on constant audit quality. This exercise assesses the likelihood that deviations from the deterrence estimates (Column 4 of Table 2) lead to an alternative conclusion. Alternative conclusions are defined by a sufficient statistic that is weakly/locally concave such that concentrated incentives are not optimal. Using the block bootstrap, I produce the joint sampling distribution of the estimated coefficients. From this, we can compute the likelihood that the jointly estimated coefficients imply a sufficient statistic with an alternative conclusion. The marginal sampling distribution of each estimated coefficient converges after bootstrapping about 10,000 draws. I bootstrap 100,000 samples to report results with 0.001% confidence.

Under the assumptions $K_{t+1} = \{\frac{1}{2}(K_{t-1} + K_t), K_t, K_{t+1}\}$, the probability of realizing an alternative conclusion is less than 0.001%.³³ If $K_{t+1} = K_{t-1}$, there's a 0.02% chance of realizing an alternative conclusion. If $K_{t+1} = K_{Trend_t}K_t$, there is a 30% chance of being indifferent between concentrated versus dispersed incentives. There is not enough information under this assumption to say whether this is a strict preference because there are only 3 points of support. If we exclude deterrence estimates capturing behavior during the month of audit ('Month of 1stAudit'), the results are similar: under assumptions $K_{t+1} = \{\frac{1}{2}(K_{t-1} + K_t), K_t, K_{t+1}\}$ the likelihood is less than 0.005%; about 0.02% likelihood if $K_{t+1} = K_{t-1}$ and 34% likelihood if $K_{t+1} = K_{Trend_t}K_t$.

³³The likelihood under $K_{t+1} = K_t$ of a local concavity is 2% on average, but the local concavity occurs at 'Anticipating2ndAudit - Year 3' where the expectations of likelihood of an audit is 96%. This means that all GPs should have the same information if the prior probability of an audit is at least that high. Given the audit agency's capacity is about 50% at it's highest, dispersed incentives would still not be optimal.

I proceed to relax the assumption that bureaucrat perceptions of audit quality are constant across years. As discussed in Section 5.1, bureaucrat responses are statistically indistinguishable across years for each variable that spans multiple years: *Post1stAnnounce*, ‘Month of *1stAudit*’, and ‘After *1stAudit*’. This supports the assumption that perceptions of audit quality were constant across years. But, to rule out the possibility of a false negative, I relax this assumption.

If the estimated parameter across years is not truly 0, then perhaps we were not powered to detect the true effect size. Given the estimates for *Post1stAnnounce*, ‘Month of *1stAudit*’, and ‘After *1stAudit*’, a power analysis provides the minimal detectable difference for each variable’s estimate by year. The analysis shows that the estimated null effects are extremely likely to be true negatives. Under standard inference rules of rejecting the null with 95% confidence and accepting the alternative with 80% confidence, the minimum detectable effect is at most 2.3 (1,000 INR) in total expenditures across all variables.

To be conservative, I model a bias of ± 5 (1,000 INR) where bias adjustments for each coefficient are assumed to be in the direction of simulating a concave function. With these adjustments, it is still unlikely that dispersed incentives are optimal (1.6% under $K_{t+1} = K_{t-1}$; $< 0.1\%$ under $\frac{1}{2}(K_{t-1} + K_t)$, K_t , K_{t+1} ; 37.5% under $K_{Trend_t}K_t$). The bias would have to be at least ± 9 (1,000 INR) to find a 15% chance of an alternative conclusion under $K_{t+1} = K_{t-1}$ while likelihoods under other belief assumptions are similar in magnitude under the bias of ± 5 (1,000 INR). While the likelihood of an alternative conclusion under $K_{Trend_t}K_t$ is not insubstantial, the magnitude of the likelihood is similar to the results without modeling an explicit bias. Since we are powered to detect differences greater than ± 5 (1,000 INR), the likelihood of a bias greater than ± 9 (1,000 INR) is unlikely.

Second, I assess deviations from assumptions on beliefs. I simulate perturbations from assumed beliefs (bound between 0 and 1) and fix deterrence estimates at their estimated mean, for each bootstrapped sample. The perturbations on beliefs are allowed to take any value in order to simulate a sufficient statistic that leads to an alternative conclusion. With these constraints, I simulate scenarios where the sufficient statistic is linear or concave. Among the bootstrapped samples, only

0.5% of the sample would yield a sufficient statistic that is linear or concave with simulated beliefs within the range of assumed beliefs for each anticipatory group (i.e. assumptions on beliefs from Table 4). Conservative estimates that exclude behavior estimated during the month of audit (coefficient captured by ‘Month of *1stAudit*’) yield a 1.2% likelihood of a sufficient statistic that is linear or concave with simulated beliefs within the range of assumed beliefs.

6.4 Counterfactual signals and welfare consequences

Using the sufficient statistic estimated in Section 6, this section estimates the welfare consequences from the perspective of the Principal if an alternative design of information of the monitoring policy were implemented over the course of the 27 months that the roll-out took place. Welfare is calculated using deterrence estimates for each bootstrapped sample, which allows us to report the variance of the welfare differences across policies. Appendix E.2 provides further details.

The results show that when comparing a monitoring policy with concentrated incentives, i.e. implementing and informing of a crackdown for some and letting others off the hook, to the actual policy of randomization without replacement, the former would on average have deterred 10% more misappropriated expenditures, relative to a total expenditure of USD 226 million under a crackdown policy during this period. Furthermore, when comparing the crackdown policy to one where incentives were dispersed and randomization occurred with replacement, the former policy would have deterred 16% more misappropriated expenditures relative to total expenditures under a crackdown policy during this period. This corresponds to a decline in misappropriated expenditures of about USD 22 million (s.e. = 0.037) when comparing the crackdown policy to actual policy of randomizing without replacement; and USD 37 million (s.e. = 0.037) when comparing the crackdown policy to dispersed incentives policy of randomizing with replacement—all without changing the Principal’s budget for audits.

If we worried about using the estimate of deterrence during the month of audit in the sufficient statistic and excluded it from the analysis, these conservative estimates show that the crackdown policy would on average have deterred 8.7% and

15.4% more misappropriated expenditures compared to the actual and dispersed incentives policies, respectively. This is equivalent to a decline in misappropriated expenditures of USD 19.8 million and USD 35 million (s.e. = 0.037) when comparing the crackdown policy to actual and dispersed incentives policies, respectively.

These potential gains are substantial, especially given wide-prevailing audit standards that it is best to not inform clients of the auditing strategy to maintain unpredictability of when an audit may be conducted.³⁴ The exercise conducted in this empirical setting makes a strong case for evaluating the possibility that atypical audit strategies, like implementing and informing of a crackdown, may yield significant returns at no additional cost to the budget for conducting audits.

7 Conclusion

The monitor’s resource constraints imply that a subset of bureaucrat activity will go unchecked. The resulting strategic behavior from bureaucrats demand that monitoring policies take into account such responses. This paper accounts for bureaucrat responses to uncertainty in timing of audit. Results show that anticipatory effects are strongest in deterring misappropriated expenditures when one is almost certain of an audit, while the response under less certainty is small. In this setting, this implies bureaucrats only respond to incentives from monitoring when the likelihood of an audit is very high. Results also show the unintended consequences of monitoring policies matter. In particular, bureaucrats substitute across time and method of misappropriating expenditures when the likelihood of an audit is substantial.

Under these empirical circumstances, designing monitoring policies which concentrate incentives at being audited or not audited with certainty yields the most deterrence. Although the result is unique to this setting, it contradicts auditing standards that advocate maintaining unpredictability among audit subjects. The findings of this paper emphasize that the best practice depends on the setting-specific elasticity of deterrence with respect to bureaucrats’ expectations of the likelihood

³⁴See as an example: U.S. auditing standards adopted and approved by the Public Company Accounting Oversight Board and the U.S. Securities and Exchange Commission to maintain unpredictability of when audits may occur.

of an audit or another policy parameter of interest in a different setting.

References

- Afridi, Farzana and Vegard Iversen (2014). “Social Audits and MGNREGA Delivery: Lessons from Andhra Pradesh”. In: *India Policy Forum* 10.1, pp. 297–341.
- Arrow, Kenneth J. (1962). “The economic implications of learning by doing”. In: *Review of Economic Studies* 29.3, pp. 155–173.
- Avis, Eric, Claudio Ferraz, and Frederico Finan (2018). “Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians”. In: *Journal of Political Economy* 126.5, pp. 1912–1964.
- Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster (2008). “Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system”. In: *Journal of the European Economic Association* 6.2-3, pp. 487–500.
- Banerjee, Abhijit V. et al. (2020). “E-Governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India”. In: *American Economic Journal: Applied Economics* (forthcoming).
- Bergemann, Dirk and Stephen Morris (Mar. 2019). “Information Design: A Unified Perspective”. In: *Journal of Economic Literature* 57.1, pp. 44–95.
- Bobonis, Gustavo J. and Frederico Finan (2009). “Neighborhood peer effects in secondary school enrollment decisions”. In: *Review of Economics and Statistics* 91.4, pp. 695–716.
- Bobonis, Gustavo J., Luis R. Cámara Fuertes, and Rainer Schwabe (2016). “Monitoring corruptible politicians”. In: *American Economic Review* 106.8, pp. 2371–2405.
- Carrillo, Paul, Dina Pomeranz, and Monica Singha (2017). “Dodging the taxman: Firm misreporting and limits to tax enforcement”. In: *American Economic Journal: Applied Economics* 9.2, pp. 144–164.
- Casaburi, Lorenzo and Ugo Troiano (2016). “Ghost-house busters: The electoral response to a large anti-tax evasion program”. In: *Quarterly Journal of Economics* 131.1, pp. 273–314.
- Centre for Sustainable Employment (2019). *State of Working India*. Tech. rep. Azim Premji University.

- Chetty, Raj (2009). “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods”. In: *Annual Review of Economics* 1.1, pp. 451–488.
- Dieye, Rokhaya, Habiba Djebbari, and Felipe Barrera-Orsorio (2014). “Accounting for Peer Effects in Treatment Response”. In: *AMSE Working Papers*.
- Duflo, Esther, Rema Hanna, and Stephen P. Ryan (2012). “Incentives work: Getting teachers to come to school”. In: *American Economic Review* 102.4, pp. 1241–1278.
- Duflo, Esther et al. (2013). “Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india”. In: *Quarterly Journal of Economics* 128.4, pp. 1499–1545.
- Eeckhout, Jan, Nicola Persico, and Petra E. Todd (2010). “A theory of optimal random crackdowns”. In: *American Economic Review* 100.3, pp. 1104–1135.
- Ferraz, Claudio and Frederico Finan (2008). “Exposing corrupt politicians: The effects of Brazil’s publicly released audits on electoral outcomes”. In: *Quarterly Journal of Economics* 123.2, pp. 703–745.
- (2011). “Electoral accountability and corruption: Evidence from the audits of local governments”. In: *American Economic Review* 101.4, pp. 1274–1311.
- Finan, F, B A Olken, and R Pande (2017). “Chapter 6 - The Personnel Economics of the Developing State”. In: *Handbook of Economic Field Experiments*. Ed. by Abhijit Vinayak Banerjee and Esther Duflo. Vol. 2. Handbook of Economic Field Experiments. North-Holland, pp. 467–514.
- Gerardino, Maria Paula, Stephan Litschig, and Dina Pomeranz (2017). “Can Audits Backfire? Evidence from Public Procurement in Chile”.
- Holmstrom, Bengt and Paul Milgrom (1991). “Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design”. In: *Journal of Law, Economics, and Organization* 7.Special Issue, pp. 24–52.
- Imbert, Clément and John Papp (2011). “Estimating Leakages in India’s Employment Guarantee Using Household Survey Data”. In: *The Battle for Employment Guarantee*. Oxford University Press.

