Optimal Monitoring and Bureaucrat Adjustments[*]

Wendy N. Wong[†]

July 18, 2023

Abstract

This paper examines the behavior of bureaucrats, implementing India's public employment program, as their expectations of being audited change. Exploiting random assignment to audit timing, I find: the rate of deterrence (in financial misappropriation) is increasing in bureaucrats' expectations; bureaucrats evade detection by adjusting the timing and type of misappropriated expenditure. I interpret these findings with a model to analyze how information communicated about audit risk should be designed. With this, I solve for the optimal signal and analyze counterfactuals. Fully-informative signals would have persuaded bureaucrats to misappropriate USD35m less compared to uninformative signals.

[*]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 their invaluable support. I am grateful for feedback from anonymous referees, Scott Ashworth, Dan Black, Oeindrila Dube, Steven Durlauf, Jeffrey Grogger, Damon Jones, Donald Moynihan, Yusuf Neggers, Guillaume Pouliot, James Robinson, Raul Sanchez de la Sierra, Shaoda Wang, Austin Wright, and 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 generously providing access and connecting me to resources in India. Any errors or expressed opinions are my own. 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.

[†]Email: wendywong@uchicago.edu.

1 Introduction

Governments adopt monitoring to address failures in the provision of public services and failures in public compliance with the law (Wilson 1991; Rose-Ackerman and Palifka 2016; Finan et al. 2017). Examples range from auditing of the administration of public programs and tax reports to patrolling for crime and civil infractions like speeding. The empirical literature shows that more extensive monitoring, such as conducting more frequent or more credible audits, leads to more compliance.[1] However, because commonplace constraints on government resources prohibit extensive monitoring, the existing literature offers limited guidance in evaluating the tradeoffs in designing the best monitoring policy. In addition, when extensive monitoring is not possible, agents may adapt their behavior to evade detection causing an unintended displacement in misconduct.[2] Monitoring policies designed to maximize the deterrence of misconduct must account for such constraints and unintended consequences.

This paper determines the optimal monitoring policy in a setting where the monitor has a choice over the information communicated to agents about their audit risk and estimates the value of information design. Given that agents adjust their behavior in response to monitoring, should governments inform them when the auditor is coming or make it hard to predict? Audits and auditing guidelines are often implemented to maintain unpredictability among audit subjects.[3] Yet, theory suggests that the answer depends on the relationship between deterrence and agents' expectations of being audited (Lazear 2006; Eeckhout et al. 2010). For example, at one extreme, if agents are only deterred when they are certain an auditor is coming, then it is best to provide this information in advance. In contrast, if they are deterred at the slightest probability of an audit, then withholding information may be better.

To advance our understanding on this question empirically, this paper examines strategic responses of bureaucrats to changing expectations of being audited. I study the monitoring policy of India's Mahatma Ghandi National Rural Employment Guarantee Scheme (NREGS), the largest public employment program in the world (Sukhtankar

1. See, for e.g., Finan et al. (2017) for a review.
2. For e.g., Yang (2008), and Gonzalez-Lira and Mobarak (2021).
3. See as an example Section 5(c): U.S. auditing standards, adopted and approved by the Public Company Accounting Oversight Board and the U.S. Securities and Exchange Commission, advocate to withhold information from audit subjects to maintain unpredictability of when/how audits may occur.

2017). First, I show that the deterrence of bureaucrats' financial misappropriation is more marginally responsive at higher than lower expectations of being audited.[4] With these results, I apply a model of Bayesian persuasion based on insights from Lazear (2006) and Kamenica and Gentzkow (2011). The model shows that a simple change in the design of information, which allows bureaucrats to predict 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. To my knowledge, this is the first empirical implementation of this theoretical construct.

Sixty-five percent of the population in India, or about 11.5% of the world population, is eligible for NREGS.[5] The program insures against income shocks by guaranteeing employment to rural households to work on public projects. Evidence in data and anecdotes reveal troubling issues along various dimensions of NREGS program performance.[6] Issues from audit reports range from participant payment delays to poor workplanning to fabricated employment and material procurement. Existing literature also documents financial misappropriation.[7]

This paper leverages a monitoring policy where audits of gram panchayats were staggered over years and randomly assigned *without* replacement. Gram panchayats (GPs) are the smallest implementing unit of NREGS.[8] The policy was implemented in the state of Jharkhand beginning in 2016. This was the first time auditing of NREGS in Jharkhand was systematically conducted by the government and done so at scale. The round of first audits across all GPs completed within three years.

The staggered implementation of audits combined with the randomization design generated random variation in bureaucrats' expectations of being audited. Bureaucrats observe who has been and is waiting to be audited. For example, the longer bureaucrats wait without being audited, the better they are able to predict their audit in advance because monitoring is without replacement. In contrast, those who were audited first anticipate their second audit will not occur until everyone's first audits have been com-

4. Throughout the paper, this notion of misappropriation intentionally works against government program goals and entails utility for bureaucrats.

5. NREGS expenditures make up about 0.4% of India's GDP.

6. See Sukhtankar (2017) for a synthesis of existing research on India's workfare program.

7. See for e.g., Khera (2011); Niehaus and Sukhtankar (2013); and Banerjee et al. (2020).

8. The 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.

pleted. Bureaucrats respond to these incentives because auditors verify the previous fiscal year's work and document issues with ongoing work. I estimate these anticipatory responses of bureaucrats by employing an event study specification around the timing of audit announcements. I do so with detailed administrative data at the GP-month level.

I use monthly program expenditures as the main outcome to proxy for changes in misappropriated spending.[9] Program expenditures are under the direct control of bureaucrats, and hard to manipulate ex-post. The frequency of measurement of program expenditures allows us to estimate the impact of anticipatory behavior around the months of announcement; whereas, the infrequent measurement from audit reports every few years do not. However, program expenditures comprise both honest and misappropriated expenditures.[10]

I provide a preponderance of evidence supporting that observed changes in expenditures are driven by changes in misappropriated spending and that it is difficult for alternative mechanisms which involve changes in honest spending to rationalize the combination of findings. First, with administrative data on other aspects of performance, I find my results are inconsistent with mechanisms where changes in expectations of being audited drive changes in honest spending. Bandiera et al. (2009) and Londoño-Vélez and Ávila-Mahecha (2021) use a similar approach using a proxy variable for unobserved behavior to study waste in government procurement and tax evasion. Second, I use data from auditor findings as a source of verification for whether expenditures were misappropriated for the changes observed during the horizons of anticipation. I also discuss the conditions under which the findings in this paper are robust to relaxing the assumption that misappropriated spending is measured without error.

For the main results, with expenditures as a proxy for misappropriated spending, I find first that the anticipatory effects of the audit on spending are substantial; and total expenditures decline (deterrence of misappropriated spending increases) more on the margin at higher than lower expectations of being audited, where these expectations are based on the audit agency's communicated rollout of audits. When expectations of being audited are high to certain, there is a 15% decline in expenditures. This is driven

9. Financial misappropriation is the type of misconduct of focus in this paper.

10. "Honest" program expenditures are those intended to contribute to program goals and do not entail personal financial gain to bureaucrats. They can include spending that is wasteful.

by a significant decline in wage expenditures. When expectations of being audited are lower (zero to medium), expenditures are statistically indistinguishable between these periods.

Second, I find that as expectations increase bureaucrats substitute misappropriation from wages to materials. When bureaucrats have medium expectations, material procurement significantly increases. This result is verified by audit reports: during this period, auditors assessed more fines related to material misappropriation. This is consistent with interpreting the estimated changes in expenditures as deterrence. Audit reports suggest that it is easier to detect wage over material misappropriation, which explains why bureaucrats adjust along this margin.

Third, intertemporal substitution diminishes the deterrence effect estimated during the audit. In particular, bureaucrats spend 11% less while auditors are present (driven by a decline in wages) and then spend 5% more in the following months once auditors leave (driven by an increase in materials). Additional results on mechanisms are consistent with changes being driven by misappropriated expenditures. For instance, alternative mechanisms such as multi-tasking issues while auditors are present do not explain the results. The findings are also consistent with real output remaining unchanged even though inputs change during this period.

If audits were instead randomized *with* replacement, would there have been greater deterrence? Under such a policy, bureaucrats would never know with certainty when they are up for an audit. When bureaucrats are deterred only under high expectations of being audited (as shown by the estimated anticipation effects), it is better to inform them about their audit in advance. I show this result is possible by modeling an information design problem and deriving conditions to determine the optimal signal. In this model, the principal is concerned with maximizing deterrence and has a choice over the information provided about the likelihood of being audited. The model shows that the function capturing the relationship between bureaucrat deterrence and their expectation of being audited is sufficient for characterizing the principal's optimal signal and for analyzing welfare under counterfactual signals. I estimate the function for this empirical setting using causally-identified moments, i.e. the estimated anticipation effects.

Estimates of the function tell us that perfectly informing a random subset of bureaucrats when they will be audited in advance and others that they will not be audited is the optimal design of information. Signals which communicate more information yield more deterrence in aggregate than those that maintain unpredictability. This implies that assigning audits randomly without replacement is better than randomly with replacement. This conclusion is robust to a series of sensitivity analyses.

Welfare estimates show that a policy providing advanced warning would have persuaded bureaucrats to misappropriate USD 19.8 and USD 35 million *less* in expenditures (or 9% and 16% of average annual expenditures from 2016-19 in Jharkhand) compared to the actual policy of randomizing without replacement and to randomizing with replacement, respectively. These 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 value of information design in other settings.

This paper provides empirical evidence on strategic responses by bureaucrats to monitoring and complements a literature spanning across multiple contexts studying the effectiveness of monitoring. Previous studies demonstrated the importance of: being told with certainty you will be audited or policed (Di Tella and Schargrodsky 2004; 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, 2011; Afridi and Iversen 2014; Bobonis et al. 2016); and having a persistent threat from monitoring over time (Avis et al. 2018).

Monitoring can be undermined when agents adapt their behavior to evade detection. For instance, targeted monitoring can lead to a displacement in crime (Yang 2008; Gonzalez-Lira and Mobarak 2021). Capturing displacement is challenging because 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). This paper contributes to this literature by leveraging administrative and audit data to measure strategic responses in combination with rich variation on the policy parameter of interest: bureaucrats' expectations of being audited. This combination is what allows us to estimate the optimal signal that should be communicated to the bureaucrat on their likelihood of being audited.

This paper shows how information disseminated about a monitoring policy, which takes into account these strategic responses, can be optimally designed with a budget-constrained policymaker in mind. This paper interprets the empirical findings with a model of information design, complementing the theoretical literature on optimal information design (e.g. Kamenica and Gentzkow 2011; Bergemann and Morris 2019; Kamenica 2019) by providing an empirical application. The intuitions from the model in this paper draw on prior theoretical work. Lazear (2006) models the trade off between informative versus uninformative signals in monitoring, and finds that the optimal signal depends on the shape of deterrence as a function of expectations of being audited. Other studies evaluate models with similar tradeoffs, but their empirical settings either require assuming an optimal policy is being implemented rather than solving for the optimum or are restricted to laboratory environments (Eeckhout et al. 2010; Chassang et al. 2020). Banerjee et al. (2019) study the optimal monitoring of drunk-driving, but their agent can choose to avoid being monitored altogether once they learn the monitor's strategy, while this is not possible for the bureaucrats in this setting. Relative to this empirical literature, this paper is novel because it studies agency issues in government and analyzes welfare under alternative communication policies.

Finally, this paper provides a novel empirical measure of the value of information. The estimated relationship between deterrence and bureaucrats' expectations of being audited provides sufficient information to determine the optimal signal and analyze welfare under counterfactual signals–all without requiring additional information. The function capturing this relationship is estimated using causally-identified parameters and not structural primitives. This approach is related to a literature in public finance that develops sufficient statistics from theoretical models (Chetty 2009). These sufficient statistics evaluate welfare from changes in tax policies as functions of reduced-form elasticities. To my knowledge, this approach has not been applied to studies motivated by the optimal design of information.

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. Participants provide manual labor on projects commissioned by the local government. They construct or maintain assets that are intended to improve rural livelihood.[11] In Jharkhand, an Indian state with 33 million people, 61 of the population relies on agricultural work, and their income is vulnerable to the volatility of seasonal agricultural output.[12] This makes NREGS an important source of reliable income. NREGS in Jharkhand has served around 7.7 million people and produced over 1 million projects.

In 2016, the government of Jharkhand began auditing gram panchayats (GPs), the smallesting implementing unit of the NREGS program. The Social Audit Unit (or audit agency) is a separate government agency that conducts the audits.[13] Competitive compensation for auditors and quality assurance mechanisms suggest it is likely the audit agency conducted audits at-scale with credibility and integrity.

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 was consistent with one of the goals of 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. During this period, 4,180 GPs were audited and informed that the previous FY's work would be a part of the audit.

Bureaucrats have been known to misappropriate NREGS public funds by over-reporting employment (Niehaus and Sukhtankar 2013; Muralidharan et al. 2016; Baner-

11. Assets include structures for water conservation and harvesting, latrines, and animal shelters.

12. Sources: Jharkhand Economic Survey 2017-18; Department of Agriculture, Jharkhand.

13. "Social audits" incorporate community and participant 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. The audit agency is funded independently of NREGS and managed by a steering committee of various stakeholders across the state government and civil society. More details on the audit agency and their processes are in Online Appendix A.2.

jee et al. 2020). Audit reports and news outlets suggest expenditures can also be misappropriated through material procurement. Example issues from audit reports include fake receipts, higher-than-expected prices for low-price goods, procured materials missing from worksites, and overreported labor for manual-labor-only projects that have been completed with machinery.

2.2 The process when auditors arrive

Auditors spend a week at the GP to verify administrative reports on labor and material expenditures from the previous FY, document other observed issues with ongoing work, and conduct public hearings of their findings. 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, verifying output at project sites (e.g. measuring dimensions of a dug pond), and interviewing households to verify past employment and document complaints.

Notably, information gathered during the audit process is also reflective of concurrent program performance. The auditors' scope of evaluation is not limited to performance from the previous FY. 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 adjudicated in a public hearing or appealed to higher government. Bureaucrats charged with financial misappropriation face paying a penalty commensurate with the amount misappropriated or risk losing their job. 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 rotates across GPs, i.e. GPs will not be selected for their second audit until all GPs receive their first audit (via randomization without replacement). In response, bureaucrats may adjust opportunistic behavior to influence the outcome of an anticipated audit. These strategic adjustments are possible because current performance may be part of an anticipated future audit (given audits are of last FY's work). 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 (the study period), all GPs received an announcement stating who has been selected for audit in each year, when the audits will happen within the year, and that the last FY's work will be audited.[14] The announcement in Year 1 stated that the audit agency plans to eventually target 50% of GPs for audit per year and that all GPs be audited regularly. 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, bureaucrats can expect that GPs take turns being audited and that there were plans to increase capacity in number of audits conducted.[15]

One's selection for audit could be anticipated, but was not perfectly predictable. Current audit capacity was observed from the announcements, but 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. 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). By the end of Year 3, the audit agency completed the round of first audits. In Year 4 (FY2019-20), they began the round of second audits.[16] Predictability of a future audit is driven by

14. The timing of the announcements varied across years, where the Year 1 announcement occurred during the last third of FY and during the beginning of the FY for Years 2 and 3 (see Online Appendix Table A.1 for details).

15. Results on the differences between Waves 1 and 2 in pre-trends or anticipation leading up to the first audit announcement (see Online Appendix B.1) along with findings from a small-sample survey conducted in 2020 on bureaucrat beliefs from 2016-19 (see Online Appendix A.5) provide evidence supporting that GP bureaucrats on average anticipated audits consistent with the turn-taking implemented via the randomization without replacement.

16. Around 300 GPs were not selected for audit under this policy for various reasons. They are described in detail in Section 3.2 and dropped from analysis. More details on notices that were publicly disseminated can be found in Online Appendix A.4.

observed changes in audit capacity (to infer future capacity) and the remaining number of GPs waiting to receive their first audit.

