THE UNIVERSITY OF CHICAGO

ESSAYS ON INCENTIVES AND DESIGN IN PUBLIC INSTITUTIONS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
WENDY WONG

CHICAGO, ILLINOIS
JUNE 2021

For my parents and grandparents,

in honor of their journey.

Happiness is when what you think, what you say, and what you do are in harmony.

-Shri Mahatma Gandhi

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

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 to my coauthors on the project *Productivity Along the Police Hierarchy*, Bocar Ba and Roman Rivera. I am grateful for feedback and support from Scott Ashworth, Dan Black, Cynthia Cook-Conley, Sakari Deichsel, Oeindrila Dube, Steven Durlauf, Jeffrey Grogger, Damon Jones, Christina Juan, Matthew Nassr, Yusuf Neggers, Guillaume Pouliot, James Robinson, Kara Camarena Ross, Raul Sanchez de la Sierra, Shaoda Wang, Austin Wright, and workshop participants at the Harris School's PhD Workshop and Student Working Group, and the University of Chicago's Development Economics Group. 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. Finally, I gratefully acknowledge financial support from the University of Chicago's Committee on Southern Asian Studies and the Becker Friedman Institute Development Economics Initiative. Any opinions expressed are my own.

ABSTRACT

Public institutions are a result of processes that gather preferences of society to formulate public mandates; that collect and distribute revenue; that determine parameters for implementing these public mandates under the guidance of law. Behind these processes are agents who uphold them. This inevitably exposes the integrity of these processes to vulnerabilities. For instance, agents may have motivations which conflict with institutional goals, or the design of institutions may be inapt in empowering agents to achieve institutional goals. These vulnerabilities can lead to a rift between the intent behind the design of our public institutions and what we observe in practice. This research documents such rifts and studies improvements in the design of our public institutions so that what we observe in practice looks more like what we intend.

Monitoring policies designed to maximize deterrence of unwanted behavior must account for attempts by agents to evade detection. Chapters 1-2 of this dissertation, *Optimal Monitoring and Bureaucrat Adjustments*, examine the strategic responses of bureaucrats, who implement India's employment guarantee program, as their expectations of being audited change. Exploiting random assignment to audit timing over multiple waves (without replacement), I find the rate of deterrence for misappropriated expenditures is increasing in bureaucrats' expectations of being audited. In addition, bureaucrats evade detection by adjusting the timing and type of expenditure to misappropriate. Applying a model of Bayesian persuasion, I analyze how information communicated on the likelihood of being audited should be designed. I estimate a sufficient statistic from the model to solve for the optimal signal and analyze welfare under counterfactuals. Concentrated incentives, i.e. notifying of audit timing in advance, would have persuaded bureaucrats to misappropriate USD 35m less in expenditures (16% of average annual expenditures) compared to dispersed incentives, i.e. messages are uninformative and audit timing is unpredictable.

Chapter 3, *Productivity along the Police Hierarchy* (with Bocar Ba and Roman Rivera), empirically shows how hierarchical structure drives productivity in organizations. We estimate the effect that bureaucrats at various managerial levels of the hierarchy have on organizational performance. Using data from the police workforce in Chicago, we decompose the dispersion in police productivity in reported crime and arrests into components resulting from hierarchical structure. We find that a substantial amount of the variation (7-50%) is driven by three components: the design of the managerial positions along the hierarchy along with competency of those who fill those positions, as well as idiosyncrasies in matches of individuals across- and within-ranks working together along the hierarchy.

CHAPTER 1

OPTIMAL MONITORING AND BUREAUCRAT ADJUSTMENTS: EMPIRICAL CONTEXT AND EVIDENCE

1.1 Introduction

Agency issues lead to inefficiencies in organizations (Milgrom and Roberts, 1992). In government, the failure to provide public services is often attributed to the misuse of authority and the misappropriation of resources by bureaucrats (Wilson, 1991; Rose-Ackerman and Palifka, 2016).[1] The potential costs of this kind of misconduct are exacerbated when populations are reliant on the state for economic development and redistribution (Shleifer and Vishny, 1993; Svensson, 2005; Olken, 2006; Olken and Pande, 2012). Governments adopt monitoring programs to identify, deter, and punish such misconduct (Finan et al., 2017). However, little is known about the tradeoffs involved in the design of monitoring policies. While monitoring is effective, budget constraints often prohibit extensive monitoring. As a result, bureaucrats may strategically adjust their behavior to minimize detection. Even less is known about what these deviations by bureaucrats imply for designing a policy that maximizes the deterrence of misconduct.

Given that bureaucrats adjust in response to monitoring, should governments inform bureaucrats when the auditor is coming or make it hard to predict? Theory suggests that the answer depends on the relationship between deterrence and bureaucrats' expectations of being audited (Lazear, 2006; Eeckhout et al., 2010). For example, at one extreme, if bureaucrats 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. Yet, auditing guidelines routinely advise to withhold information from audit subjects to maintain unpredictability.[2]

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

2. See as an example Section 5(c): U.S. auditing standards, adopted and approved by the Public

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, 2017). First, I show that the deterrence of bureaucrats' misconduct is more marginally responsive at higher than lower expectations of being audited. With these results, I apply a model of Bayesian persuasion (Kamenica and Gentzkow, 2011). I then show 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.[3] 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.[4] Issues from audit reports range from participant payment delays to poor workplanning to fabricated employment and material procurement. Existing literature also documents financial misappropriation (e.g. Khera, 2011; Niehaus and Sukhtankar, 2013; Muralidharan et al., 2016; Banerjee et al., 2020).

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.[5] 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

Company Accounting Oversight Board and the U.S. Securities and Exchange Commission, advocate for maintaining unpredictability of when audits may occur.

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

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

5. 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.

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 completed. Bureaucrats respond to these incentives because auditors verify both 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 announcements. I do so with detailed administrative data at the GP-month level.

The main outcome is program expenditures. Program expenditures are measured monthly, under the direct control of bureaucrats, and hard to manipulate ex-post. They focus the study of misconduct on financial misappropriation. However, they comprise both honest and misappropriated expenditures.[6] I take two approaches to disentangle whether incentives from the monitoring policy are driving changes in honest or misappropriated expenditures. Using other administrative data, I test hypotheses to check whether the results are consistent with potential confounding mechanisms that could affect honest expenditures. Bandiera et al. (2009) and Londoño-Vélez and Ávila-Mahecha (2020) use a similar approach to study waste in government procurement and tax evasion. In addition, I use data from audit reports as a source of verification for the estimated changes in program expenditures.

First, I find that the anticipatory effects of the audit on spending are substantial; declines in total expenditures are more responsive at higher (i.e. almost certain one will be audited) than lower expectations of being audited. When expectations of being audited are lower, expenditures are statistically indistinguishable between these periods. When expectations of being audited are high, there is a 15% decline in expenditures.

6. "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.

This is driven by a significant decline in wage expenditures.

Second, I find that bureaucrats substitute misappropriation in wages for misappropriation in materials. I observe this when bureaucrats have moderate expectations of being audited and material procurement significantly increases. This result is verified by audit reports. Audit reports show an increase in fines related to material misappropriation during this period, which is consistent with the expenditures being misappropriated. Audit reports suggest that material misappropriation may be harder to detect relative to wage misappropriation, which explains why bureaucrats adjust along this margin.

Third, intertemporal substitution diminishes the deterrence 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 once auditors leave (driven by an increase in materials). Results to disentangle mechanisms are consistent with changes being driven by misappropriated expenditures. In particular, alternative mechanisms such as multitasking 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 this policy, bureaucrats would never know with certainty when they are up for an audit. When bureaucrats only respond 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 relationship between bureaucrat deterrence and their expectation of being audited is a sufficient statistic for characterizing the principal's optimal signal and for analyzing welfare under counterfactual signals. The shape of this function is an empirical question and determined by the net benefit to bureaucrats from misappropriating additional expenditures. I use reduced-form parameters (the estimated anticipation effects) to estimate this sufficient statistic.

Estimates of the sufficient statistic tell us that informing a random subset of bureaucrats when they will be audited in advance (i.e. concentrated incentives) 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 (i.e. dispersed incentives where likelihood of audit is equal). The conclusion remains unchanged in a series of robustness checks.

Welfare estimates show that a policy with such 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 possibility in other settings that changes in the design of information may yield significant returns.

Contributions to the Literature

This paper (1) provides empirical evidence on strategic responses by bureaucrats to monitoring; (2) 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; and (3) provides a novel empirical measure of the value of information.

This paper complements the literature studying the effectiveness of various rules of monitoring on deterrence. Previous studies demonstrated the importance of: knowing with certainty that your audit will occur (Olken, 2007); having a reliable monitor that can discover and accurately report findings (Banerjee et al., 2008; Duflo et al., 2012; Duflo et al., 2013); having a reliable system for imposing penalties when infractions are found, including an informed electorate (Ferraz and Finan, 2008; Ferraz and Finan, 2011; Afridi and Iversen, 2014; Bobonis et al., 2016); and having a persistent threat from monitoring over time (Avis et al., 2018).

Less is known about strategic responses to a monitoring policy when audit subjects believe their actions are likely to go undetected by auditors. The challenge is that measures of verified performance beyond audits are limited. The tax literature has made use of third-party data to examine how tax reports change with the threat of audit (Casaburi and Troiano, 2016; Carrillo et al., 2017; Londoño-Vélez and Ávila-Mahecha, 2020). Other studies have made use of administrative and audit data, and find audits lead to strategic adjustments in procurement by bureaucrats (Gerardino et al., 2017; Lichand and Fernandes, 2019). This paper leverages administrative and audit data to measure strategic responses in combination with rich variation on the policy parameter of interest: bureaucrats' expectations of being audited. This combination is what allows us to determine the optimal design of information.

This paper is an empirical application of a model of information design, which complements the theoretical literature on optimal information design (e.g., Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Kamenica, 2019). The intuitions from the model in this paper draw on prior theoretical work. Lazear (2006) models the trade off between the provision of concentrated (e.g. informing of audits in advance) versus dispersed (e.g. randomizing with replacement) incentives through monitoring, and finds that the optimal signal depends on the shape of deterrence as a function of expectations of being audited. Eeckhout et al. (2010) study a similar model and examine monitoring of speeding vehicles, but constraints in their empirical setting require assuming an optimal policy is being implemented rather than solving for the optimum. Banerjee et al. (2019) study monitoring of drunk-driving, but their decision-maker solves a different problem than the one in this paper. Drivers can choose to avoid being monitored altogether once they learn the monitor's strategy, 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.

The estimated relationship between deterrence and bureaucrats' expectations of being audited is a sufficient statistic from the theoretical model. Without requiring additional information, the sufficient statistic allows us to determine the optimal signal and analyze welfare under counterfactual signals. The sufficient statistic is estimated using reduced-form parameters. 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 and not structural primitives. To my knowledge, the sufficient statistics approach has not been applied to studies motivated by the optimal design of information. This approach can be applied to other settings where communication on monitoring can play a role to improve governance. Examples range from oversight for fraud in campaign financing, tax returns, and the allocation of social insurance to police use-of-force.

1.2 Background

1.2.1 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. Wages set by the state government are generally below the minimum wage for manual work in agriculture and other industries.[7] Participants construct or maintain assets that are intended to improve rural livelihood. Assets include structures for water conservation and harvesting, homes, latrines, and animal shelters.

Jharkhand is a state in the eastern part of India. The population is close to 33 million with 76% of people living in rural areas. Sixty-one percent of the population relies on

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

agricultural work and are vulnerable to fluctuations in income changing with seasonal agricultural output.[8] NREGS in Jharkhand has served around 7.7 million people and produced over 1 million projects. In 2016, the state government of Jharkhand began auditing GPs implementing the NREGS program.

The Social Audit Unit (hereinafter referred to as the audit agency) is a separate government agency that conducts the audits.[9] The audit agency is funded independently of NREGS and managed by a steering committee of various stakeholders across the state government and civil society. Competitive compensation for auditors and quality assurance mechanisms suggest it is likely the audit agency conducted audits at-scale with credibility and integrity. More details on the audit agency and their processes are in Appendix 1.A.

The goal of the audit agency was to audit all GPs for the first time before selecting GPs for audit for the second time. This goal was consistent with the NREGS national act to ensure regular auditing of all implementing bodies. To do so, the audit agency randomly selected GPs *without* replacement for audit from 2016-2019 until all GPs were audited. This paper focuses on the effect this monitoring policy had on bureaucrat behavior during these 3 fiscal years (FY) when GPs were receiving their first audit.[10] During this period, 4,180 GPs were audited and informed that the previous FY's work would be a part of the audit.

1.2.2 The process when auditors arrive

Auditors spend a week at the GP to verify administrative reports from the previous FY, document other observed issues with ongoing work, and conduct public hearings of their

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

9. "Social audits" incorporate community and participant feedback, hence the name. As noted in Section 1.2.2, 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.

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

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, checking adherence of job advertisements to program guidelines, verifying output at project sites (e.g. measuring dimensions of a dug pond), and interviewing households to verify past employment and document complaints often about recent employment.

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

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

The threat of punishment to bureaucrats from the audit is credible. Issues identified from the audit are presented and adjudicated in a public hearing with auditors, bureaucrats, and participants in attendance.[11] Issues are resolved when evidence is provided

11. 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.

to show the issue is unfounded or when those culpable agree to take corrective action. Bureaucrats charged with financial misappropriation face paying a penalty commensurate with the amount misappropriated or risk losing their job.[12] Unresolved issues are escalated to be adjudicated at a higher-level public hearing. The audit agency reports around $1.5 million USD (11 crore INR) have been recovered through the audits, which is around 0.8% of Jharkhand's NREGS expenditures in FY2018-19.

1.2.3 The information environment and bureaucrat incentives

Bureaucrats can anticipate their next audit when selection for audit is predictable. In this setting, selection is based on whether all GPs have received their first audit before selecting GPs for their second 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 past performance is part of the audit evaluation. This section summarizes what bureaucrats knew and what we can infer about their incentives in response to the monitoring policy.

Every year from FY2016-2019 (the study period), all GPs received an announcement that stated who has been selected for audit, that last year's administrative reports would be part of the audit, and the audit dates. The announcement in Year 1 stated that the audit agency plans to eventually target 50% of GPs for audit every year and that all GPs be audited regularly.[13] It also stated that this fell short of the benchmark in the 2006 NREGS Act Section 17 requiring all GPs be audited twice a year. With this information, it would have been reasonable to expect that being audited for a second time would only occur after all GPs received their first audit.

One's selection for audit could be anticipated, but was not perfectly predictable.

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

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

Current audit capacity was observed, future audit capacity was unknown. So, the number of years it would take to complete the round of first audits was not known ex-ante. Even the leadership of the audit agency was uncertain about future audit capacity. In Year 1 (FY 2016-17), 548 GPs were audited; 1,495 GPs in Year 2 (FY 2017-18); and 2,137 GPs in Year 3 (FY 2018-19). 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.[14] Anticipation in each year comes from the observed changes in both audit capacity and GPs waiting to receive their audit.

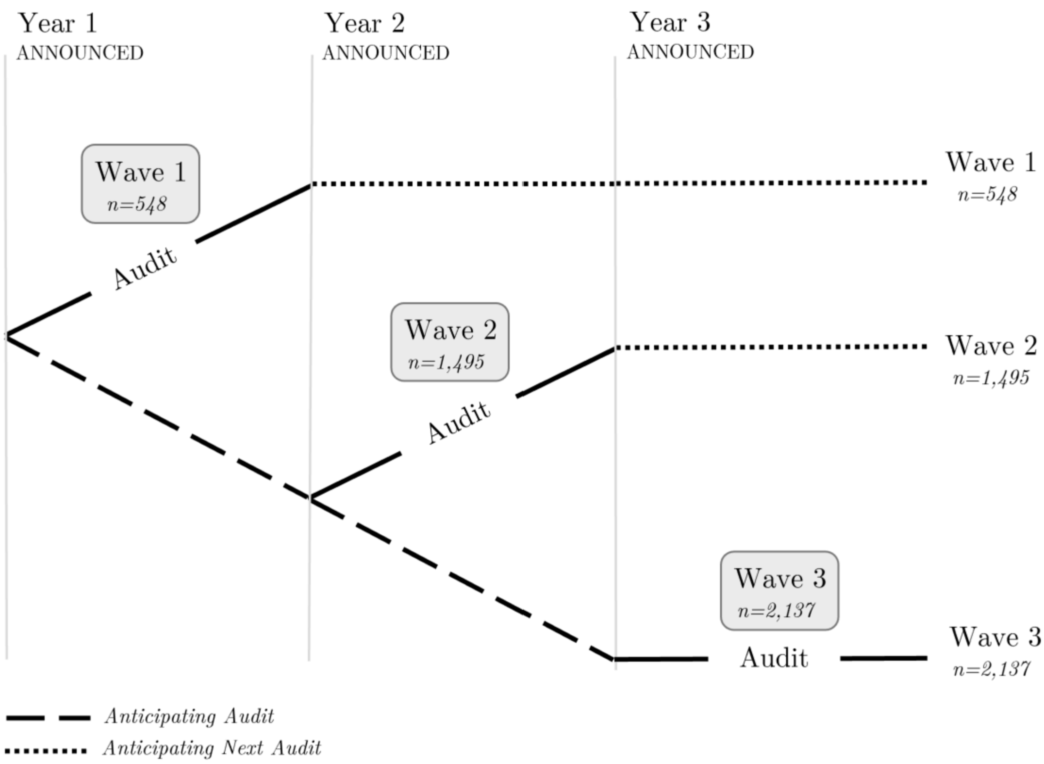


Figure 1.1: Roll-out of the round of first audits and evolution of beliefs for each wave. Dashed and dotted lines denote horizons of anticipating the first and second audits, respectively. Wave 1 follows the path on the top; Wave 2 follows the path in the middle; and Wave 3 follows the path on the bottom. 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 1.1 illustrates how expectations of a future audit for each wave may evolve over the three years of roll-out. Dashed and dotted lines

14. Around 300 GPs were not selected for audit under this policy for various reasons. They are described in detail in Section 1.3.2 and dropped from analysis. More details on notices that were publicly disseminated can be found in Appendix 1.A.

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. The timing of the announcements varied across years, where in Year 1 the announcement occurred during the last third of FY and during the beginning of the FY for Years 2 and 3 (see Table 1.A1 for details).

As we move from Year 1 to Years 2 and 3, bureaucrats who have not had their first audit have increasing expectations that this year's performance will be audited next year. E.g. Wave 3's expectations that current performance will be audited increases over time (i.e. horizon denoted by the dashed lines). While Wave 1 bureaucrats believe the likelihood they will be audited a second time in Year 2 is very low since there are a sufficient number of other GPs waiting for their first audit (i.e. horizon denoted by the dotted line during Year 2). On the other hand, as bureaucrats observe the completion of the round of first audits in Year 3, Waves 1 and 2 bureaucrats have a very high expectation of receiving a second audit in Year 4 (i.e. horizons denoted by the dotted lines during Year 3). About 84% of Waves 1 and 2 GPs compared to 18% of Wave 3 GPs were audited in Year 4.

Additionally, bureaucrats also have an incentive to adjust behavior while their audit is occurring. During the audit, auditors work and sleep at the GP office where NREGS administrative matters take place. And part of the information gathered during the audit process is reflective of concurrent program performance, as discussed in Section 1.2.2.