- Jha, Manish K. and B. R. Ambedkar (2019). *Annihilation of Caste (The Annotated Critical Edition)*. Vol. 54. 1, pp. 158–163.
- Kamenica, Emir (2019). “Bayesian Persuasion and Information Design”. In: *Annual Review of Economics* 11, pp. 249–272.
- Kamenica, Emir and Matthew Gentzkow (2011). “Bayesian Persuasion”. In: *American Economic Review* 101.6, pp. 2590–2615.
- Khera, Reetika (2011). *The Battle for Employment Guarantee*. Oxford University Press, p. 264.
- Lalive, Rafael and M. Alejandra Cattaneo (2009). “Social interactions and schooling decisions”. In: *Review of Economics and Statistics* 91.3, pp. 457–477.
- Lazear, Edward P. (2006). “Speeding, terrorism, and teaching to the test”. In: *Quarterly Journal of Economics* 121.3, pp. 1029–1061.
- Lichand, Guilherme and Gustavo Fernandes (2019). “The Dark Side of the Contract: Do Government Audits Reduce Corruption in the Presence of Displacement by Vendors?” Working Paper.
- Manski, Charles F. (1993). “Identification of endogenous social effects: The reflection problem”. In: *Review of Economic Studies* 60.3, pp. 531–542.
- Milgrom, Paul and Chris Shannon (1994). “Monotone Comparative Statics”. In: *Econometrica* 62.1, pp. 157–180.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar (2016). “Building state capacity: Evidence from biometric smartcards in India”. In: *American Economic Review* 106.10, pp. 2895–2929.
- Narayanan, Rajendran, Sakina Dhorajiwala, and Rajesh Golani (2019). “Analysis of Payment Delays and Delay Compensation in MGNREGA: Findings Across Ten States for Financial Year 2016–2017”. In: *Indian Journal of Labour Economics* 62.1, pp. 113–133.
- Niehaus, Paul and Sandip Sukhtankar (2013). “Corruption dynamics: The golden goose effect”. In: *American Economic Journal: Economic Policy* 5.4, pp. 230–269.
- Olken, Benjamin A. (2006). “Corruption and the costs of redistribution: Micro evidence from Indonesia”. In: *Journal of Public Economics* 90.4-5, pp. 853–870.

- Olken, Benjamin A. (2007). “Monitoring corruption: Evidence from a field experiment in Indonesia”. In: *Journal of Political Economy* 115.2, pp. 200–249.
- Olken, Benjamin A and Rohini Pande (2012). “Corruption in Developing Countries”. In: *Annual Review of Economics* 4.1, pp. 479–509.
- Rose-Ackerman, Susan and Bonnie J. Palifka (2016). *Corruption and Government: Causes, Consequences, and Reform*. 2nd. Cambridge University Press, p. 642.
- Shleifer, Andrei and Robert W Vishny (Aug. 1993). “Corruption”. In: *The Quarterly Journal of Economics* 108.3, pp. 599–617.
- Sukhtankar, Sandip (2017). “India’s National Rural Employment Guarantee Scheme: What Do We Really Know about the World’s Largest Workfare Program?” In: *Brookings-NCAER India Policy Forum*. Vol. 13, pp. 231–86.
- Svensson, Jakob (2005). “Eight questions about corruption”. In: *Journal of Economic Perspectives* 19.3, pp. 19–42.
- Syversen, Chad (2011). “What determines productivity?” In: *Journal of Economic Literature* 49.2, pp. 326–365.
- Vazquez-Bare, Gonzalo (2017). “Identification and Estimation of Spillover Effects in Randomized Experiments”. Working Paper.
- Wilson, James Q. (1991). *Bureaucracy: What Government Agencies Do And Why They Do It*. Basic Books, p. 468.

Appendix

A Background

A.1 Details on implementation of NREGS in Jharkhand at the GP

At the gram panchayat (GP), there are several bureaucrats responsible for operating NREGS. The president of a GP (*mukhiya*) is an elected official. The president facilitates the process for selecting projects to fund and oversees the allocation of job cards. The secretary of the GP (*panchayat sachiv*) provides job cards, manages employment allocation and wage payments, and uploads administrative data to the NREGS database. Outside of their NREGS responsibilities, the president and secretary also manage other programs and matters at the GP office. The NREGS employment assistant (*gram rozgar sewak*) provides project work to workers, pays wages, and manages NREGS projects. Engineers at a higher administrative level ensure the quality of public projects across the GPs they oversee. The state government appoints the secretary, NREGS employment assistant, and engineer. Finally, the GP hires direct supervisors, often beneficiaries, of project sites. These supervisors take attendance for the payroll and help beneficiaries apply for work.³⁵

A.2 Details on the audit agency

The Social Audit Unit was founded and funded in May 2016. Audits were piloted in 49 GPs in June 2016; then the first wave of audits took place from December 2016 until March 2017 (the end of the fiscal year). The national act describes audits as an important component of NREGS. But, since the program began in 2006, limited resources kept state governments like Jharkhand from implementing a credible audit program. Before the creation of the audit agency, audits were conducted by civil society organizations on an ad-hoc basis or conducted by the bureaucrats who themselves were the object of audit interest. It was not until the

³⁵See the MGNREGA Operational Guidelines 2013 for details on roles and responsibilities.

creation of this audit agency that a credible and systematic audit process was in place for public welfare programs, like NREGS, in the state of Jharkhand.

The audit agency is funded independently of NREGS and managed by a steering committee of various stakeholders across the state government and civil society. There are around 7 stakeholders on the steering committee. The steering committee is largely removed from NREGS implementation. One of the members is the state commissioner responsible for implementing NREGS in Jharkhand; however, there are no reasons to suspect that his participation in the steering committee would compromise quality of audits. As the highest manager accountable for the state's performance in NREGS, he has an incentive to root out corruption with monitoring.

Due to hiring practices and quality assurance mechanisms, it is likely that the audits were conducted at-scale by the audit agency with credibility and integrity. First, audit managers and auditors are hired at competitive salaries where compensation for the lowest-ranked auditor is at least 2 times the minimum wage (from 550 INR per day to a salary of 35,000 INR per month for district-level managers). These salaries are comparable to what research agencies in India pay their surveyors and field research managers. These rates are also competitive compared to auditor rates in other government agencies. While we cannot assume there were no auditors corrupted by bureaucrats, we know that the potential loss of the job from being fired is not insignificant. There is the obvious loss of salary, but also potential hardship in finding another comparable job. This is especially salient for positions requiring higher education levels because those with a high school education or higher made up about 60% of the unemployed working age population in 2018 in India (Centre for Sustainable Employment, 2019).

Furthermore, the audit agency has multiple mechanisms to ensure audit quality. First, they audit at least 5% of the audits they conduct for quality assurance.³⁶ Second, auditors cannot be assigned to audit their home region to prevent potential conflicts of interest. Third, strict guidelines are in place for seeking accommodations and provisions during the week of their stay to audit the GP. In particular, they

³⁶Five-percent is based on set intentions, but there is no data available at this point of the back-checked audits. GPs selected for back-checked audits are determined by field reports of collusion and a random sampling.

do not rely on local bureaucrats to facilitate the logistical aspects of their stay. The auditors setup a home-base during the period of audit at the GP government office, organize their own transportation, and even rent cookware to cook their own food. When they are not conducting audit verification fieldwork, they use the local government office to work, eat, and sleep. This is an intentional feature of the audit process emphasized by the audit agency. Some of the GPs for audit are in very remote areas where it may be hard to find options for lodging and meals. This helps minimize any leverage a local bureaucrat may have by offering their resources and currying favor with auditors. Just as importantly, these guidelines are put in place for fear of tarnishing the integrity of the audit especially as it may be perceived by local beneficiaries, whose incentives to report during the audit can be affected.

A.3 Details on the roll-out schedule and public notices about the audits

Table A.2 summarizes details of the audit schedule for each year of the roll-out of audits.

Table A.2: Audit Schedule, 2016-2019

Fiscal year	Announcement date	# Audits	Duration of audit calendar		
2016-17	29-Dec-16	548	17-Dec-16	–	29-Mar-17
2017-18	2-May-17	1,495	9-May-17	–	21-Mar-18
2018-19	23-Mar-18	2,137	13-Apr-18	–	14-Mar-19

The following describe the formal notices that were publicly disseminated and what could be learned by all bureaucrats:

- Audit agency created, May 2, 2016 - Notice on creation of the audit agency and personnel to be hired to staff the agency.
- Year 1 announcement, December 29, 2016 - Commencement of 548 audits for Wave 1.

Official notice was disseminated on guidelines for conducting the audits and announcement of GPs in Wave 1 of audits to be conducted in the remainder of Year 1. The notice also states that the goal of the audit agency is to eventually be able to audit 50% of GPs every year. This is against the benchmark stated in the 2006 NREGS Act Section 17 that requires all GPs to be audited twice a year.³⁷

The notice does not mention that GPs are being randomly selected. Furthermore, given the notice and 2006 NREGS Act, it is reasonable to expect that audits will be rolled out to all GPs before one can expect to be audited again. Roll-out without replacement is discussed in steering committee meeting minutes which are made public. It is *not* known that the roll-out would take 3 years to complete. In fact, from meeting minutes of the steering committee, there is considerable uncertainty even among the committee about future audit capacity driven by annual budget approval processes contingent on current performance and the audit agency's future capacity to recruit and train a workforce of auditors.

- Year 2 announcement, May 2, 2017 - Commencement of 1,495 audits for Wave 2.
- Year 3 announcement, March 23, 2018 - Commencement of 2,137 audits for Wave 3.

Remaining GPs that have not been audited are being completed in Year 3. It is highly likely that those not selected for Wave 3 audits will be selected for audit in Year 4. About 80% of GPs audited in Year 4 were not audited in Year 3.

Every announcement across the three years states that part of the audit will involve a verification of administrative reports on employment and public projects from the previous FY.