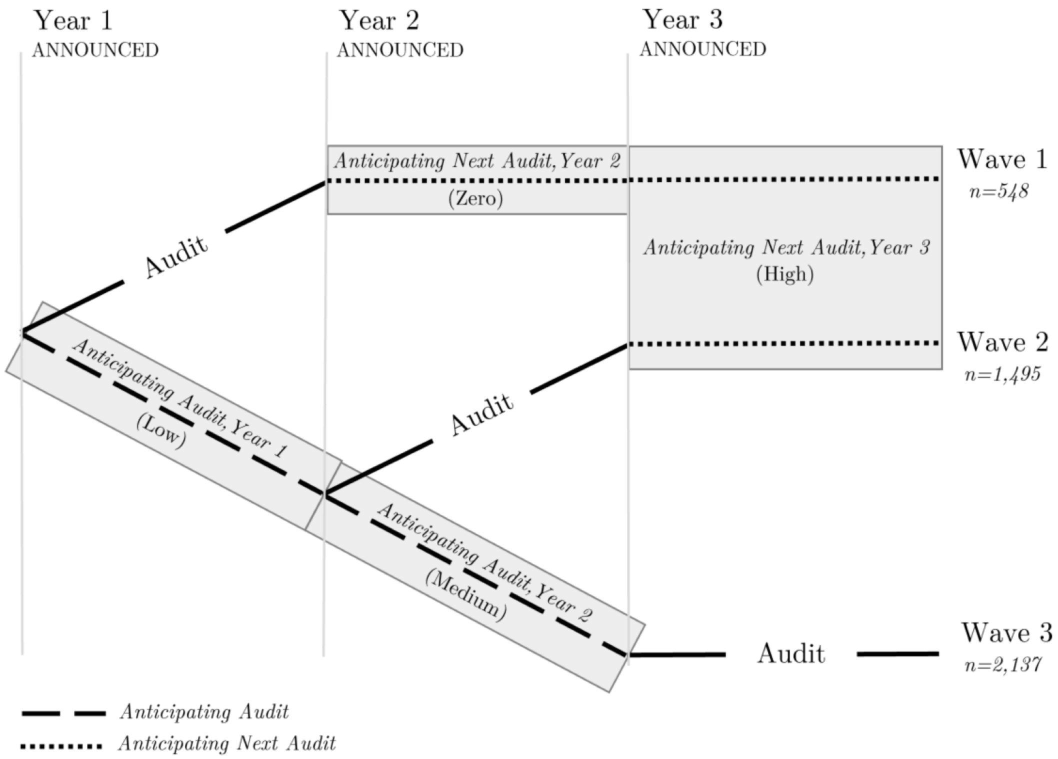


Figure 1: Roll-out of the round of first audits and evolution of beliefs for each wave. Dashed and dotted lines denote horizons of anticipating the first and second audits, respectively. Wave 1 follows the path on the top; Wave 2 follows the path in the middle; and Wave 3 follows the path on the bottom. Anticipation horizons in shaded grey areas are crudely referred to as the 'Zero', 'Low', 'Medium', 'High' probability groups for ease of interpretation in later results. Timing of the announcements is not drawn to scale.

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.

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. In particular, Wave 3's expectations that current performance will be audited increases over time. This is represented by the horizon denoted by the dashed lines in Figure 1, crudely referred to as the 'Low' and 'Medium' probability horizons in Year 1 and 2, respectively, for ease of interpretation of later results. While 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 or ‘Zero’ probability horizon). 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 or ‘High’ probability horizon). About 84% of Waves 1 and 2 GPs compared to 18% of Wave 3 GPs were audited in Year 4, consistent with the national act requiring all GPs to be audited regularly.[17]

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 take place. And part of the information gathered during the audit process is reflective of concurrent program performance, as discussed in Section 2.2. We refer to the Month of Audit as the probability ‘One’ horizon.

Note, these crude labelings of the anticipation horizons as ‘Low’, ‘Medium’, etc. are not assertions of bureaucrats beliefs, they are to ease interpretation of regression estimates. I remain agnostic to what bureaucrats beliefs are until Section 6.1, where I make a range of assumptions about beliefs over next year’s audit capacity to apply a model to determine the optimal communication of audit risk.

3 Empirical Strategy

This section discusses the main specification and assumptions to estimate behavior in anticipation of an audit; and the approach for determining whether the use of changes in program expenditures in response to the monitoring policy, the main outcome measure, is a reasonable proxy for changes in misappropriated spending.

3.1 Estimating equation and identification assumptions

Figure 1 shows that there are five parameters to be estimated on the impact of expectations of being audited on expenditures corresponding to behavior during the horizons of

17. The Year 4 announcement also broadened the scope of performance subject to audit to expenditures from the last three fiscal years. To my knowledge, the announcement was the first time this detail was publicly released.

anticipation: two parameters when bureaucrats are anticipating the first audit in Years 1 and 2; two parameters when anticipating the next audit in Years 2 and 3; and 1 parameter aggregating behavior during the month of audit across Years 1 through 3. These parameters correspond to the main regressors in Equation 1. I estimate the impact of these horizons on program expenditures using an event study specification with 'events' defined by relative time to announcement, see Figure 2. We are interested in identification of each 'event' to understand how bureaucrats respond as expectations of being audited change.

$$\begin{aligned} y_{it} = {} & \alpha_i + \alpha_{dt} + AnticipatingAudit_{it}'\beta + \underbrace{\delta_1 \text{Before}Audit_{it} + \delta_2 \text{Monthof}Audit_{it}}_{\text{post-announcement period, disaggregated}} \\ & + \underbrace{\delta_3 \text{After}Audit_{it} + AnticipatingNextAudit_{it}'\gamma}_{\text{post-announcement period, disaggregated (continued)}} + \varepsilon_{it} \end{aligned} \tag{1}$$

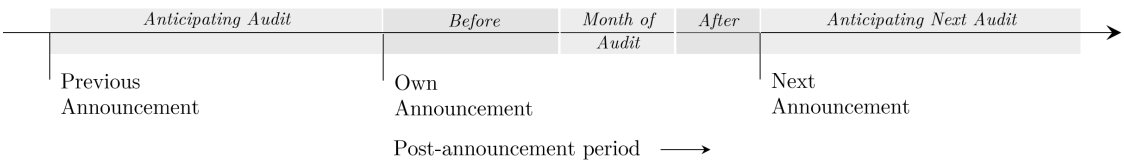


Figure 2: This timeline figure shows how the main regressors in Equation 1 correspond to the time relative to one's own announcement.

The main outcome, denoted as y_{it}, is total expenditures measured as the sum of wage and material expenditures.[18] I also show results for wage and material expenditures separately. Section 3.3 discusses why program expenditures are reasonable in this particular setting for making inferences on bureaucrat changes in financial misappropriation as expectations of being audited change. Results from Sections 4.2 and 4.3 provide further evidence that estimated changes in this outcome measure reflect changes in misappropriation.

α_i denotes fixed effects controlling for time-invariant unobservables for each GP i; α_{dt} denotes fixed effects controlling for GP-invariant unobservables for each month

18. The excluded component of expenditures 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).

t within a district d such as district-level seasonality; ε_{it} is the residual error term.[19] Standard errors are clustered by block—one administrative level above the GP and the unit of stratification for being selected for audit each year.

Identification of each horizon of bureaucrat behavior captured by the remaining variables in Equation 1 will each be discussed in turn.

$AnticipatingAudit_{it}$ is a vector of two dummy variables which capture bureaucrat anticipation of the first audit announcement, separately in Years 1 (β_0) and 2 (β_1), corresponding to the 'Low' and 'Medium' probability horizons referenced in Figure 1. Given random assignment to audit without replacement each year, bureaucrat expectations of their first audit are also randomly assigned, allowing us to identify β. Appendix Table 1 shows overall balance on observable characteristics across waves.

The remaining variables in Equation 1 capture behavior after a GP receives the announcement that they are up for their first audit, i.e. the post-announcement period. Identification of collective behavior during this post-announcement period requires parallel trends across waves leading up to the announcement. I assume parallel trends leading up to the announcement conditional on controlling for anticipatory behavior, so $AnticipatingAudit_{it}$ is not only a parameter of interest but also serves as a control variable for identifying behavior during the post-announcement period. Identification of behavior in the post-announcement period is not violated if we can control for the period of anticipation and assume that parallel trends holds during the pre-period that excludes periods of anticipation.[20] Appendix Figure 1 provides evidence consistent with parallel trends. [21]

19. Districts are two administrative levels above the GP. The 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 $\leq$ 10 GPs; 42% of blocks had $\leq$ 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.

20. See Sun and Abraham (2020) for example.

21. 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. But, without controlling for anticipation, we should not expect parallel trends for subsequent waves because bureaucrats may anticipate their turn for an audit.

The post-announcement period is disaggregated into mutually exclusive dummy variables as follows. The variable Before$Audit_{it}$ captures the period once one learns of their first audit but before the first audit occurs. $MonthofAudit_{it}$ captures the month of audit, crudely referred to as the probability 'One' horizon. $AfterAudit_{it}$ captures the months following the first audit, but prior to learning information from the next announcement. The estimated coefficients for these variables capture the effect of an active audit on bureaucrat behavior as well as any persistent effects of having experienced an audit. While GPs were randomly assigned for audit each year, the order in which audits occur within each year is non-random. Based on the audit agency's account of how the within-year audit schedule was determined, I assume that district-month fixed effects account for the non-random scheduling of audits within the year.[22]

Finally, consider $AnticipatingNextAudit_{it}$, a vector of two dummy variables which capture horizons during which bureaucrats are anticipating their *second* audit in Years 2 (γ_0) and 3 (γ_1), corresponding to the 'Zero' and 'High' probability horizons referenced in Figure 1. The coefficient estimates of $AnticipatingNextAudit_{it}$, γ, are parameters of interest because they capture strategic responses by bureaucrats anticipating their second audits as the round of first audits progress. γ is identified if it is not confounded with persistent effects of the audit. I test for persistent effects with After$Audit_{it}$. I also estimate a by-month event study and test whether behavior after the audit (controlling for variation from $AnticipatingNextAudit_{it}$) is statistically different from behavior before the audit (result reported in Section 4.3).

All regressions use '$AnticipatingAudit$ - Year 1' as the reference group. [23]

Interestingly, we find evidence of this in Online Appendix Figure B.1.

Additionally, the specification in Equation 1 controls for behavior of GPs who are *already treated* with the announcement and first audit, which addresses concerns documented in the recent literature on implementing two-way fixed effects estimators with staggered adoption and heterogeneous treatment effects (Goodman-Bacon 2018; Sun and Abraham 2020; Callaway and Sant'Anna 2021). The main results and findings in this paper are qualitatively consistent when using heterogeneity-robust estimators (Sun and Abraham 2020). Intuitively, the main results are not sensitive to these estimators because behavior when bureaucrats are already treated is controlled for with $AnticipatingNextAudit$. When behavior during these horizons is separately estimated, variation from already-treated cohorts is not used to estimate the treatment effects around the period of the announcement.

22. According to the audit agency, they designed the schedule to complete audits within a district in time for higher-level hearings (for unresolved issues) and to be logistically practical.

23. Otherwise, the parameters of interest would be collinear with a linear combination of time fixed effects. The reason is the horizons of anticipation are defined by when annual announcements for audit are released, and this affects all GPs within a wave simultaneously. Where for every announcement, a GP is either anticipating their first audit or has received an announcement for their first audit.

3.2 Data sources and sample restrictions

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 (Ministry of Rural Development 2019). MIS is a national government data portal that tracks detailed information on program implementation in each GP.[24]

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 through the program. Anecdotal evidence from interviews with government officials suggest that once expenditures have been paid, they cannot be manipulated in administrative data by bureaucrats. Panel datasets constructed with these outcomes are by GP-month from April 2014 to March 2019.

Audit characteristics and outcomes from audit reports include share of portfolio audited, number of auditors, documented issues, and fines assessed for these issues. The audit reports are from MIS and used to construct a GP-level dataset. This is used to compare audit performance across waves with differing anticipatory behavior. Currently, only a subset of audit reports are available for analysis, and are only from Waves 2 and 3 of the audit.[25]

Sample restrictions By the end of Year 3, audits were conducted in 4,180 GPs 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 at the request of upper-level

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

25. Among audits conducted in Waves 2 and 3, 77% of audit reports are available for analysis. In tests comparing pre-audit GP characteristics of Waves 2 and 3 with available audit reports, there is balance on GP population characteristics, number of auditors, and total expenditures under audit. There are statistically significant differences in wage expenditures and the number of projects under audit, which will be included as controls in the regressions with audit reports data.

government officials[26]; 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, 112 GPs were selected for audit twice over the audit roll-out. A subset of these 112 GPs were audited twice because they were also selected for special audit; for the remainder, they were 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, and 3,897 GPs in the sample for analysis in the balanced panel. All analyses use the balanced panel.

3.3 Inferring deterrence from administrative data

The main analysis of this paper estimates the effect of the monitoring policy on total expenditures (the main outcome measure) and not misappropriated expenditures (the ideal outcome, but a measure that is unobservable).[27] I provide evidence in Section 4.2-4.3 that the estimated changes in total expenditures ($\hat{\Delta}_y$)—caused by changes in expectations of being audited resulting from random assignment to audit timing—are a reasonable proxy for changes in misappropriated expenditures (Δ_m, or the deterrence effect). Bandiera et al. (2009) and Londoño-Vélez and Ávila-Mahecha (2021) use a similar approach to study waste in government procurement and tax evasion.

Changes in total expenditures (Δ_y) decompose into changes in honest (Δ_h) and misappropriated (Δ_m) spending. That is, $\Delta_m = a\Delta_y$ where $a \in [0,1]$ and $\Delta_h = (1-a)\Delta_y$.

26. Requests 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.

27. The outcome measures 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. See Online Appendix A.1 for description of bureaucrats at the GP level.

We can make inferences on the possible scenarios of comovement in honest and misappropriated spending based on our estimated $\hat{\Delta}_y$. For instance, a decline in total expenditures ($\hat{\Delta}_y < 0$) can only be explained by a decline in both or one of Δ_h and Δ_m. I consider the possible changes in honest and misappropriated spending that may explain the estimated $\hat{\Delta}_y$ by drawing on additional data and analyses. I argue with findings in Section 4.2-4.3 that the estimated changes in total spending, $\hat{\Delta}_y$, are driven by changes in misappropriated spending and that it is difficult for alternative mechanisms which involve changes in honest spending to rationalize the combination of findings. Finally, in Section 6.2, I discuss the conditions under which the findings in this paper are robust to relaxing this assumption that misappropriated spending is measured without error.

Could we have done better with a different outcome measure? The monthly administrative data on expenditures that I use in this paper are the only measures of GP *monthly* performance available. A frequent (monthly) independent verification of employment and material procurement over time would be the ideal measure of the deterrence effect from the impact of frequently-occurring changes in expectations of being audited.[28] But, it is expensive to conduct audits with this kind of intensity and inefficient if optimal deterrence can be achieved with audits of less frequency. In this setting, a GP's performance is being verified through audits once every few years.

4 The Impact of Changing Expectations

This section shows that the deterrence effect from anticipation is strongest on the margin when one has high expectations of being audited, while under less certainty there is no detectable effect. I provide evidence we can interpret the results as deterrence effects (changes in misappropriated expenditures). I also show bureaucrats strategically adjust the way they misappropriate spending, particularly they substitute spending across time and type of expenditure.

28. Online Appendix B.2 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 from administrative data or audit reports are too coarse to learn about bureaucrats' responses to the monitoring policy, compared to what we learn from monthly administrative data.

4.1 Bureaucrats' response to changing likelihood of audit

Table 1 estimates the event study specification (Equation 1) for total, wage, and material expenditures. As discussed in Section 3.1, the reference group or horizon of anticipation is $AnticipatingAudit$ - Year 1 (Low).