1.3 Empirical Strategy

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

1.3.1 *Estimating equation and identification assumptions*

We are interested in estimating how bureaucrats respond during time t given their expectations of the likelihood that performance in time t will be audited. I use a differences-in-differences model to estimate bureaucrat responses as expectations change during various stages of the monitoring policy. The following specification estimates the intent-to-treat from learning one has been selected for their first audit:

$$y_{it} = \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \delta_0 Post1stAnnounce_{it} + \epsilon_{it} \tag{1.1}$$

where 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 t within a district d; ϵ_{it} is the residual error term.[15] Standard errors are clustered by block—one administrative level above the GP and the unit of stratification for being selected for audit each year. y_{it} denotes the GP performance outcome. The main outcome is total expenditures, measured as the sum of wage and material expenditures. The excluded component 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). Each component of bureaucrat behavior captured by the remaining variables in Equation 1.1 and their identification will be discussed in turn.

First, consider $Post1stAnnounce_{it}$, a dummy variable capturing the period after each GP learns from announcement of their selection for the first audit. This vari-

15. 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 ≤ 10 GPs; 42% of blocks had ≤ 10 GPs left to be audited by 2017-18). Using the announcement data to check, an average of 98% of GPs within these blocks were audited. For the 2018-19 audit, the remainder of unaudited GPs were selected for audit. In these waves of the audit, we would expect to have independence in observed and unobserved variables between the treatment and control groups conditional on block fixed effects, which also controls for the number of GPs in a given block during this time.

able captures the intent-to-treat of informing bureaucrats of their audit. The effect of $Post1stAnnounce_{it}$ on GP performance y_{it} would be identified under the assumption of parallel trends between those announced for their first audit and not yet announced.

However, being able to anticipate one's first audit is a potential threat to the parallel trends assumption needed to identify $Post1stAnnounce_{it}$. With the random assignment to audit in Year 1, we should expect parallel trends between Wave 1 and non-Wave 1 groups prior to announcement.[16] But, we should not expect parallel trends for subsequent waves because random assignment without replacement under the information environment can lead to anticipation. See the discussion of this anticipatory behavior in Section 1.2.3.

We can control for this horizon of anticipation with $Anticipating1stAudit_{it}$ in order to identify $Post1stAnnounce_{it}$. $Anticipating1stAudit_{it}$ is a vector of two dummy variables which capture anticipation for those waiting for their first audit separately as they observe the announcements in Years 1 and 2. Importantly, the estimates of $Anticipating1stAudit_{it}$ are also parameters of interest because they capture strategic responses by bureaucrats as expectations of being audited vary. Given random assignment without replacement to audit each year, bureaucrat expectations of when their first audit will occur are also randomly assigned, allowing us to identify $Anticipating1stAudit_{it}$.

The following is the main specification of this paper, which disaggregates behavior during $Post1stAnnounce_{it}$ and estimates the profile of bureaucrat responses as expectations of being audited change during the roll-out. This disaggregation into periods capturing

16. To illustrate, under the standard difference-in-differences estimator with two periods in the potential outcomes framework, we have that

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

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

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

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

changing expectations of being audited is effectively an event study specification.

$$y_{it} = \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \underbrace{\delta_1 \text{Before}1stAudit_{it} + \delta_2 \text{Monthof}1stAudit_{it}}_{Post1stAnnounce\text{, disaggregated}} \tag{1.2}$$

$$+ \underbrace{\delta_3 \text{After}1stAudit_{it} + Anticipating2ndAudit'_{it}\gamma}_{Post1stAnnounce\text{, disaggregated (continued)}} + \epsilon_{it}$$

$Anticipating1stAudit_{it}$ is as described above. The remaining variables are mutually exclusive groups and are defined such that $Post1stAnnounce_{it} =$ $\text{Before}1stAudit_{it} + Monthof1stAudit_{it} + After1stAudit_{it} + Anticipating2ndAudit_{it}$. The variable $\text{Before}1stAudit_{it}$ captures the period once one learns they will be receiving their first audit but before the first audit occurs. $Monthof1stAudit_{it}$ captures the period during the month of audit. $After1stAudit_{it}$ 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 experiencing an audit. While GPs were randomly assigned for audit each year, the order in which audits occur within each year is non-random. According to the audit agency, they designed the schedule to complete audits within a district in time for higher-level hearings (for unresolved issues) and to be logistically practical. The district-specific time fixed effects accounts for this timing of audits.

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

$Anticipating2ndAudit_{it}$) is statistically different from behavior before the audit (result reported in Section 1.4.3).

All regressions use '$Anticipating1stAudit$ - Year 1' as the reference group; otherwise, $Anticipating1stAudit$ + $Post1stAnnounce$ would be collinear with a linear combination of time fixed effects. The reason is periods captured by $Anticipating1stAudit$ and $Post1stAnnounce$ are defined by when annual announcements for audit are released, and this affects all GPs within a wave simultaneously. For every announcement, a GP is either anticipating their first audit or has received an announcement for their first audit. This implies that the sum of $Post1stAnnounce$ and $Anticipating1stAudit$ variables are collinear with a linear combination of the time fixed effects. So, using '$Anticipating1stAudit$ - Year 1' as the reference group allows us to estimate the coefficients in Equation 1.2.

1.3.2 Data sources and sample restrictions

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

Annual audit announcements The audit agency provided documentation on the announcements from 2016-2020. Announcements detail GPs selected for audit and audit dates. Together with the announcement dates, this information helps us capture the effect during periods of anticipation discussed in Section 1.2.3; and the effect of learning of an upcoming audit and experiencing an audit.

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

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

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

Audit characteristics and outcomes from audit reports include share of portfolio audited, number of auditors, documented issues, and 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. [18]

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 government officials[19]; 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

18. 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.

19. Requests for audits can be submitted to the audit agency by higher-level government officials; they are referred to as apecial 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.

an administrative boundary change. Furthermore, 42 GPs were selected for audit twice over the audit roll-out. A subset of these 42 GPs were audited twice because they were also selected for special audit; for the remainder, they 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.

1.3.3 Tests for violations of identifying assumptions

Using census and administrative data, statistical tests show balance on observable characteristics across waves (Table 1.1). Except there are statistical differences between Waves 2 and 3 Scheduled Tribes population, share of person-days allocated to females in FY 2015-16, and share of person-days scheduled caste and days of delayed payment in FY 2014-15. The number and extent of these differences are consistent with arising by chance. Overall, the differences are small (4% difference in Scheduled Tribe population and a 3% and 6% difference in share of person-days scheduled caste and female, respectively), and differences in variation in demographic parameters that tend to be stable over time can be accounted for with GP fixed effects in our main specification.

To test for parallel trends, Figure 1.2 shows an event study of total expenditures with lags and leads around the month of announcement and including only '*Anticipating2ndAudit* - Year 2' as a control for anticipatory behavior.[20] There is no statistically distinguishable trend during the months before the announcement (p-value = 0.45). This lends credibility to our difference-in-differences approach. During the months following the announcement, expenditures decline which will be explored in Section 1.4. Additionally, there is no evidence of pre-trends for wage and material expenditures (p-

20. Recall that '*Anticipating2ndAudit* - Year 1' is the reference group. See discussion in Section 1.3.1.

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

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

Table 1.1: Tests of balance in observables across waves

value = 0.19 and 0.99, respectively; see Appendix Figure 1.B2).

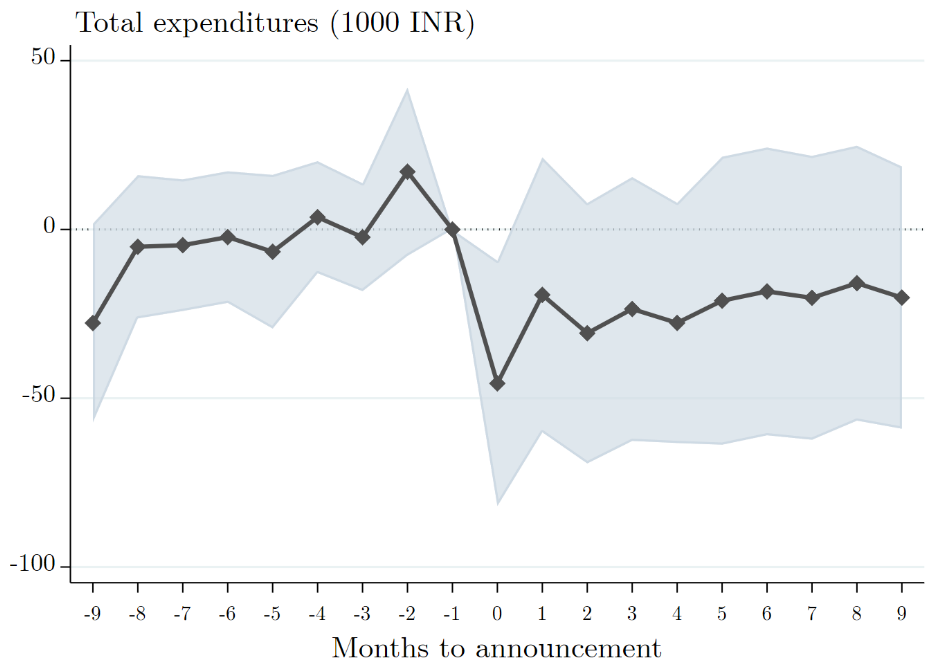


Figure 1.2: Announcement event study for total expenditures (1000 INR). The omitted category in this regression is the one month lead before the announcement. The regression includes GP and district month-year fixed effects. Standard errors are clustered by block.

For robustness, I estimate the same event study excluding variables capturing potential anticipatory behavior of the first audit and also cannot reject the null (Appendix 1.B). Due to potential anticipatory behavior in subsequent waves, we should not expect parallel trends between GPs in Wave 2 and Wave 3 without controlling for their horizons of anticipation. Analyzing the event studies by wave, there is evidence of pre-trends leading up to the Year 2 announcement. This supports our primary difference-in-differences specification in Equation 1.1 which accounts for horizons of anticipating the first audit. These findings also validate our interest in estimating anticipatory behavior.

1.3.4 Inferring deterrence from administrative data

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

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

An independent verification of employment and material procurement over time would be the ideal measure of the effect of the monitoring policy. It would allow us to compare reported versus actual outcomes, and measure overreporting. But, this information is ex-post unavailable, as are third-party sources of verification over time. In addition, Appendix 1.B emphasizes the importance of using monthly data to study the response of bureaucrats to the monitoring policy over time. It shows that measures of annual performance are too coarse to draw conclusions about bureaucrats' strategic adjustments to the monitoring policy from year to year. To my knowledge, the administrative data are the only data that provides measures of GP by month performance for the duration of the study period.

Other studies construct proxies for over-reported employment by estimating the share of fake households under NREGS (Niehaus and Sukhtankar, 2013; Banerjee et al., 2020; Muralidharan et al., 2016). Banerjee et al. (2020) do so at scale by comparing administrative records of working households with households documented in the 2012 Socioeconomic Caste Census (SECC). [22] A similar exercise in our empirical setting would not

21. E.g. see news article here describing how corruption prevents participants accessing benefits of the program.

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

work for two reasons. First, during the beginning of our study period, almost all (99.95%) wage payments were made directly to bank accounts. This follows a program shift from paying wages in cash to direct deposits in bank accounts. With challenges to opening fake bank accounts, this suggests that payment of overreported wages is less likely due to fake workers and more likely due to collusion with or coercion of participants to over-report employment. Second, population changes since 2012 would make the comparison less reliable. For these reasons, estimating the share of fake households with the SECC is not suitable here.

To address these challenges, I study changes in monthly expenditures in response to the monitoring policy. Bandiera et al. (2009) and Londoño-Vélez and Ávila-Mahecha (2020) use a similar approach to study waste in government procurement and tax evasion. Random assignment of audits isolates the effect of incentives created by the monitoring policy on expenditures. To be clear, let c_p denote the true share of total expenditures that is misappropriated under a given audit policy p. Suppose $p \in \{0, 1\}$ where 0 denotes no audit policy and 1 denotes some audit policy. Our goal is to make inferences about $c_1 - c_0$ using administrative reports on expenditures. Let γ_p represent expenditures from the administrative data under policy p. γ_p aggregates honest and misappropriated expenditures under the program.

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

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

1.4 The Impact of Changing Expectations

This section shows that the anticipatory effects are strongest on the margin in deterring misappropriated expenditures when one is almost certain of an audit, while the response under less certainty is statistically indistinguishable from zero. All results examine the effect of the policy on total expenditures alongside the effect on wage and material expenditures for insights into the response of bureaucrats. I provide evidence we can interpret these effects as changes in misappropriated expenditures. On the other hand, deterrence does not come without substitution across time and type of expenditure to misappropriate.

1.4.1 Bureaucrats' response to changing likelihood of audit

Table 1.2 presents the fixed effects regressions: Panel A shows Equation 1.1 and Panel B shows Equation 1.2 disaggregating behavior during $Post1stAnnounce$ into mutually exclusive groups. The outcomes for both panels are total, wage, and material expenditures. As discussed in Section 1.3.1, the reference group for regressions in both panels is the horizon of anticipating the first audit in Year 1 ($Anticipating1stAudit$ - Year 1).

There is a decline in expenditures after learning of selection for audit. Table 1.2 Panel A Column 1 shows a 7% decline (-17.9/269.5, p-value=0.054) relative to baseline total expenditures once a GP learns about selection for their first audit ($Post1stAnnounce$). This effect is driven by a drop in wage expenditures (Column 2).[24,25]

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

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

25. Qualitatively, the estimated decline in total expenditures during $Post1stAnnounce$ is robust to relaxing the assumption that the treatment effects across waves and time are homogeneous, which can

Panel A: Difference-in-differences

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

Panel B: Event study with *Post1stAnnounce* disaggregated

	Expenditures (1,000 INR):		
	(1)	(2)	(3)
	Total	Wages	Materials
Anticipating1stAudit - Year 2	13.87	-5.98	19.84**
	(11.33)	(6.05)	(8.58)
Post1stAnnounce, disaggregated:			
Before 1*stAudit*	-18.83**	-10.21**	-8.62
	(9.46)	(5.04)	(7.14)
Month of 1*stAudit*	-42.07***,†	-35.33***,†	-6.74
	(11.68)	(5.99)	(9.30)
After 1*stAudit*	-8.42	-17.54***,†	9.12
	(11.26)	(5.62)	(8.70)
Anticipating2ndAudit - Year 2	-4.78	-14.54*	9.76
	(13.83)	(7.67)	(9.07)
Anticipating2ndAudit - Year 3	-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 1.2: Effect of stages of the monitoring policy on Bureaucrats' response in program expenditures. This table estimates the main differences-in-differences specification for three outcome variables: total (wages + materials), wage, and material expenditures. The regressions in Panel A estimate Equation 1.1 and the regressions in Panel B estimate Equation 1.2 which includes variables disaggregating behavior during *Post1stAnnounce* into mutually exclusive groups. All regressions include district-month-year and GP fixed effects. Standard errors are clustered by block. The omitted category is the horizon of anticipating one's first audit during Year 1 (*Anticipating1stAudit* - Year 1). The baseline is the mean from the beginning of the panel (two years prior to first audits) up to and including the period captured by *Anticipating1stAudit* - Year 1. This longer period is included in the baseline to average out seasonal variation in expenditures. *** p<0.01, ** p<0.05, * p<0.1. With a Bonferroni correction for the family-wise error rate, we reject the null hypothesis when † p<0.025 for Panel A and p<0.008 for Panel B.

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

Expenditures decline substantially when GPs anticipate with high likelihood that they will be audited. As the roll-out completes in Year 3 and the likelihood of a second audit in Year 4 is extremely likely ('$Anticipating2ndAudit$ - Year 3'), there is a 15% (p-value=0.003) drop in total expenditures. This is driven by a 15% (p-value¡0.0001) decline in wage expenditures (or 14% decline in person-days of employment) and a 13% (p-value=0.24) statistically insignificant decline in material expenditures relative to the reference group.[26] Similarly, when the audit is actually occurring and concurrent performance is subject to monitoring ('Month of $1stAudit$'), there is a 16% (p-value¡0.0001)

lead to negative weights on the average treatment effect on the treated in the difference-in-differences comparisons. I compute the weights following Chaisemartin and D'Haultfœuille (2020), and find that of 72,773 ATTs, 17% have a negative weight and the negative weights sum to -0.4. Employing the estimator derived in Chaisemartin and D'Haultfœuille (2020) which addresses the negative weights and allows for treatment heterogeneity, I find that the coefficient on $Post1stAnnounce$ is negative and greater: -96 with a standard error of 25.

26. We also estimate Equation 1.2 where post-audit anticipatory behavior in Year 3 ('$Anticipating2ndAudit$ - Year 3') is disaggregated by Wave 1 and Wave 2's responses. I find that the average decline in total expenditures is 10% and 15% for Waves 1 and 2, respectively, and their responses are not statistically distinguishable from one another.

decline in total expenditures largely driven by a decline in wage expenditures. These results are robust to tests for spillover effects during the month your neighbors receive an audit (Appendix 1.C). The estimates for '$Anticipating2ndAudit$ - Year 3' and 'Month of $1stAudit$' are both statistically different from anticipatory behavior captured by '$Anticipating1stAudit$ - Year 2' and '$Anticipating2ndAudit$ - Year 2'. This suggests the effect from the policy on deterring misappropriated expenditures is more responsive when there is greater certainty about the likelihood of an audit.

Of note, the effects of anticipating the second audit on total expenditures are identified when there are no persistent effects from the audit (discussed in Section 1.3.1). The effects on total and material expenditures during the months after the audit ('After $1stAudit$') are insignificant and small relative to the reference group. On the other hand, there is a significant decline in wage expenditures during this period, but this effect is being driven by the decline in wage expenditures experienced 1 month after the audit. After parsing out this effect, there is no effect 2+ months after the audit on wage expenditures that is statistically distinguishable from 0 (p-value = 0.25). What happens the months following the audit is explored further in Section 1.4.3. These results suggests the estimates for $Anticipating2ndAudit$ on total expenditures are not confounded with persistent effects from the audit.

Altogether, the results from Table 1.2 Panel B show that bureaucrats response in program expenditures are more responsive when expectations of being audited are high. Can we interpret these estimates as changes in misappropriated expenditures? The following set of results show that decreases in expenditures can be interpreted as deterrence and increases in expenditures can be interpreted as increased misappropriation.

1.4.2 Deterrence and substitution in anticipation

Anticipatory behavior captured by '$Anticipating1stAudit$ - Year 2' in Table 1.2 Panel B can be interpreted as changes in misappropriated expenditures. The anticipatory behavior captured by '$Anticipating1stAudit$ - Year 2' is measured relative to '$Anticipating1stAudit$

	Issue fine amount (1000 INR), by issue type:							
	Wage misappropriation				Material receipts misappropriation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wave 3 - Wave 2	-25.881^{***}	-12.681	-12.674	18.413	30.943^{***}	32.701^{**}	56.186^{***}	64.490^{***}
	(7.499)	(7.982)	(11.453)	(17.436)	(10.730)	(13.253)	(20.181)	(17.601)
Audit manager experience				-3.638^{*}				-0.972
				(1.863)				(2.136)
Wave 2 Mean	68.75	68.75	73.48	73.48	21.58	21.58	22.08	22.08
Controls		X	X	X		X	X	X
Manager FE			X	X			X	X
Observations	2,360	2,360	1,771	1,771	2,361	2,361	1,772	1,772
Adjusted R^2	0.147	0.163	0.185	0.187	0.031	0.031	0.034	0.033

Table 1.3: Differences is audit issue fines across waves is 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 fixed effects. 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.