³⁷The commencement of Wave 1 audits started with this notice, along with a press conference and video-conference with all district officials to discuss and disseminate the notice.

B Model of Information Design and Deterrence

B.1 Using the static model to evaluate dynamic settings

Using features of the empirical setting and making reasonable assumptions, we can use the single-period model to evaluate dynamic settings and treat each bilateral sender-receiver game independently. I discuss these assumptions here. Appendix B.2 sets up the dynamic setting that can be evaluated with the static model.

First, in the empirical setting, the announcement is public and all bureaucrats receive the same information. So, the Principal sends a public signal which allows us to treat each bilateral sender-receiver game independently (Kamenica, 2019). In this case, the Principal considers the vector of Bureaucrats’ best responses when determining the optimal signal .

Furthermore, when analyzing the model, we will examine the game between the Principal and an arbitrary Bureaucrat. In particular, we will not enumerate model specifications that capture heterogeneity among Bureaucrats, like ability or propensity to be corrupt; the only parameter that matters is the Bureaucrat’s posterior beliefs. In other words, the difference in Bureaucrat response can only be explained by differences in posterior beliefs (holding all else equal) because the signal structure chosen by the Principal randomly assigns signals to each receiver. This maps well to our empirical setting where assignment to audit and anticipatory beliefs were also randomly assigned. Section 6.1 provides further discussion of mapping the model to the empirical context.

Second, the model assumes that the state space (whether audited or not) and action space (expenditures misappropriated by Bureaucrat) are invariant across time, which allows us to consider a static model (Kamenica, 2019). It is reasonable to assume a time-invariant action space because the scope of the Bureaucrats’ authority and responsibilities are unlikely to change over time. In the context of a monitoring policy, we can also expect the state space to remain unchanged as the relevant output of a monitoring policy is to monitor or not.

B.2 Setup: Dynamic model with multiple receivers

There are N receivers (or Bureaucrats) responsible for the implementation of the workfare program over the course of some finite length of time T , where the set of time periods is $\mathcal{T} = \{1, \dots, T\}$ and indexed by t . The set of receivers is $\mathcal{I} = \{1, \dots, N\}$, indexed by i , and $|\mathcal{I}| = N$. Every period each receiver oversees some amount of expenditures to be allocated denoted by x_{it} and has the technology to extract private rents from x_{it} .

The sender (or Principal) always wants as much of $X_t = \sum_i x_{it}$, the total program expenditures, to go towards realizing the goals of the workfare program as possible. The Principal uses monitoring as a policy to discipline Bureaucrats' behavior where significant penalties are imposed as punishment for being caught extracting private rents. However, the Principal is budget constrained and can only conduct $K < N$ audits every period.

There are two states of the world for Bureaucrat i in period t : x_{it} will be audited (1) or not (0), where $\omega_{it} \in \{0, 1\}$ denotes an element of the state space for i in t . The state space of the game is $\Omega = \{0, 1\}^{NT}$, where we can think of an element of the state space, $\omega \in \Omega$, as $\omega = \{\omega_{1t}, \dots, \omega_{Nt}\}_{\forall t \in \mathcal{T}}$.

Every period, the Principal decides and commits to a signal structure or policy, $\pi : \Omega \rightarrow \Delta(S)$ where $S = \Delta(\Omega)$ is the set of signal realizations over the state space. Each Bureaucrat takes an action $a_{it} \in A = [0, 1]$, which is the share of x_{it} to misappropriate for private gains, conditional on posterior beliefs induced by their signals.

B.3 Proofs

Proposition 1 (Persuasion Works). *The Principal benefits from sending an informative signal (concentrating incentives) if and only if $U(q_0) < \hat{U}(q_0)$. Moreover, if the function is convex at $U(q_0)$, then $U(q_0) < \hat{U}(q_0)$ and an informative signal is better. (Kamenica and Gentzkow, 2011)*

Proof. Let's start with the first statement of the proposition. Suppose that the Principal benefits from sending information, then there exists a signal structure π

that induces a distribution of posteriors which yields payoff $\mathbb{E}_{q \sim \pi(\cdot)}[U(q)] = \hat{U}(q_0)$. We can construct π particularly to yield payoff $\hat{U}(q_0) > U(q_0)$.

Now, suppose that $U(q_0) < \hat{U}(q_0)$, then there is a distribution of posteriors induced by a signal π that achieves payoff $\hat{U}(q_0)$ different from an uninformative signal which achieves $U(q_0)$. Since $U(q_0) < \hat{U}(q_0)$, the Principal prefers to send an informative signal.

In the second statement of the proposition, suppose that $U(q)$ is strictly convex when evaluated at q_0 . Let $q_0 = \lambda q_1 + (1 - \lambda)q_2$ where $\lambda \in (0, 1)$ and q_1, q_2 are in the domain of $U(\cdot)$. By Jensen's inequality, $U(q_0) = U(\lambda q_1 + (1 - \lambda)q_2) < \lambda U(q_1) + (1 - \lambda)U(q_2)$. Since $\lambda U(q_1) + (1 - \lambda)U(q_2)$ is in the set $co(U(q))$ it is an achievable payoff for the Principal. Now, pick q_1, q_2 such that $\lambda U(q_1) + (1 - \lambda)U(q_2) = \hat{U}(q_0) > U(q_0)$. \square

Proposition 2. *The Principal prefers concentrated over dispersed incentives if $U(q)$ is globally convex. The converse is true if $U(q)$ is globally concave.*

Proof. Apply Jensen's inequality as in the above proof. Let $q_0 = \lambda q_1 + (1 - \lambda)q_2$ where $\lambda \in (0, 1)$ and q_1, q_2 are in the domain of $U(\cdot)$. By Jensen's inequality, $U(q_0) = U(\lambda q_1 + (1 - \lambda)q_2) < \lambda U(q_1) + (1 - \lambda)U(q_2)$. Since $\lambda U(q_1) + (1 - \lambda)U(q_2)$ is in the set $co(U(q))$ it is an achievable payoff for the Principal. Now, pick q_1, q_2 such that $\lambda U(q_1) + (1 - \lambda)U(q_2) = \hat{U}(q_0) > U(q_0)$. So the Principal prefers to concentrate incentives by sending signals that induce posteriors q_1 and q_2 , rather than leaving everyone with the prior q_0 where everyone has the same likelihood of audit (dispersed incentives). \square

B.4 Motivating example in detail

The Principal is deciding between randomly selecting N Bureaucrats for audit without replacement over two periods (dispersed incentives) or randomly selecting Bureaucrats for audit with replacement over two periods (concentrated incentives). In every period, the Principal only has the capacity to conduct $M < N$ audits or share $p = \frac{M}{N} \in (0, 1)$ of Bureaucrats. Suppose Bureaucrats have perfect information on the Principal's audit capacity (M) every period.

Let $U(q)$ be deterred financial misappropriation in a given period (reflecting Bureaucrats' optimal adjustments in behavior) as a function of Bureaucrats' expectations of the likelihood of an audit (q). The more likely they are to be audited (when q is higher), the more Bureaucrats adjust. Consider $U(q)$ to be the Principal's value function as a function of Bureaucrats' best response. Assume for simplicity that the Principal is patient. Under randomization without replacement, the Principal's expected utility given Bureaucrats' beliefs and best response over the two periods is:

$$U(p) + \underbrace{pU(0)}_{(a)} + \underbrace{(1-p)U\left(\frac{p}{1-p}\right)}_{(b)}$$

where the first term reflects Bureaucrat behavior in the first period and the last two terms are based on Bureaucrat behavior in the last period. The share p audited in the first period will not be audited in the second period, they know this, and behave accordingly (term (a)). While those waiting to be audited (share $1-p$) believe they will be audited with probability $\frac{p}{1-p}$ and adjust their behavior accordingly (term (b)). On the other hand, if randomizing with replacement, the Principal's expected utility is:

$$2U(p)$$

This gives us:

$$U(p) \leq \underbrace{pU(0) + (1-p)U\left(\frac{p}{1-p}\right)}_{=\tilde{U}}$$

Whether the Principal prefers to randomize with or without replacement depends on this inequality.

Corollary B.2. *When deciding between randomizing with (concentrated incentives) or without replacement (dispersed incentives), the Principal prefers randomization with replacement if $U(q)$ is globally concave. Conversely, the Principal prefers randomization without replacement if $U(q)$ is globally convex.*

Proof. If $U(q)$ is globally concave, then

$$U(p) > pU(0) + (1-p)U\left(\frac{p}{1-p}\right)$$

by definition of a concave function and randomization with replacement is preferred. If $U(q)$ is globally convex, then the reverse is true. \square

B.5 Sufficiency of $U(q)$ for analyzing welfare changes and determining the optimal signal

Let W_1, \dots, W_n be a sample from the probability distribution of the Principal's welfare (as measured by levels of bureaucrat financial misappropriation), $f(w|\theta)$, where θ is a vector of parameters that determine the bureaucrat's decision to misappropriate finances. The sample $\left((W_1, q_1), \dots, (W_n, q_n)\right)$ from the joint distribution $f(w, q)$ where q is bureaucrat expectations of the likelihood of an audit and is randomly assigned. The distribution $f(w|q)$ is sufficient for θ . That is, given information on q , θ provides no additional information on the Principal's welfare and consequently, the optimal signal. This is because random assignment of q from the signal structure π holds all other pay-off relevant parameters in θ equal.

In practice, I estimate statistics for $f(w|q)$ for observed q . This will be used to estimate $U(q)$. With this and the assumption that $U(\cdot)$ is monotonic in q , we can assess changes in welfare as q changes and we can also construct the optimal signal.

C Empirical Strategy

C.1 Parallel Trends and Anticipation By Year

Figure C.2 plots a time series of the average person-days of employment for each wave. It suggests parallel pre-trends between the treatment and control groups leading up to the first announcement (denoted by the first dashed line).³⁸ Furthermore, those being audited experience slight declines in employment as audits are implemented throughout the year.

There are two events for every wave of the audit where the GPs potentially have an incentive to change their behavior as a response: 1) the announcement of the

³⁸The spike in employment from March-June 2016 is likely attributed to severe droughts earlier in the year that led to increased demand for employment during a weak harvest. Since this event occurs prior to the roll-out of audits, I do not consider it a threat to identification.