Total expenditures slightly increase, though statistically insignificant, when GPs have Medium expectations. Notably, the estimate is driven by a statistically insignificant decline in wage expenditures and a 24% significant increase in material expenditures, suggesting bureaucrats substitute from wage to material expenditures. Similarly, when expectations are Zero (the year after Wave 1 receives their audit) there is no detectable change in total expenditures and this is coupled with an average decline in wage and increase in material expenditures. This suggests bureaucrats resume business as usual overall when they have lower expectations of being audited.

Expenditures decline substantially when expectations of being audited are High. As the roll-out completes in Year 3 and the likelihood of a second audit in Year 4 is extremely likely, there is a 15% drop in total expenditures. This is driven by a 15% decline in wage expenditures and a 13% statistically insignificant decline in material expenditures. Similarly, during the month of audit when concurrent performance is subject to monitoring and when expectations are One, there is a 16% decline in expenditures largely driven by a decline in wage expenditures. The estimates for expectations High and One are both statistically different from behavior when expectations are lower captured by Medium and Zero expectations.

	Expenditures (1,000 INR):		
	(1)	(2)	(3)
	Total	Wages	Materials
AnticipatingAudit - Year 2 [Medium]	13.87	-5.98	19.84**
	(11.33)	(6.05)	(8.58)
PostAnnounce, disaggregated:			
Before *Audit*	-18.83**	-10.21**	-8.62
	(9.46)	(5.04)	(7.14)
Month of *Audit* [One]	-42.07***,†	-35.33***,†	-6.74
	(11.68)	(5.99)	(9.30)
After *Audit*	-8.42	-17.54***,†	9.12
	(11.26)	(5.62)	(8.70)
AnticipatingNextAudit - Year 2 [Zero]	-4.78	-14.54*	9.76
	(13.83)	(7.67)	(9.07)
AnticipatingNextAudit - Year 3 [High]	-39.38***,†	-28.06***,†	-11.32
	(13.30)	(7.28)	(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 1: Effect of stages of the monitoring policy on Bureaucrats' response in program expenditures. This table estimates the main event study specification (Equation 1) for three outcome variables: total (wages + materials), wage, and material expenditures. The first term in brackets refers to the estimated coefficient; the second term in brackets refers to the probability horizons referenced in Section 2.3. 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 (*AnticipatingAudit* - 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 *AnticipatingAudit* - Year 1. This longer period is included in the baseline to average out seasonal variation in expenditures. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. With a Bonferroni correction for the family-wise error rate, we reject the null hypothesis when † $p<0.008$.

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 3.1). The effects on total and material expenditures during the months after the audit (After*Audit*) are insignificant and small. 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. 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 4.3. These results suggests the estimates for *AnticipatingNextAudit* are not confounded with persistent effects from the audit.

4.2 Deterrence and substitution in anticipation

Recall that Table 1 shows that anticipatory behavior when expectations are Medium relative to Low (which is effectively the difference in behavior between Waves 2 and 3 in the year prior to their respective audits) does not lead to a detectable change in expenditures. That is, we cannot rule out $\hat{\Delta}_y = 0$ (using the terminology described in Section 3.3). When disaggregated, we see a significant increase in material spending and an average decrease in wage spending.

Does this result reflect changes, if at all, in honest or misappropriated spending (i.e. changes in Δ_h or Δ_m, respectively)? Bureaucrats anticipating an audit may be deterred from misappropriating expenditures, they may be motivated to improve their performance leading to changes in honest spending, or they may not react at all. I use auditor reports from Years 2 and 3 to estimate differences in bureaucrat performance as a way to verify whether spending during this period was honest or misappropriated. I do so with auditor reported issues related to bureaucrat incompetencies, which I interpret to reflect changes in honest bureaucrat behavior and consequently honest spending,[29] and issues related to misappropriated spending in wages and materials. When examining issues related to lack of bureaucrat competency, I disaggregate them into the quality of exerted effort (poor planning and quality) and whether effort was exerted at all (negligence of tasks). Table 2 tests for differences in auditor detection of these issues between Waves 2 and 3, who had Low and Medium expectations, respectively, during the year prior to their audit which was subject to auditor scrutiny.

Columns 1 and 2 of Table 2 show there is no detectable difference between Waves 2 (Low) and 3 (Medium) in fine amount or number of issues related to lack of competency, suggesting changes in honest spending are not driving the measured effect on total spending for the Medium group (relative to Low group) from Table 1. We also find no difference between these groups in bureaucrat efficiency in making timely wage payments (Appendix Table 2). Timeliness of wage payments is an important measure that higher-level government officials use to regularly monitor GP bureaucrat performance, and not known to be possible mechanism for bureaucrats to leverage to misappropriate

29. Honest behavior is that which was intended to contribute to program goals and do not entail personal financial gain to bureaucrats. They can include behavior that is wasteful, possibly due to a lack of skill or training.

spending. In fact, there are null effects on timeliness of wage payments from all horizons of anticipation. These findings are consistent with: the estimated changes in total spending are not driven by changes in honest spending ($\Delta_h \approx 0$).

Columns 3 and 4 of Table 2 show the difference between Waves 2 (Low) and 3 (Medium) in fines and number of issues related to wage misappropriation is statistically indistinguishable from 0. This effect is consistent with the estimate for the Medium group (relative to Low) of no detectable change in spending on wages (Table 1). Additionally, Wave 3 experiences a threefold increase in fines related to issues of material receipt misappropriation compared to Wave 2, and this effect is highly statistically significant. This is consistent with the results that Wave 3 spent 24% more on materials during this period of anticipation, showing evidence of strategic adjustments (and not pure deterrence) as expectations of being audited increase.

These results lend credibility to our interpretation that the estimated differences in bureaucrat behavior during periods of anticipation are interpretable as a deterrence effect. 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 the type of expenditure to misappropriate.

The shift from misappropriating wages to materials is specific to this context and is perhaps explained by lower detection rates of material relative to wage misappropriation. Audit reports show that issues related to wage misappropriation are more than twice as likely to be documented compared to material misappropriation (16% vs 6%, respectively), but materials are also a smaller share of total expenditures. Bureaucrats may be choosing material over wage expenditures because they believe material misappropriation is easier to hide from auditors, as wage misappropriation can be discovered by auditors during household interviews–evidence supporting this is discussed in Online Appendix C.2.

4.3 Deterrence and substitution during the audit

The following results support that the estimated decline in total expenditures during the 'Month of *Audit*' (Probability One relative to the Low group) in Table 1 is driven by

	Issue type:			
	Lack of competency		*Misappropriated spending*	
	Poor planning and quality	Negligence of tasks	Wage misappropriation	Material receipts misappropriation
	(1)	(2)	(3)	(4)
	Panel A: Issue fine amount (1000 INR)			
Wave 3 - Wave 2	6.973	-24.253	18.413	64.510***
	(12.051)	(17.884)	(17.436)	(17.605)
Wave 2 Mean	42.24	14.43	73.48	22.08
Adjusted R^2	0.195	0.061	0.187	0.033
	Panel B: Number of issues			
Wave 3 - Wave 2	0.242	0.419	0.040	1.046***
	(0.416)	(0.438)	(0.516)	(0.256)
Wave 2 Mean	3.5	2.75	3.79	0.4
Adjusted R^2	0.223	0.139	0.563	0.288
Observations	1,772	1,772	1,772	1,772

Table 2: Differences in audit issue fines across waves are consistent with interpreting differences in wage and material expenditures while anticipating the first audit as misappropriated. Unit of observation is GP. All regressions include block and audit manager fixed effects, and a vector of control variables: number of employed households and works to verify; number of auditors, and audit manager experience (number of audits conducted to date) . Standard errors are clustered by block. *p<0.1; **p<0.05; ***p<0.01.

changes in misappropriated spending. That is, $\hat{\Delta}_y < 0$, using the terminology described in Section 3.3, is driven by $\Delta_m < 0$. The results also provide evidence that bureaucrat's strategically substitute misappropriated spending across time and expenditure type as expectations of being audited change.

The experience of undergoing an audit could lead to changes in both honest and misappropriated spending. Bureaucrats may be deterred from misappropriating expenditures because concurrent and past performance is being evaluated by auditors (see discussion in Section 2.2). However, an ongoing audit could also be disruptive to business-as-usual activities leading to changes in honest spending if bureaucrats are multi-tasking between usual work and audit-related activities (Holmstrom and Milgrom 1991), even though these audits are not designed to be disruptive. For instance, employment may be halted and worksites may be closed. Additionally, the experience of an audit could help bureaucrats learn how to do their jobs better leading to changes in honest spending, and improvements in productivity (Arrow 1962; Syverson 2011).

To disentangle these potential mechanisms and examine whether the decline in total spending is driven by changes in honest or misappropriated spending, I estimate event studies to examine month-to-month responses around the time of audit. Figure 3 presents the event studies for total, wage, and material expenditures. Relative to 10 and more months prior to audit (the omitted category), total expenditures decrease by 9-12% during the month of the audit, proceed to increase by 3-6% 2-4 months after the audit though not statistically significant, and revert to pre-audit levels afterward. This effect is driven by a 17% decline in wage expenditures during the month the audit and an 24-30% increase in material expenditures 2-4 months after the audit.

Bureaucrats learning to do their jobs better as a result of experiencing an audit is unlikely. Anything learned should persist, but the observed changes are specific to the months around the time of audit. F-tests for each outcome cannot rule out the hypothesis of equal trends during the months before and after the period where bureaucrats respond to 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, and why the disaggregated event studies in Figure 3 reflect changes in misappropriated spending (Δ_m).

First, while the presence of auditors affects employment and material procurement, it does not affect the timeliness of wage payments (Appendix Table 2). This matters because if multi-tasking explains the decline in employment during an audit, then it may also affect other tasks, like making timely payments. Local bureaucrats are responsible for submitting payrolls of time worked to higher administrative offices who then make the payments.

Second, bureaucrats' responses in the months around the audit are unlikely to have impacted real program output, providing further evidence that multi-tasking issues are less likely. 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). This suggests real output depends on inputs from both labor and materials.

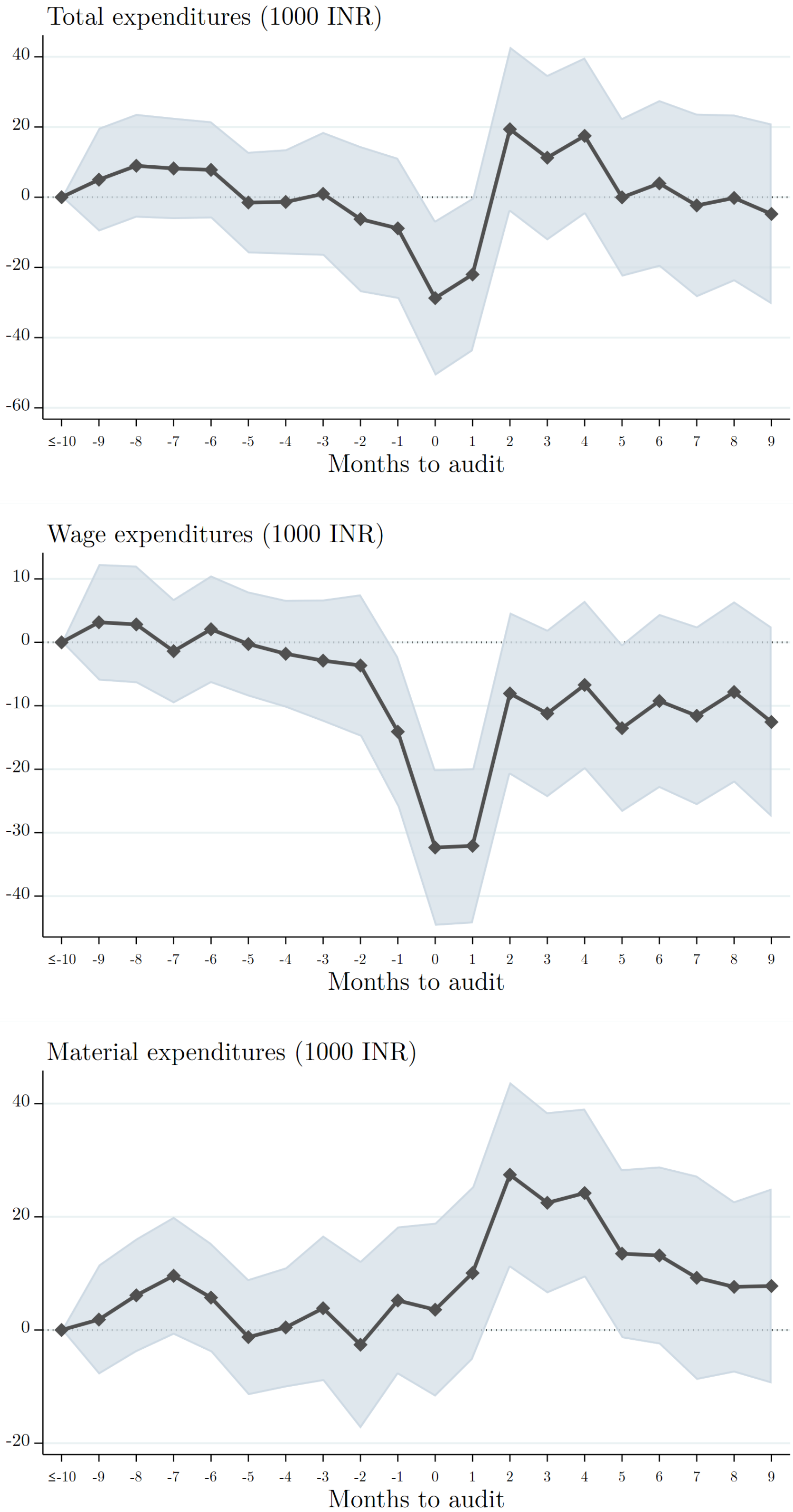


Figure 3: Changes in expenditures around the time of audit. Event studies around the month of audit are estimated with: $y_{it} = \alpha_i + \alpha_{dt} + AnticipatingAudit'_{it}\beta + \sum_{k\in\tau} \delta^k Audit^k_{it} + \varepsilon_{it}$ where $Audit^k_{it}$ is an indicator taking the value 1 if i is k months from audit at time t. 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. 95% confidence intervals are captured by the shaded region. Standard errors are clustered by block.

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. But Figure 3 shows no such decline and instead we observe an increase in procured materials occurring the month after the audit and after the decline in employment.

Substitution from labor to material inputs in a way that affected real output is also unlikely to have been induced by the audit. To spend (more) on labor relative to materials is an indicator of NREGS program performance. Engaging in projects that required more materials than labor is not an indication of better performance.[30] Furthermore, the observed decline in employment in the event studies is observed both across projects that require materials and projects that only require labor (see Appendix Figure 2). Substitution between inputs in a way that changes real output is not consistent with this combination of findings.

Taken together, these results show that the substitution between wage to material expenditures is driven by changes in misappropriated rather than honest spending, and bureaucrats strategically substitute misappropriated spending across time and expenditure type as expectations of being audited change.[31] Bureaucrat adjustments around the time of audit were short-lived at best; lost rents were recovered later as bureaucrats substitute their behavior across time and type of expenditure (from wage to material misappropriation). These adjustments are consistent with the observed behavior of bureaucrats anticipating their first audit (findings from Section 4.2).

30. The 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 manual 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.

31. These results suggest that false payments equivalent to 18% of the level of baseline employment were averted during the month of 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.

5 A Model of Information Design

With insights into how bureaucrats respond with changing expectations, what is the best policy for the audit agency in communicating expectations of being audited? This section presents an example that illustrates the main intuition of a model applied to the results from Table 1 to determine the Principal's optimal signal. This example 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 of when an audit will occur when selection is 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 $q_0 = \frac{M}{N}$ of Bureaucrats. Let $U(q)$ be the amount of deterrence of misappropriated expenditures for some likelihood of audit q. In the first period, a likelihood q_0 of audit for all N Bureaucrats leads to a deterrence of $U(q_0)$. In the second period, deterrence is $U(q_0)$ under randomization with replacement. Under randomization without replacement, the deterrence in the second period is $q_0 U(0) + (1-q_0)U\left(\frac{q_0}{1-q_0}\right)$. The share q_0 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 $1-q_0$) believe they will be audited with probability $\frac{q_0}{1-q_0}$ and adjust their behavior accordingly (second term). This gives us $U(q_0) \lesseqgtr q_0 U(0) + (1-q_0)U\left(\frac{q_0}{1-q_0}\right)$. Jensen's inequality tells us that the convexity of $U(\cdot)$ is a necessary and sufficient condition for determining when auditing without replacement yields more deterrence than with replacement.[32] The shape (or convexity) of $U(\cdot)$ is determined by the relative benefit to bureaucrats from misappropriating additional expenditures considering the costs from being caught as q changes.