- Year 1' (the reference group). This is effectively the difference in expenditures between Wave 2 and 3 in the year prior to their respective audits. We can verify the estimated differences in their wage and material expenditures by comparing audit reports from Waves 2 and 3. Table 1.3 tests for differences in audit performance between Wave 2 and Wave 3. Results are presented on the amount fined for issues related to wage and material expenditures.

The difference in fines related to issues of wage misappropriation between Waves 2 and 3 is statistically indistinguishable from 0 (after controlling for potential confounders). This effect is consistent with our estimate for '$Anticipating1stAudit$ - Year 2' that spending on wages is statistically indistinguishable from zero (Table 1.2 Panel B). 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 regardless of the control variables included. This is consistent with the estimate for '$Anticipating1stAudit$ - Year 2' that Wave 3 spent 25% (p-val = 0.02) more on materials during this period of anticipation. When examining the number and share of issues identified for wage and material misappropriation, the results are qualitatively similar (Appendix Table 1.C2 and 1.C3)

These results lend credibility to our interpretation that the estimated differences in bureaucrat behavior during periods of anticipation are interpretable as changes in misappropriated expenditures. Moreover, an average decrease in wage expenditures and employment misappropriation issues coupled with an average increase in material expenditures and material misappropriation issues shows that bureaucrats substitute across 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 to adjust wage for material expenditures because they believe misappropriation of these expenses are easier to hide from auditors. Misappropriation of wage expenditures can be discovered by auditors during interviews with participating households. 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, Appendix Table 1.C4 shows that Wave 3 GPs have a greater number of issues and fines related to the refusal to provide records for auditors.

1.4.3 Deterrence and substitution during the audit

This section shows how the decline in expenditures driven by a decline in employment during the months around the audit, estimated by 'Month of $1stAudit$' in Table 1.2 Panel B, can be interpreted as changes in misappropriated expenditures. Bureuacrats may be deterred from misappropriating expenditures because concurrent performance is being evaluated by auditors (see discussion in Section 1.2.2). However, the actual act of experiencing an audit introduces potential confounders to interpreting bureaucrat behavior as changes in misappropriated expenditures. Two alternative mechanisms potentially confound our interpretation and are tested: a disruptive audit leads to difficulties multi-tasking where usual tasks cannot be completed (Holmstrom and Milgrom, 1991); and the audit helps bureaucrats learn and improve productivity (Arrow, 1962; Syverson, 2011).[27]

There are two aspects to the multi-tasking issue while auditors are present: (1) a disruptive audit keeps bureaucrats from getting honest work done; and (2) a disruptive audit keeps bureaucrats from misappropriating expenditures as they try to minimize detection while auditors are present.[28] While the audit process is designed to prevent the audit from being disruptive to honest work (see discussion in Section 1.2.2), the former is still a potential concern as a confounder. The latter is not because it is an aspect of the bureaucrat's multi-tasking problem that contributes to changes in misappropriated expenditures.

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

$$y_{it} = \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \sum_{k \in \tau} \delta^k 1stAudit^k_{it} + \epsilon_{it} \tag{1.3}$$

where $1stAudit^k_{it}$ is an indicator taking the value 1 if i is k months from audit at time t. As discussed in Section 1.3.1, the set of fixed effects account for potentially endogenous timing of audits due to scheduling constraints of the audit agency.

Figure 1.3 presents the leads and lags of the regression specified in Equation 1.3 for

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

28. Field interviews with participants confirm that this is potentially a widespread problem.

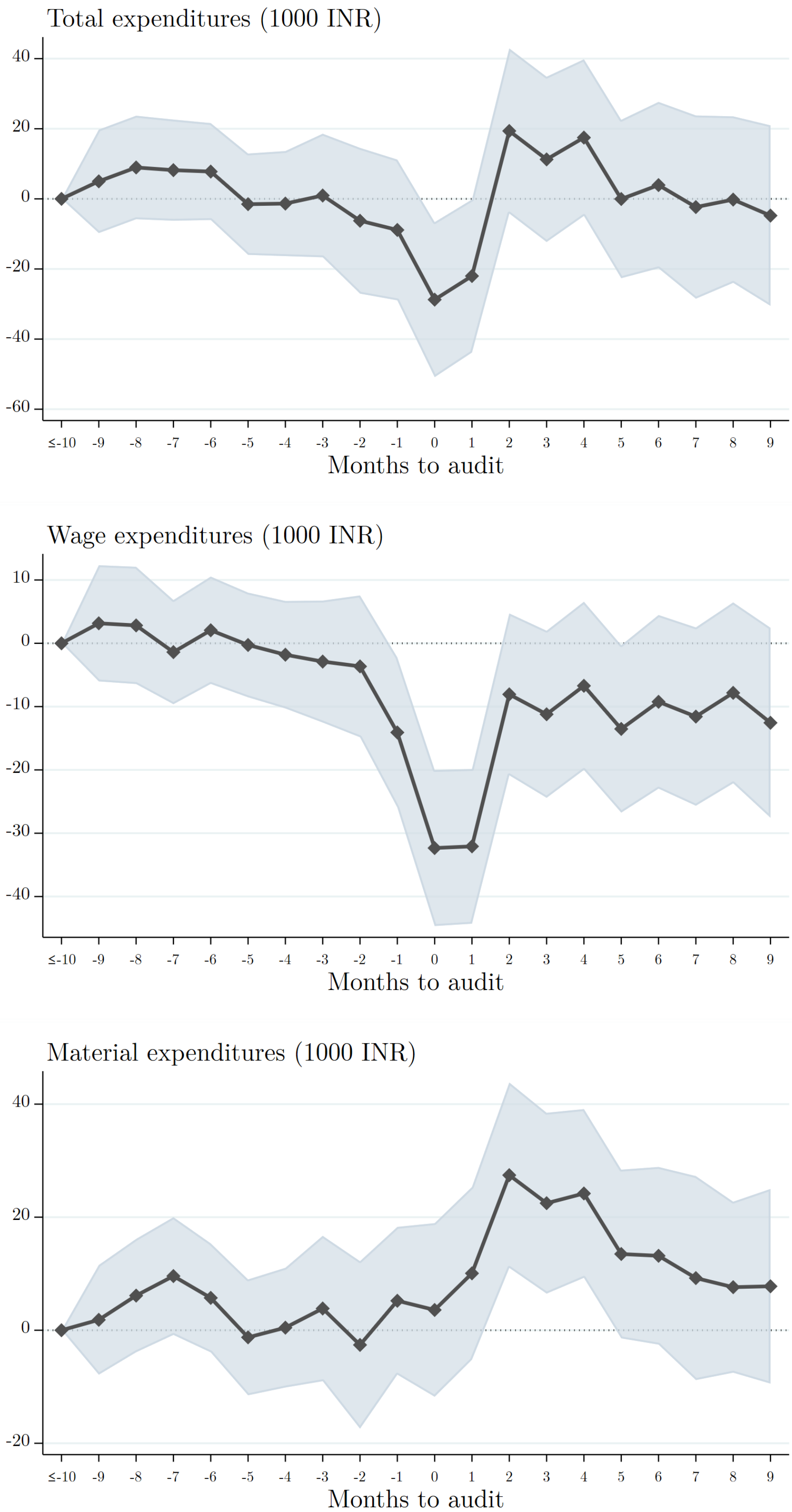


Figure 1.3: Changes in expenditures around the time of audit (Equation 1.3). The omitted category is 10 or more months before the audit. The raw mean of the omitted category is 272, 187, and 85 (1,000 INR) for total, wage, and material expenditures, respectively. The regressions include GP and district month-year fixed effects. Standard errors are clustered by block.

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

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

There are several reasons why multi-tasking issues do not explain the decline in expenditures during the month of the audit. First, while the presence of auditors affects employment and material procurement, it does not affect efficiency in making wage payments. If multi-tasking explains the decline in employment, then it may also affect other tasks. Delays are a low-cost task for bureaucrats to shirk while auditors are present. They are a lagged indicator of performance and current measures of delay are not detectable by auditors. Studying delays in making wage payments also allows us to disentangle GP bureaucrat effort from incentives to be corrupt.[29] Delayed payments are often cited to reflect a lack of a core competency of the government in targeting resources[30], and not a known mechanism through which bureaucrats misappropriate finances.

There is no statistically significant change in average days of delayed payment per household during the month around the time of audit; the average increase during the month of audit is less than one day of delayed payment per household (Appendix Fig-

29. Days of delayed payment is the number of extraneous days it takes to process a payroll (Banerjee et al., 2020; Narayanan et al., 2019). Once the payroll is processed, the central government deposits wages to participant bank accounts. Appendix 1.C explores delays in greater detail, and the conclusion that audits do not affect delays remains unchanged.

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

ure 1.C4). This provides a piece of evidence that multi-tasking issues during the audit are not driving the measured decline in employment and increase in material procurement.

Second, bureaucrats' responses in the months around the audit are unlikely to have impacted program output, providing further evidence that multi-tasking is not a concern. If we were worried that the decline in labor input during the month of the audit was a decline in real employment and thus would lead to a decline in real output, then we would also expect a corresponding decline in material input.[31] During the period prior to the first wave of audits, a 1% increase in material expenditures is associated with a 0.06% increase in person-days of work provided (t-stat = 8.12). If material and labor inputs are positively correlated, then real output will tend to depend on both inputs.

Figure 1.3 shows no decline in materials procured around the time of audit corresponding to the observed decline in employment. If procurement of materials were not in sync with employment, we would expect a corresponding decline in materials procured some months before or after the decline in employment (if this were a decline in real employment). However, a increase in procured materials is the only observed response occurring the month after the decline in employment. This suggests that real output did not decline during the month of audit and multi-tasking while auditors were present was not an issue.

The findings are consistent when examining employment by whether the project worked on required materials or only required labor. Figure 1.4 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 1.3) does not correspond to an increase in employment on projects requiring materials as shown in Figure 1.4 (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.

31. 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 bureuacrats to fake the works funded on the payroll by paying wages for fabricated work and having machines complete the work.

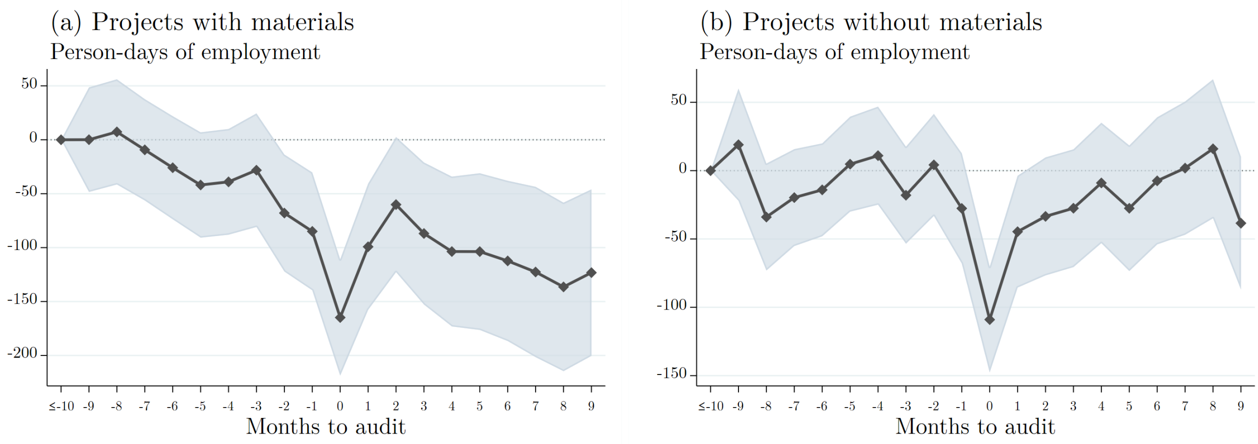


Figure 1.4: Person-days of employment around the time of audit, by whether the project requires materials. The omitted category is 10 or more months before the audit was conducted. The raw mean of the omitted category is 855 and 310 work-days for projects requiring and not requiring materials, respectively. The regression includes GP and district month-year fixed effects as specified under Equation 1.3. Standard errors are clustered by block.

These results are also unlikely to be a result of second order effects from the audit, e.g. bureaucrats are transferred or fired as a punishment, and the GP's workflow disrupted. Punishments from audits are enforced in higher-level hearings and these hearings can happen within the same month or months after the audit. If we disaggregate the event study by whether the higher-level hearing occurred within the month or months after the audit, we observe no differences in trends between these groups.

Taken together, these results show that the changes in expenditures in the months around the audit left program output unaffected and can be interpreted as changes in misappropriated expenditures. Furthermore, the decline then increase in expenditures around the time of audit suggests substitution by bureaucrats 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 (Section 1.4.2).

Interpreting the measured decline in employment (Figure 1.3) as a decline in misappropriated employment, the audit averted false payments equivalent to 18% of the level of employment in the months before the audit. This decline is consistent in magnitude with what Imbert and Papp (2011) measure in Khera (2011). They estimate overreported employment in person-days spent on public works using household survey data to be between 20-29% of the reported person-days of NREGS employment across the states in their sample.

The results from this section suggest that bureaucrat adjustments through decreases in misappropriated expenditures around the time of audit are short-lived at best. But, results also suggest lost rents are recovered later as bureaucrats substitute their behavior across time and type of expenditure.

The empirical evidence from this chapter will be used to inform the theoretical analysis in Chapter 2. The theoretical analysis will characterize the optimal design of information communicated to bureaucrats about the likelihood of audit. With these results, I also analyze welfare implications of alternative communication policies.

Chapter 1 Appendix

1.A Background

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.[32]

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

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

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 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.[33] Second, auditors

33. 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.

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, 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.

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

Table 1.A1 summarizes details of the audit schedule for each year of the roll-out of audits.

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

Table 1.A1: Audit Schedule, 2016-2019

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

- Audit agency created, May 2, 2016 - Notice on creation of the audit agency and personnel to be hired to staff the agency.

- Year 1 announcement, December 29, 2016 - Commencement of 548 audits for Wave 1.

  Official notice was disseminated on guidelines for conducting the audits and announcement of GPs in Wave 1 of audits to be conducted in the remainder of Year 1. The notice also states that the goal of the audit agency is to eventually be able to audit 50% of GPs every year. This is against the benchmark stated in the 2006 NREGS Act Section 17 that requires all GPs to be audited twice a year.[34]

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

- Year 2 announcement, May 2, 2017 - Commencement of 1,495 audits for Wave 2.

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

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

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

34. 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.

Parallel Trends and Anticipation By Year

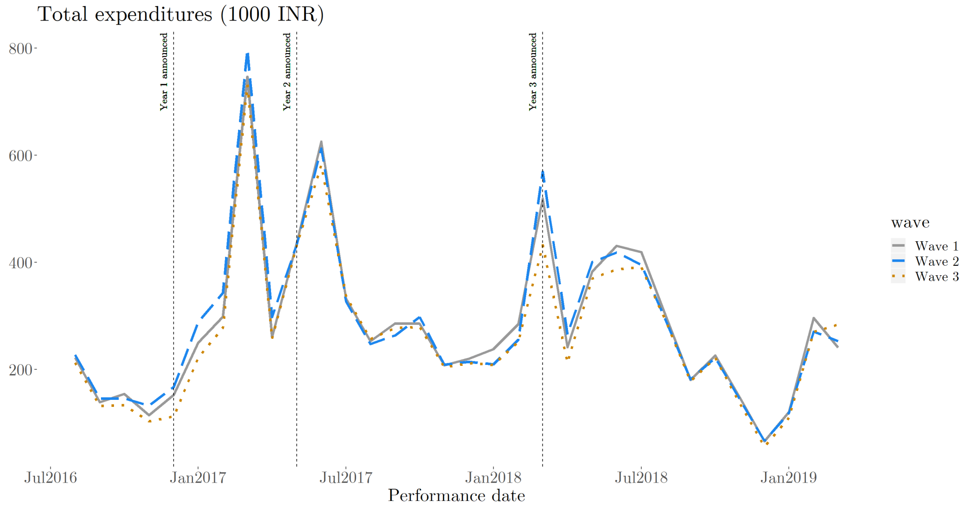


Figure 1.B1: Raw means of monthly total expenditures, by wave

Figure 1.B1 plots a time series of the average person-days of employment for each wave. It suggests parallel pre-trends between the treatment and control groups leading up to the first announcement (denoted by the first dashed line).[35] Furthermore, those being audited experience slight declines in employment as audits are implemented throughout the year.

There are two events for every wave of the audit where the GPs potentially have an incentive to change their behavior as a response: 1) the announcement of the audit schedule which is simultaneously distributed across all GPs; and 2) the audit itself. Once informed of the audit, bureaucrats in their respective audit group may respond differentially in anticipation of the audit. When estimating the effect of the audit itself, we should not expect parallel pre-trends in the treatment and control group if one group reacts in anticipation of the audit while the other is operating business as usual.

As discussed in Section 1.3.1, we should expect parallel pre-trends between the groups prior to the Year 1 announcement of Wave 1 audits and can expect parallel pre-trends

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

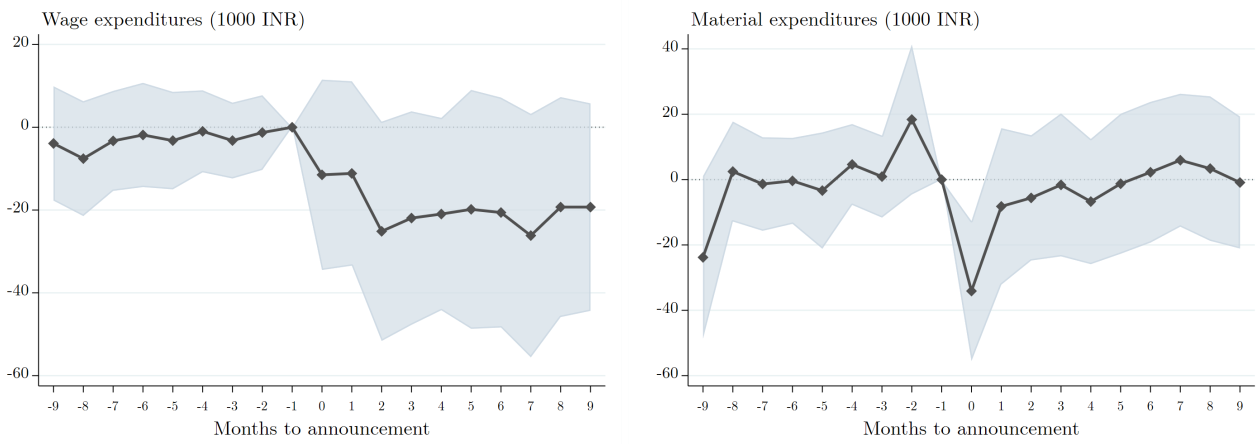


Figure 1.B2: Announcement event study for wage and material expenditures. The omitted category is the month before the announcement. The raw mean of the omitted category is 179 and 69 (1000 INR) and the p-values of no pre-trends are 0.99 and 0.11 for wage and material expenditures, respectively. The regression includes GP and district month-year fixed effects. Standard errors are clustered by block.

prior to the announcement in subsequent years after accounting for the horizons of pre-audit anticipation. We can test for violations of parallel trends by estimating the following event study around the time of announcement:

$$y_{it} = \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \sum_{k \in \tau} \delta^k Announce^k_{it} + \epsilon_{it} \tag{1.B1}$$

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

In the months before the announcement, the total expenditures do not seem to be following a trend and are statistically indistinguishable from behavior the month before the announcement (p-value = 0.45, Figure 1.2). This lends credibility to our difference-in-differences approach. Results look similar for wage and material expenditures (p-value = 0.19 and 0.99, respectively) captured in Figure 1.B2.

The same event study, excluding variables capturing potential pre-audit anticipatory behavior, lead to the same conclusion—that we cannot reject that there are no pre-trends.