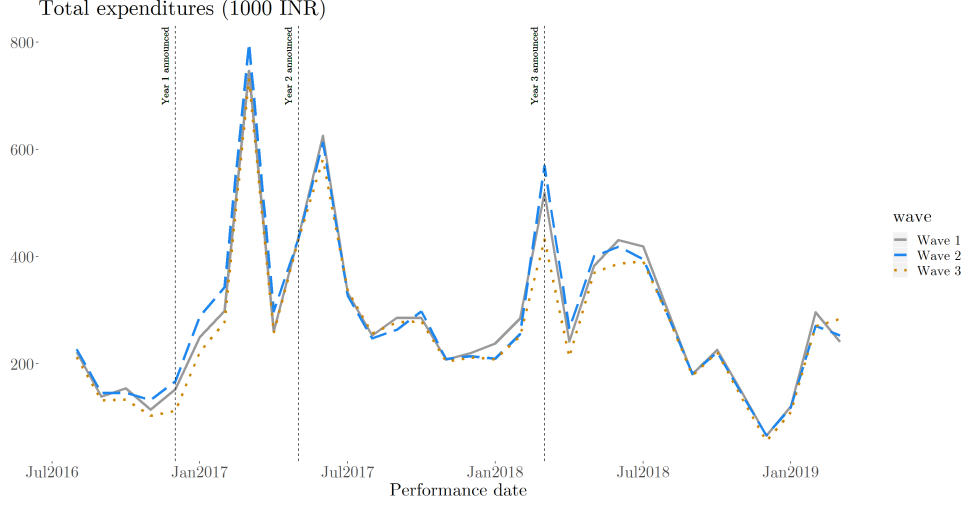


Figure C.2: Raw means of monthly total expenditures, by wave

audit schedule which is simultaneously distributed across all GPs; and 2) the audit itself. Once informed of the audit, bureaucrats in their respective audit group may respond differentially in anticipation of the audit. When estimating the effect of the audit itself, we should not expect parallel pre-trends in the treatment and control group if one group reacts in anticipation of the audit while the other is operating business as usual.

As discussed in Section 4.1, we should expect parallel pre-trends between the groups prior to the Year 1 announcement of Wave 1 audits and can expect parallel pre-trends prior to the announcement in subsequent years after accounting for the horizons of pre-audit anticipation. We can test for violations of parallel trends by estimating the following event study around the time of announcement:

$$y_{it} = \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \sum_{k \in \tau} \delta^k Announce^k_{it} + \epsilon_{it} \quad (C.2)$$

where $Announce^k_{it}$ is a dummy variable taking a 1 if a GP is k months from learning when they will be audited at time t . This vector of dummy variables, indexed by k , comprise our lags and leads to the announcement.

In the months before the announcement, the total expenditures do not seem

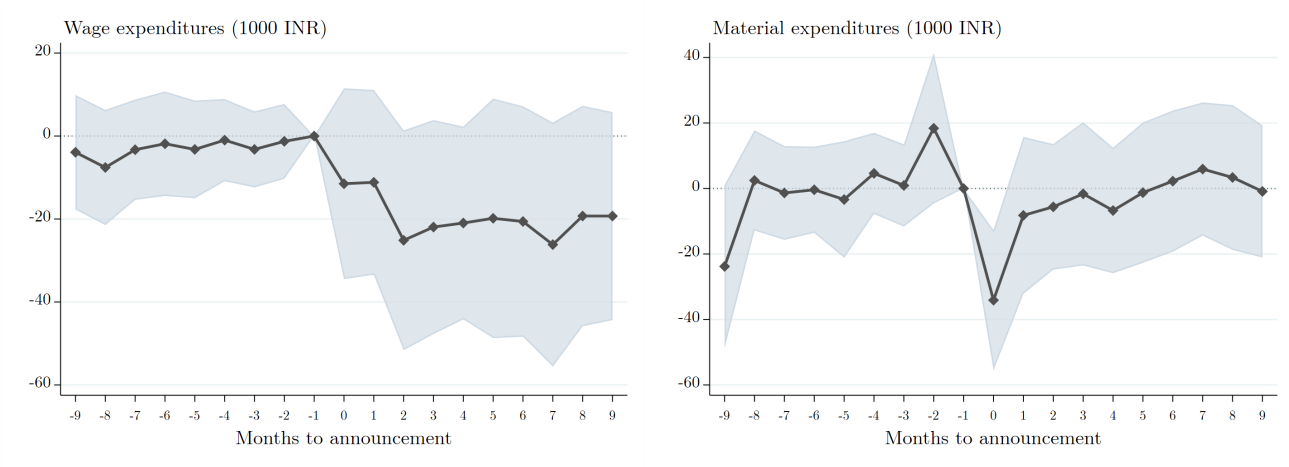


Figure C.3: Announcement event study for wage and material expenditures. The omitted category is the month before the announcement. The raw mean of the omitted category is 179 and 69 (1000 INR) and the p -values of no pre-trends are 0.99 and 0.11 for wage and material expenditures, respectively. The regression includes GP and district month-year fixed effects. Standard errors are clustered by block.

to be following a trend and are statistically indistinguishable from behavior the month before the announcement (p -value = 0.45, Figure 3). This lends credibility to our difference-in-differences approach. Results look similar for wage and material expenditures (p -value = 0.19 and 0.99, respectively) captured in Figure C.3.

The same event study, excluding variables capturing potential pre-audit anticipatory behavior, lead to the same conclusion—that we cannot reject that there are no pre-trends. However, breaking up the estimation by wave, there is evidence of pre-trends leading up to the Year 2 announcement for Wave 2 audits. This supports our primary difference-in-differences specification in Equation 1. It accounts for horizons of anticipatory behavior to not only provide credible estimates of the effect of the announcement and the audit, but also validates our interest in estimating the parameters of anticipatory behavior.

Figure C.4 estimates Equation C.2 but excludes the variables capturing the horizons where GPs potentially anticipate their first audit. We are led to the same conclusion that we cannot reject there is no pre-trend prior to each wave’s announcement (p -value = 0.41).

But, if we examine the same event study on total expenditures by comparing

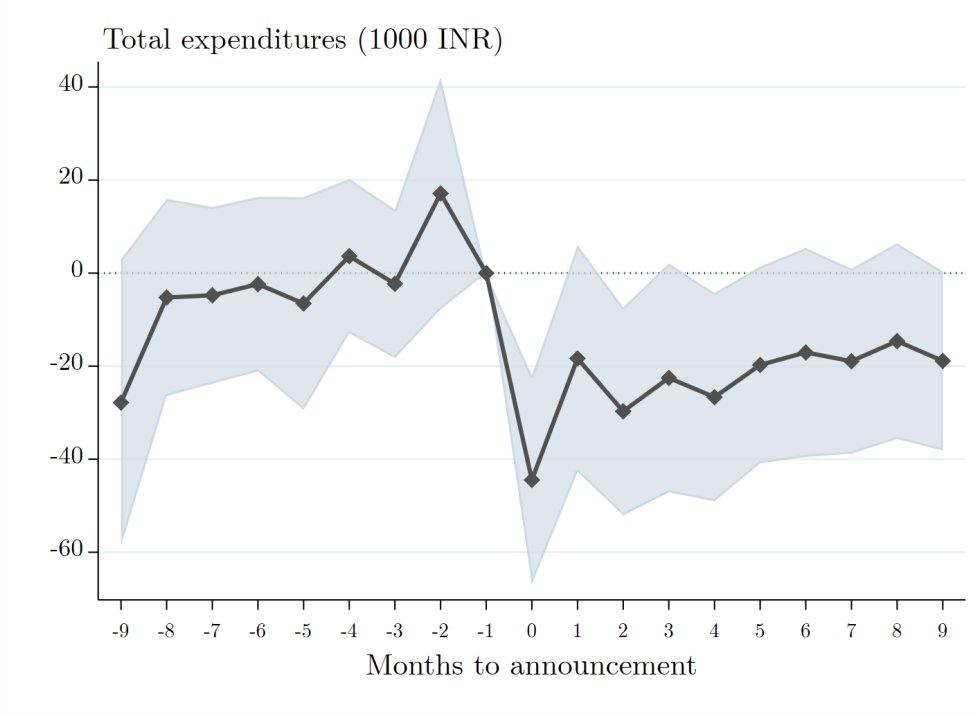


Figure C.4: Announcement event study for total expenditures without controlling for anticipatory behavior. The omitted category in this regression is the one month lead before the announcement. The raw mean of the omitted category is 248 (1000 INR) and the p -values of no pre-trends is 0.34. The regression includes GP and district month-year fixed effects. Standard errors are clustered by block.

wave 1 to those not audited in Wave 1, and Wave 2 to those not yet audited in Wave 2, we observe that we cannot reject a test of pre-trends in the Wave 1 comparison (p -value = 0.35) but we reject pre-trends in the Wave 2 comparison (p -value = 0.01). Since our balance checks support that the GPs selected for audit in each wave were randomly selected, then it is more likely the case that the observed pre-trends in the Wave 2 comparison are attributed to differences in the beliefs over being audited in the next fiscal year leading up to each wave's respective announcement of audit. For this reason, our preferred specification through this paper is to account for horizon of anticipation of one's first audit as specified in Equations 1- 3.

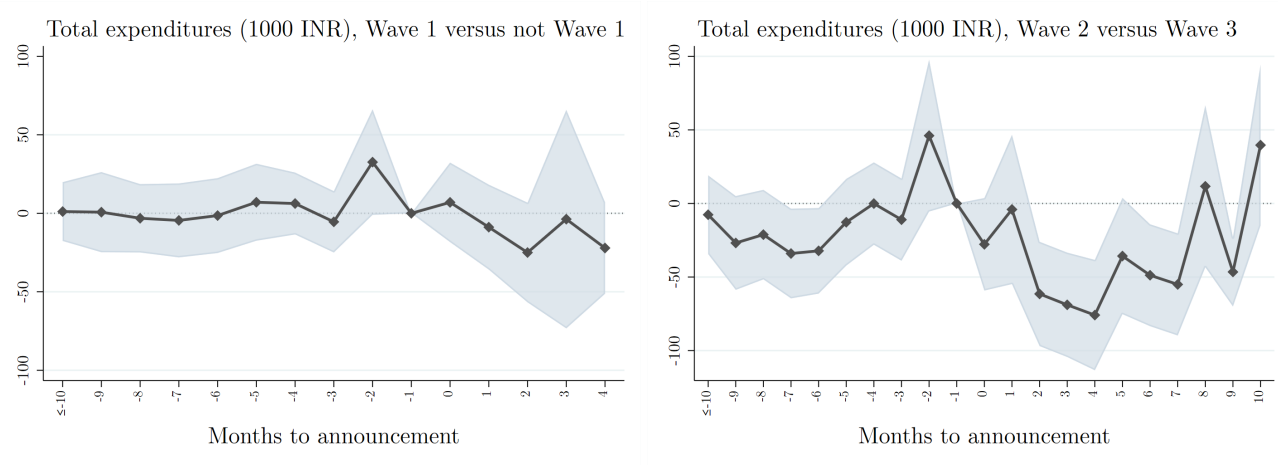


Figure C.5: Announcement event study by wave for total expenditures. The omitted category in this regression is the one month lead before the announcement. It includes GP and district month-year fixed effects as specified under Equations C.2 and 3, and observations are weighted by the inverse of number of households in the GP. The figure on the left compares those selected for audit in Wave 1 to those not audited and the panel of data is truncated at the month prior to the announcement of the second wave audit. The figure on the right compares those selected for audit in Wave 2 to those not yet audited (and will be audited in Wave 3) and the panel of data is truncated at the month prior to the announcement of the third wave audit. So, for instance, the announcement pre-periods in the figure on the left include data from both Waves 2 and 3 while the announcement post-periods include data only from Wave 2.