In the model, which is based on insights and results from Lazear (2006) and Kamenica and Gentzkow (2011), the Principal (e.g. audit agency) communicates infor-

32. This simple 2-period example can be easily extended to match our 3-period empirical setting and would yield the same conclusions.

mation about the monitoring policy, which influences the Bureaucrat's decision on the amount of expenditures to misappropriate. The shape of the relationship between the Bureaucrat's choice with respect to their expectations of being audited, i.e. $U(q)$ is the object of interest. This relationship is sufficient for determining the Principal's optimal signal and for evaluating how alternative signals affect the Principal's welfare. The following proposition from the model tells us under what conditions for $U(q)$ is providing more information to the bureaucrat better. The exposition of the full model is in Online Appendix D.

Proposition 1 (When providing more information is better). *If $U(q)$ is convex at q_0, then $U(q_0) < \hat{U}(q_0)$ and an informative signal is preferred. This result implies the following about the optimal signal given the shape of $U(q)$:*

(i) The Principal prefers an informative over uninformative signal for any q_0 if $U(q)$ is convex for all q. The Principal prefers an uninformative signal for any q_0 when $U(q)$ is concave for all q.

(ii) An informative signal is still preferred if $U(q)$ is convex at q_0 and concave elsewhere for some q.

A geometric interpretation of this result means that a Principal prefers to send an informative signal if $U(q)$ is convex 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. Online Appendix Figure D.1 provides illustrations of when informative versus uninformative signals are better.

Ultimately, the shape of $U(q)$ in a particular setting must be learned empirically. Estimating $U(q)$ is sufficient for determining the optimal signal and analyzing the Principal's welfare under alternative signals. If $U(q)$ is convex at prior q_0, then communicating more information about the likelihood of being audited is better than maintaining uncertainty over when the audit will occur. With the random selection of GPs for audit without replacement in this empirical setting, I estimate $U(q)$ with causally-identified empirical moments in the data and apply the results from Proposition 2 to determine the best policy on communicating the likelihood of being audited (Section 6).

6 Information Design and Counterfactuals

Section 5 tells us that the shape of bureaucrats' deterrence given their expectations of being audited is sufficient for determining the optimal signal to communicate audit risk and evaluating the Principal's welfare under alternative signals. We can estimate this relationship with results on total expenditures from Table 1 by leveraging the changing expectations of bureaucrats from random assignment to audit without replacement. We can do so under reasonable assumptions (described in Section 6.1) and without needing to specify the underlying primitives of the choice problem, like parameters for preferences and constraints.

I find the estimated relationship (between bureaucrats' deterrence and their expectations of being audited) is largely convex. This implies, in this particular setting, that policies which communicate full information about audit risk will yield more deterrence than policies which maintain uncertainty (Section 6.2). Online Appendix E.1 provides robustness checks and analyzes the sensitivity of the conclusion by relaxing assumptions. Section 6.3 uses the estimated relationship to simulate welfare consequences to the Principal under alternative signals.

6.1 Assumptions for estimating the optimal signal

I make two assumptions for estimating $U(q)$ to determine the optimal signal.

Assumption 1. *The deterrence response of bureaucrats depends only on their current expectations of being audited.*

First, Assumption 1 implies that the estimated differences in anticipatory behavior (results from Table 1) are only explained by differences in bureaucrats' expectations of being audited. This assumes the results are not confounded with other changes to the environment over time which may also affect deterrence, like bureaucrats' changing perceptions of audit quality or credibility.[33]

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

We can test for violations of this assumption. Appendix Table 3 shows that there are no significant differences across years in audit quality, measured as differences in audit inputs. This result together with the fact that the rules of the audit were unchanged support that the quality of audit implementation did not change over time. If the perceived quality or credibility of the audit is changing, then one might expect the bureaucrat response to vary by wave cohort. But, there are also no significant differences across waves during the month of audit and when multiple waves experience the same horizon of anticipation (Appendix Table 4). Furthermore, if bureaucrat perceptions and learning of the audit were informed by peer experiences, Appendix Table 5 shows that the anticipatory and direct effects of the audit are unaffected after accounting for concentration of audits within one's local peer network (spillover effects). These results provide additional evidence that the anticipatory effects are driven by changes in expectations of being audited rather than changing bureaucrat perceptions about the audit. For robustness, this assumption is later relaxed in a sensitivity analysis by assuming a bias in our estimate of the deterrence response conditional on expectations of being audited (Online Appendix E.1).

Second, Assumption 1 implies that the bureaucrat's response only depends on their current beliefs and not the history of actions or beliefs leading up to the current period. This also means that the bureaucrat's response depends only on their current beliefs and not the policy that generated those beliefs. That is, conditional on their expectations of being audited, 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.

Assumption 2. *Based on information communicated by the audit agency, bureaucrats perceived that GPs would take turns being audited and were uncertain about next year's audit capacity.*

To infer bureacrat expectations over the likelihood that this year's work will be audited (q), we consider a flexible range of beliefs over next year's audit capacity as follows: Let $K_{\tau+1}$ denote bureaucrats beliefs about next year's audit capacity where time τ is the current year. Assume $K_{\tau+1}$ could be equal to: (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)* $Trend_{\tau} \times K_{\tau}$*, 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.*

	Likelihood this year's work will be audited, q				
	Assumptions on next year's audit capacity, $K_{\tau+1} =$				
	$K_{\tau-1}$	$\frac{1}{2}(K_{\tau-1}+K_{\tau})$	K_{τ}	$\text{Trend}_{\tau} \times K_{\tau}$	$K_{\tau+1}$
AnticipatingAudit - Year 1 (Low)	0.00	0.07	0.14	0.14	0.39
AnticipatingAudit - Year 2 (Medium)	0.24	0.44	0.65	1.00	0.92
Month of *Audit* (One)	1.00	1.00	1.00	1.00	1.00
AnticipatingNextAudit - Year 2 (Zero)	0.00	0.00	0.00	0.00	0.04
AnticipatingNextAudit - Year 3 (High)	0.67	0.82	0.96	1.00	1.00

Table 3: Bureaucrats' beliefs over the likelihood this year's work will be audited based on Assumption 2. $Trend_{\tau} \times K_{\tau}$ assumes next year's capacity is based on the growth rate in audit capacity from last year to this year. Beliefs do not vary across assumptions for 'Month of *Audit*' because concurrent performance is susceptible to detection by auditors. So, while an audit is happening, I assume bureaucrats believe today's performance will be audited with probability 1. Beliefs largely do not vary across assumptions for '*AnticipatingNextAudit* - Year 2' because Wave 1 GPs in the year following their audits (Year 2) believe they will not be audited a second time until the round of first audits has completed.

Assumption 2 helps us pin down bureaucrats changing expectations of being audited during the round of first audits.

First, Assumption 2 states that bureaucrats made decisions based on the information made available to them by the audit agency from 2016-19; that is, they have rational expectations. Thus, they had knowledge of and acted according to GPs being audited in turns leading to anticipatory behavior (achieved by the audit agency's randomization without replacement described in Section 2.3). The best supporting evidence of this is the difference in bureaucrat anticipation in the months leading up to the announcement in their respective waves (see Appendix Section B.1). In particular, when restricting the sample to periods prior to the second announcement, there is no evidence of differing pre-trends leading up to each wave's respective announcements. In contrast, when comparing only Waves 2 and 3 GPs in the periods before the third announcement, we observe declines in spending showing differences in anticipation leading up to Waves 2 and 3's respective announcements.

In an attempt to better support this assumption, I implemented a small phone survey in 2020 to measure bureaucrat perceptions 4 years after the audits began. But during the height of pandemic lockdowns in India, I was only able to elicit responses from 23 GPs, making it hard to confidently draw conclusions from this survey data about bureaucrats' past perceptions. What I find among the small sample of responses is that a majority of bureaucrats were either ex-post unsure about how GPs were selected or believed that GPs took turns getting audited. This suggests that rational expectations may be violated for some bureaucrats. However, given random assignment (without replacement) to wave, I argue that the implications are minimal, especially considering that the best evidence supporting this assumption is the abovementioned differences in anticipation across waves using data from *all* GPs in the study sample. In particular, when a bureaucrat deviates from having rational expectations, their decisions do not depend on the information communicated by the audit agency; thus, the behavior driven by these deviations is likely balanced across audit waves, since waves were randomly assigned and determined by the information communicated by the audit agency. This implies that bureaucrat responses stemming from deviations from rational expectations would likely add noise to the main effects estimated across the horizons of anticipation. See Online Appendix A.5 for more details on survey interpretation.

Second, Assumption 2 also states that bureaucrats were uncertain about next year's audit capacity, as this was not announced. Bureaucrats expectations that this year's work gets audited is a function of the number of audits to be conducted in the next year (which is unknown) as a share of the remaining GPs to be audited (which is known from the announcements) because of the randomization without replacement. To be agnostic about how beliefs about next year's audit capacity are formulated, I make a flexible range of assumptions on bureaucrats' beliefs about next year's audit capacity (Table 3).[34]

34. Probabilities for *AnticipatingAudit* - Year 1, *AnticipatingAudit* - Year 2, *AnticipatingNextAudit* - 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 *AnticipatingAudit* - 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. The special audit in FY2017-18 audited an additional 175 GPs for the first time. Likewise, 21 GPs

6.2 The optimal design of information

We can estimate the function $U(q)$ to determine the optimal signal by mapping the points of support on bureaucrat beliefs from Table 3 ("x" values) to their associated estimates of deterrence in total expenditures from Table 1 ("y" values). Based on the model in Section 5, we can interpret this graph of $U(q)$ as the Principal's expected utility as a function of the Bureaucrats' beliefs about being audited. Greater declines in expenditure means more deterrence and higher expected utility for the Principal. Figure 4 plots $U(q)$ under each beliefs assumption.

Proposition 2 tells us that 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. We evaluate this at the audit agency's capacity, q_0, which ranged from 14-53% during the study period. That is, rather than leaving bureaucrats with their prior expectations of being audited (better when concavities present), could the Principal do better by informing those selected for audit they are high risk and those not selected that they are low risk (better when convexities present)?

If we consider the average deterrence response (black curves in Figure 4), then under all belief assumptions the curves are convex at q_0 (the observed range in capacity to conduct audits each year, 14-53%).[35] In other words, with a constrained budget, the Principal's optimal policy is to provide information on whether one is at high risk of being audited or not. In particular, when $K_{\tau+1} = \{K_\tau, Trend_\tau K_\tau, K_{\tau+1}\}$, the optimal signal is to communicate whether GPs have been selected for audit or not, providing full information. When $K_{\tau+1} = \{K_{\tau-1}, \frac{1}{2}(K_{\tau-1} + K_\tau)\}$, the optimal signal is to communicate to those selected for audit that they are high risk (67% likely when $K_{\tau-1}$ and 82% likely when $\frac{1}{2}(K_{\tau-1} + K_\tau)$) and those not selected that they will not be audited. Sensitivity analyses in Online Appendix E.1 show this result, that $\hat{U}(q)$ is convex at q_0, is robust to statistical noise and relaxing Assumptions 1 and 2.

We have at maximum five points of support to interpolate the shape of $U(q)$ using

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.

35. 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.

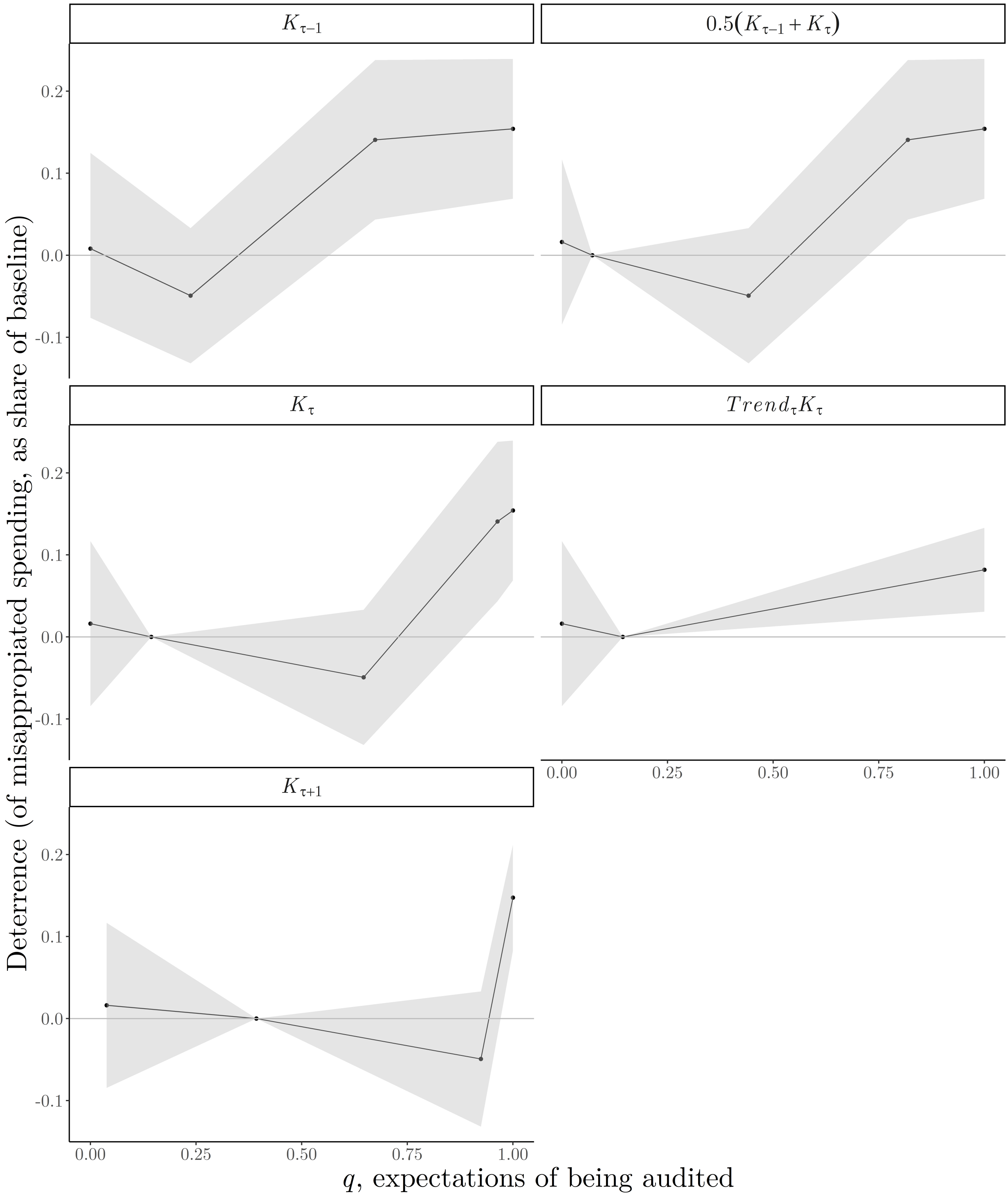


Figure 4: Principal's expected utility as a function of Bureaucrats' beliefs, under different beliefs assumptions. Black dotted line plots the mean of regression estimates with grey shaded area representing the 95% confidence interval for each parameter estimate. The deterrence estimates are relative to the reference group, '*AnticipatingAudit* - Year 1', and reported as a share of the baseline mean. 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.

a piece-wise linear function. A limitation of this approach is that this interpolation may smooth out any local concavities that make an uninformative signal optimal. If we are correct about our theoretical assumption in Section D.1 that bureaucrat deterrence is monotone in expectations of being audited, then this is of limited concern.