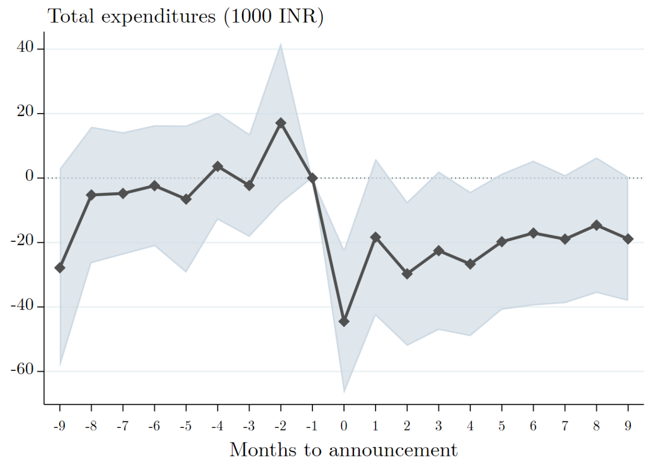


Figure 1.B3: Announcement event study for total expenditures without controlling for anticipatory behavior. The omitted category in this regression is the one month lead before the announcement. The raw mean of the omitted category is 248 (1000 INR) and the p-values of no pre-trends is 0.34. The regression includes GP and district month-year fixed effects. Standard errors are clustered by block.

However, breaking up the estimation by wave, there is evidence of pre-trends leading up to the Year 2 announcement for Wave 2 audits. This supports our primary difference-in-differences specification in Equation 1.1. It accounts for horizons of anticipatory behavior to not only provide credible estimates of the effect of the announcement and the audit, but also validates our interest in estimating the parameters of anticipatory behavior.

Figure 1.B3 estimates Equation 1.B1 but excludes the variables capturing the horizons where GPs potentially anticipate their first audit. We are led to the same conclusion that we cannot reject there is no pre-trend prior to each wave's announcement (p-value $= 0.41$).

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

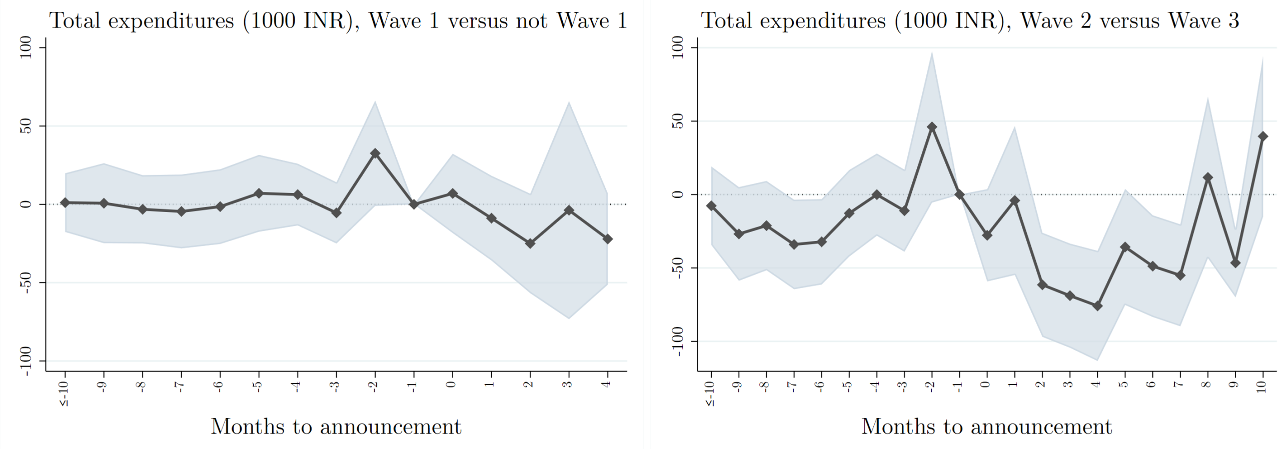


Figure 1.B4: Announcement event study by wave for total expenditures. The omitted category in this regression is the one month lead before the announcement. It includes GP and district month-year fixed effects as specified under Equations 1.B1 and 1.3, and observations are weighted by the inverse of number of households in the GP. The figure on the left compares those selected for audit in Wave 1 to those not audited and the panel of data is truncated at the month prior to the announcement of the second wave audit. The figure on the right compares those selected for audit in Wave 2 to those not yet audited (and will be audited in Wave 3) and the panel of data is truncated at the month prior to the announcement of the third wave audit. So, for instance, the announcement pre-periods in the figure on the left include data from both Waves 2 and 3 while the announcement post-periods include data only from Wave 2.

audit as specified in Equations 1.1- 1.3.

Alternatives to the baseline specification

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

Table 1.B1 shows the results for the specification in Equation 1.1 using only month-year fixed effects and not district-month-year fixed effects in the first column. The remaining columns show the specification in Equation 1.1 for additional outcome variables. Results are qualitatively similar for other measures of employment.

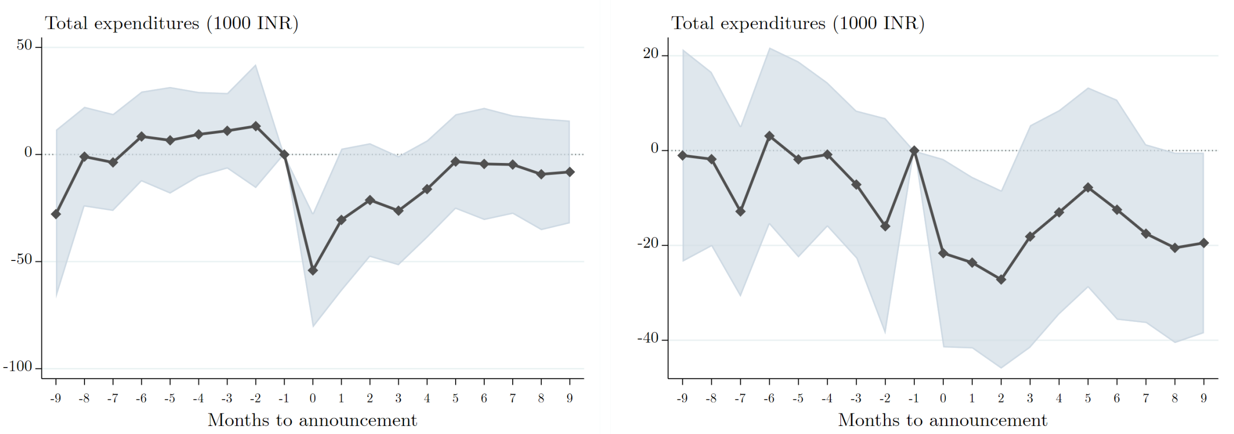
Figure 1.B5 shows the specification in Equation 1.B1 using month-year and block-month-year fixed effects. Qualitatively, the results are similar and we cannot reject that there are no pre-trends (p-value $= 0.36$ and 0.53, respectively). I use district$\times$month FE over block $\times$ month-year fixed effects because for Wave 2 of the audit, all GPs within a block were assigned to be audited if they were considered to be small blocks. In this case,

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

Table 1.B1: Effect of stages of the monitoring policy on additional outcomes and specifications. The first regression uses month-year FE for the main outcome variable, total expenditures; the remaining regressions show the effect of the policy on other performance outcomes using district-month-year FE. Standard errors are clustered by block. The omitted category is the horizon of anticipating one's first audit during Year 1. *** p<0.01, ** p<0.05, * p<0.1

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

Finally, the preferred specification for the outcome variables on expenditures is levels of expenditures, rather than log-transformed for example. Several reasons inform this decision. First, the results are qualitatively robust under a log transformation of the outcome variable. Second, zero expenditures in wages and materials by month are not uncommon. Taking the log transformation of total, wage, and material expenditures would drop different observations in regressions across each outcome variable. Using levels of expenditures ensures that regressions are performed on the same sample for each outcome variable. Finally, a Box-Cox test, which tests goodness of fit of transformations of the outcome variable, was also performed. Residuals from regressions with the transformation of best fit (including an inverse hyperbolic sine transformation) were bimodal, and not normally distributed. This would violate our fixed effects regression assumption that the residuals be normally distributed.



(a) includes GP and month-year FE

(b) includes GP and block-month-year FE

Figure 1.B5: Announcement event study with month-year and block-month-year fixed effects. Standard errors are clustered by block.

The importance of using monthly performance data

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

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

	Total Expenditures (1000 INR)				
	(1)	(2)	(3)	(4)	(5)
	FY1415	FY1516	FY1617	FY1718	FY1819
			$Year1$	$Year2$	$Year3$
Treatment group mean by fiscal year					
$Wave1$	2,465	3,154	4,382	3,906	3,024
$Wave2$	2,405	3,196	4,301	3,768	2,868
$Wave3$	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 1.B2: 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

expenditures on materials and labor.

1.C The Impact of Changing Expectations

Testing for Spillover Effects of the Audit

This section estimates whether the direct effects from the audit are confounded with spillover effects from GPs within your own block. Random assignment of GPs to audit means that the concentration of audits within a block is also random. So, we can estimate the spillover effects of the audit from GPs within the same block. Results show spillover effects from being audited are not a concern for identification of the direct and anticipatory effects of the audit policy.

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

When treatment is randomized, spillover effects can be estimated in a reduced-form linear-in-means specification by including a control for share of GPs being audited within one's block (Manski, 1993; Bobonis and Finan, 2009; Lalive and Cattaneo, 2009; Dieye et al., 2014):

$$\begin{aligned} y_{it} =& \alpha_i + \alpha_{dt} + Anticipating1stAudit'_{it}\beta + \delta_1 Post1stAnnoounce\text{Before}1stAudit_{it} \\ &+ \delta_2 1stAudit_{it} + Anticipating2ndAudit'_{it}\gamma + \eta ShareBlockAudited_{it} + \epsilon_{it} \end{aligned} \tag{1.C1}$$

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

First, an event study around time to announcement including $ShareBlockAudited$ as a control shows that the results do not change (Figure 1.C1). Furthermore, an F-test (p = 0.84) shows support for parallel pre-trends.

Table 1.C1 Column 1 shows the main specification from Equation 1.1 without spillover effects. Column 2 includes $ShareBlockAudited$ as a control and shows that the estimates on both anticipatory and direct audit effects are unaffected. Column 3 interacts $ShareBlockAudited$ with an indicator for the month of audit to provide evidence for assumption (i) that we cannot reject the spillover effects, if any, are equal across audited and not audited groups during the month of audit. Column 4 includes a quadratic term for $ShareBlockAudited$ and provides evidence for assumption (ii) where the coefficient on the quadratic terms is insignificant. This tells us we cannot reject the spillover effects are linear in the share of group receiving treatment.

Furthermore, Table 1.C1 shows overall that the estimated anticipatory and direct

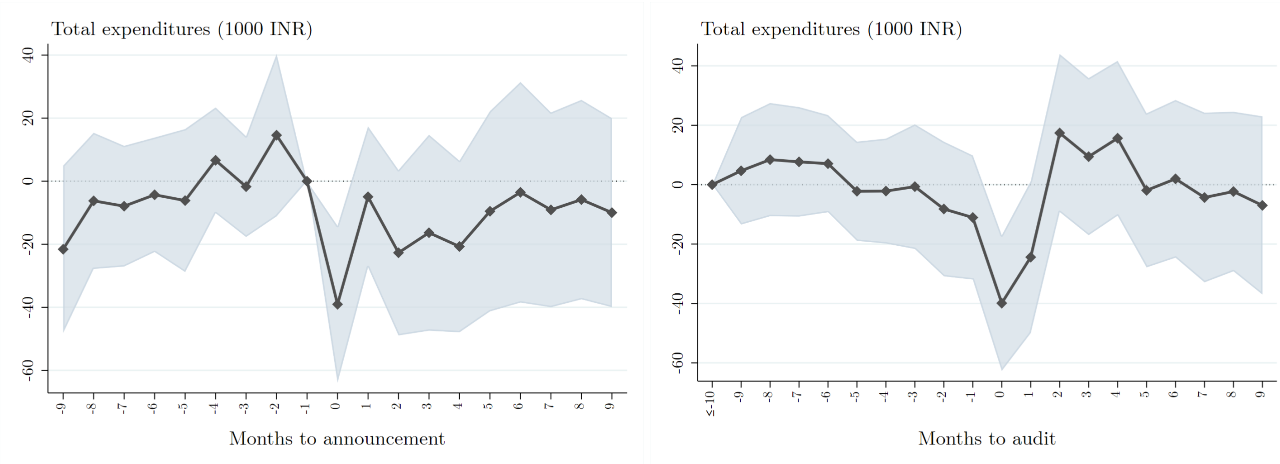


Figure 1.C1: Event studies around time to announcement and audit are unaffected by controlling for spillovers. The omitted category is 10 or more months before the announcement and audit, respectively. The raw mean of the omitted category is 250 and 272 (1,000 INR) for the announcement and audit event studies, respectively. The regressions include GP and district-month-year fixed effects. Standard errors are clustered by block.

effects of the audit are largely unchanged after accounting for concentration of audits within one's block. This provides additional evidence that the anticipatory effects are driven by changes in expectations of being audited rather than perceptions about the audit driven by peer experiences.

Figure 1.C1 shows an event study around month of audit (controlling for '$Anticipating1stAudit$ - Year 2' as in the main specification) and the results remain unchanged when controlling for spillovers. As before, we cannot reject the hypothesis of no pre-trends prior to one's announcement (p-value $= 0.5$). Altogether, the estimates in the main analysis are robust to spillover concerns.

	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

Table 1.C1: Tests for spillover effects of the audit. All regressions include district-month-year and GP fixed effects. Standard errors are clustered by block. The omitted category is the horizon of anticipating one's first audit during Year 1 (*Anticipating1stAudit* - Year 1). The baseline is the mean from the beginning of the panel (two years prior to first audits) up to and including the period captured by '*Anticipating1stAudit* - Year 1'. *** p<0.01, ** p<0.05, * p<0.1.

Additional results with audit report data

	Number of issues, by type:							
	Wage misappropriation				Material receipts misappropriation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wave 3 - Wave 2	-0.495	-0.344	-0.049	-0.246	1.023^{***}	1.015^{***}	0.887^{***}	1.047^{***}
	(0.330)	(0.299)	(0.357)	(0.547)	(0.133)	(0.157)	(0.164)	(0.256)
Audit manager experience				0.023				-0.019
				(0.044)				(0.019)
Wave 2 Mean	3.67	3.67	3.79	3.79	0.42	0.42	0.40	0.40
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	0.392	0.417	0.535	0.535	0.266	0.267	0.289	0.289

Table 1.C2: Differences is audit issue counts across waves is consistent with differences in wage and material expenditures while anticipating the first audit. Unit of observation is GP. All regressions include block fixed effects. 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.

	Share of total issues, by type:								
	Wage misappropriation			Material receipts misappropriation			Project non-existent		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Wave 3 - Wave 2	-0.055^{***}	-0.043^{***}	-0.032^{*}	0.036^{***}	0.037^{***}	0.037^{***}	-0.005	-0.007	-0.026
	(0.010)	(0.010)	(0.017)	(0.005)	(0.006)	(0.011)	(0.011)	(0.011)	(0.021)
Audit manager experience			-0.0002			-0.001			0.002
			(0.001)			(0.001)			(0.002)
Wave 2 Mean	0.17	0.17	0.17	0.04	0.04	0.04	0.09	0.09	0.09
Controls		X	X		X	X		X	X
Manager FE			X			X			X
Observations	2,361	2,361	1,772	2,361	2,361	1,772	2,361	2,361	1,772
Adjusted R^2	0.247	0.265	0.356	0.224	0.224	0.233	0.238	0.242	0.301

Table 1.C3: Audit performance across waves is consistent with differences in behavior while anticipating the first audit. Unit of observation is GP. All regressions include block fixed effects and the following 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.

	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 1.C4: 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.

	Number of issues, 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	1.336***	1.119***	1.309***	0.634**	0.112***	0.103***	0.144***	0.121**
	(0.192)	(0.190)	(0.261)	(0.314)	(0.033)	(0.034)	(0.045)	(0.052)
Audit manager experience				0.079***				0.003
				(0.030)				(0.005)
Wave 2 Mean	1.82	1.82	1.83	1.83	0.13	0.13	0.13	0.13
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	0.216	0.222	0.316	0.318	0.058	0.057	0.158	0.158

Table 1.C5: Differences is audit issue counts 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. Unit of observation is GP. All regressions include block fixed effects. 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.

	Number of issues, by type:							
	Machine completed work				Project non-existent			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wave 3 - Wave 2	-0.336^{***}	-0.289^{***}	-0.273^{***}	-0.320^{***}	0.719^{***}	0.609^{***}	0.696^{**}	0.120
	(0.113)	(0.106)	(0.066)	(0.092)	(0.233)	(0.228)	(0.299)	(0.453)
Audit manager experience				0.005				0.067^{**}
				(0.009)				(0.033)
Wave 2 Mean	0.44	0.44	0.40	0.40	1.73	1.73	1.77	1.77
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	0.024	0.033	0.166	0.165	0.227	0.252	0.308	0.309

Table 1.C6: Differences is audit issue counts across waves for issues related to completing works with machines and project non-existent. Unit of observation is GP. All regressions include block fixed effects. 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.

	Issue fine amount (1000 INR), by issue type:							
	Machine completed work				Project non-existent			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wave 3 - Wave 2	−13.273**	−10.439**	−10.962**	−15.436*	8.142	17.466	32.700**	33.809
	(5.266)	(4.813)	(5.452)	(7.842)	(9.092)	(11.249)	(16.355)	(23.737)
Audit manager experience				0.524				−0.130
				(0.518)				(1.847)
Wave 2 Mean	16.45	16.45	17.69	17.69	54.45	54.45	55.42	55.42
Controls		X	X	X		X	X	X
Manager FE			X	X			X	X
Observations	2,360	2,360	1,772	1,772	2,361	2,361	1,772	1,772
Adjusted R^2	0.014	0.019	0.179	0.179	0.169	0.213	0.260	0.260

Table 1.C7: Differences is audit issue fines across waves for issues related to completing works with machines and project non-existent. Unit of observation is GP. All regressions include block fixed effects. 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.

Testing the Assumption that Perception of Audit Quality is Constant

This section tests the assumption in Section 1.3 that bureaucrat perceptions of audit quality was constant across years. With this assumption, we can attribute differences in bureaucrat responses to the monitoring policy across years to changing expectations of being audited. To test this assumption, we can examine whether audit quality or bureaucrats' responses to the audit differ across years.

Using data from audit reports on inputs and outputs of audits, we can test for differences in audit implementation to see if audit quality was comparable across years. Workload per auditor and audit expenditures serve as proxy variables for audit quality. Workload per auditor is measured as the sum of households and projects to be verified divided by number of auditors. Only a subset of data from Waves 2 and 3 are available. Table 1.C8 shows no statistically significant differences between Wave 2 and 3 in workload

	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

Table 1.C8: Audit quality by year of audit. Unit of observation is the GP. Omitted group is Wave 2. Outcome variables are: workload per auditor measured as the number of households and projects to verify per auditor; and total audit expenditures. Control variables include: Number of employed households and works to verify; and number of auditors. Standard errors are clustered by block. *p<0.1; **p<0.05; ***p<0.01

per auditor and audit expenditures.