C.2 Alternatives to the baseline specification

This section presents alternatives to the baseline specification and shows that results from our preferred specification are robust.

Table C.2 shows the results for the specification in Equation 1 using only month-year fixed effects and not district-month-year fixed effects in the first column. The remaining columns show the specification in Equation 1 for additional outcome variables. Results are qualitatively similar for other measures of employment.

Figure C.6 shows the specification in Equation C.2 using month-year and block-month-year fixed effects. Qualitatively, the results are similar and we cannot reject that there are no pre-trends (p -value = 0.36 and 0.53, respectively). I use district \times month FE over block \times month-year fixed effects because for Wave 2 of the audit, all GPs within a block were assigned to be audited if they were considered

	(1) Total expenditures (1000 INR)	(2) # HHs provided employment	(3) Person-days of work generated	(4) Delay in days
<i>Anticipating1stAudit</i> - Year 2	14.34 (11.81)	-0.540 (2.539)	-26.64 (39.95)	-0.554 (0.766)
<i>Post1stAnnounce</i> , disaggregated:				
Before <i>1stAudit</i>	-11.31 (10.46)	-3.472 (2.273)	-64.97* (35.68)	0.0817 (0.654)
Month of <i>1stAudit</i>	-42.48*** (12.61)	-16.32*** (2.410)	-273.1*** (37.42)	0.543 (0.724)
After <i>1stAudit</i>	-17.43 (11.67)	-4.350* (2.466)	-84.82** (38.67)	0.295 (0.751)
<i>Anticipating2ndAudit</i> - Year 2	-2.338 (15.64)	-4.080 (3.708)	-78.75 (57.25)	-0.0239 (1.278)
<i>Anticipating2ndAudit</i> - Year 3	-39.07*** (13.51)	-10.50*** (3.172)	-169.4*** (47.91)	0.787 (1.264)
Observations	233,760	233,760	233,760	206,795
Time FE	monyr	district-monyr	district-monyr	district-monyr
Baseline mean	269.5	87.48	1197	21.12
Adj. R-squared	0.34	0.55	0.50	0.10

Table C.2: Effect of stages of the monitoring policy on additional outcomes and specifications. The first regression uses month-year FE for the main outcome variable, total expenditures; the remaining regressions show the effect of the policy on other performance outcomes using district-month-year FE. Standard errors are clustered by block. The omitted category is the horizon of anticipating one’s first audit during Year 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

to be small blocks. In this case, incorporating block month-year fixed effects would absorb variation being generated by the announcement and audit for blocks that met this criteria.

C.3 The importance of using monthly performance data

To show the importance of using more-frequent, monthly data on bureaucrat performance for our main analysis, I compare performance across waves with annual data. I construct annual means by treatment group, where treatment groups are defined by wave of audit.

We should expect that the difference in means by treatment group should be statistically indistinguishable from zero during the pre-audit periods which include FY 2014-15 and FY 2015-16. This is the true for all comparisons except when we compare total expenditures in FY 2014-15 for those in Wave 1 to those in Wave 3,

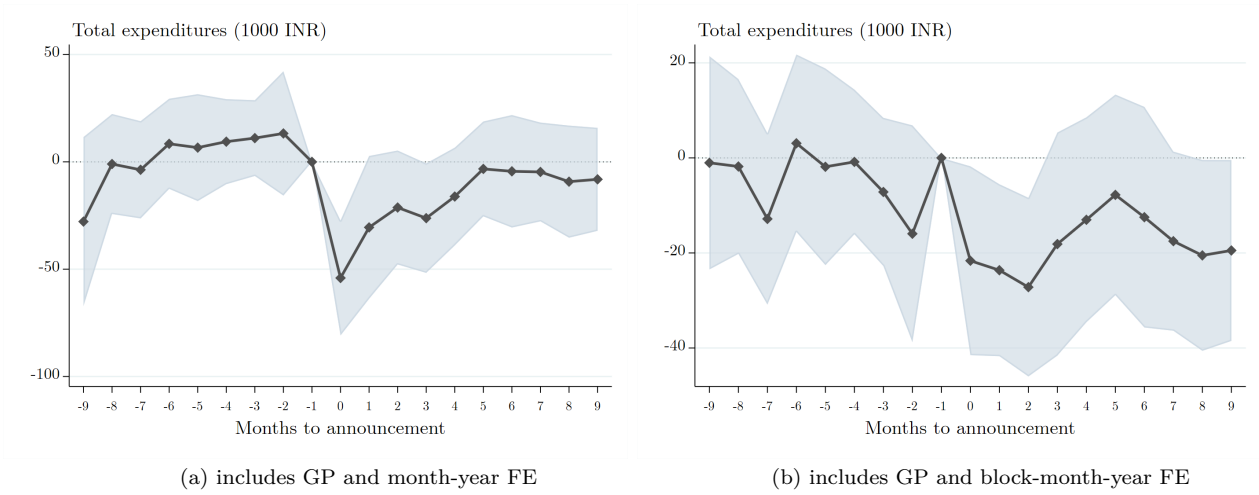


Figure C.6: Announcement event study with month-year and block-month-year fixed effects. Standard errors are clustered by block.

which could be due to chance occurrence.

From Year 1 onward, performance is statistically distinguishable performance across waves. Using data at the annual level is too coarse to detect any changes in behavior in the various stages of the response to the audit policy, especially since horizons of anticipation do not overlap perfectly with fiscal years. This further supports our approach to using available data on frequent measures of bureaucrat performance to infer bureaucrat adjustments to the audit policy especially during periods of performance outside the scope of audit. Frequent measures of verification of bureaucrat performance are infeasible during periods outside the scope of audit, and in this particular context third-party sources of verification are either unavailable or not possible for checking measures correlated with expenditures on materials and labor.

D Results on the Impact of Anticipation

D.2 Testing for Spillover Effects of the Audit

This section estimates whether the direct effects from the audit are confounded with spillover effects from GPs within your own block. Random assignment of GPs to

	Total Expenditures (1000 INR)				
	(1)	(2)	(3)	(4)	(5)
	FY1415	FY1516	FY1617	FY1718	FY1819
			<i>Year1</i>	<i>Year2</i>	<i>Year3</i>
Treatment group mean by fiscal year					
<i>Wave1</i>	2,465	3,154	4,382	3,906	3,024
<i>Wave2</i>	2,405	3,196	4,301	3,768	2,868
<i>Wave3</i>	2,296	3,046	4,224	3,866	3,025
t-tests of differences in means across treatment groups					
<i>Wave2 – Wave1</i>	-60.13 [0.503]	41.95 [0.693]	-81.09 [0.195]	-138.5 [0.382]	-156 [0.219]
<i>Wave3 – Wave1</i>	-169.3 [0.0441]**	-108 [0.257]	-157.9 [0.574]	-40.54 [0.768]	0.0873 [0.114]
<i>Wave2 – Wave3</i>	109.2 [0.154]	149.9 [0.257]	76.80 [0.479]	-97.98 [0.338]	-156.1 [0.999]
Adj. R-squared	0.37	0.38	0.40	0.29	0.28
<i>N</i>	3,896	3,896	3,896	3,896	3,896

Table C.3: Annual difference in means in total expenditures (1000 INR). Regressions include block fixed effects to account for the randomization design. *p*-values are in brackets and reflect tests of difference in estimated coefficients for each Wave. Standard errors are clustered by block. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

audit means that the concentration of audits within a block is also random. So, we can estimate the spillover effects of the audit from GPs within the same block. Results show spillover effects from being audited are not a concern for identification of the direct and anticipatory effects of the audit policy.

Spillover effects can be a concern because communication among peers in other GPs within your block may effect your own performance. Block managers describe how performance and administrative matters at lower administrative units are often discussed as a group. Group texting applications like *WhatsApp* are often used to communicate information. This suggests a free flow of information across GPs within the same block, where group performance and administrative matters like the audit are discussed.

When treatment is randomized, spillover effects can be estimated in a reduced-form linear-in-means specification by including a control for share of GPs being

audited within one's block (Manski, 1993; Bobonis and Finan, 2009; Lalive and Cattaneo, 2009; Dieye et al., 2014):

$$y_{it} = \alpha_i + \alpha_{dt} + \textit{Anticipating1stAudit}'_{it}\beta + \delta_1 \textit{Post1stAnnoounceBefore1stAudit}_{it} + \delta_2 \textit{1stAudit}_{it} + \textit{Anticipating2ndAudit}'_{it}\gamma + \eta \textit{ShareBlockAudited}_{it} + \epsilon_{it} \quad (\text{D.2})$$

where *ShareBlockAudited* denotes the share of GPs being audited in *i*'s block at time *t*. The spillover effect through the linear-in-means specification is identified when (i) spillover effects are equal across audited and not audited groups, and (ii) the spillover effects are linear in share of group being audited (Vazquez-Bare, 2017).

First, an event study around time to announcement including *ShareBlockAudited* as a control shows that the results do not change (Figure D.2). Furthermore, an *F*-test ($p = 0.84$) shows support for parallel pre-trends.

Table D.2 Column 1 shows the main specification from Equation 1 without spillover effects. Column 2 includes *ShareBlockAudited* as a control and shows that the estimates on both anticipatory and direct audit effects are unaffected. Column 3 interacts *ShareBlockAudited* with an indicator for the month of audit to provide evidence for assumption (i) that we cannot reject the spillover effects, if any, are equal across audited and not audited groups during the month of audit. Column 4 includes a quadratic term for *ShareBlockAudited* and provides evidence for assumption (ii) where the coefficient on the quadratic terms is insignificant. This tells us we cannot reject the spillover effects are linear in the share of group receiving treatment.

Furthermore, Table D.2 shows overall that the estimated anticipatory and direct effects of the audit are largely unchanged after accounting for concentration of audits within one's block. This provides additional evidence that the anticipatory effects are driven by changes in expectations of the likelihood of an audit rather than perceptions about the audit driven by peer experiences.