Finally, under certain conditions, this result holds if we relax the assumption that we can use program expenditures as a proxy for changes in misappropriated spending (Δ_m)–without error coming from changes in honest spending (Δ_h). This result does not change if the change in honest spending with respect to expectations of being audited is constant, i.e. $\frac{\partial \Delta_h}{\partial q} = 0$. This would lead to a uniform shift in the estimated $U(q)$ in Figure 4 without changing its convexity, and thus without a change to the optimal signal. This result also does not change if $\frac{\partial \Delta_h}{\partial q} > 0$, where honest spending increases as expectations increase leading to greater declines in misappropriated spending as expectations increase. This would preserve the convexity in the estimated $U(q)$ in Figure 4. This result would fail to hold for specific cases of a nonlinear change in honest spending as expectations increase, which would violate our theoretical assumption of monotonicity in bureaucrat behavior as expectations change, and specific cases where honest spending decreases as expectations increase.

6.3 Counterfactual signals and welfare consequences

This section estimates the welfare consequences to the Principal under alternative signals on communicated expectations of being audited. Welfare under each policy is calculated over the course of the 27 months that the round of first audits took place. Welfare is estimated using the joint sampling distribution of deterrence behavior under the various horizons of anticipation shown for various assumptions of $K_{\tau+1}$ in Figure 4. This joint sampling distribution is produced with a block bootstrap,[36] and $U(q)$ is estimated for each bootstrapped set of estimates. This approach allows us to report the standard errors of the welfare estimates using causally-identified empirical moments in the data. Online Appendix E.2 provides further details on the calculations.

Assuming $K_{\tau+1} = K_\tau$, the results show that spending under the full information

36. This is the same block bootstrap implemented for the sensitivity analyis (Online Appendix E.1). 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.

signal during this period would have been USD 226 million (s.e. = 12). Compared to the partial information signal (modeled as the actual policy implemented randomizing *without* replacement), the full information signal would have deterred 10% more in misappropriated expenditures on average, which is equivalent to USD 22 million (p-value<0.001). Compared to the uninformative signal (randomization *with* replacement), the full information signal would have deterred 16% more in misappropriated expenditures, which is equivalent to USD 37 million (p-value<0.001)–all without changing the Principal's budget for audits. The conclusions are qualitatively similar under other belief assumptions for $K_{\tau+1}$ (see Appendix Table 6) and if we exclude the month of audit from the analysis.

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. The exercise conducted in this empirical setting makes a strong case for evaluating the possibility that atypical audit strategies, like implementing and informing of an audit in advance, may yield significant returns at no additional cost.

7 Conclusion

The monitor's resource constraints imply that a subset of bureaucrat activity will go unchecked, potentially allowing bureaucrats to adapt. Monitoring policies designed to maximize deterrence must account for bureaucrat attempts to evade detection. This paper provides empirical evidence on strategic responses by bureaucrats to monitoring. Taking into account these strategic responses, this paper also shows how information disseminated about the likelihood of audit can be optimally designed with a budget-constrained policymaker in mind.

Results show that the deterrence of misappropriated expenditures is strongest when one is almost certain of an audit, while the response under less certainty is statistically indistinguishable from zero. In addition, the unintended consequences of monitoring policies matter. When expectations of being audited increase, bureaucrats substitute across time and type of expenditure to misappropriate.

Interpreting these findings with a model of Bayesian persuasion, I find that designing monitoring policies which inform bureaucrats in advance would yield the most deterrence. Signals which provide more information are better. This implies randomization without replacement is better than with replacement. I arrive at this result by estimating a function ($U(q)$) from the model, which provides sufficient information for us to solve for the optimal signal and analyze welfare under counterfactuals. I estimate $\hat{U}(q)$ using causally-identified moments and a set of flexible assumptions on bureaucrats' beliefs.

This paper provides a novel empirical measure of the value of information. In this setting, up to USD 35m (16% of average annual program expenditures) could have been saved. This case study contradicts auditing standards that advocate maintaining unpredictability among audit subjects. The findings of this paper emphasize that the best practice depends on the relationship between deterrence and bureaucrats' expectations of being audited (or another policy parameter of interest in a different setting).

Author affiliations: Wendy Wong is a Postdoctoral Scholar with the Development Innovation Lab at the University of Chicago's Becker Friedman Institute for Economics.

References

Afridi, Farzana, and Vegard Iversen. 2014. “Social Audits and MGNREGA Delivery: Lessons from Andhra Pradesh.” *India Policy Forum* 10 (1): 297–341.

Arrow, Kenneth J. 1962. “The economic implications of learning by doing.” *Review of Economic Studies* 29 (3): 155–173.

Avis, Eric, Claudio Ferraz, and Frederico Finan. 2018. “Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians.” *Journal of Political Economy* 126 (5): 1912–1964.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. “Active and passive waste in government spending: Evidence from a policy experiment.” *American Economic Review* 99 (4): 1278–1308.

Banerjee, Abhijit, Esther Duflo, and Rachel Glennerster. 2008. “Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system.” *Journal of the European Economic Association* 6 (2-3): 487–500.

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande. 2020. “E-Governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India.” *American Economic Journal: Applied Economics (forthcoming).*

Banerjee, Abhijit, Esther Duflo, Daniel Keniston, and Nina Singh. 2019. “The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India.” Working Paper.

Bergemann, Dirk, and Stephen Morris. 2019. “Information Design: A Unified Perspective.” *Journal of Economic Literature* 57 (1): 44–95.

Bobonis, Gustavo J., Luis R.Cámara Fuertes, and Rainer Schwabe. 2016. “Monitoring corruptible politicians.” *American Economic Review* 106 (8): 2371–2405.

Callaway, Brantly, and Pedro H. C. Sant’Anna. 2021. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics (Forthcoming).*

Carrillo, Paul, Dina Pomeranz, and Monica Singha. 2017. “Dodging the taxman: Firm misreporting and limits to tax enforcement.” *American Economic Journal: Applied Economics* 9 (2): 144–164.

Casaburi, Lorenzo, and Ugo Troiano. 2016. “Ghost-house busters: The electoral response to a large anti-tax evasion program.” *Quarterly Journal of Economics* 131 (1): 273–314.

Chassang, Sylvain, Lucia Del Carpio, and Samuel Kapon. 2020. “Making the Most of Limited Government Capacity: Theory and Experiment.” *NBER Working Paper* (Cambridge, MA).

Chetty, Raj. 2009. “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods.” *Annual Review of Economics* 1 (1): 451–488.

Di Tella, Rafael, and Ernesto Schargrodsky. 2004. “Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack.” *American Economic Review* 94 (1): 115–133.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. “Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india.” *Quarterly Journal of Economics* 128 (4): 1499–1545.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. “Incentives work: Getting teachers to come to school.” *American Economic Review* 102 (4): 1241–1278.

Eeckhout, Jan, Nicola Persico, and Petra E. Todd. 2010. “A theory of optimal random crackdowns.” *American Economic Review* 100 (3): 1104–1135.

Ferraz, Claudio, and Frederico Finan. 2011. “Electoral accountability and corruption: Evidence from the audits of local governments.” *American Economic Review* 101 (4): 1274–1311.

———. 2008. “Exposing corrupt politicians: The effects of Brazil’s publicly released audits on electoral outcomes.” *Quarterly Journal of Economics* 123 (2): 703–745.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2017. “Chapter 6 - The Personnel Economics of the Developing State.” In *Handbook of Economic Field Experiments,* edited by Abhijit Vinayak Banerjee and Esther Duflo, 2:467–514. Handbook of Economic Field Experiments. North-Holland.

Gonzalez-Lira, Andres, and Ahmed Mushfiq Mobarak. 2021. “Slippery Fish: Enforcing Regulation when Agents Learn and Adapt.” *NBER Working Paper* (Cambridge, MA).

Goodman-Bacon, Andrew. 2018. “Difference-in-Differences With Variation in Treatment Timing.”

Holmstrom, Bengt, and Paul Milgrom. 1991. “Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *Journal of Law, Economics, and Organization* 7 (Special Issue): 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.

Kamenica, Emir. 2019. “Bayesian Persuasion and Information Design.” *Annual Review of Economics* 11:249–272.

Kamenica, Emir, and Matthew Gentzkow. 2011. “Bayesian Persuasion.” *American Economic Review* 101 (6): 2590–2615.

Khera, Reetika. 2011. *The Battle for Employment Guarantee.* Oxford University Press.

Lazear, Edward P. 2006. “Speeding, terrorism, and teaching to the test.” *Quarterly Journal of Economics* 121 (3): 1029–1061.

Londoño-Vélez, Juliana, and Javier Ávila-Mahecha. 2021. “Enforcing Wealth Taxes in the Developing World: Quasi-experimental Evidence from Colombia.” *American Economic Review: Insights* 3 (2): 131–48.

Ministry of Rural Development. 2019. *The Mahatma Ghandi National Rural Employment Guarantee Act of 2005 Data Registers.* https://nrega.nic.in/netnrega/home.aspx(AccessedonMarch9,2019andAugust12,2020).

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. “Building state capacity: Evidence from biometric smartcards in India.” *American Economic Review* 106 (10): 2895–2929.

Niehaus, Paul, and Sandip Sukhtankar. 2013. “Corruption dynamics: The golden goose effect.” *American Economic Journal: Economic Policy* 5 (4): 230–269.

Olken, Benjamin A. 2007. “Monitoring corruption: Evidence from a field experiment in Indonesia.” *Journal of Political Economy* 115 (2): 200–249.

Rose-Ackerman, Susan, and Bonnie J. Palifka. 2016. *Corruption and Government: Causes, Consequences, and Reform.* 2nd. Cambridge University Press.

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,* 13:231–86.

Sun, Liyang, and Sarah Abraham. 2020. “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.” *Journal of Econometrics.*

Syverson, Chad. 2011. “What determines productivity?” *Journal of Economic Literature* 49 (2): 326–365.

Wilson, James Q. 1991. *Bureaucracy: What Government Agencies Do And Why They Do It.* Basic Books.

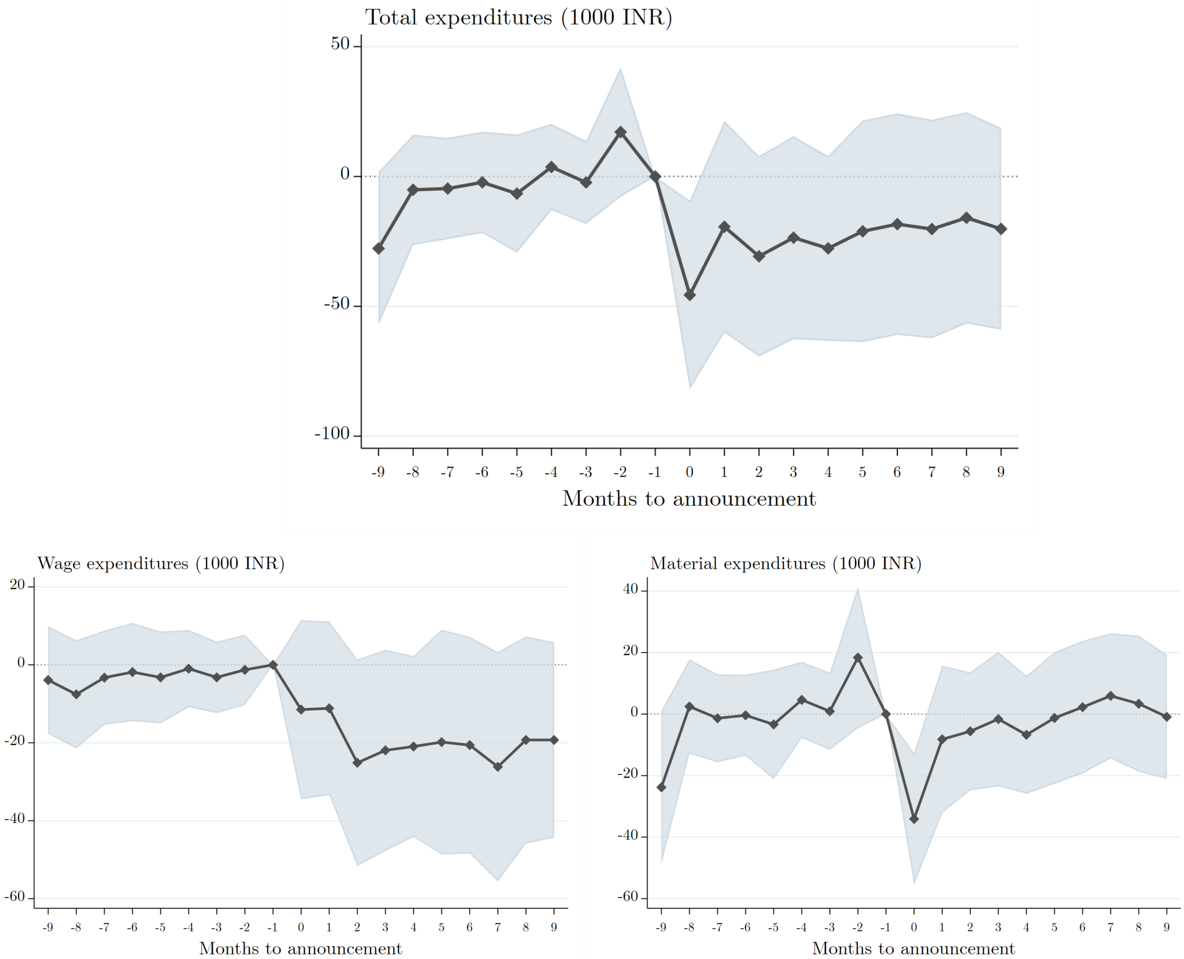
Yang, Dean. 2008. “Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines.” *The Review of Economics and Statistics* 90 (1): 1–14.