While this tells us whether the audit treatment was uniformly applied across years, more information is required to understand whether the bureaucrats responded differentially over time. If the perceived quality or credibility of the audit is changing, then one might expect the response to vary by year; if constant, then one might expect estimates to be same across years. To test for this we can extend the estimates from Table 1.2 Panel B by parsing the treatment effects by year for total expenditures, the main outcome of interest for estimating the sufficient statistic in Section 2.2. This tells us whether bureaucrats had notable differences in their responses to the monitoring policy over time.

Table 1.C9 shows that the treatment effects for ('Month of 1*stAudt*', and 'After 1*stAudit*') across Year are on average different but not statistically distinguishable for total expenditures. We cannot reject the hypothesis that the true difference between their responses is zero. Particularly, there are no distinguishable effects across years during the month of audit; and during the months following the audit (outside of post-audit

	Total expenditures (1,000 INR)	
	(1)	(2)
Anticipating1stAudit - Year 2	13.87	14.33
	(11.33)	(11.76)
Post1stAnnounce, disaggregated:		
Before *1stAudit*	-18.83**	-17.72**
	(9.46)	(8.56)
Month of *1stAudit*	-42.07***	
	(11.68)	
Month of *1stAudit* - Year 1		-54.61**
		(23.46)
Month of *1stAudit* - Year 2		-45.10***
		(15.50)
Month of *1stAudit* - Year 3		-33.04**
		(14.54)
After *1stAudit*	-8.42	
	(11.26)	
After *1stAudit* - Year 1		26.69
		(30.58)
After *1stAudit* - Year 2		-10.08
		(13.39)
After *1stAudit* - 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 *1stAudit*, H_0 : Year 1 = Year 2 = Year 3 (p-val)		0.67
After *1stAudit*, 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

Table 1.C9: Effect of stages of the monitoring policy disaggregated by year. This table estimates the main differences-in-differences specification breaking down 'Month of *1stAudit*' and 'After *1stAudit*' by year and '*Anticipating2ndAudit* - Year 3 by Wave' for total expenditures. Regressions include district-month-year and GP fixed effects. Standard errors are clustered by block. The omitted category is the horizon of anticipating one's first audit during Year 1 (*Anticipating1stAudit* - Year 1). The baseline is the mean from the beginning of the panel (two years prior to first audits) up to and including the period captured by *Anticipating1stAudit* - Year 1. This longer period is included in the baseline to average out seasonal variation in expenditures. *** p<0.01, ** p<0.05, * p<0.1

anticipatory behavior). If the null is true, then the assumption that Bureaucrat beliefs over credibility of audit are constant across years is credible. And the estimated changes bureaucrats' responses are driven by changing expectations of being audited. So, we can proceed with Section 2.2.2's estimates of the sufficient statistic to determine the optimal design of information.

On the other hand, if the parameter is *not* truly zero, then it could be that we are not powered to detect small differences given the sample size. And the estimated response of bureaucrats is confounded with changing perceptions of audit quality. Section 2.2.3 addresses this possibility by conducting robustness checks which relax this assumption.

Delayed payments as a measure of effort

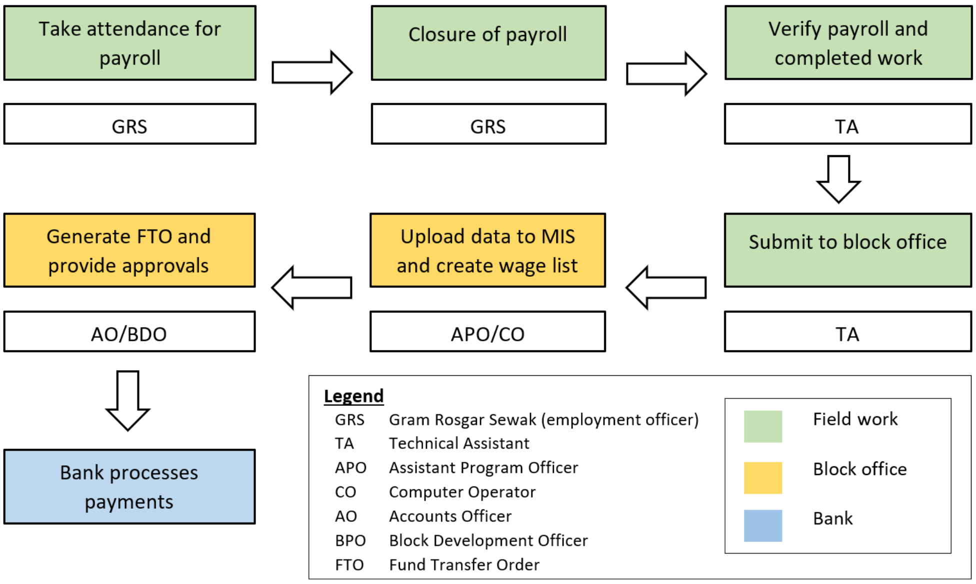


Figure 1.C2: Example of stages to the NREGS payment process. White boxes beneath steps present the officials responsible for implementing that step. FTO denotes fund transfer order. Gram Rozgar Sewak (GRS) is one of the key GP-level personnel and also responsible for allocating days of work. Source: 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 1.C2. 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.[36] According to the NREGS National Act, workers should be compensated 0.05% of unpaid wages for each day of delay in wage payment.[37]

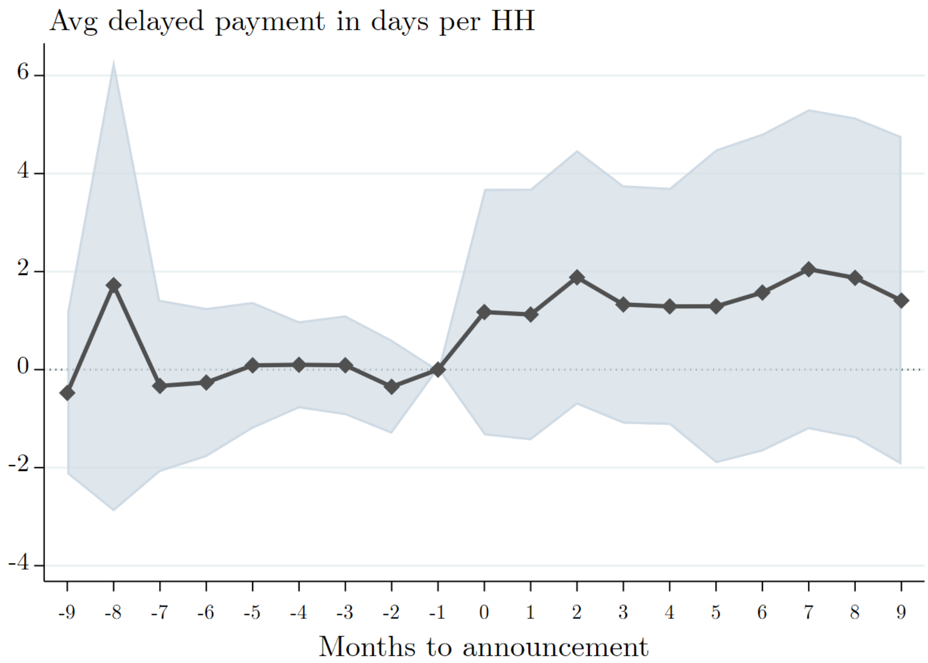


Figure 1.C3: Announcement event study on mean days of delay per household. This omitted category in this regression is the one month lead before the announcement and its mean is 1.9 days of delay per household. The p-value on a test of no pre-trends is 0.97. Standard errors are clustered by block.

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

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

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

delays reflect negatively on performance and thus costly for bureaucrats.

This section complements the analysis in Section 1.4.3 on whether delays are affected around the time of audit. Figure 1.C3 shows that prior to a month before the announcement, mean delayed payments in days per household do not follow a statistically distinguishable trend (p-value = 0.97). Furthermore, there are are also no statistically distinguishable effects following the announcement, although there is an average increase in delays post-announcement.

Figure 1.C4 plots the event study around month of audit and shows the audit does not affect delays in making wage payments.

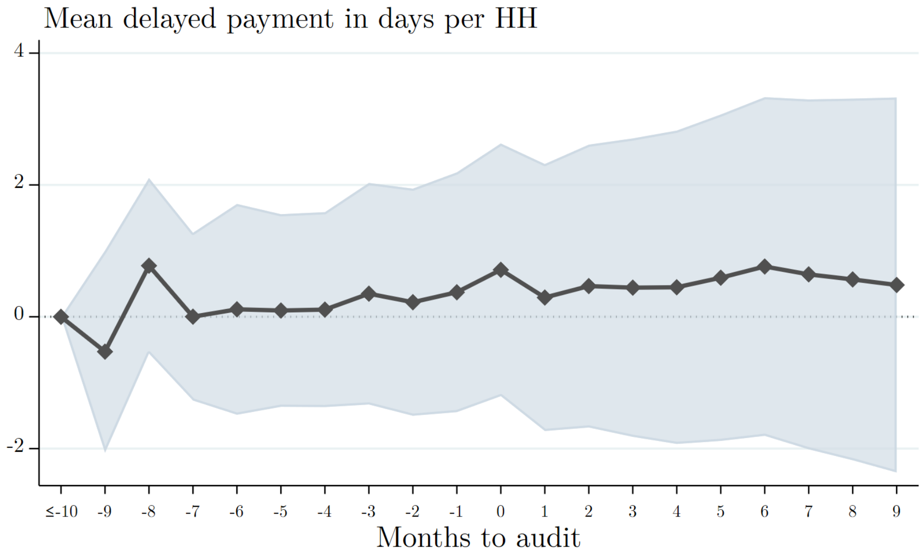


Figure 1.C4: Audit event study on delays in making wage payments. The omitted category is 10 or more months before the audit was conducted and its mean is 20 mean delayed payment in days per household. The regression includes GP and district month-year fixed effects as specified under Equation 1.3. Standard errors are clustered by block.

CHAPTER 2

OPTIMAL MONITORING AND BUREAUCRAT ADJUSTMENTS: THEORY AND WELFARE ANALYSIS

2.1 A Model of Information Design and Deterrence

This chapter presents a model of information design used to evaluate the empirical evidence documented about the NREGS monitoring policy in Jharkhand studied in Chapter 1. In the model, the Principal (e.g. audit agency or citizens) communicates information about the monitoring policy that allows the Bureaucrat to learn about the likelihood of being audited. This information affects the Bureaucrat's choice on expenditures to misappropriate. The Bureaucrat's choice affects payoffs to both the Principal and the Bureaucrat. Section 2.1.2 provides conditions for when the Principal would prefer to send informative signals, i.e. information that helps the Bureaucrat better predict when the audit will occur, and solves for the optimal signal. The relationship (or rate of change) of the Bureaucrat's choice with respect to their expectations of being audited is the object of interest. This relationship is a sufficient statistic for determining the Principal's optimal signal and for evaluating how alternative signals affect the Principal's welfare. This sufficient statistic is estimated in Section 2.2.

The following example illustrates the main intuitions of the model and provides conditions under which selecting bureaucrats for audit randomly with or without replacement is better.

Example (*To randomize with or without replacement?*)**.** Consider a Principal who is deciding between randomly selecting N Bureaucrats for audit with or without replacement over two periods. There is greater predictability of when an audit will occur when auditing without replacement because Bureaucrats observe who is audited and waiting-to-be-audited. The Principal is interested in deterring misappropriated expenditures by Bureaucrats. In every period, the Principal only has the capacity to conduct $M < N$ audits or share $q_0 = \frac{M}{N}$ 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. 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.

2.1.1 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 with N Bureaucrats and T periods under reasonable assumptions (see Appendix 2.A for details).[1] The Bureaucrat privately benefits from misappropriated expenditures, while the Principal is made worse off.

The Principal chooses a communication policy, π, that conveys to the Bureaucrat the likelihood of the audit. Formally, the Principal announces π, a chosen probability distribution over likelihoods of audit . 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.

1. This means we can think of the Bureaucrat's response given their expectations of being audited or given their expectations of being audited in a particular period as the same. In other words, the Bureaucrat's uncertainty over whether they are audited or when they will be audited mean the same thing in this model. The application of this assumption to the empirical setting is discussed further in Section 2.2.1.

The Principal is budget-constrained and can only audit M Bureaucrats, where $M < N$. The Bureaucrat knows this. If the Principal provides no information, each Bureaucrat is expected to be monitored with equal likelihood, given the Principal's capacity constraint. That is, the Bureaucrat has a prior about the likelihood of an audit, $q_0 = \frac{M}{N}$. Upon receiving a signal, the Bureaucrat forms the posterior belief about the likelihood of an audit, q, according to Bayes' rule. Using updated information, the Bureaucrat chooses an action $a \in [0, 1]$, 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 misappropriating a and $c(a)$ is the punishment for choosing a. 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. 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, i.e. $u'_a = -k < 0$. 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 given 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, $a^*(q)$.

Example (*A signal that concentrates incentives*)**.** For example, the Principal may consider a π that concentrates (a signal that provides information about the audit in advance) or disperses incentives (uninformative signal where everyone knows the prior). The prior, q_0, is the unconditional likelihood of being audited and determined by the capacity for audits announced every year. To implement a π that concentrates incentives at learning about an audit for sure or not, the Principal can choose a signal space where the Bureaucrat could draw signal $s \in \{H, L\}$: being selected for audit with high (H) or low (L) likelihood. The Principal can design π so that:

$$\pi(H|\text{Audit}) = 1 \quad \pi(H|\text{No Audit}) = 0$$

$$\pi(L|\text{Audit}) = 0 \quad \pi(L|\text{No Audit}) = 1$$

This design of π perfectly informs Bureaucrats in advance whether they will be audited or not. By Bayes' Rule, the share q_0 of Bureaucrats who drew signal H would have posterior beliefs of the likelihood of an audit of $q_H = 1$ and the remaining $1 - q_0$ who drew signal L would have posterior belief $q_L = 0$.

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.

2.1.2 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.[2]

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 $q_0 = \lambda_1 q_1 + ... + \lambda_n q_n$.[3] Define the largest payoff possible to the Principal as

2. 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.

3. 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 $U(q)$.

$\hat{U}(q_0) = \max\{C\}$. 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 (concentrating incentives) if and only if $U(q_0) < \hat{U}(q_0)$.

All proofs are in Appendix 2.A. This condition for sending an informative signal in Proposition 1 was derived in Kamenica and Gentzkow (2011) and reflects the theoretical findings in Lazear (2006) and Eeckhout et al. (2010).

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 of maintaining uncertainty. The following result further develops the intuition of Proposition 1.

Proposition 2. 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 concentrated over dispersed incentives for any q_0 if $U(q)$ is convex for all q. The Principal prefers dispersed incentives 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.

To construct the optimal signal, we can find the set of distributions of posterior beliefs which achieves payoff $\hat{U}(q_0)$. With the posterior distribution of interest, we can

$\{u|(q_0, u) \in co(U(q))\}$ 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 $\hat{U}(q_0) = \max\{u|(q_0, u) \in co(U(q))\}$.

backout a signal that induced it. 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.

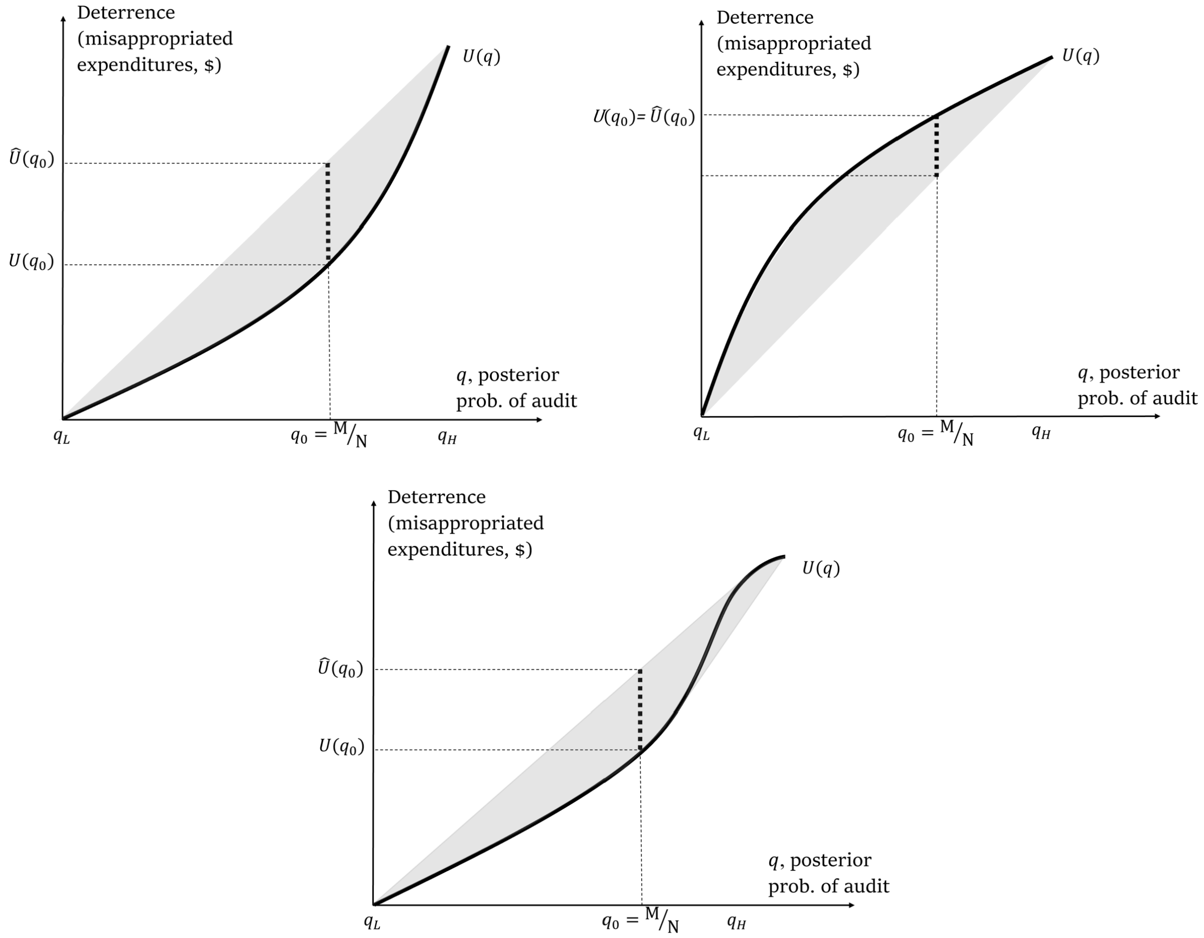


Figure 2.1: When does the Principal prefer concentrated versus dispersed incentives? These figures represent the Principal's utility as a function of the Bureaucrat's posterior beliefs. When the function is convex at q_0 (e.g. top left panel), the Principal prefers to concentrate incentives and communicate additional information about the probability of audit. When the function is concave for all q (e.g. top right panel), the Principal prefers to disperse incentives and 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 concentrate incentives if the function is also concave at q_0.

Figure 2.1 provides illustrations of when concentrated versus dispersed incentives are better. The bold dotted line represents the set C, the payoffs attainable by the Principal. The top-left panel of Figure 2.1 shows that $\hat{U}(q_0)$, i.e. concentrating incentives by telling 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. dispersed incentives when everyone believes they will be audited with probability q_0. To achieve utility $\hat{U}(q_0)$, our signal must place weight on posterior beliefs q_H and q_L such that $q_0 = Pr(H)q_H + (1 - Pr(H))q_L$. To solve for $Pr(H)$, we can then use Bayes' Rule to deduce the optimal signal by solving for $\pi(H|\text{Audit})$ and $\pi(L|\text{Audit})$. If $U(q)$ is concave for all q (e.g. top-right panel of Figure 2.1), a signal where posteriors equal the prior q_0 is optimal. Finally, if the function is convex only for some q but concave elsewhere (e.g. bottom panel of Figure 2.1), then the Principal prefers to concentrate incentives if the function is also concave at q_0.