Figure D.2 shows an event study around month of audit (controlling for '*Anticipating1stAudit* - Year 2' as in the main specification) and the results remain unchanged when con-

	Total expenditures (1,000 INR)			
	(1)	(2)	(3)	(4)
<i>Anticipating1stAudit</i> - Year 2	13.87 (11.33)	15.52 (11.27)	14.54 (11.28)	15.42 (11.28)
<i>Post1stAnnounce</i> , disaggregated:				
Before <i>1stAudit</i>	-18.83** (9.46)	-17.00* (9.43)	-17.93* (9.47)	-16.97* (9.44)
Month of <i>1stAudit</i>	-42.07*** (11.68)	-49.06*** (11.59)	-42.42** (17.25)	-49.17*** (11.61)
After <i>1stAudit</i>	-8.42 (11.26)	-6.44 (11.25)	-7.47 (11.17)	-6.40 (11.27)
<i>Anticipating2ndAudit</i> - Year 2	-4.78 (13.83)	-3.47 (13.85)	-4.47 (13.65)	-3.50 (13.83)
<i>Anticipating2ndAudit</i> - Year 3	-39.38*** (13.30)	-37.76*** (13.30)	-38.97*** (13.33)	-37.69*** (13.31)
<i>ShareBlockAudited</i>		25.64 (18.96)		40.35 (65.87)
<i>ShareBlockAudited</i> ²				-21.10 (87.61)
Month of <i>1stAudit</i> = 0 \times <i>ShareBlockAudited</i>			32.71 (22.70)	
Month of <i>1stAudit</i> = 1 \times <i>ShareBlockAudited</i>			10.85 (31.92)	
Observations	233,760	233,760	233,760	233,760
Baseline mean	269.5	269.5	269.5	269.5
Adj. R-squared	0.40	0.40	0.40	0.40
$(H_0: \text{Month of } 1stAudit = 0 \times ShareBlockAud) - (\text{Month of } 1stAudit = 1 \times ShareBlockAud), p\text{-val} = 0.57$				

Table D.2: Tests for spillover effects of the audit. All regressions include district-month-year and GP fixed effects. Standard errors are clustered by block. The omitted category is the horizon of anticipating one’s first audit during Year 1 (*Anticipating1stAudit* - Year 1). The baseline is the mean from the beginning of the panel (two years prior to first audits) up to and including the period captured by ‘*Anticipating1stAudit* - Year 1’. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

trolling for spillovers. As before, we cannot reject the hypothesis of no pre-trends prior to one’s announcement (p -value = 0.5). Altogether, the estimates in the main analysis are robust to spillover concerns.

D.3 Testing the Assumption that Perception of Audit Quality is Constant

This section tests the assumption in Section 4 that bureaucrat perceptions of audit quality was constant across years. With this assumption, we can attribute differ-

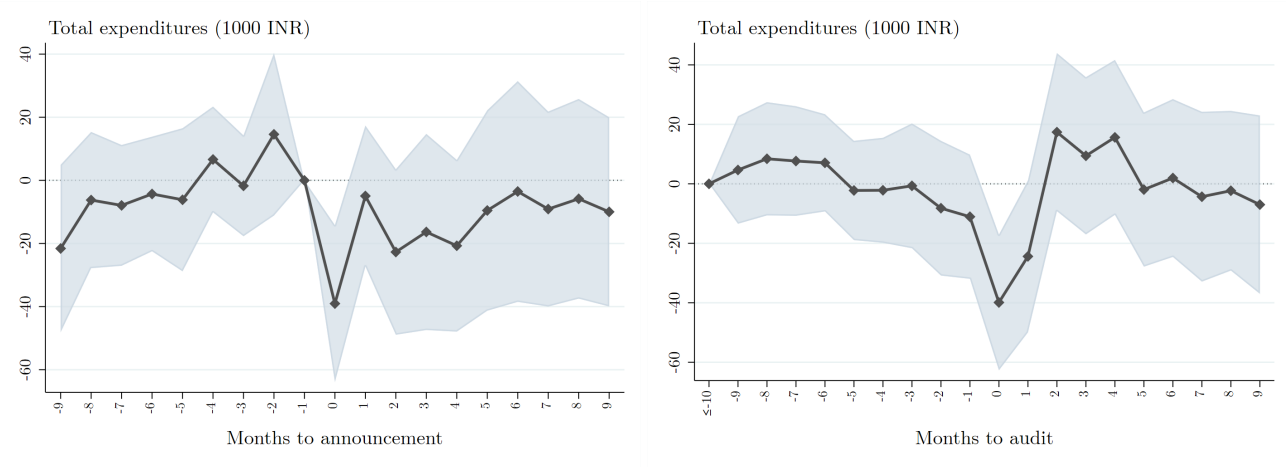


Figure D.2: Event studies around time to announcement and audit are unaffected by controlling for spillovers. The omitted category is 10 or more months before the announcement and audit, respectively. The raw mean of the omitted category is 250 and 272 (1,000 INR) for the announcement and audit event studies, respectively. The regressions include GP and district-month-year fixed effects. Standard errors are clustered by block.

ences in bureaucrat responses to the monitoring policy across years to changing expectations of the likelihood of an audit. To test this assumption, we can examine whether audit quality or bureaucrats' responses to the audit differ across years.

Using data from audit reports on inputs and outputs of audits, we can test for differences in audit implementation to see if audit quality was comparable across years. Workload per auditor and audit expenditures serve as proxy variables for audit quality. Workload per auditor is measured as the sum of households and projects to be verified divided by number of auditors. Only a subset of data from Waves 2 and 3 are available. Table D.3 shows no statistically significant differences between Wave 2 and 3 in workload per auditor and audit expenditures.

While this tells us whether the audit treatment was uniformly applied across years, more information is required to understand whether the bureaucrats responded differentially over time. If the perceived quality or credibility of the audit is changing, then one might expect the response to vary by year; if constant, then one might expect estimates to be same across years. To test for this we can extend the estimates from Table 2 Panel B by parsing the treatment effects by year for

	Workload per auditor		Total audit expenditures (Rs)	
	(1)	(2)	(3)	(4)
Wave 3	-7.326 (10.066)	-1.897 (4.540)	-1,660.990 (1,127.538)	-1,584.749 (1,013.885)
Num. HHs to check		0.432*** (0.019)		-0.191 (2.201)
Num. works to check		0.422*** (0.032)		9.910** (4.051)
Num. Auditors		-44.329*** (3.221)		1,137.143 (738.204)
Mean of Dep Var	249.69	249.69	27206.45	27206.45
Observations	2,445	2,445	2,676	2,445
R ²	0.496	0.949	0.363	0.412
Adjusted R ²	0.441	0.944	0.298	0.346

Table D.3: Audit quality by year of audit. Unit of observation is the GP. Omitted group is Wave 2. Outcome variables are: workload per auditor measured as the number of households and projects to verify per auditor; and total audit expenditures. Control variables include: Number of employed households and works to verify; and number of auditors. Standard errors are clustered by block. *p<0.1; **p<0.05; ***p<0.01

total expenditures, the main outcome of interest for estimating the sufficient statistic in Section 6. This tells us whether bureaucrats had notable differences in their responses to the monitoring policy over time.

Table D.4 shows that the treatment effects for (‘Month of *1stAudit*’, and ‘After *1stAudit*’) across Year are on average different but not statistically distinguishable for total expenditures. We cannot reject the hypothesis that the true difference between their responses is zero. Particularly, there are no distinguishable effects across years during the month of audit; and during the months following the audit (outside of post-audit anticipatory behavior). If the null is true, then the assumption that Bureaucrat beliefs over credibility of audit are constant across years is credible. And the estimated changes bureaucrats’ responses are driven by changing expectations of the likelihood of an audit. So, we can proceed with Section 6.2’s estimates of the sufficient statistic to determine the optimal design of information.

	Total expenditures (1,000 INR)	
	(1)	(2)
<i>Anticipating1stAudit</i> - Year 2	13.87 (11.33)	14.33 (11.76)
<i>Post1stAnnounce</i> , disaggregated:		
Before <i>1stAudit</i>	-18.83** (9.46)	-17.72** (8.56)
Month of <i>1stAudit</i>	-42.07*** (11.68)	
Month of <i>1stAudit</i> - Year 1		-54.61** (23.46)
Month of <i>1stAudit</i> - Year 2		-45.10*** (15.50)
Month of <i>1stAudit</i> - Year 3		-33.04** (14.54)
After <i>1stAudit</i>	-8.42 (11.26)	
After <i>1stAudit</i> - Year 1		26.69 (30.58)
After <i>1stAudit</i> - Year 2		-10.08 (13.39)
After <i>1stAudit</i> - Year 3		-6.95 (12.43)
<i>Anticipating2ndAudit</i> - Year 2	-4.78 (13.83)	-1.74 (13.68)
<i>Anticipating2ndAudit</i> - Year 3	-39.38*** (13.30)	
<i>Anticipating2ndAudit</i> - Year 3, Wave 1		-28.12** (13.98)
<i>Anticipating2ndAudit</i> - Year 3, Wave 2		-41.24*** (12.63)
Observations	233,760	233,760
Baseline mean	269.5	269.5
Adj. R-squared	0.397	0.400
Month of <i>1stAudit</i> , H_0 : Year 1 = Year 2 = Year 3 (p-val)		0.67
After <i>1stAudit</i> , H_0 : Year 1 = Year 2 = Year 3 (p-val)		0.54
<i>Anticipating2ndAudit</i> - Year 3, H_0 : Wave 1 = Wave 2 (p-val)		0.22

Table D.4: Effect of stages of the monitoring policy disaggregated by year. This table estimates the main differences-in-differences specification breaking down ‘Month of *1stAudit*’ and ‘After *1stAudit*’ by year and ‘*Anticipating2ndAudit* - Year 3 by Wave’ for total expenditures. Regressions include district-month-year and GP fixed effects. Standard errors are clustered by block. The omitted category is the horizon of anticipating one’s first audit during Year 1 (*Anticipating1stAudit* - Year 1). The baseline is the mean from the beginning of the panel (two years prior to first audits) up to and including the period captured by *Anticipating1stAudit* - Year 1. This longer period is included in the baseline to average out seasonal variation in expenditures. *** p<0.01, ** p<0.05, * p<0.1

On the other hand, if the parameter is *not* truly zero, then it could be that we are not powered to detect small differences given the sample size. And the estimated response of bureaucrats is confounded with changing perceptions of audit quality. Section 6.3 addresses this possibility by conducting robustness checks which relax this assumption.