Appendix

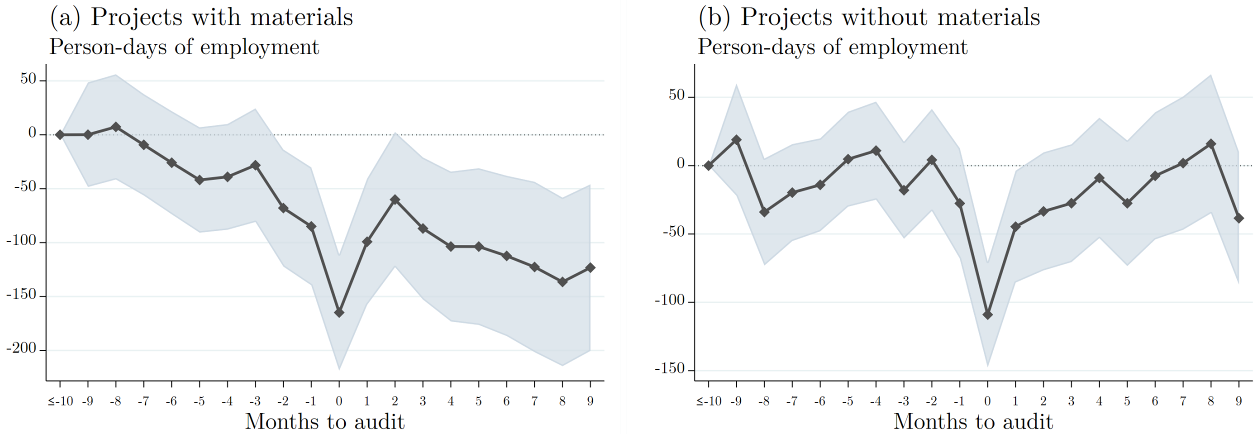
A Appendix Tables and Figures

	$Wave1$	$Waves2-1$	p-value	$Waves3-1$	p-value	$Waves2-3$	p-value	Observations
Panel A: 2011 Village Census								
Number of households	1153.85	13.46	0.35	13.12	0.35	0.34	0.98	3,684
Total population	6072.11	52.01	0.50	64.54	0.38	-12.53	0.86	3,684
Scheduled castes population	760.63	-5.42	0.82	-3.45	0.87	-1.97	0.93	3,684
Scheduled tribes population	2035.97	88.00	0.13	-8.13	0.88	96.13	0.02**	3,684
Literate population	3026.02	52.80	0.20	50.83	0.23	1.97	0.96	3,684
Total working population	2662.77	-6.30	0.86	-2.41	0.94	-3.89	0.90	3,684
Main working population	1210.52	-10.43	0.73	12.65	0.62	-23.07	0.33	3,684
Main working population, cultivation	515.71	-14.67	0.35	-8.79	0.60	-5.88	0.62	3,684
Main working population, agriculture	297.51	0.06	1.00	6.74	0.58	-6.68	0.51	3,684
Main working population, household industries	49.73	-3.69	0.49	1.61	0.75	-5.30	0.23	3,684
Marginal working population	1452.25	4.13	0.89	-15.06	0.55	19.18	0.46	3,684
Marginal working population, cultivation	430.35	13.48	0.42	11.88	0.49	1.60	0.91	3,684
Marginal working population, agriculture	791.96	-7.07	0.73	-23.08	0.22	16.00	0.34	3,684
Marginal working population, household industries	48.38	-2.78	0.43	-1.53	0.65	-1.25	0.63	3,684
Total geographical area (sq. km.)	1914.76	-1.39	0.98	-28.25	0.58	26.87	0.53	3,694
Forest area (hectares)	508.05	-21.29	0.56	-29.52	0.37	8.23	0.76	3,694
Barren, uncultivable land area (hectares)	85.03	-1.17	0.84	-4.48	0.46	3.31	0.45	3,694
Permanent pastures/grazing land area (hectares)	30.66	2.62	0.41	1.78	0.46	0.85	0.73	3,694
Total unirrigated land area (hectares)	587.54	15.86	0.57	20.44	0.48	-4.59	0.87	3,694
Wells and tubewells area (hectares)	32.93	-3.25	0.15	5.58	0.36	-8.83	0.14	3,694
Tanks and lakes area (hectares)	33.64	6.79	0.09*	14.41	0.07*	-7.62	0.29	3,694
p-value of F-test of joint orthogonality			0.61		0.72		0.97	
Panel B: MGNREGA MIS 2014-16, annual average								
Number of households provided employment	1003.86	-23.93	0.41	-43.15	0.12	19.23	0.43	3,854
Labor expenditures (1000 INR) per household	2.09	-0.07	0.27	-0.09	0.21	0.02	0.54	3,852
Material expenditures (1000 INR) per household	0.91	0.00	0.93	-0.01	0.79	0.02	0.59	3,852
% Labor expenditures of Total	71.62	-1.08	0.04**	-0.28	0.56	-0.80	0.09*	3,854
Days of delayed payment per household	27.78	2.24	0.23	0.62	0.70	1.61	0.37	3,852
% SC	12.52	-0.15	0.79	0.65	0.20	-0.80	0.07*	3,845
% ST	35.64	0.25	0.79	-0.32	0.74	0.57	0.46	3,845
% Women	30.93	0.50	0.26	-0.06	0.90	0.56	0.15	3,845
% Work completed	64.96	-0.22	0.70	-0.40	0.48	0.18	0.68	3,845
p-value of F-test of joint orthogonality			0.36		0.15		0.21	
p-value of F-test of joint orthogonality on all covariates			0.57		0.84		0.93	
p-value of Likelihood ratio test on multinomial logit with and without all covariates	0.59							

Appendix Table 1: Tests of balance in observables across waves. Using census and administrative data, statistical tests show balance on observable characteristics across waves. Except there are statistical differences between Waves 2 and 3 Scheduled Tribes population, share of person-days of labor allocated to Scheduled Caste, and differences across waves in percent of expenditures in labor. The number and extent of these differences are consistent with arising by chance. Overall, the differences are small (5% difference in Scheduled Tribe population; 6% difference in percentage of person-days Scheduled Caste; and less than or around a percentage point difference in percent of expenditures in labor), 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.



Appendix Figure 1: Announcement event study for total expenditures (1000 INR). This figure provides evidence consistent with parallel trends. It shows an event study of total expenditures with lags and leads around the month of announcement and controlling for anticipatory behavior. 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. There is no statistically distinguishable trend during the months before the announcement (p-value = 0.45). Additionally, there is no evidence of pre-trends for wage and material expenditures (p-value = 0.19 and 0.99, respectively)



Appendix Figure 2: Person-days of employment around the time of audit, by whether the project required materials or only required labor. It shows that the decline in employment is observed across projects of both types. Morever, the observed increase in materials 2-4 months after the audit (third figure of Figure 3) does not correspond to an increase in employment on projects requiring materials as shown in Figure 2 (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. 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. Standard errors are clustered by block.

	(1)	(2)
	Delay in days	
Anticipating1stAudit - Year 2	111.23	119.55
	(105.04)	(129.63)
Post1stAnnounce	0.32	
	(67.79)	
Month of 1*stAudit*		-32.00
		(71.36)
Anticipating2ndAudit - Year 2		-14.93
		(144.06)
Anticipating2ndAudit - Year 3		-201.88
		(161.01)
Before 1*stAudit*		-0.23
		(72.96)
After 1*stAudit*		-77.45
		(73.99)
Observations	233,760	233,760
Baseline mean	1869	1869
Adj. R-squared	0.233	0.230

Appendix Table 2: Null effects from changing expectations of being audited on Bureaucrats' efficiency in making wage payments as a measure of effort affecting changes in honest expenditures. This table estimates the Equation 1 with days of delayed payment as the outcome. More details on delays as a measure of honest effort is in Online Appendix Section C.3. 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 (*AnticipatingAudit* - 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 *AnticipatingAudit* - Year 1. *** p<0.01, ** p<0.05, * p<0.1 *** p<0.01, ** p<0.05, * p<0.1.

	Workload per auditor		Total audit expenditures (INR)	
	(1)	(2)	(3)	(4)
Wave 3	−7.326	−1.897	−1,660.990	−1,584.749
	(10.066)	(4.540)	(1,127.538)	(1,013.885)
Num. HHs to check		0.432***		−0.191
		(0.019)		(2.201)
Num. works to check		0.422***		9.910**
		(0.032)		(4.051)
Num. Auditors		−44.329***		1,137.143
		(3.221)		(738.204)
Mean of Dep Var	249.69	249.69	27,206.45	27,206.45
Observations	2,445	2,445	2,676	2,445
Adjusted R^2	0.441	0.944	0.298	0.346

Appendix Table 3: Audit inputs are indistinguishable by year of audit, suggesting bureaucrats perception of audit quality overtime should remain unchanged with respect to perceived audit inputs (an empirical test for a violation of Assumption 1). 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. 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 (1,000 INR)	
	(1)	(2)
Anticipating1stAudit - Year 2	13.87	14.33
	(11.33)	(11.76)
Post1stAnnounce, disaggregated:		
Before 1*stAudit*	-18.83**	-17.72**
	(9.46)	(8.56)
Month of 1*stAudit*	-42.07***	
	(11.68)	
Month of 1*stAudit* - Year 1		-54.61**
		(23.46)
Month of 1*stAudit* - Year 2		-45.10***
		(15.50)
Month of 1*stAudit* - Year 3		-33.04**
		(14.54)
After 1*stAudit*	-8.42	
	(11.26)	
After 1*stAudit* - Year 1		26.69
		(30.58)
After 1*stAudit* - Year 2		-10.08
		(13.39)
After 1*stAudit* - Year 3		-6.95
		(12.43)
Anticipating2ndAudit - Year 2	-4.78	-1.74
	(13.83)	(13.68)
Anticipating2ndAudit - Year 3	-39.38***	
	(13.30)	
Anticipating2ndAudit - Year 3, Wave 1		-28.12**
		(13.98)
Anticipating2ndAudit - 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 1*stAudit*, H_0 : Year 1 = Year 2 = Year 3 (p-val)		0.67
After 1*stAudit*, H_0 : Year 1 = Year 2 = Year 3 (p-val)		0.54
Anticipating2ndAudit - Year 3, H_0 : Wave 1 = Wave 2 (p-val)		0.22

Appendix Table 4: Effect of stages of the monitoring policy disaggregated by year. This table estimates the main differences-in-differences specification breaking down 'Month of *Audit*' and 'After *Audit*' by year and '*AnticipatingNextAudit* - 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 (*AnticipatingAudit* - 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 *AnticipatingAudit* - 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

	Total expenditures (1,000 INR)			
	(1)	(2)	(3)	(4)
Anticipating1stAudit - Year 2	13.87	15.52	14.54	15.42
	(11.33)	(11.27)	(11.28)	(11.28)
Post1stAnnounce, disaggregated:				
Before 1*stAudit*	-18.83**	-17.00*	-17.93*	-16.97*
	(9.46)	(9.43)	(9.47)	(9.44)
Month of 1*stAudit*	-42.07***	-49.06***	-42.42**	-49.17***
	(11.68)	(11.59)	(17.25)	(11.61)
After 1*stAudit*	-8.42	-6.44	-7.47	-6.40
	(11.26)	(11.25)	(11.17)	(11.27)
Anticipating2ndAudit - Year 2	-4.78	-3.47	-4.47	-3.50
	(13.83)	(13.85)	(13.65)	(13.83)
Anticipating2ndAudit - Year 3	-39.38***	-37.76***	-38.97***	-37.69***
	(13.30)	(13.30)	(13.33)	(13.31)
ShareBlockAudited		25.64		40.35
		(18.96)		(65.87)
$ShareBlockAudited^2$				-21.10
				(87.61)
Month of $1stAudit = 0 \times ShareBlockAudited$			32.71	
			(22.70)	
Month of $1stAudit = 1 \times ShareBlockAudited$			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: (Month of $1stAudit = 0 \times ShareBlockAud$)– (Month of $1stAudit = 1 \times ShareBlockAud$), p-val = 0.57

Appendix Table 5: Bureaucrat responses in anticipation and during the audit are unchanged when controlling for potential spillover behavior from peer bureaucrats. This provides additional evidence that the anticipatory effects are driven by changes in expectations of being audited rather than perceptions about the audit influenced by peer experiences. 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. See Online Appendix C.1 for justification of these tests. 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 (*AnticipatingAudit* - 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 '*AnticipatingAudit* - Year 1'. *** p<0.01, ** p<0.05, * p<0.1.

Beliefs Assumption, $K_{\tau+1} =$	Mean Inform. (Mil. USD)	Inform.− Some Info.	Std. Error	Inform.− Uninform.	Std. Error
	(1)	(2)	(3)	(4)	(5)
$K_{\tau-1}$	226	21.17	0.038	32.3	0.038
$\frac{1}{2}(K_{\tau-1}+K_{\tau})$	226	21.75	0.038	38.1	0.038
K_{τ}	226	21.75	0.037	36.8	0.037
$\text{Trend}_{\tau} \times K_{\tau}$	240	7.87	0.037	2.8	0.037
$K_{\tau+1}$	227	20.78	0.035	32.5	0.035
Estimates excluding Month of Audit:					
$K_{\tau-1}$	228	19.24	0.038	30.3	0.038
$\frac{1}{2}(K_{\tau-1}+K_{\tau})$	228	19.82	0.038	36.2	0.038
K_{τ}	228	19.82	0.037	34.9	0.037
$\text{Trend}_{\tau} \times K_{\tau}$	247	0.93	0.037	2.0	0.037
$K_{\tau+1}$	228	19.82	0.036	31.6	0.036

Appendix Table 6: Welfare analysis by beliefs assumption showing that the full information signal is preferred by the Principal compared to the uninformative and some information signals. Welfare is measured as differences in expenditures (Columns 2 and 4), which we interpret as differences in misappropriate expenditures driven by changing expectations of being audited. The rows present the welfare calculations under each assumption on beliefs about next year's audit capacity, with conservative calculations excluding behavior during $Month of Audit$ showing qualitatively consistent results. Bootstrapped standard errors for the differences in means are provided in Columns 3 and 5.

Online 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 participants, of project sites. These supervisors take attendance for the payroll and help participants apply for work.[37]

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 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

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

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.[38] 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 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,

38. Five-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.

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 participants, whose incentives to report during the audit can be affected.

A.3 Details on the issues identified during audit

Among 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 discrepancies with the observed and recorded features of the constructed project, including the project being non-existent; and 12% of issues are related to officials refusing to cooperate in some way with the audit process.

While the audit reports data 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 of issues uncovered during the audit.

A.4 Details on the roll-out schedule and public notices about the audits

Table A.1 summarizes details of the audit schedule for each year of the roll-out of audits.

Table A.1: 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 regularly audited twice a year.[39]

  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.

39. The 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.

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 84% of Waves 1 and 2 GPs compared to 18% of Wave 3 GPs were audited in Year 4.

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.

A.5 Survey of bureaucrats' knowledge on audit policy

The rotation of audits across GPs is implied by the information from the Year 1 announcement and 2006 NREGS Act that all GPs be audited regularly. But, it may not have been explicitly recognized by GP bureaucrats. This could pose challenges to inferences we make on bureaucrats' expectations of being audited across the horizons of anticipation.

From September-November 2020, a remote survey was conducted by phone interview and SurveyMonkey to understand bureaucrats' knowledge of the audit policy from 2016-19. Piloting of the survey was conducted from February-March 2020, and continued rollout of the survey was paused from April-August 2020 during restrictions to address the surge in COVID-19 in India.

We randomly sampled 83 GPs, stratified by wave and whether the share of scheduled caste or scheduled tribe population was above average. In each GP, we attempted to reach multiple bureaucrats and most often received responses from the head bureaucrat (*mukhiya*) and the NREGS-dedicated bureaucrat (*panchayat rozgar sewak*). Among the sampled GPs, we received responses from 25 GPs (30% of the 83 sampled GPs) and 7 GPs not in the original sample.

In a multiple choice question, bureaucrats were asked how GPs were selected for audit of NREGS each year during 2016-2020. They could respond: GPs were selected randomly (regardless of whether they were recently audited); GPs were selected at the discretion of state government officials (regardless of whether they were recently audited); GPs took turns being audited (E.g. All GPs received their 1st audit before anyone was selected for their 2nd audit); or 'I'm not sure'. The responses from the survey should be interpreted with caution for a couple of reasons: first, we were only able to obtain responses from a small sample of GPs, meaning survey sample means may not be consistent with study population means, especially by wave; second, eliciting beliefs about the past audit program is challenging if memory is imperfect, although many bureaucrats were able to ballpark the number of audits conducted in each year starting from 2016.

We consider the responses of the most senior bureaucrats in each GP. Among them, 38% responded 'I'm not sure',[40] 29% responded that GPs took turns being audited, 24% responded that GPs were selected at the discretion of state government officials (which is not inaccurate since a small subset of GPs were adhoc selected for audit after state officials suspected corrupt behavior, see discussion in Section 3.2), and 9% responded that GPs were selected randomly (this could have been randomly with or without replacement).

For bureaucrats who responded 'I'm not sure' or that GPs were selected at the discretion of state government officials, their responses would not be affected by the horizons of anticipation described in Figure 1 since their beliefs do not depend on the announcements of GPs selected for audit. If we assume that these types of bureaucrats are balanced across waves of the audit given random assignment to audit timing, then we can treat the impact of their behavior on the results in Table 1 as noise, and that the results in Table 1 are driven by bureaucrats who believed GPs took turns being audited, the modal response among bureaucrats who had views on how GPs were selected for audit.

For the 9% (or 2 responses) of bureaucrats who said that GPs were selected randomly (and perhaps also thought it was random with replacement, but were not asked about this explicitly), we would expect no evidence of anticipation across waves leading up to the announcement. Because there is evidence of anticipation when examining pre-announcement behavior among Waves 2 and 3 GPs (see results from Appendix Section B.1 all GPs in the study sample), we can conclude with confidence that bureaucrats

40. Respondents who were interviewed by phone were much more likely to report 'I'm not sure', which could be indicative systematic differences in responding when speaking with an enumerator over the phone. For instance, differences in the way questions may have been prompted like emphasis on certain parts of the question or perceptions that the enumerator was working on behalf of state government officials.

holding this belief are very unlikely to be driving the main effects from the horizons of anticipation observed among the full sample.