Ultimately, the shape of $U(q)$ is an empirical question. It is a sufficient statistic 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 this sufficient statistic and apply the results from Proposition 2 to determine the best policy on communicating the likelihood of being audited (Section 2.2). I estimate this sufficient statistic without needing to specify additional primitives underlying the Principal's choice problem.

2.2 Information Design and Counterfactuals

This section provides conditions under which the estimates of anticipation (or interpretable as estimates of deterrence, given evidence in Section 1.4) and assumptions on Bureaucrats' beliefs can be used to estimate the sufficient statistic from the model in Section 2.1 and determine the optimal monitoring policy. Results show that designing monitoring policies which concentrate incentives at being audited with certainty or not will deter more corrupt behavior than policies that maintain uncertainty of the likelihood of an audit (Section 2.2.2). Section 2.2.3 analyzes the sensitivity of the conclusions on the optimal monitoring policy by relaxing assumptions. Section 2.2.4 estimates welfare consequences to the Principal under alternative signals.

2.2.1 Assumptions for estimating the sufficient statistic

The main results from the model in Section 2.1 tell us that we care about the shape of bureaucrats' deterrence given their expectations of being audited. This is a sufficient statistic for evaluating the Principal's welfare and determining the optimal design of information. Because of bureaucrats' changing expectations generated by being selected for audit without replacement, the sufficient statistic can be estimated in this empirical setting using the estimates of deterrence from Table 1.2 Panel B. We can do so without needing to specify additional underlying primitives that determine preferences and constraints of the choice problem.

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

First, comparing bureaucrats' anticipatory responses given differences in their expectations of being audited may be confounded with changing perceptions of the audit each year. Interpreting the differences in anticipatory and direct responses to be a result of differing expectations requires assuming: bureaucrat perceptions of audit quality remains unchanged across years. This assumption means that the estimated parameters for anticipatory and direct effects of the audit on total expenditures are being driven by varying expectations of being audited. And they are not confounded with perceived changes in quality or credibility of the audit each year.[4]

Empirical tests show it is reasonable to make this assumption (See Appendix 1.C). In particular, there are no significant differences in measures of audit quality or bureaucrat response to the stages of the monitoring policy across years. Furthermore, the test

4. 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.

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

for spillover effects in Appendix 1.C shows anticipatory and direct effects of the audit on total expenditures are unaffected after accounting for concentration of audits within one's block. This provides additional evidence that the anticipatory effects are driven by changes in expectations of being audited rather than perceptions about the audit to the extent they are informed by peer experiences. For robustness, this assumption is relaxed in a sensitivity analysis (Section 2.2.3) when determining the optimal policy in Section 2.2.2.

Second, I assume the estimated response of bureaucrats from Table 1.2 Panel B: the only payoff-relevant parameter for the bureaucrat for deciding today's action is their current expectation over the likelihood today's performance will be monitored. This has two implications. First, the Bureaucrat's response only depends on their current beliefs and not the policy that generated these current beliefs. That is, conditional on expectations of 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. Second, the Bureaucrat's response only depends on their current beliefs and not the history of actions or beliefs leading up to the current period. That is, the Bureaucrat's best response has a Markov property and dynamics of the Bureaucrat's response are completely explained by their beliefs.

Furthermore, the random assignment of expectations of being audited in the empirical setting parallels the model. That is, the signal structure chosen by the Principal randomly assigns signals to each receiver, so the difference in Bureaucrat response can only be explained by differences in posterior beliefs (holding all else equal).

Finally, assumptions need to be made on bureaucrats' beliefs. The deterrence estimate for each anticipatory group from Table 1.2 Panel B are based on varying expectations of being audited. To approximate Bureaucrat beliefs on the likelihood of an audit, I draw on the information Bureaucrats' received through audit agency announcements to infer next year's audit capacity and the remaining number of GPs to be selected for audit.

Table 2.1 presents the range of beliefs for each anticipatory group under the various assumptions. Assuming a range in beliefs over next year's audit capacity allows us to be more agnostic about Bureuacrats' expectations. Let $K_{\tau+1}$ denote next year's audit capacity where time τ is the current year. Assume that $K_{\tau+1}$ can be equal to any of the following: (i) $K_{\tau-1}$, last year's audit capacity; (ii) $\frac{1}{2}(K_{\tau-1} + K_{\tau})$, the average of last year's and this year's audit capacity; (iii) K_{τ}, this year's audit capacity; (iv) $\text{Trend}_{\tau} \times K_{\tau}$, this year's audit capacity multiplied by recent growth in capacity; and (v) $K_{\tau+1}$, beliefs about next year's audit capacity are ex-post consistent.

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

Table 2.1: Assumptions on Bureaucrats' beliefs on likelihood today's work will be audited.

Probabilities for *Anticipating1stAudit* - Year 1, *Anticipating1stAudit* - Year 2, *Anticipating2ndAudit* - Year 3 are calculated as follows: The denominator is the remaining number of GPs yet to be audited after learning who is selected for audit that year. The numerator is based on the assumption of future audit capacity ($K_{\tau+1}$) using information on past audit capacity. E.g. when $K_{\tau+1} = K_{\tau}$, expectations for *Anticipating*1*stAudit* - 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.[5] $Trend_{\tau} \times K_{\tau}$ assumes next year's capacity is based on the growth rate in audit capacity from last year to this year.

These assumptions affect assumed beliefs for the '*Anticipating*1*stAudit* - Year 1', '*Anticipating*1*stAudit* - Year 1', and '*Anticipating* 2*ndAudit* - Year 3' groups. Beliefs do not vary for 'Month of 1*stAudit*' 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 do not vary for '*Anticipating*2*ndAudit* - 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.

We will estimate the sufficient statistic under these various assumptions on beliefs from Table 2.1 and their associated estimates of deterrence from Table 1.2 Panel B. Put simply, the sufficient statistic is a graph of the beliefs ("x" values) against their associated estimates of deterrence ("y" values) from Table 1.2 Panel B.

2.2.2 The optimal design of information

Figure 2.2 plots the deterrence estimate for each beliefs assumption. Using the sufficient statistic from the model in Section 2.1, we can interpret the graph as a plot of the Principal's expected utility as a function of the Bureaucrats' beliefs about being audited. The deterrence estimates from Table 1.2 Panel B are relative to the reference group, '*Anticipating*1*stAudit* - Year 1', and reported as a share of the baseline mean. We can interpret this transformation as the relative decline in misappropriated expenditures

5. The special audit in FY2017-18 audited an additional 175 GPs for the first time. Likewise, 21 GPs from Wave 1 were selected for their second audit in Year 3, reportedly by accident. Under assumption $K_{\tau+1} = K_{\tau+1}$, Wave 1 would have anticipated this, hence their beliefs of 0.04. The special and duplicate audit GPs are accounted for when constructing expectations the denominator, but not included in the regressions as described in sample restrictions in Section 1.2.

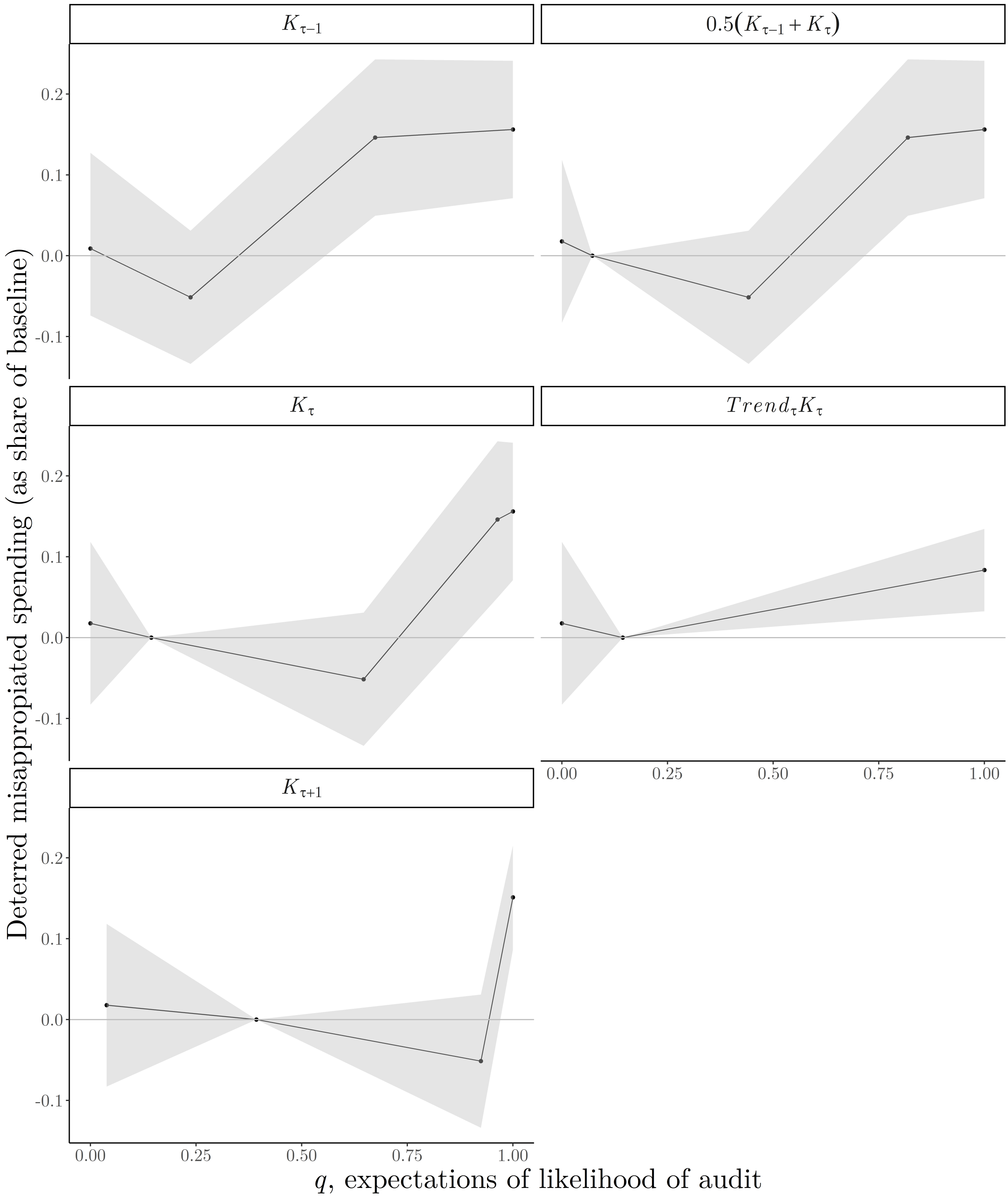


Figure 2.2: Principal's expected utility as a function of Bureaucrats' beliefs, under different beliefs assumptions. Black dotted line plots the mean of regression estimates with grey shaded area representing the 95% confidence interval for each parameter estimate. Under assumptions $K_{\tau+1} = \{K_{\tau-1}, Trend_\tau K_\tau, K_{\tau+1}\}$, some anticipatory groups are assumed to have the same beliefs. So, the mean of those groups' deterrence estimates and associated standard errors are plotted.

per month as a share of the reference group level of total expenditures. Higher levels of a decline in expenditures means more deterrence and higher expected utility for the Principal.

As illustrated in the model in Section 2.1, 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 communicating some fixed probability of audit for all Bureaucrats (better when concavities present), could the Principal do better be randomly allocating some to a higher probability group and others to a lower probability group (better when convexities present)?

If we considered the average deterrence response (black curves in Figure 2.2), then under all belief assumptions the curves are convex at q_0 (the observed range in capacity to conduct audits each year, 14-53%). In other words, the Principal's optimal monitoring policy is to communicate in advance whether one is selected for audit or not. The average elasticity under each assumption ranges from 0.07-0.14, where the elasticity for the increasing part of the graphs with more than 3 points of support range from 0.64-5.7. With a constrained budget, this implies that the optimal signal must place weight on communicating the event that some will be monitored with certainty.

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

2.2.3 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 the assumptions made on bureaucrat beliefs and constant perceptions of audit quality in Section 2.2.1. How wrong would our assumptions have to be in order for us to conclude that no information or a dispersed incentives policy would have been better? Technically, we want to know under what belief and deterrence parameters would the sufficient statistic be a line (i.e. Principal indifferent) or weakly/locally concave at q_0 (i.e. Principal prefers dispersed incentives)? Overall, the results provide greater certainty that the sufficient statistic is a convex curve as shown in Figure 2.2 and that the optimal design of information is one which concentrates incentives through the communicated likelihood of being audited.

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 a sufficient statistic with an alternative conclusion. The marginal sampling distribution of each estimated coefficient converges after bootstrapping about 10,000 draws. I bootstrap 100,000 samples to report results with 0.001% confidence.

First, I assess deviations from the mean of deterrence estimates and then relax the assumption on constant audit quality. This exercise assesses the likelihood we observe an alternative conclusion using the bootstrapped joint sampling distribution of the estimated coefficients (Column 4 of Table 1.2) . 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%.[6] If $K_{t+1} = K_{t-1}$, there's a 0.02% chance of realizing an alternative conclusion. If $K_{t+1} = K_{Trend_t} K_t$, there is a 30% chance of being indifferent between concentrated versus dispersed incentives. There is not enough information under this assumption to say whether this is a strict preference because there are only 3 points of support. If we exclude deterrence estimates capturing behavior during the month of audit ('Month of $1stAudit$'), the results are similar: under assumptions $K_{t+1} = \{\frac{1}{2}(K_{t-1}+K_t), K_t, K_{t+1}\}$ the likelihood is less than 0.005%; about 0.02% likelihood if $K_{t+1} = K_{t-1}$ and 34%

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

likelihood if $K_{t+1} = Trend_t K_t$.

Second, I relax the assumption that bureaucrat perceptions of audit quality are constant across years. As discussed in Section 1.4.1, bureaucrat responses are statistically indistinguishable across years for each variable that spans multiple years: $Post1stAnnounce$, 'Month of $1stAudt$', and 'After $1stAudit$'. This supports the assumption that perceptions of audit quality were constant across years. But, to rule out the possibility of a false negative, I relax this assumption.

If the estimated parameter across years is not truly 0, then perhaps we were not powered to detect the true effect size. Given the estimates for $Post1stAnnounce$, 'Month of $1stAudt$', and 'After $1stAudit$', 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). With these adjustments, it is still unlikely that dispersed incentives are optimal (1.6% under $K_{t+1} = K_{t-1}$; $<$ 0.1% under $\frac{1}{2}(K_{t-1} + K_t)$, K_t, K_{t+1}; 37.5% under $Trend_t K_t$). 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).

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 in order to simulate a sufficient statistic that is linear or concave. Among these perturbations, only 0.5% of the sample would yield a sufficient statistic that is linear or concave with simulated beliefs within the range of assumed beliefs for each anticipatory group (i.e. assumptions on beliefs from Table 2.1). Conservative estimates that exclude behavior estimated during the month of audit (coefficient captured by 'Month of $1stAudit$') yield a 1.2% likelihood of a sufficient statistic that is linear or concave with simulated beliefs within the range of assumed beliefs.

2.2.4 Counterfactual signals and welfare consequences

Using the sufficient statistic estimated in Section 2.2, this section estimates the welfare consequences from the perspective of the Principal if an alternative design of information of the monitoring policy were implemented over the course of the 27 months that the roll-out took place. Welfare is calculated for each policy using the deterrence estimates for each bootstrapped sample, which allows us to report the variance of the estimated welfare. Appendix 2.B provides further details.

The results show that spending under concentrated incentives (i.e. informing of an audit in advance) during this period would have been USD 225.6 million (s.e. = 12). Compared to the actual policy of randomization without replacement, concentrated incentives would have deterred 10% more misappropriated expenditures on average. Compared to a policy where incentives are dispersed and no information was provided (randomization occurred with replacement), concentrated incentives would have deterred 16% more misappropriated expenditures. This corresponds to a decline in misappropriated expenditures of about USD 22 million (p-value¡0.001) when comparing the concentrated incentives policy to the actual policy of randomizing without replacement; and USD 37 million (p-value¡0.001) when comparing the concentrated incentives policy to dispersed incentives policy of randomizing with replacement–all without changing the Principal's budget for audits.

If we worried about using the estimate of deterrence during the month of audit in the sufficient statistic and excluded it from the analysis, these conservative estimates show that the concentrated incentives policy (total expenditures of USD 227.6, s.e. = 12.3) would on average have deterred 8.7% and 15.4% more misappropriated expenditures compared to the actual and dispersed incentives policies, respectively. This is equivalent to a decline in misappropriated expenditures of USD 19.8 million and USD 35 million (p-values¡0.001) when comparing the concentrated incentives policy to actual and dispersed incentives policies, respectively.

These potential gains are substantial, especially given wide-prevailing audit standards that it is best to not inform clients of the auditing strategy to maintain unpredictability of when an audit may be conducted. 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 to the budget for conducting audits.

2.3 Conclusion

The monitor's resource constraints imply that a subset of bureaucrat activity will go unchecked. Monitoring policies designed to maximize deterrence must account for bureaucrat attempts to evade detection. Chapter 1 provides empirical evidence on strategic responses by bureaucrats to monitoring. Taking into account these strategic responses, Chapter 2 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.

Under these empirical circumstances, designing monitoring policies which inform bureaucrats in advance yields 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 sufficient statistic from a model of Bayesian persuasion that allows us to solve for the optimal signal and analyze welfare under counterfactuals. I estimate the sufficient statistic using reduced-form parameters and a set of flexible assumptions on bureaucrats' beliefs.

These chapters provide 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 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).

Chapter 2 Appendix

2.A A Model of Information Design and Deterrence

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 2.A 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 2.2.1 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.

Setup: Dynamic model with multiple receivers

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

The sender (or Principal) always wants as much of $X_t = \sum_i x_{it}$, the total program expenditures, to go towards realizing the goals of the workfare program as possible. The Principal uses monitoring as a policy to discipline Bureaucrats' behavior where significant penalties are imposed as punishment for being caught extracting private rents. However, the Principal is budget constrained and can only conduct $M < N$ 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 $\omega = \{\omega_{1t}, ..., \omega_{Nt}\}_{\forall t \in \mathcal{T}}$.

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.

Proofs

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

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.

$\square$

Proposition 2. If $U(\cdot)$ is convex at $U(q_0)$, then $U(q_0) < \hat{U}(q_0)$ and an informative signal is better. Moreover, the Principal prefers concentrated over dispersed incentives if $U(q)$ is convex for all q. The converse is true if $U(q)$ is concave for all q.

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 concentrated over dispersed incentives for any q_0 if $U(q)$ is convex for all q. The Principal prefers dispersed incentives for any q_0 when $U(q)$ is concave for all q.

(ii) An informative signal is still preferred if $U(\cdot)$ 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 to concentrate incentives 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 (dispersed incentives). 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 dispersed incentives and will send an uninformative signal where posterior beliefs equal the prior belief q_0. □

Motivating example in detail

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

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

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

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

$$2U(p)$$

This gives us:

$$U(p) \lesseqgtr \underbrace{pU(0) + (1-p)U\Big(\frac{p}{1-p}\Big)}_{=\tilde{U}}$$

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

Corollary 2.A1. When deciding between randomizing with (concentrated incentives) or without replacement (dispersed incentives), the Principal prefers randomization with replacement if $U(q)$ is concave for all q. Conversely, the Principal prefers randomization without replacement if $U(q)$ is convex for all q.