D.4 Delayed payments as a measure of effort

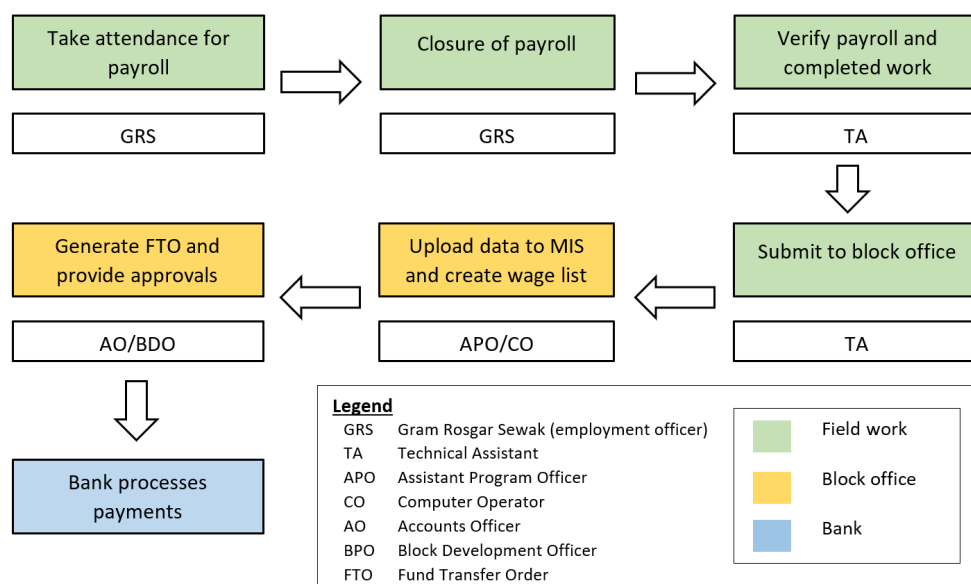


Figure D.3: Example of stages to the NREGS payment process. White boxes beneath steps present the officials responsible for implementing that step. FTO denotes fund transfer order. Gram Rozgar Sewak (GRS) is one of the key GP-level personnel and also responsible for allocating days of work. Source: Evidence for Policy Design, September 2015 presentation to the Ministry of Rural Development of the Government of India.

The problem of delayed payments is well-documented in the literature (Banerjee et al., 2020; Narayanan et al., 2019). Delays are counted as days over the 15 day maximum for processing payment from the time of the closure of the payroll (or muster roll). There are several steps in the administrative process between attendance for work and processing the wage payment through the bank as shown in Figure D.3. Delays tend to occur during the closure of the payroll to entering the data into their MIS at the GP level; between data entry to generation of the

wagelist at the block level; and between the first signature of the fund transfer order to the second signature at the block level.³⁹ According to the NREGS National Act, workers should be compensated 0.05% of unpaid wages for each day of delay in wage payment.⁴⁰

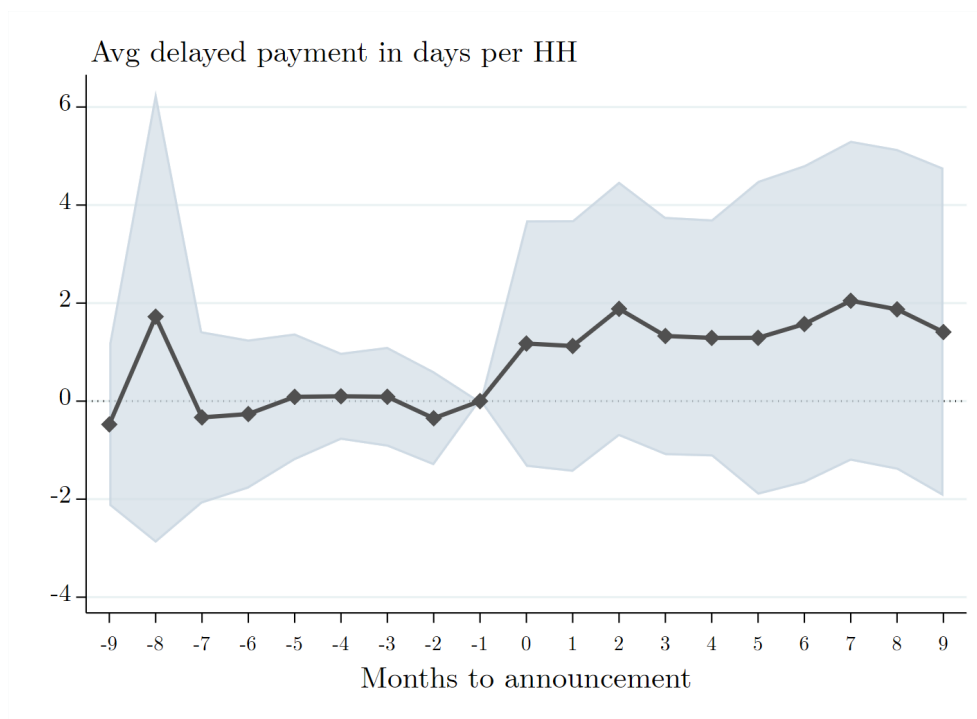


Figure D.4: Announcement event study on mean days of delay per household. This omitted category in this regression is the one month lead before the announcement and its mean is 1.9 days of delay per household. The p -value on a test of no pre-trends is 0.97. Standard errors are clustered by block.

Delaying payments has not been documented as a means for misappropriating government resources and is typically discussed as a measure of quality of program implementation. When examining the effect of the audit on delayed payments, I consider the possibility that these outcomes are reflective of the amount of effort exerted by the bureaucrat. Although, in theory, it is possible that withholding payments for honest work could be used by bureaucrats as leverage. But, it is an unlikely strategy since increased delays reflect negatively on performance and thus

³⁹According to the Evidence for Policy Design, September 2015 presentation to the Ministry of Rural Development of the Government of India.

⁴⁰The Mahatma Gandhi National Rural Employment Guarantee Act, 2005.

costly for bureaucrats.

This section complements the analysis in Section 5.3 on whether delays are affected around the time of audit. Figure D.4 shows that prior to a month before the announcement, mean delayed payments in days per household do not follow a statistically distinguishable trend (p -value = 0.97). Furthermore, there are also no statistically distinguishable effects following the announcement, although there is an average increase in delays post-announcement.

Figure D.5 plots the event study around month of audit and shows the audit does not affect delays in making wage payments.

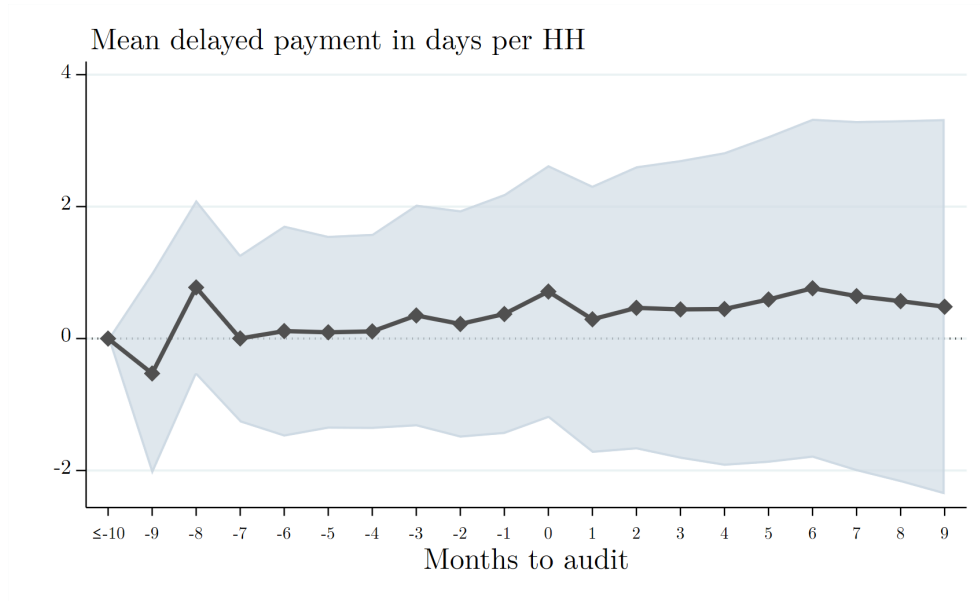


Figure D.5: Audit event study on delays in making wage payments. The omitted category is 10 or more months before the audit was conducted and its mean is 20 mean delayed payment in days per household. The regression includes GP and district month-year fixed effects as specified under Equation 3. Standard errors are clustered by block.

E The Optimal Design of Information and Counterfactual Signals

E.1 Procedures for sensitivity analyses

This section describes the procedures for conducting the sensitivity analyses discussed in Section 6.3. I perturb either the beliefs or the deterrence parameters, each in turn. For ease of interpretation, when discussing perturbations of the belief and deterrence parameters, I will refer to them as perturbations of the x - and y -values, respectively. Note also that we have a maximum of 5 points of support to estimate this sufficient statistic, i.e. five (x, y) pairs, as seen in Figure 7 under the different beliefs assumptions.

Overall, when simulating perturbations from x -values, I fix y -values at its mean and simulate perturbations of x -values (bound between 0 and 1) that require us to draw a different conclusion from the one made in Section 6 that the curve is convex. The most probable worst-case scenario to challenge our conclusion is a simulated sufficient statistic that takes the shape of a line, as the distance from a globally convex function to a linear function is weakly less than its distance to a globally concave function (assuming these hypothetical functions share the same endpoints). So, I always start with a simulated line and progress from there to more concave functions.

When allowing for deviations from the mean y -values, I perform simulations fixing beliefs (for each set of assumed beliefs) where y -values can deviate from the mean of parameters estimated in Column 5 of Table 2. Based on the distributions of the estimated y -value parameters from the regression, I calculate the likelihood of each simulation leading to an alternative conclusion (i.e. the sufficient statistic is linear or concave) which gives us a distribution of probabilities that we have drawn the wrong conclusion.

Deviations from assumptions on beliefs

Let \tilde{x} refer to simulated x -values. To simulate deviations from beliefs assumptions (y -values fixed at mean), I run the following procedure:

- (i) Take the minimum and maximum of y -values, assign them \tilde{x} -value 0 and 1, respectively. Selecting the minimum and maximum of y -values and then horizontally “stretching” the values out to 0 and 1 ensures that we do not end up with impossible x -values or probabilities, and maximizes likelihood that remaining y -values fall within possible range of x -values or beliefs assumed in Table 4.
- (ii) Construct the line, $f(\tilde{x})$, that goes through these endpoints.
- (iii) Find where the remaining interior y -values fall on the line $f^{line}(\tilde{x})$ and backout their \tilde{x} -values. Let $\tilde{\mathbf{x}}^{line}$ be the vector of \tilde{x} -values s.t. $y = f^{line}(\tilde{x})$.
- (iv) To simulate a strictly concave curve, the vector of simulated interior x -values must jointly satisfy: $\tilde{\mathbf{x}}^{concave} < \tilde{\mathbf{x}}^{line}$.