Although only administered among a small sample of respondents, these survey findings along with the differences by wave in pre-trends or anticipation leading up to audit announcement provide evidence supporting that the main results in Section 4.1 reflect the behavior of bureaucrats who thought GPs took turns being audited, as achieved by the randomization without replacement.

B Empirical Strategy

B.1 Parallel Trends and Anticipation By Year

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. But, without controlling for anticipation, we should not expect parallel trends for subsequent waves because bureaucrats may anticipate their turn for an audit.

Disaggregating the event study around the announcement by wave, there is evidence of pre-trends leading up to the Year 2 announcement for Wave 2 audits. We cannot reject a test of pre-trends in the Wave 1 versus not Wave 1 comparison (p-value = 0.35, where $H_0 =$ periods prior to announcement are jointly equal to 0) but we reject pre-trends in the Wave 2 versus Wave 3 comparison (p-value = 0.01). This supports our primary specification (Equation 1), which 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.

B.2 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

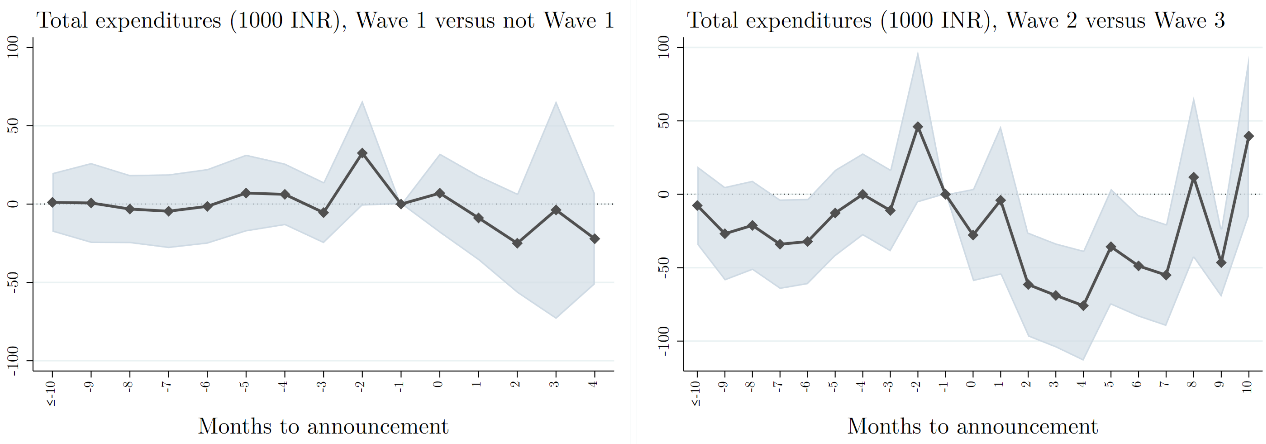


Figure B.1: Announcement event study by wave for total expenditures. The omitted category in this regression is the one month lead before the announcement. It includes GP and district month-year fixed effects. The figure on the left compares those selected for audit in Wave 1 to those not audited and 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.

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, 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.

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

Table B.1: 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

C The Impact of Changing Expectations

C.1 Testing for Spillover Effects of the Audit

This section describes the rationale for the tests for spillover effects from GPs within the same block in Appendix Table 5. Random assignment of GPs to 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):

$$\begin{aligned} y_{it} =& \alpha_i + \alpha_{dt} + AnticipatingAudit_{it}'\beta + \delta_1 \text{Before}Audit_{it} + \delta_2 \text{Monthof}Audit_{it} \\ &+ \delta_3 \text{After}Audit_{it} + AnticipatingNextAudit_{it}'\gamma + \eta ShareBlockAudited_{it} + \varepsilon_{it} \end{aligned} \tag{C.2}$$

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).

C.2 Additional results with audit report data

As discussed in Section 4.2, bureaucrats may be choosing material over wage expenditures because they believe material misappropriation is easier to hide from auditors, as wage misappropriation can be discovered by auditors during household interviews. Additionally, bureaucrats have been known to refuse providing program registers and receipts to auditors for verification. With this possibility, they can hide fake receipts and prevent the verification of material procurement. However, refusal to cooperate with the audit comes at a cost and is an issue documented and fined by auditors. Indeed, Table C.1 shows that Wave 3 GPs have a greater number of issues and fines related to the refusal to provide records for auditors.

	Issue fine amount (1000 INR), by type:							
	Records not provided to auditors				Bills/Vouchers not provided to auditors			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wave 3 - Wave 2	154.967*	252.967***	273.056***	434.863**	23.038***	22.106**	23.039*	36.734***
	(81.643)	(60.431)	(87.711)	(172.601)	(7.428)	(9.337)	(12.766)	(10.527)
Audit manager experience				−18.943				−1.603
				(12.720)				(1.406)
Wave 2 Mean	147.43	147.43	150.67	150.67	9.93	9.93	9.62	9.62
Controls		X	X	X		X	X	X
Manager FE			X	X			X	X
Observations	2,361	2,361	1,772	1,772	2,361	2,361	1,772	1,772
Adjusted R^2	0.127	0.174	0.210	0.211	0.051	0.055	0.142	0.143

Table C.1: Differences is audit issue fine amounts across waves for issues related to not providing any records to auditors for verification. The first outcome variable pools issues where no registers, bills, or vouchers are provided. The second outcome presents only issues related to bills or vouchers not being provided. Control variables: number of employed households and works to verify; 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.

C.3 Delayed payments as a measure of effort

Studying delays in making wage payments allows us to disentangle GP bureaucrat effort from incentives to misappropriate. Delayed payments are often cited to reflect a lack of a core competency of the government in targeting resources[41], and not a known mechanism through which bureaucrats misappropriate finances. 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.

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

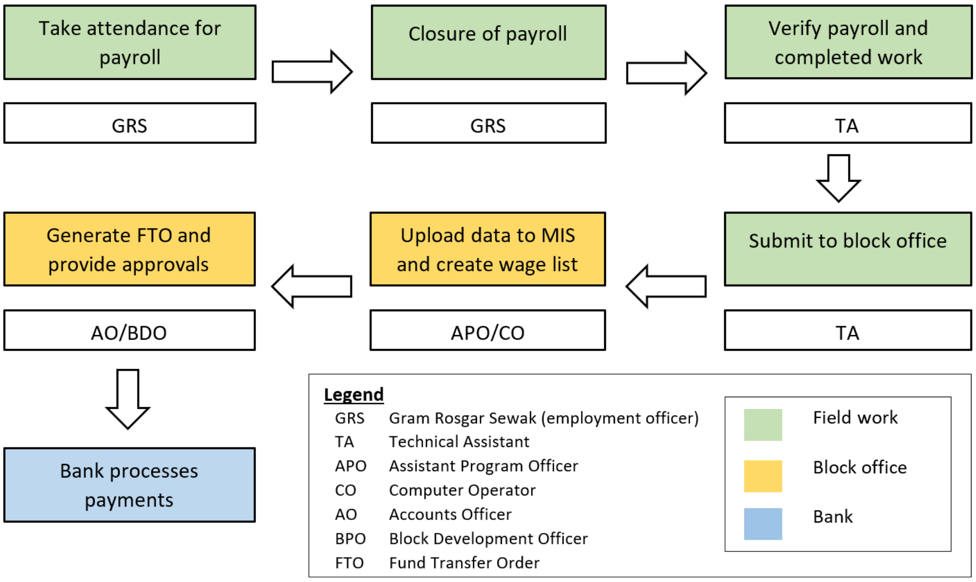


Figure C.1: Example of stages to the NREGS payment process. White boxes beneath steps present the officials responsible for implementing that step. FTO denotes fund transfer order. Gram Rozgar Sewak (GRS) is one of the key GP-level personnel and also responsible for allocating days of work. Source: Evidence 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 C.1. 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.[42] According to the NREGS National Act, workers should be compensated 0.05% of unpaid wages for each day of delay in wage payment.[43]

42. According to the Evidence for Policy Design, September 2015 presentation to the Ministry of Rural Development of the Government of India.
43. The Mahatma Gandhi National Rural Employment Guarantee Act, 2005.

D Model of Information Design and Deterrence

D.1 Setup

There is a Principal who oversees implementation of a government program by N Bureaucrats. Consider the Principal's interaction with a single, arbitrary Bureaucrat. This model is static and two-player but accommodates settings, such as this empirical setting, with N Bureaucrats and T periods under reasonable assumptions.[44] 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 an audit. The Bureaucrat is uncertain whether current performance will be audited. Their expectations of being audited inform their choice on the amount of expenditures to misappropriate. Formally, the Principal announces π, a chosen probability distribution over likelihoods of audit. I 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.

The Principal is budget-constrained and can only audit M Bureaucrats, where $M < N$. The Bureaucrat knows this and internalizes the budget constraint as a prior on the likelihood of an audit. If the Principal provides no information, each Bureaucrat is expected to be monitored with equal likelihood, $q_0 = \frac{M}{N}$, given the Principal's budget constraint. Upon receiving a signal, the Bureaucrat forms the posterior belief about the likelihood of an audit, q, according to Bayes' rule. Then, the Bureaucrat chooses an action $a \in [0,1]$, which is the 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 $q' > q$, $a' > a$, if $V(a',q) - V(a,q) \leq 0$ then $V(a',q') - V(a,q') \leq 0$. This assumption means that if a Bureaucrat who expects the audit with probability q prefers a (where $a < a'$), he would make the same choice if the probability of audit was q' (where $q' > q$). For example, $V(a,q) = v(a) - qc(a)$ where $v(a)$ is the net benefit from misappropriat-

44. See Online Appendix D.2 for the assumptions. The application of these assumptions to the empirical setting is discussed in Section 6.1.

ing a and $c(a)$ is the punishment for choosing a. Note that since $1-a$ represents the remaining share of expenditures spent on honest activites, then $V(1-a,q)$ also satisfies the single-crossing property. We abstract away from modeling heterogeneity in the Bureaucrat's propensity to misappropriate (e.g. ability to misappropriate finances) because assignment to a signal from π is random. This accords with the actual monitoring policy where expectations of being audited are randomly assigned. We can think of the Bureaucrat's response, $a(q)$, given belief q, as holding all other factors affecting $a(\cdot)$ equal.

The Bureaucrat chooses an action that maximizes expected utility:

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

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. The Principal's utility is net of the costs for conducting M audits. Naturally, the Principal's utility is decreasing in a. This utility function reflects a Principal who values deterrence on its own, whether it comes from only some GPs or equitably across all GPs. The change in the Principal's utility with respect to q, is negatively proportional to the change in $a(q)$ with respect to q. This means that the convexity of $u(a(q))$ is determined by the (negative) convexity of $a(q)$. This is important because we can interpret $u(a(q))$ as the deterrence of misappropriated expenditures (scaled and shifted by constants k and b, respectively).

The Principal's expected utility, $\mathbb{E}_{q \sim \pi(\cdot)}[u(a(q))]$, sums over the likelihood of posterior beliefs q induced by the set of signals that can be drawn from π. The Principal will choose a communication policy, π^*, in order to maximize this expected utility given the Bureaucrat's best response, $a^*(q)$.

The timing of this game is as follows:

(1) Principal commits to communication policy, π, about the likelihood of an audit.

(2) Nature draws a signal for Bureaucrat from distribution 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.

D.2 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 D.3 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 Bureacrat'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 Bureacrats' 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.

D.3 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 $\mathscr{T} = \{1, ..., T\}$ and indexed by t. The set of receivers is $\mathscr{I} = \{1, ..., N\}$, indexed by i, and $|\mathscr{I}| = N$. 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 $X_t = \sum\limits_i x_{it}$, 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 $M < N$ 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 $\omega = \{\omega_{1t}, ..., \omega_{Nt}\}_{\forall t \in \mathscr{T}}$.

Every period, the Principal decides and commits to a signal structure or policy, $\pi : \Omega \to \Delta(S)$ 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.

D Analysis

The optimal policy depends both on the rate of change of the Bureaucrats' response to their expectations of being audited and the Principal's capacity to conduct audits.

Since $V(a, q)$ satisfies the single-crossing property, then we know that $a^*(q)$, the Bureaucrat's best response for some given posterior belief q, is weakly decreasing in q (Milgrom and Shannon 1994). With this and the fact that $u'_a < 0$, then $u(a^*(q))$ is weakly increasing in q.

For notational convenience, let $U(q) = u(a^*(q))$. $U(q)$ is the Principal's value function given $a^*(q)$. The profile of payoffs represented by $U(q)$ and weighted combinations of these payoffs are achievable using π. Recall that π can be designed to place weight on a chosen subset of posterior beliefs as long as Bayes' Rule is satisfied. That is, if the Principal designed π to place weight on posterior beliefs q' and q'', then their payoff is a weighted combination of $U(q')$ and $U(q'')$, so long as the mean of posterior beliefs is equal to the prior q_0.[45]

Recall that the Principal will choose a communication policy, π^*, that solves:

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

For a given q_0, define the set C to be all convex combinations of the values of $U(q)$ such that the mean of the posterior beliefs in the convex combination is equal to q_0. In particular, let $\lambda_1, ..., \lambda_n$ be a vector such that $\lambda_1 + ... + \lambda_n = 1$ and $q_1, ..., q_n$ is the vector of posterior beliefs, then each convex combination in the set C must be such that $q_0 = \lambda_1 q_1 + ... + \lambda_n q_n$.[46] Define the largest payoff possible to the Principal as $\hat{U}(q_0) = \max\{C\}$. The Principal's optimal signal, π^*, achieves the payoff $\hat{U}(q_0)$.

Proposition 1 (Persuasion Works)**.** *The Principal benefits from sending an informative signal if and only if $U(q_0) < \hat{U}(q_0)$.*

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 $\mathbb{E}_{q \sim \pi(\cdot)}[U(q)] = \hat{U}(q_0)$. Since an informative signal is better for the Principal than an uninformative one, with payoff $U(q_0)$, we can construct π such that $\hat{U}(q_0) > U(q_0)$.

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.

45. Bayes' 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: $q_0 = Pr(\text{Audit}) = \pi(\text{Audit}|H)Pr(H) + \pi(\text{Audit}|L)Pr(L) = \pi(H|\text{Audit})Pr(\text{Audit}) + \pi(L|\text{Audit})Pr(\text{Audit})$. The second equality follows from the law of total probability and the third equality follows from Bayes' Rule.

46. We could have alternatively defined the set of attainable payoffs using convex hulls. 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. The largest payoff possible to the Principal for some given prior q_0 is $\hat{U}(q_0) = \max\{u|(q_0, u) \in co(U(q))\}$.

□

The condition in Proposition 1 for sending an informative signal was derived in Kamenica and Gentzkow (2011). This model uses this result to provide a more general approach to solving for the models in Lazear (2006) and Eeckhout et al. (2010); the theoretical findings are consistent.

Proposition 1 says that the optimal signal depends on the elasticity of the Bureaucrat's response with respect to their expectations of being audited. For example, if the Bureaucrat is only responsive to the incentives from monitoring when their expectations of being audited are very high, then it would be better to inform them of audits in advance than to maintain uncertainty. In contrast, if the Bureaucrat is more responsive on the margin when there is uncertainty around the likelihood of being audited compared to knowing for sure, then the Principal is better off maintaining uncertainty. The following result further develops the intuition of Proposition 1.

Proposition 2 (When providing more information is better)**.** *If $U(q)$ is convex at q_0, then $U(q_0) < \hat{U}(q_0)$ and an informative signal is preferred. This result implies the following about the optimal signal given the shape of $U(q)$:*

(i) The Principal prefers an informative over uninformative signal for any q_0 if $U(q)$ is convex for all q. The Principal prefers an uninformative signal for any q_0 when $U(q)$ is concave for all q.

(ii) An informative signal is still preferred if $U(q)$ is convex at q_0 and concave elsewhere for some q.