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

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

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

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

2.B Information Design and Counterfactuals

Sensitivity analyses

Table 2.B1 provides details on the results with assumed bias in Section 2.2.3. In this exercise, I relax the assumption on constant perceptions of audit quality and incorporated a bias in deterrence estimates.

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

Table 2.B1: Likelihood of alternative conclusion relaxing assumptions on constant perceptions of audit quality and assuming bias in estimates. The bias incorporated purposely increased the likelihood of drawing an alternative conclusion. So, for e.g., ±5 bias was incorporated for each anticipatory group such that the sufficient statistic was closer to being concave. The bias was modeled separately for each beliefs assumption. The probability of an alternative conclusion indicates the likelihood that the bootstrapped joint distribution of estimated coefficients formed a sufficient statistic that dispersed incentives was optimal (weakly/locally concave).

Welfare calculations

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

I analyze welfare for three policies: (1) the actual implemented policy of randomizing without replacement; (2) dispersed incentives where all GPs have the same expectations of being audited (equivalent to randomizing with replacement); and (3) concentrated incentives where all GPs are perfectly informed in advance of when they will be audited or not.

For each policy, we will examine total expenditures over the course of the 27 months that it took for the round of first audits to roll-out under the original monitoring policy. We will examine how total expenditures changes as beliefs evolve under counterfactual policies. Each policy will be assessed relative to a baseline capturing behavior prior to the implementation of the audits. The baseline is total annualized level of expenditures during the 27 months using the average of monthly expenditures during the 2 years prior to audits and including the anticipatory period in Year 1. This is the same baseline used in Table 1.2 and is the preferred baseline since all deterrence parameters are estimated relative to anticipatory behavior in Year 1. For each policy and the baseline, I exclude the year during which a GP experiences an audit to exclude behavior after one learns about their audit prior to receiving the audit and behavior during the months after the audit but still within the audit year. We can think of each calculation as excluding behavior during this period which may capture other phenomena of GPs responding to the policy not related to expectations of being audited, e.g. perceived salience of audits holding expectations fixed.

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

(1) Randomization without replacement (or p_1)

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

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 expenditures as a function of bureaucrats' posterior beliefs, q, on the likelihood of an audit. Under this policy, $U(q_{Year1})$ corresponds to the '*Anticipating*1*stAudit* - Year 1'; $U(q_{Year2})$ corresponds to '*Anticipating*1*stAudit* - Year 2'; $U(q'_{Year2})$ corresponds to '*Anticipating*2*ndAudit* - Year 2'; and $U(q'_{Year3})$ corresponds to '*Anticipating*2*ndAudit* - Year 3'.

(2) Dispersed incentives (or 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. For the main welfare analysis, I assume that $K_{\tau+1} = K_{\tau}$, i.e. GPs believe next year's audit capacity is equivalent to what they observe about this year's audit capacity.

(3) Concentrated incentives (or (p_3))

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

In the main analysis, I assume that deterrence under $U(1)$ is equivalent to behavior when the auditors are present (estimates provided by 'Month of $1stAudit$'. The reported conservative estimates assume that deterrence under $U(1)$ is equivalent to behavior in Year 3 when Waves 1 and 2 believe with very high likelihood they will be audited in Year 4 (estimates provided by '$Anticipating2ndAudit$ - Year 3'. These are conservative estimates because first, they address concerns that some other behavioral phenomena may be driving the response when auditors are present. Second, under some beliefs assumptions about tomorrow's audit capacity (see Table 2.1), the expectations of being audited captured by '$Anticipating2ndAudit$ - Year 3' can be less than 1. This makes the estimate conservative because the deterrence response from bureaucrats under $U(1)$ could be greater that what is estimated with '$Anticipating2ndAudit$ - Year 3'.

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

CHAPTER 3

PRODUCTIVITY ALONG THE POLICE HIERARCHY

This project is joint work with Bocar Ba and Roman Rivera.

3.1 Introduction

Hierarchies structure the production of output in organizations. They structure the way organizations approach the task, decision, or problem at hand by dividing tasks and directing flows of information among members of the organization (Garicano and Van Zandt, 2012). Organizational structures matter for output. Firm organizational structure, as captured by the degree of vertical integration, for example, is an important driver of productivity (Syverson, 2011).

Hierarchies define the layers of management within the organization, where managers acquire and process information from subordinates. Empirical work showing that managers affect the delivery of public services and performance in firms (e.g. Bertrand and Schoar (2003), Bloom et al. (2015), Tsai et al. (2015), and Rasul and Rogger (2018)) suggest that hierarchical management structures matter for productivity, too. Yet, little remains known empirically about how the structure of layers of management within a hierarchy affect productivity in organizations. Theoretical work shows when some hierarchical structures may be efficient (e.g. Radner et al. (1992), Garicano (2000), and Hart and Moore (2005)), implying that understanding the implications of the design of a hierarchy is of practical importance.

This paper focuses on how hierarchies affect the performance of public institutions in an empirical setting. We study a police agency and the performance of the agency's geographical units called districts. We estimate the effect of layers of management on performance within the district. Our empirical strategy leverages switches of managers across geographic units within the police agency to identify individual manager fixed effects by rank. We use the estimates of the individual manager-rank fixed effects to understand how the structure of the hierarchy affects police performance. To do so, we carry out a decomposition of the dispersion in productivity of district performance explained by managers across different ranks.

Our analysis uses longitudinal employee data of a police agency linked to district time-varying outcomes to map a team within a district to the district's performance. The employee data allows us to track managers of different ranks along the chain of command. We focus on managers within the police district which includes commanders, lieutenants, and sergeants and excludes police officers. Reported crime and arrests are the two main outcomes of police performance.[1] These outcomes are disaggregated into violent, property, and non-index crimes.[2]

For the variance decomposition of productivity dispersion we estimate three components that capture features of the hierarchy. The first component (heterogeneity) captures the heterogeneous effect of officers in the various managerial positions along the hierarchy. These are positions which, by intent and design, determine the division of tasks and flow of information. This effect captures the heterogeneity across individuals by rank who have taken up these positions, and its contribution to the variation in district productivity.

The second component (matching across ranks) captures idiosyncrasies in the matching of managers of different ranks working together along the hierarchy in a district. For instance, idiosyncrasies in matches of managers along the hierarchy may lead to more or less frictions in the flow of information. These frictions may depend on how well the managers worked together. Relatedly, the third component (matching within ranks) captures idiosyncrasies in the matching of managers of the same rank working together in a

1. We treat levels of crime as the objective that police agencies work to minimize, and proxy levels of crime with reported crime as one of the main outcomes. We study arrests as one of the main outputs of the police agency, recognizing that there are a host of other tasks implemented by the police agency that work to minimize crime. For instance, while we tend to believe that crime tends to decrease with arrests, an officer that has made fewer arrests in conjunction with other crime-reducing tasks may have a greater effect on decreasing crime than an officer that has made more arrests. This is the reason why we also study reported crime as an outcome measure of police performance. We discuss this in more detail in Section 3.3.1.

2. Non-index crimes, as categorized by the FBI, include offenses of the following type: drug, traffic, prostitution, fraud, weapons, loitering, and vandalism.

district.

We find that all three components explain a substantial portion of the dispersion in productivity for reported crime and arrests for violent, property, and non-index crimes. These components collectively explain 7-23% of dispersion for violent crimes and arrests, 32-35% of dispersion for property crimes and arrests, and 44-50% of the dispersion for non-index crimes and arrests. The matching across ranks component explains more of the variation than either of the heterogeneity and matching within ranks components, but the difference is marginal. We also find that these three components explain more of the dispersion in productivity for property and non-index crimes than violent crime. This seems reasonable because violent crime can be more idiosyncratic and harder to predict or manage. We also find that these components explain more of the dispersion in productivity for reported violent and non-index crime than compared to arrests for violent and non-index crime.

These findings speak to how hierarchical structure affects police agency performance. Was it the careful design of positions that allowed the agency to perform better? Or is it that good decisions have been made on assigning teams within- and across-ranks within the district? Our findings show that these managerial positions play an important role in explaining the performance of the police agency. They also show that decisions managers have authority over, which affect who gets promoted (as captured by the heterogeneity component) and who works with whom (as captured by the matching components) have an important affect on performance as well.

Contributions to the literature

This paper contributes to two strands of literature.

The first focuses on the value of managers to organizational productivity. Bloom et al. (2015), Tsai et al. (2015), and Rasul and Rogger (2018) collect detailed qualitative data to show how aspects of managerial practices lead to better or worse outcomes. Bertrand and Schoar (2003) and Fenizia (2020) take a more reduced-form approach and estimate the fixed effects of chief executives in firms and office managers in government, without the level of detail on management practices explored in the preceding set of papers. Another set of papers within this literature focuses on quantifying the informational advantage of managers. Rogger and Somani (2018) find that managers errors' are reduced with evidence briefings. Dal Bó et al. (2021) find that middle managers have a significant informational advantage in disseminating agricultural information through extension workers.

This paper contributes to this literature by estimating the fixed effect of managers by rank across the hierarchy. The previous literature focuses on the effect of a single layer of management. Our estimation of manager-rank fixed effects allows us to understand how managers of different ranks drive unexplained productivity in the police agency's performance outcomes. This paper also studies manager effects on productivity in the setting of a police agency, a context where organizational drivers of output and agency objectives are difficult to study. The difficulty with studying productivity in a public agency is that organizational objectives and the translation of organizational outputs into their objectives are often hard to measure or define. For this reason, we focus two main outcomes: reported crime as a measure of the agency's objectives and arrests as a measure of the agency's output.

Our paper also contributes to the literature on crime policy.[3] Our study is most closely related to recent research focused on the impact of managerial directives on police enforcement and crime (Mas, 2006; Mummolo, 2018; Rivera and Ba, 2019). Furthermore, the detection of non-index crimes is sensitive to managerial policy choices such as broken-windows policing or stop, question, and frisk (Kelling and Wilson, 1982). This places importance on understanding the impact of managerial practices on non-index offenses. Unlike index crimes that tend to rely on victims' reporting, reporting of non-index crimes is more reliant on officers' discretion in enforcing the law (Lum and Nagin, 2017). In

3. Most of the economics literature has focused on the responsiveness of crime to the quantity of policing via the number of police officers (Levitt, 1997; McCrary, 2002; Evans and Owens, 2007; Mello, 2019) or police tactics (Eeckhout et al., 2010; Adda et al., 2014; Mastrobuoni, 2020).

practice, most arrests are for non-index offenses, and crimes cleared by arrests tend to be higher for these less-severe offenses. This paper estimates the impact of managers at varying ranks on violent, property, and non-index crimes. It allows us to compare the impact of managers across these types of crime. To the best of our knowledge, the variance decomposition provides the first empirical estimates of the ways in which the chain of command drive productivity dispersion in reported crime and arrests for violent, property, and non-index offenses.

3.2 Why should the hierarchy affect police agency performance?

Police departments are designed to "protect the lives, property, and rights of all people"; maintain order; and enforce the law.[4] Execution of this mission requires neighborhood-level monitoring to deter, incapacitate, and report crime. To do so, police departments organize a chain of command, or hierarchy, that divides the tasks and directs the flow of information among the workforce.

In the Chicago Police Department (CPD), there are currently 22 geographic district offices under the Bureau of Patrol.[5] This paper focuses on the Chicago Police Department personnel working in these districts. Each district is headed by a commander. Lieutenants are subordinate to commanders, sergeants are subordinate to lieutenants, and police officers are subordinate to sergeants.

The commander of a district provides high-level oversight and ensures that their district meets the goals of the police agency.[6] They aggregate local information provided by subordinates to effectively manage operations at the district level. Their responsibilities includes monitoring the performance of supervisory personnel through the chain of command, allocating the talent of subordinates effectively, disseminating information on

4. Chicago Police Directives: Vision, Mission Statement, and Core Values

5. Before 2012, there were 25 geographic district offices due to the closure of three district offices.

6. Chicago Police Directives: District Commander

crime-related problems, and securing training needs.

The main tasks of district lieutenants include overseeing sergeant supervision of patrols, directing responses to incidents in the field, monitoring radio dispatches, providing continuity of operations across patrol shifts, ensuring compliance with policy and procedures for tasks like processing arrests and using body worn cameras.[7]

The main tasks of district sergeants include designating the assignment of police officers to their patrols, training and mentoring police officers on their duties and responsibilities, respond to field incidents, conducting roll calls, monitor radio dispatches, reviewing compliance with policy and procedures, ensuring proper maintenance of equipment, and completing personnel evaluations.[8]

From lieutenants to sergeants, these managers have more specialized knowledge of criminal activity and how it is being addressed. They provide more direct oversight and make decisions on local activity within their district.

By design, these managerial positions along the chain of command have direct effects on the performance of a police district in achieving their mission. For example, a commander can misallocate talent of managers, a lieutenant's loose supervision can breed weak performance of officers, a sergeant can designate patrol assignments to officers that lack preparation for a particular locality.

It may also be that managers in these positions along the hierarchy perform better collectively when there are good working relationships between superiors and subordinates. Good working relationships can mitigate information frictions. For example, frictions related to the ease of communicating across ranks driven by levels of trust or comfort. This can lead to greater accuracy of information flowing through the hierarchy. Good working relationships can also mean that supervisors delegate tasks more effectively. For example, supervisors who know their subordinates well may be more likely to delegate tasks in a way that limits agency issues.

7. Chicago Police Directives: Watch Operations Lieutenant and District Field and Tactical Lieutenants

8. Chicago Police Directives: District Station Supervisor and District Field Sergeants

In this paper, we aim to capture the potential ways in which managers along the chain of command can affect district performance individually through the design of their positions and as a team working together along the hierarchy.

3.3 Empirical Strategy

This section starts by introducing a conceptual framework underlying our estimation approach. It then discusses the identification of our estimation and our approach for the variance decomposition of the productivity dispersion in performance outcomes.

3.3.1 Conceptual Framework

Consider the following production function of the police agency:

$$Y = AF(K, L)$$

where Y is the agency's output resulting from a combination of total factor productivity (TFP), A, a factor-neutral shifter in output (e.g. talent and effort of managers and individuals, organizational structures and processes, technological innovations); and capital (K) and labor (L) inputs. We can think of $Y = (Y_1, Y_2, .., Y_m)$ as a vector of outputs that directly affect crime. This vector capture the tasks officers work on related to law enforcement and maintenance of community safety with the goal of minimizing crime. These tasks help avert and incapacitate crime. They include patrolling a beat, responding to radio calls, and making arrests and stops. The police agency's objective is to minimize crime, subject to a production technology which we assume is Cobb-Douglas:

$$\min\ C(Y) \quad \text{subject to } Y = AK^{\beta}L^{\gamma}$$

where C denotes true levels of crime, a function of the agency's output Y.

There are several difficulties with measuring police objectives and linking their output (Y) to their objectives. First, true levels of crime are unobserved. The best observable proxy for true levels of crime is reported crime. The quality of reported crime as a proxy will depend on the crime and the likelihood that any incident of that type of crime is reported. So, this makes it difficult to assess how outputs of a police agency achieve their intended goal of minimizing crime.

Second, police agencies can affect levels of crime (through Y) by incapacitating attempts at criminal activity and deterring potential criminal activity through the threat of law enforcement. The former is observable, but the latter, deterred crime that might have happened, is fundamentally unobserved. This presents another difficulty in measuring how policing efforts may affect crime.

These measurement challenges in combination with multiple outputs contributing to decreased crime make it difficult to understand how the agency's output affects crime. To illustrate, suppose that Y_1 represents arrests and that $\frac{\partial C}{\partial Y_1} \leq 0$. $\frac{\partial C}{\partial Y_1}$ can be heterogeneous across officers or districts. That is, higher numbers of arrests in one district may not imply greater declines in crime than the other district. This implies that some districts, with fewer arrests per period, can have a greater effect on decreasing crime than others because reported crime does not capture unreported or averted crime, and activities other than arrests may affect levels of crime.

In this paper, we are interested in how the structure of the hierarchy, captured by the effect of managers across ranks, affects police performance. We focus on two outcome measures of performance varying by district-month: reported crime and arrests. Arrests capture a direct output by the police agency.[9] However, for the reasons described above, arrests do not provide a complete picture of how the agency's output may affect their objective, to minimize crime. This is the reason why we also study the effect of managers on reported crime.

3.3.2 Estimation and Identification

For estimation, we start with the following:

9. Future directions for research will include data on stops made by officers during patrols.

$$Y_{dt} = A_{dt}(K_{dt})^{\beta} L_{dt}^{\gamma}$$

$$C_{dt} = A_{dt}(K_{dt})^{\beta} L_{dt}^{\gamma}$$

where d indexes geographical district within the Chicago Police Department and t indexes the month.

We assume capital allocated per officer is fixed. Officers are allocated the same resources, e.g. car, weapons, uniform, etc. So, let K_{dt} / $L_{dt} = k_t$ where k_t is a time-varying constant fixed across districts that captures physical capital per officer. We also assume that the number of officers allocated per district is commensurate with the district's size or demands given the criminal environment. This implies that the workload per officer in any district is on par. Then our equations become:

$$Y_{dt} = A_{dt}(k_t L_{dt})^{\beta} L_{dt}^{\gamma} = A_{dt} k_t^{\beta} L_{dt}^{\beta+\gamma}$$

$$\ln Y_{dt} = \ln A_{dt} + \beta \ln k_t + (\beta + \gamma) \ln L_{dt}$$

where $\ln A_{dt} = \alpha_{ir(d,t)} + \alpha_d + \alpha_t + D_{dt} + \epsilon_{dt}$ captures TFP. TFP is decomposed into α_d, a district fixed effect; α_t, a monthly fixed effect; D_{dt} is a vector of time-varying variables on district demographic characteristics (population and racial composition), which are strong predictors of levels of crime within a geographic area; $\alpha_{ir(d,t)}$ captures an officer i fixed effect disaggregated by their rank r within CPD over the course of their career, where (d, t) maps the district-months to the officer-rank[10]; ϵ_{dt} captures idiosyncrasies to district-level performance over time. We estimate the police agency's production function with the following OLS specification:

10. The mapping to (d, t) will become useful for the variance decomposition to help determine when officers worked together in the same district

$$\ln Y_{dt} = \alpha_{ir(d,t)} + \alpha_d + \underbrace{\alpha_t + \beta \ln k_t}_{\eta_t} + \underbrace{D_{dt} + (\beta + \gamma) \ln L_{dt}}_{X'_{dt}\delta} + \epsilon_{dt}$$

$$= \alpha_{ir(d,t)} + \alpha_d + \eta_t + X'_{dt}\delta + \epsilon_{dt} \tag{3.2}$$

And, correspondingly, using crime as an outcome, we estimate:

$$\ln C_{dt} = \alpha_{ir(d,t)} + \alpha_d + \eta_t + X'_{dt}\delta + \epsilon_{dt}$$

Our analysis focuses on officers who have a management role. These managers are above the "police officer" rank and are Sergeants, Lieutenants, and Commanders of a police district. Practically, we exclude the "police officer" rank from analysis to avoid over-fitting because there are over 9,000 employees distributed across 25 districts observed over 108 months.

The parameter of interest, $\alpha_{ir(d,t)}$, helps us understand how managers by rank across the hierarchy affect district performance. We estimate this parameter, even though our outcomes measure team output at the district-month level, following Bonhomme (2021). All standard errors are clustered at the district level.