Deviations from the mean of deterrence estimates

Table E.2 provides details on the results with assumed bias in Section 6.3. In this exercise, I relax the assumption on constant perceptions of audit quality and incorporated a bias in deterrence estimates.

abs(Bias)	Beliefs Assumption, $K_{\tau+1} =$	Prob. Alt. Conclusion
2	$K_{\tau-1}$	0.001
2	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	0.0001
2	K_{τ}	0.00004
2	$K_{\tau+1}$	0
2	$\text{Trend}_{\tau} \times K_{\tau}$	0.352
5	$K_{\tau-1}$	0.016
5	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	0.001
5	K_{τ}	0.0005
5	$K_{\tau+1}$	0
5	$\text{Trend}_{\tau} \times K_{\tau}$	0.375
7	$K_{\tau-1}$	0.055
7	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	0.006
7	K_{τ}	0.001
7	$K_{\tau+1}$	0
7	$\text{Trend}_{\tau} \times K_{\tau}$	0.375
9	$K_{\tau-1}$	0.148
9	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	0.021
9	K_{τ}	0.004
9	$K_{\tau+1}$	0
9	$\text{Trend}_{\tau} \times K_{\tau}$	0.375
15	$K_{\tau-1}$	0.711
15	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	0.294
15	K_{τ}	0.091
15	$K_{\tau+1}$	0.0002
15	$\text{Trend}_{\tau} \times K_{\tau}$	0.375
20	$K_{\tau-1}$	0.840
20	$\frac{1}{2}(K_{\tau-1} + K_{\tau})$	0.747
20	K_{τ}	0.476
20	$K_{\tau+1}$	0.007
20	$\text{Trend}_{\tau} \times K_{\tau}$	0.375

Table E.2: Likelihood of alternative conclusion relaxing assumptions on constant perceptions of audit quality and assuming bias in estimates. The bias incorporated purposely increased the likelihood of drawing an alternative conclusion. So, for e.g., ± 5 bias was incorporated for each anticipatory group such that the sufficient statistic was closer to being concave. The bias was modeled separately for each beliefs assumption. The probability of an alternative conclusion indicates the likelihood that the bootstrapped joint distribution of estimated coefficients formed a sufficient statistic that dispersed incentives was optimal (weakly/locally concave).

In addition to the approach taken in Section 6.3, I also simulate possible deviations from the mean of deterrence estimates and assess their likelihood of occurring. We arrive at the same conclusions. When allowing for deviations from the mean of deterrence estimates, I perform simulations for each beliefs assumption. So, while beliefs or x -values are fixed for given a particular beliefs assumption, we are still examining perturbations of y -values across the range of beliefs dictated by the beliefs assumptions.

Let \tilde{y} denote simulated y -values. The procedure is as follows:

- (i) Take the estimates of y associated with $\min(x)$ and $\max(x)$ for a given set of beliefs. We are going to use these two estimates as endpoints of simulated lines. The lines will be simulated along the estimated 95% confidence intervals of these two y estimates. Denote the 95% confidence interval for each estimate as $y_{x_{min}}^{95\%CI} = [y_{x_{min}}^{2.5\%ile}, y_{x_{min}}^{97.5\%ile}]$ and $y_{x_{max}}^{95\%CI} = [y_{x_{max}}^{2.5\%ile}, y_{x_{max}}^{97.5\%ile}]$.
- (ii) Let $i \in \{0, 1, \dots, I\}$ index the number of steps taken along $y_{x_{min}}^{95\%CI}$, and let $j \in \{0, 1, \dots, J\}$ index the number of steps taken along $y_{x_{max}}^{95\%CI}$. Eefine the step size along each interval as: $\Delta_i = \text{range}(y_{x_{min}}^{95\%CI})/I$ and $\Delta_j = \text{range}(y_{x_{max}}^{95\%CI})/J$.
- (iii) For a given i and j construct the line, $f_{(i,j)}(x)$, such that the following two points fall on the line:

$$\left(x_{min}, y_{x_{min}}^{2.5\%ile} + i\Delta_i\right), \left(x_{max}, y_{x_{max}}^{2.5\%ile} + j\Delta_j\right)$$

- (iv) Find where the remaining interior fixed x -values fall on the line and calculate $\tilde{y} = f_{(i,j)}(x)$. Now we have a set of 4-5 (depending on the beliefs assumption) simulated points on the line $f_{(i,j)}(x)$.
- (v) Calculate the probability of realizing the simulated line $f_{(i,j)}(x)$ or a function that is more concave. Notice that to construct a more concave function departing from the line $f_{(i,j)}(x)$, the \tilde{y} -values of the endpoints are weakly decreasing from their simulated points while the \tilde{y} -values of the interior points are weakly increasing from their simulated points. So, using the regression estimates,

calculate the joint probability of those events occurring. In particular, we are interested in calculating the following probability:

$$Pr\left(y_{x_{min}} \leq y_{x_{min}}^{2.5\%ile} + i\Delta_i \text{ and } \mathbf{y}_{x_{int}} \geq \tilde{\mathbf{y}}_{x_{int}} \text{ and } y_{x_{max}} \leq y_{x_{max}}^{2.5\%ile} + j\Delta_j\right)$$

Denote this probability as $q_{(i,j)}$ and denote the set of probabilities for all i and j as Q . We can estimate $q_{(i,j)}$ by bootstrapping to estimate the empirical joint distribution of each point of the sufficient statistic. In practice, a single draw for the block bootstrap is the panel data for randomly selected GP with replacement. Within a bootstrapped sample, the number of draws from each wave reflect actual number of audits in each wave. This simulates the information environment which is important for estimating anticipatory behavior. Bootstrapped estimates converge after 6,000 iterations.

- (vi) Repeat these steps for all i and j . Let $N = IJ$ denote the total number of simulations. Obtain Q^k for a given set of beliefs k .
- (vii) Repeat above steps for each set of beliefs.

We are interested in the statistic Q^k obtained from simulations under each beliefs assumption k . We would like to know the likelihood of the perturbations of estimated deterrence parameters (or y -values) that led to alternative conclusions about the optimal monitoring policy. The likelihood of an alternative conclusion under $K_{t+1} = \{K_{t-1}, \frac{1}{2}(K_{t-1} + K_t), K_t, K_{t+1}\}$, the mean probability of realizing an alternative conclusion is less than 0.001%. If $K_{t+1} = Trend_t K_t$, there's a mean 0.1% chance of realizing an alternative conclusion with a max of 3.6%.

E.2 Welfare calculations

This section lays out in detail how welfare across the information policies were computed.

I analyze welfare for three policies: (1) the actual implemented policy of randomizing without replacement; (2) dispersed incentives where all GPs have the same

expectation of the likelihood of an audit (equivalent to randomizing with replacement); and (3) concentrated incentives where all GPs are perfectly informed of when they will be audited or not (equivalent to a crackdown).

For each policy, we will examine total expenditures over the course of the 27 months that it took for the round of first audits to roll-out under the original monitoring policy. We will examine how total expenditures changes as beliefs evolve under counterfactual policies. Each policy will be assessed relative to a baseline capturing behavior prior to the implementation of the audits. The baseline is total annualized level of expenditures during the 27 months using the average of monthly expenditures during the 2 years prior to audits and including the anticipatory period in Year 1. This is the same baseline used in Table 2 and is the preferred baseline since all deterrence parameters are estimated relative to anticipatory behavior in Year 1. For each policy and the baseline, I exclude the year during which a GP experiences an audit to exclude behavior after one learns about their audit prior to receiving the audit and behavior during the months after the audit but still within the audit year. We can think of each calculation as excluding behavior during this period which may capture other phenomena of GPs responding to the policy not related to expectations of the likelihood of an audit, e.g. perceived salience of audits holding expectations fixed.

The calculation of counterfactual expenditures for each policy estimates anticipatory behavior using the deterrence estimates from Year 1 to Year 3. The calculation for each policy is discussed in detail below:

(1) Randomization without replacement (or p_1)

$$\begin{aligned} DeterredExpenditures_{p_1} = & n_{Months, Year1} \left[(n_{GPs, Wave2} + n_{GPs, Wave3}) U(q_{Year1}) \right] \\ & + n_{Months, Year2} \left[n_{GPs, Wave3} U(q_{Year2}) + n_{GPs, Wave1} U(q'_{Year2}) \right] \\ & + n_{Months, Year3} \left[(n_{GPs, Wave1} + n_{GPs, Wave2}) U(q'_{Year3}) \right] \end{aligned}$$

where $n_{GPs, Wave x}$ corresponds to the number of GPs in Wave x , $n_{Months, Year y}$

corresponds to the number of months during Year y that the monitoring policy was in place, i.e. 4 months in Year 1, 11 months in Year 2, and 12 months in Year 3. $U(q)$ corresponds to the amount of deterred misappropriated expenditures as a function of bureaucrats' posterior beliefs, q , on the likelihood of an audit. Under this policy, $U(q_{Year1})$ corresponds to the '*Anticipating1stAudit* - Year 1'; $U(q_{Year2})$ corresponds to '*Anticipating1stAudit* - Year 2'; $U(q'_{Year2})$ corresponds to '*Anticipating2ndAudit* - Year 2'; and $U(q'_{Year3})$ corresponds to '*Anticipating2ndAudit* - Year 3'.

(2) Dispersed incentives (or p_2)

$$\begin{aligned} DeterredExpenditures_{p_2} = & n_{Months, Year1} \left[(n_{GPs, Wave2} + n_{GPs, Wave3}) U\left(\frac{K_{Year2}}{N}\right) \right] \\ & + n_{Months, Year2} \left[(n_{GPs, Wave2} + n_{GPs, Wave3}) U\left(\frac{K_{Year3}}{N}\right) \right] \\ & + n_{Months, Year3} \left[(n_{GPs, Wave1} + n_{GPs, Wave2}) U\left(\frac{K_{Year4}}{N}\right) \right] \end{aligned}$$

where N is the total number of GPs; $K_{Year y}$ is the assumption made on next year's audit capacity. For the main welfare analysis, I assume that $K_{\tau+1} = K_{\tau}$, i.e. GPs believe next year's audit capacity is equivalent to what they observe about this year's audit capacity.

(3) Concentrated incentives (or (p_3))

$$\begin{aligned} DeterredExpenditures_{p_3} = & n_{Months, Year1} \left[n_{GPs, Wave2} U(0) + n_{GPs, Wave3} U(1) \right] \\ & + n_{Months, Year2} \left[n_{GPs, Wave1} U(0) + n_{GPs, Wave3} U(1) \right] \\ & + n_{Months, Year3} \left[n_{GPs, N-Wave4} U(0) + n_{GPs, Wave4} U(1) \right] \end{aligned}$$

In the main analysis, I assume that deterrence under $U(1)$ is equivalent to behavior when the auditors are present (estimates provided by '*Month of 1stAudit*'). The reported conservative estimates assume that deterrence under $U(1)$ is equivalent to behavior in Year 3 when Waves 1 and 2 believe with very high likelihood

they will be audited in Year 4 (estimates provided by ‘*Anticipating2ndAudit - Year 3*’. These are conservative estimates because first, they address concerns that some other behavioral phenomena may be driving the response when auditors are present. Second, under some beliefs assumptions about tomorrow’s audit capacity (see Table 4), the expectations of the likelihood of an audit captured by ‘*Anticipating2ndAudit - Year 3*’ can be less than 1. This makes the estimate conservative because the deterrence response from bureaucrats under $U(1)$ could be greater than what is estimated with ‘*Anticipating2ndAudit - Year 3*’.

I provide confidence intervals of the welfare calculations using the bootstrapped estimates of the deterrence parameters.