Proof. Let's start with the first statement of the proposition. Suppose that $U(q)$ is 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 $U(\cdot)$ where $U(\cdot)$ is convex around q_0. By Jensen's inequality, $U(q_0) = U(\lambda q_1 + (1-\lambda)q_2) < \lambda U(q_1) + (1-\lambda)U(q_2)$. Let $\tilde{q} = \lambda U(q_1) + (1-\lambda)U(q_2)$ where $\tilde{q} \in C$ by definition of the set C. $\tilde{q}$ is an achievable payoff for the Principal. We've established that $\tilde{q} > U(q_0)$ and we know that $\hat{U}(q_0) = \max\{C\}$. So, $\hat{U}(q_0) \geq \tilde{q} > U(q_0)$. We've established that whenever $U(q_0)$ is convex at q_0, we have that $\hat{U}(q_0) > U(q_0)$. This also shows that statement *(ii)* holds. However, it is not always true that when $U(q)$

is concave at $U(q_0) > \hat{U}(q_0)$ and an uninformative signal is preferred. A counterexample to consider is if there is a local concavity at q_0, but convexity in the function $U(q)$ elsewhere.

Let's turn to *(i)* of the proposition. 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 C 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 an informative signal when $U(q)$ is convex for all q 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 (uninformative signal). If $U(q)$ is concave for all q, then Jensen's inequality tells us that $U(q_0) = U(\lambda q_1 + (1-\lambda)q_2) > \lambda U(q_1) + (1-\lambda)U(q_2)$. By Proposition 1, the Principal prefers an uninformative signal and will send an uninformative signal where posterior beliefs equal the prior belief q_0. $\square$

A geometric interpretation of this solution means that a Principal prefers to send an informative signal if $U(q)$ is convex 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 D.1 provides illustrations of when informative versus uninformative signals are better. The bold dotted line represents the set C, the payoffs attainable by the Principal. The top-left panel of Figure D.1 shows that $\hat{U}(q_0)$, i.e. informing some they will be selected for audit with a very high likelihood (signal H leading to posterior q_H) and others they will not be audited (signal L leading to posterior q_L), is greater than $U(q_0)$, i.e. uninformative signal when everyone believes they will be audited with probability q_0. If $U(q)$ is concave for all q (e.g. top-right panel of Figure D.1), a signal where posteriors equal the prior q_0 is optimal. Finally, if the function is partially convex (e.g. bottom panel of Figure D.1), then the Principal prefers an informative signal if the function is also concave at q_0.

Ultimately, the shape of $U(q)$ in a particular setting must be learned empirically. Estimating $U(q)$ is sufficient for determining the optimal signal and analyzing the Principal's welfare under alternative signals. If $U(q)$ is convex at prior q_0, then communicating more information about the likelihood of being audited is better than maintaining

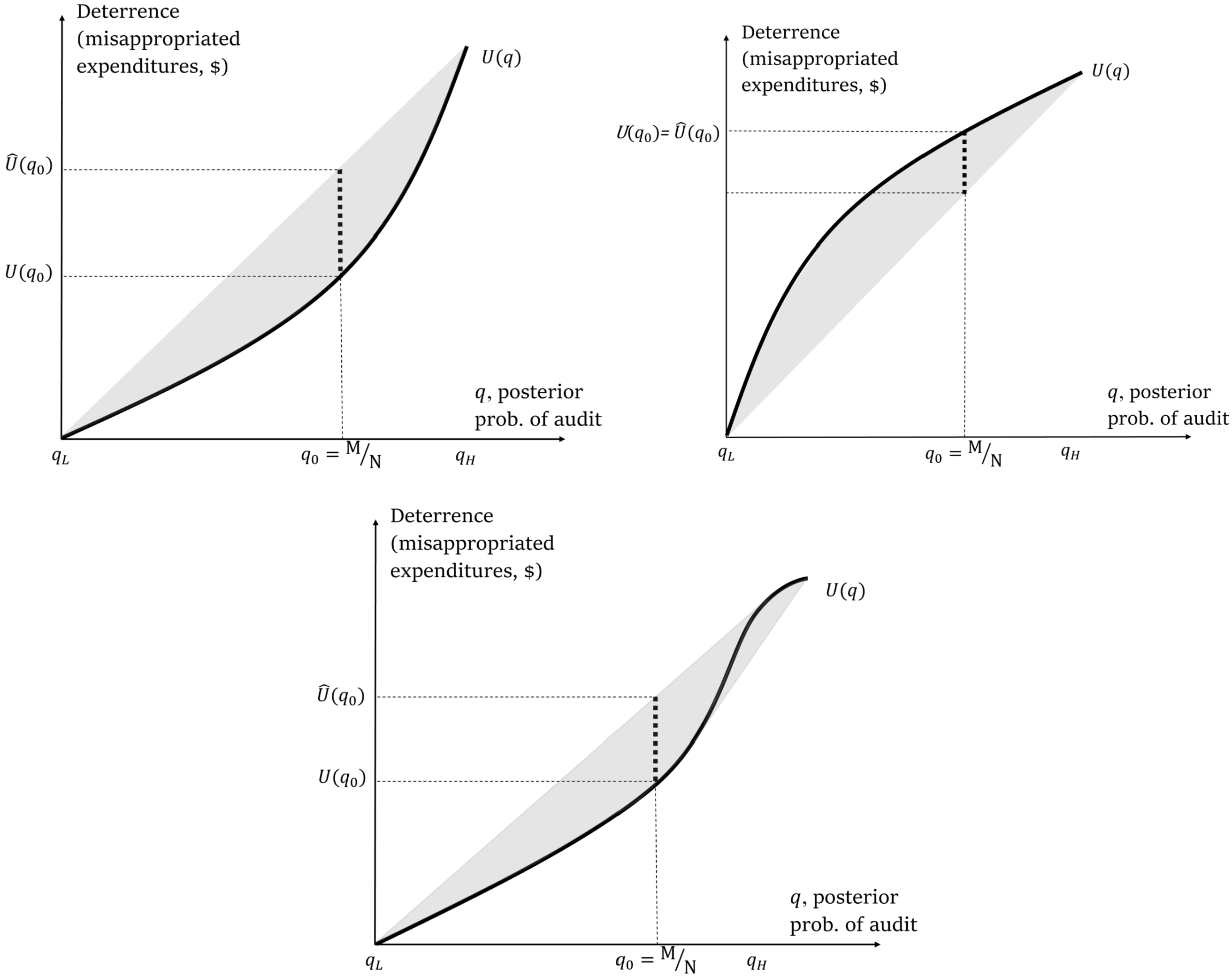


Figure D.1: When does the Principal prefer an informative over uninformative signal? These figures represent the Principal's utility as a function of the Bureaucrat's posterior beliefs. When the function is convex at q_0 (e.g. top left panel), the Principal prefers to communicate additional information about the probability of audit. When the function is concave for all q (e.g. top right panel), the Principal prefers to communicate no additional information about the probability of audit. If the function is convex only for some q but concave elsewhere (e.g. bottom panel), then the Principal prefers to communicate more information if the function is also concave at q_0.

uncertainty over when the audit will occur. With the random selection of GPs for audit without replacement in this empirical setting, I estimate $U(q)$ with causally-identified empirical moments in the data and apply the results from Proposition 2 to determine the best policy on communicating the likelihood of being audited (Section 6).

D.1 Sufficiency of $U(q)$ for analyzing welfare changes and determining the optimal signal

Let $U_1, ..., U_n$ be a sample from the probability distribution of the Principal's welfare (as measured by levels of bureaucrat financial misappropriation), $f(u|\theta)$, where θ is a vector of parameters that determine the bureaucrat's decision to misappropriate finances. The sample $\Big((U_1, q_1), ..., (U_n, q_n)\Big)$ from the joint distribution $f(u,q)$ where q is bureaucrat expectations of being audited and is randomly assigned. The distribution $f(u|q)$ 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 approximate $f(u|q)$ using the causally-identified moments corresponding to the 5 horizons of anticipation (q). This is the estimate of $U(q)$. 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.

E The Optimal Design of Information and Counterfactual Signals

E.1 Sensitivity analysis and robustness of conclusions

This section evaluates the sensitivity of our conclusions about the optimal monitoring policy to: inherent noise from the regression estimates, and Assumptions 1 and 2. How wrong would our assumptions have to be in order for us to conclude that an uninformative signal would have been better? Technically, we want to know under what belief and deterrence parameters would the function $U(q)$ be linear (i.e. Principal indifferent) or weakly/locally concave at q_0 (i.e. Principal prefers uninformative signal)? The results from this section provide greater certainty that the $U(q)$ is a convex curve as shown in

Figure 4 and that a signal informative about audit risk is optimal.

For these sensitivity checks, I produce the joint sampling distribution of the estimated coefficients using the block bootstrap. With this, we can compute the likelihood that the jointly estimated coefficients lead to an estimated $U(q)$ leading us to draw 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.

First, I assess deviations from the mean of deterrence estimates. This exercise assesses the likelihood we observe an alternative conclusion using the bootstrapped joint sampling distribution of the estimated coefficients (Column 4 of Table 1) . 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%.[47] 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 an informative and uninformative signal. 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 *Audit*'), 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 $K_{t+1} = Trend_t K_t$.

Second, I relax the first implication of Assumption 1 that bureaucrat perceptions of audit quality are constant across years. As shown in Appendix 4, bureaucrat responses are statistically indistinguishable across years for each variable that spans multiple years: 'Month of *Audit*', 'After *Audit*', and *AnticipatingNextAudit* - Year 3. 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. And the estimated response of bureaucrats conditional on expectations of being audited is confounded with changing perceptions of audit qual-

47. The likelihood under $K_{t+1} = K_t$ of a local concavity is 2% on average, but the local concavity occurs at '*AnticipatingNextAudit* - 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, an uninformative signal would still not be optimal.

ity. Given the estimates for *PostAnnounce*, 'Month of *Audit*', and 'After *Audit*', a power analysis provides the minimal detectable difference for each variable's estimate by year. 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 constant bias of ± 5 (1,000 INR) where bias adjustments for each coefficient are assumed to be in the direction of simulating a concave function (the alternative conclusion). Results are presented in Table E.1. With these adjustments, it is still unlikely that an uninformative signal is 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 $Trend_t K_t$). 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 when assuming a bias of ± 5 (1,000 INR). While the likelihood of an alternative conclusion under $Trend_t K_t$ is not insubstantial, as before, there is not enough information to conclude whether there's a strict preference (given 3 points of support).

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.1: 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 estimated $U(q)$ 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 $U(q)$ that an uninformative signal was optimal (weakly/locally concave).

Third, I assess deviations from assumptions on beliefs. I simulate perturbations from assumed beliefs (bound between 0 and 1) and use the deterrence estimates for each bootstrapped sample. The perturbations can take any value, within the range of assumed beliefs for each anticipatory group (i.e. assumptions on beliefs from Table 3), in order to simulate an estimated $\hat{U}(q)$ that is linear or concave. Among these perturbations, only 0.5% of the sample would yield a $\hat{U}(q)$ that is linear or concave. Conservative estimates that exclude behavior estimated during the month of audit (coefficient captured by 'Month of *Audit*') yield a 1.2% likelihood of a $\hat{U}(q)$ that is linear or concave with simulated beliefs within the range of assumed beliefs.

E.2 Welfare calculations

This section describes how welfare across the information policies were computed.

I analyze welfare for three policies: (1) a partially-informative signal which is the actual policy that was implemented (randomizing without replacement); (2) an uninformative signal where all GPs have no additional information over the prior and have the same expectations of being audited (randomizing with replacement); and (3) a fully informative signal where all GPs are perfectly informed in advance about when they will be audited or not.

To compare welfare under these policies, I estimate what total expenditures might have been under each policy during the 27-month period when the round of first audits were conducted. To estimate the counterfactual expenditures, we know that each policy induces a different set of beliefs for bureaucrats and we also know from Figure 4 how bureaucrats change their spending as their expectations of being audited change. I compare the projected changes in spending under each counterfactual communication policy to estimate the differences in welfare attributable to the design of communicated expectations of being audited.

The following formulas describe how I estimate expenditures under the counterfactual policies. For each policy, I exclude the year when a GP is announced for and experiences an audit because this period may capture behavior that is not driven by changing expectations of being audited. For example, the experience of being announced for an audit later in the year may increase the perceived salience of audits holding expectations fixed. Standard errors of the estimates are calculated with the bootstrapped estimates of $U(q)$, the estimates of the change in bureaucrat behavior as a response to changing expectations (the main results from Table 1.

(1) Partially informative signal (or randomization without replacement, p_1)

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

where $n_{GPs,Wavex}$ corresponds to the number of GPs in Wave x, $n_{Months,Yeary}$ 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 spending as a function of bureaucrats' posterior beliefs, q, on the likelihood of an audit. Under this signal capturing the actual policy that took place, $U(q_{Year1})$ corresponds to '*AnticipatingAudit* - Year 1'; $U(q_{Year2})$ corresponds to '*AnticipatingAudit* - Year 2'; $U(q'_{Year2})$ corresponds to '*AnticipatingNextAudit* - Year 2'; and $U(q'_{Year3})$ corresponds to '*AnticipatingNextAudit* - Year 3'.

(2) Uninformative signal (or randomization with replacement, p_2)

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

where N is the total number of GPs; $K_{Year\,y}$ is the assumption made on next year's audit capacity. When $\frac{K_{Year\,y}}{N}$ is equal to a probability corresponding to one of the anticipatory groups, then $U\big(\frac{K_{Year\,y}}{N}\big)$ is the corresponding regression estimate. When $\frac{K_{Year\,y}}{N}$ is not equal to a probability corresponding to one of the anticipatory groups, then $U\big(\frac{K_{Year\,y}}{N}\big)$ is estimated by linear interpolation between the two nearest anticipatory group estimates.

(3) Fully informative signal (or p_3)

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

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 *Audit*'. 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 '*AnticipatingNextAudit* - 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 3), the expectations of being audited captured by '*AnticipatingNextAudit* - Year 3' can be less than 1. This makes the estimate conservative because the deterrence response from bureaucrats under $U(1)$ could be greater that what is estimated with '*AnticipatingNextAudit* - Year 3'.

Finally, when we assume $K_{\tau+1} = K_{\tau+1}$, I estimate $U(0)$ using '*AnticipatingNextAudit* - Year 2' where the beliefs for that group are 0.04. I do so because this is the minimum probability under this assumption, with no other point of support to interpolate what $U(0)$ might be. This approach provides conservative estimates because we would expect the gains in welfare to be smaller with $U(0.04)$, provided that we have assumed $U(q)$ is monotonically increasing in q.

I provide standard errors for the welfare calculations using the bootstrapped estimates of the deterrence parameters.

References

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande. 2020. "E-Governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India." *American Economic Journal: Applied Economics (forthcoming).*

Bobonis, Gustavo J., and Frederico Finan. 2009. "Neighborhood peer effects in secondary school enrollment decisions." *Review of Economics and Statistics* 91 (4): 695–716.

Centre for Sustainable Employment. 2019. *State of Working India.* Technical report. Azim Premji University.

Eeckhout, Jan, Nicola Persico, and Petra E. Todd. 2010. "A theory of optimal random crackdowns." *American Economic Review* 100 (3): 1104–1135.

Kamenica, Emir. 2019. "Bayesian Persuasion and Information Design." *Annual Review of Economics* 11:249–272.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615.

Lalive, Rafael, and M. Alejandra Cattaneo. 2009. "Social interactions and schooling decisions." *Review of Economics and Statistics* 91 (3): 457–477.

Lazear, Edward P. 2006. "Speeding, terrorism, and teaching to the test." *Quarterly Journal of Economics* 121 (3): 1029–1061.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies* 60 (3): 531–542.

Milgrom, Paul, and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1): 157–180.

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." *Indian Journal of Labour Economics* 62 (1): 113–133.

Vazquez-Bare, Gonzalo. 2017. "Identification and Estimation of Spillover Effects in Randomized Experiments." Working Paper.