Identification of officer-rank fixed effects comes from the reassignment of officers across districts over time. Without movement of officers across districts, we would not be able to disentangle time-invariant effects of officers from districts on district-level police performance (Abowd et al., 1999). Furthermore, since there are only 22 districts within CPD, all 22 districts are included in the connected set through officer switches over time.[11]

Our approach provides an unbiased estimate of officer-rank and district fixed effects if officers do not sort across districts based on officer-district match effects or transitory

11. Following Abowd et al. (1999), the connected set represents districts linked by officers officers who have worked in these districts. Fixed effects estimated among the connected set are comparable; whereas, fixed effects across two sets of districts that are not connected are not directly comparable because they are normalized differently.

shocks at the district level.

$$\mathbb{E}[\epsilon_{dt}\alpha_{ir}|X_{dt}, \alpha_d, \alpha_t] = 0 \tag{3.3}$$

In other words, identification requires that the time-varying residual component of the performance measure must be uncorrelated with officers' reassignment across districts. This means that sorting based on permanent characteristics of either the district or officer does not violate the assumption. For example, this assumption allows sorting of senior officers to categorically safe districts.

In this setting, this is a reasonable assumption because transfers of officers across districts are facilitated by a bidding process, which is generally based on seniority (Rim et al., 2021). To check for violations of the identification assumption, we test for correlations between pre-determined district characteristics that may be varying across time and future manager fixed effects.

3.3.3 Variance Decomposition

We decompose the total variation in the performance measure into variation explained by the officer-rank fixed effect. We begin by residualizing the data from other variables in Equation 3.2, including time-varying district characteristics and year-month fixed effects.

$$\tilde{Y}_{dt} = Y_{dt} - (\alpha_d + \eta_t + X'_{dt}\delta) \tag{3.4}$$

We then estimate the effect of officer-rank fixed effects on the remaining dispersion in productivity:

$$\tilde{Y}_{dt} = \tilde{\alpha}_{ir(d,t)} + \epsilon_{dt}$$

where $\tilde{\alpha}_{ir(d,t)}$ denotes the manager-rank fixed effects residualized of the same variables. We then implement the following variance decomposition on the remaining dispersion in productivity:

$$
\begin{aligned}
\mathrm{Var}(\tilde{Y}_{dt}) = \quad & \underbrace{\sum_i \sum_r \mathrm{Var}(\tilde{\alpha}_{ir})}_{\text{Heterogeneity in ranks component}} \quad + \underbrace{2 \sum_{i,r} \sum_{i' \neq i, r' \neq r} \mathrm{Cov}(\tilde{\alpha}_{ir}, \tilde{\alpha}_{i'r'})}_{\text{Matching component, across ranks}} \\
& + \underbrace{2 \sum_{i,r} \sum_{i' \neq i, r' = r} \mathrm{Cov}(\tilde{\alpha}_{ir}, \tilde{\alpha}_{i'r'})}_{\text{Matching component, within rank}} + \underbrace{\mathrm{Var}(\epsilon_{dt})}_{\text{Residual component}} \qquad (3.5)
\end{aligned}
$$

The first term captures heterogeneity among officers by their rank in their ability to affect police performance. This term also captures the extent to which various ranks, by the nature and design of these positions, affect police performance.

The second term captures the covariance between officers across different ranks. We weight the covariance calculation by number of months officers (i, r) and (i', r') have worked together within a district at the same time. The weighted covariance captures the share of unexplained productivity dispersion attributable to the matching of individuals working together along the hierarchy.[12]

The third term captures the covariance between officers within the same rank. We weight the covariance calculation by number of months officers (i, r) and (i', r) have worked together within a district at the same time. The weighted covariance captures the share of unexplained productivity dispersion attributable to the matches of individuals working together as peers of the same rank. Finally, the fourth term captures the residual variance due to, for example, transitory shocks.

Estimates of the variance-covariance components may be biased due to limited mobility of officers across districts. We provide bias-corrected estimates following the procedure in Andrews et al. (2008) and Bonhomme (2021).

12. Future research prospects using beat-level data would allow us to estimate time-varying officer-rank fixed effects to compute the matching components of the covariance terms corresponding to the same periods in time that coworkers worked together.

3.4 Data and Summary Statistics

3.4.1 Data sources

For analysis, we combine CPD administrative data of police officers and their district assignments with district time-varying data on levels of crime and arrests. The administrative data was collected in collaboration with the Invisible Institute. These data were obtained by Freedom of Information Act (FOIA) requests from CPD and the Chicago Department of Human Resources, and through court-ordered releases.

Our administrative data on police officers span 2007-2015. Officer demographic data include officer race, sex, birth year, and appointment date. Officer human resources data contain information on the officer's rank and district assignments over time. We focus our analysis on officers assigned to CPD's 22 geographic districts managed by the Bureau of Patrol. We also restrict attention to managers at the district-level, i.e. sergeants, lieutenants, and commanders. We do not consider the role of police officers in the current analysis. This leaves us with 64,874 officer-district-month level observations for analysis

Data from the US Census and the American Community Survey are merged at the district-year level. District demographics from these surveys, such as population and racial composition, are used as control variables in our analysis.

Our main outcome measures to measure performance of the district managerial team are violent, property, and non-index crimes and arrests at the district-month level. These data come from the arrests dataset. Reported crime and arrests at the district-month level are counted per 100,000.

3.4.2 Summary Statistics

The mean population of a CPD district is 116,525, where on average 42% of the district population is Black, 29% White, 22% Hispanic, and 7% other. In a typical month, districts on average experience about 106, 395, and 741 incidents per 100,000 of violent, property, and non-index crime, respectively. In a typical month, districts on average make 19, 40, and 297 arrests per 100,000 for incidents of violent, property, and non-index crime, respectively.

During 2007-15, there are on average 7,390 sworn officers per year under the Bureau of Patrol. The managers we study, make up on average about 9% of the workforce per year. Table 3.1 describes additional summary statistics of officers by their manager rank. Managers tend to have at least 20 years of experience, with 2-4 years average difference in years of experience between Sergeants, Lieutenants, and Commanders. Managers are also majority white and male.[13] And while Black officers make up 8-13% of Sergeants and Lieutenants, they make up 35% of Commanders across districts.[14] Reassignment of managers across districts are infrequent on average, occurring up to 1 time on average during the study period. However, we do observe some managers being reassigned at least 3 times during the study period.

	Manager rank		
	Sergeant	Lieutenant	Commander
Number of officers	985	121	28
Mean age	46	49	50
Mean tenure (years)	20	24	26
Mean reassignments across districts	0.32	1	0.68
Max reassignments across districts	4	8	3
Share Female	0.16	0.17	0.085
Share Black	0.13	0.078	0.35
Share White	0.71	0.84	0.55
Share Hispanic	0.14	0.073	0.096

Table 3.1: Officer summary statistics by manager rank. Statistics for officers in a managerial position who were employed from 2007-2015.

13. As a share of the total police workforce, white officers make up 50% of the workforce on average.

14. As a share of the total police workforce, Black officers make up 25% of the workforce on average.

3.4.3 Tests for violations of sorting

We test for evidence of endogenous mobility of officers across districts that may violate our sorting assumption (Equation 3.3). We are concerned that officers may sort based on district-specific trends. For instance, officers attempt to be reassigned to certain districts based on the district's projected trends in crime over time.

One way to test for this is to check for correlations between pre-determined, monthly-varying district characteristics on the average estimated fixed effect for a future group of managers. The future group of managers is defined by an impending change in managerial team composition in the next month, so in these regressions we focus on periods of time prior to officer reassignments across districts. In our regressions, we take lagged district characteristics such as reported violent, property, and non-index crimes, as well as the size of the district workforce. Our lagged variables include the mean of the previous 4 months ('LagMean') and the growth rate from the previous 4 months ('LagGrowth').

	Future manager FE on reported crime:		
	Violent	Property	Non-index
	(1)	(2)	(3)
LagMean_Violent	−0.004	0.002	−0.006
	(0.004)	(0.003)	(0.004)
LagMean_Property	0.0001	0.001***	−0.0001
	(0.001)	(0.0002)	(0.001)
LagMean_NonIndex	0.001	−0.001	0.001
	(0.001)	(0.0005)	(0.001)
LagMean_Labor	−0.00001	−0.00003*	−0.00000
	(0.00003)	(0.00002)	(0.00003)
LagGrowth_Violent	−0.012	−0.017	−0.008
	(0.014)	(0.012)	(0.014)
LagGrowth_Property	−0.002	0.002	0.0005
	(0.004)	(0.003)	(0.003)
LagGrowth_NonIndex	0.002	0.0001	−0.00002
	(0.003)	(0.002)	(0.003)
LagGrowth_Labor	0.00000	−0.0002	0.0001
	(0.0002)	(0.0002)	(0.0001)
Observations	2,447	2,447	2,447
Adjusted R^2	0.508	0.618	0.425

Table 3.2: Tests for sorting. This table estimates the effect of lagged, pre-determined, monthly-varying district characteristics on the average estimated fixed effect for a future group of managers. The future group of managers is defined by an impending change in managerial team composition in the next month. We use the fixed effects estimated using reported violent, property, and non-index crimes as the outcome. The 'LagMean' variables take the mean of the lagged reported violent, property, and non-index crime, as well as total number of officers in the district (captured by 'Labor') in the previous 4 months. The 'LagGrowth' variables capture the growth rate in the previous 4 months for the same variables. All regressions are estimated interacting the lag variables with whether a reassignment happens among the managerial team and include district and year-month fixed effects. The reported lag variables are those interacted with the occurrence of an impending reassignment. Standard errors are clustered at the district level.

Table 3.2 presents the results of the regressions for the average future manager fixed effect estimated for the three outcome variables on reported crime. We do not observe evidence of sorting for violent and non-index crimes. But there is potential sorting on reported property crime, where LagMean for property crime and total number of officers significantly predict changes in the future manager fixed effect on property crime. The predicted changes from potential sorting range from 16-99% of a standard deviation of the future manager fixed effect, for the mean to max of LagMean for property crime. And 38-55% of a standard deviation of the future manager fixed effect, for the mean to max of LagMean for total officers.

One way to address the potential bias from temporal sorting (discussed in Section 3.6.2) is to leverage the switches of officers across districts as quasi-experimental variation to use a differences-in-differences estimation of the manager fixed effect. A test for parallel pre-trends to support the differences-in-differences strategy would also provide evidence of the existence of sorting. The estimated effect can be used to characterize the potential bias by comparing it with the fixed effects estimation we outline (Section 3.3.2) and implement.

3.5 Productivity along the Police Agency Hierarchy

In this section we start by documenting the residual dispersion in productivity as we progressively add the regressors in Equation 3.2. Then, given the estimates of the manager-rank fixed effects, we proceed with the results from the variance decomposition described in Section 3.5.2.

3.5.1 Dispersion in Productivity

Manager-rank fixed effects explain a considerable portion of the remaining dispersion in productivity measured by reported crime and arrests. This is after discounting the variation in outcomes explained by environmental variables that predict a district's potential for criminal activity, unrelated to the resources committed by the police agency. To assess the amount of dispersion in police productivity attributed to the hierarchical structure, we compare the adjusted R^2 of models nested in Equation 3.2. Model 1 includes district demographic and district fixed effect variables (environment variables). Model 2 additionally includes year-month fixed effects and logged total number of officers (base variables). Model 3 additionally includes our explanatory variables of interest, manager-rank fixed effects variables.

When we incorporate manager-rank fixed effects, the adjusted R-squared increases by 0.1-0.9%, depending on the outcome. The changes in explanatory power when incorporating manager-rank fixed effects are small at face value. However, a substantial amount of the variation on reported crime and arrests is explained by characteristics of a district location that make it more or less prone to criminal activity. District population, racial composition, and district fixed effects explain about 76-90% of the variation in outcomes on reported crime and arrests.

We consider the amount of residual dispersion explained by manager-rank fixed effects net of the variation explained by variables capturing the district environment, $(R_3^2 - R_2^2)/(1 - R_1^2)$. Manager-rank fixed effects explain 0.5-5% of the remaining dispersion. This is similar in approach and magnitude to what Bertrand and Schoar (2003) and Fenizia (2020) find for managers of firms and government welfare agencies, respectively. And if we examine the remaining dispersion net of the additional variation captured by the base variables, then manager-rank fixed effects explain about 2.5-15% of the remaining dispersion.

	District-level monthly outcomes					
	(Log) Reported crime			(Log) Arrests		
	Violent	Property	Non-index	Violent	Property	Non-index
Model 1: Environment						
R_1^2	0.89	0.80	0.90	0.76	0.79	0.89
Model 2: Environment + Baseline						
R_2^2	0.957	0.948	0.973	0.830	0.847	0.951
Model 3: Environment + Baseline + Manager-rank						
R_3^2	0.960	0.952	0.977	0.834	0.858	0.957
Residual dispersion explained by Manager-rank						
$R_3^2 - R_2^2$	0.002	0.003	0.004	0.001	0.009	0.005
$(R_3^2 - R_2^2)/(1 - R_1^2)$	0.019	0.018	0.038	0.005	0.043	0.049
$(R_3^2 - R_2^2)/(1 - R_2^2)$	0.064	0.086	0.150	0.025	0.075	0.124

Table 3.3: Dispersion in productivity at the police district measured by reported crimes and arrests outcomes. Outcomes are measured at the district-month level and measured per 100,000 population within the district. Model statistics are reported for 3 models each with successive inclusion of the explanatory variables in Equation 3.2. Model 1 includes district demographic and district fixed effect variables (Environment variables). Model 2 additionally includes year-month fixed effects and logged total number of officers (Baseline specification variables). Model 3 additionally includes our explanatory variables of interest, manager-rank fixed effects variables. All reported R^2 are adjusted for the number of regressors in the model.

From here, we are interested in explaining what drives the remaining, unexplained dispersion in productivity. The goal is to understand how the hierarchical structure affects the residual dispersion in outcomes. Particularly, how do ranks within the hierarchy or matches of managers working together along the hierarchy affect productivity?

3.5.2 Variance Decomposition

The decomposition of the remaining dispersion in productivity into the three components described in Equation 3.5 is described in Table 3.4. After residualizing the data from other variables in Equation 3.4, we find that the variation from all three components driven by manager-rank fixed effects explain 7-51%, a substantial portion, of the remaining dispersion in reported crime and arrests. The three components collectively explain 7-23% of dispersion for violent crimes and arrests, 32-35% of dispersion for property crimes and arrests, and 44-50% of the dispersion for non-index crimes and arrests.

We find that the heterogeneity and matching across- and within-ranks components explain a substantial portion of the residual dispersion for property and non-index crimes and arrests, and less so for violent crimes and arrests. This seems reasonable because violent crime can be more idiosyncratic and harder to predict or manage. After correcting for limited mobility bias, we find that the matching across-rank components explains more of the dispersion than the other components, though marginally. This means that idiosyncrasies in well-matched managers along the hierarchy explains more of the dispersion than officer-rank fixed effects and matches of officers within the same rank. We also find that these components explain more of the dispersion in reported violent and non-index crime than for violent and non-index arrests. This result suggests that aspects of managerial police work other than arrests are important for reducing crime.

	District-level monthly outcomes					
	(Log) Reported crime			(Log) Arrests		
	Violent	Property	Non-index	Violent	Property	Non-index
Total Variance	0.025	0.012	0.01	0.092	0.057	0.03
Variance Decomposition:						
Bias uncorrected						
Heterogeneity	0.0021	0.0014	0.0018	0.0031	0.0061	0.0047
	(8%)	(12%)	(18%)	(3%)	(11%)	(16%)
Matching (across ranks)	0.0021	0.0015	0.0018	0.0024	0.008	0.0048
	(8%)	(13%)	(18%)	(3%)	(14%)	(16%)
Matching (within ranks)	0.0019	0.0011	0.0015	0.0026	0.0069	0.0042
	(8%)	(9%)	(15%)	(3%)	(12%)	(14%)
Residual	0.0189	0.008	0.0049	0.0839	0.036	0.0163
	(76%)	(67%)	(49%)	(91%)	(63%)	(54%)
Bias corrected						
Heterogeneity	0.0017	0.0013	0.0017	0.0015	0.0052	0.0042
	(7%)	(11%)	(17%)	(2%)	(9%)	(14%)
Matching (across ranks)	0.0021	0.0015	0.0018	0.0024	0.0081	0.0048
	(8%)	(13%)	(18%)	(3%)	(14%)	(16%)
Matching (within ranks)	0.0019	0.0011	0.0015	0.0026	0.0069	0.0042
	(8%)	(9%)	(15%)	(3%)	(12%)	(14%)
Residual	0.0193	0.0081	0.005	0.0855	0.0368	0.0168
	(77%)	(68%)	(50%)	(93%)	(65%)	(56%)

Table 3.4: Variance Decomposition with manager-rank fixed effects. Outcomes are measured at the district-month level and measured per 100,000 population within the district. Total variance captures the remaining dispersion in the outcome variables after estimating the residuals from a model with the other regressors described in Equation 3.4. The components of the variance decomposition are as described in Equation 3.5. The residual component is whats remaining when we subtract the variation explained by the heterogeneity and matching across- and within-rank components from the total variance.

3.6 Conclusion

3.6.1 Discussion

This paper shows how components of a hierarchy can affect performance of public institutions in an empirical setting. We focus on studying a police agency where our data allow us to track managers of different ranks along the chain of command. Our decomposition of the dispersion in productivity aims to capture: the heterogeneous effect of officers in the various managerial positions along the hierarchy; and idiosyncrasies in the matching of managers of the same and different ranks working together along the hierarchy.

Our findings show that all three components explain a substantial portion of the dispersion in productivity for reported crime and arrests for violent, property, and non-index crimes. These components collectively explain 7-23% of dispersion for violent crimes and arrests, 32-35% of dispersion for property crimes and arrests, and 44-50% of the dispersion for non-index crimes and arrests. The matching across ranks component explains more of the variation than either of the heterogeneity and matching within ranks components, but the difference is marginal. We also find that these components explain more of the dispersion in productivity for reported violent and non-index crime than compared to arrests for violent and non-index crime.

These findings show how hierarchical structure can affect police agency performance. They show that managerial positions and the way they are organized into a chain of command play an important role in explaining the performance of the police agency. We show the extent to which residual productivity can be attributed to both the design of positions as well as ability of individuals, and the careful composition of teams working together within- and across-ranks in a district. These findings also place importance on the the decisions that managers have authority over, which affect who gets promoted (as captured by the heterogeneity component) and who works with whom (as captured by the matching components).

3.6.2 Directions for future research

In this section we lay out directions for future research that may improve upon the variance decomposition we estimate. First, using higher-frequency performance data can help us identify time-varying manager-rank fixed effects. This would allow for more precise estimates of the components of the variance decomposition. It would also imply that the identification assumption could allow for sorting based on temporal variation at the district level. Second, we can conduct event studies leveraging quasi-experimental variation in officer switches to help assess the bias in our estimates due to sorting. This approach would also allow us to explore the mechanisms in which competent managers improve performance under their watch.

District-watch data

Using district-month varying outcomes as measures of the team output of a police district can be constraining for understanding how hierarchical structure affects performance. Ideally, the matching across- and within-rank covariance components of the variance decomposition (Equation 3.5) would be estimated with manager-rank fixed effects that are disaggregated across time. Recall that the matching covariance components capture idiosyncrasies in the extent to which matched individuals work well together along the hierarchy. This allows us to estimate the covariance of different manager-rank effects on performance during periods where they worked in the same district.

With outcomes that vary only at the district-month, manager-rank-month fixed effects are not identified. So, the approach we take in Section 3.5.2 estimates the matching covariance components using the number of months two managers worked together within a district as weights.

To obtain a more precise estimate of the matching covariance components, one could use higher-frequency performance data at the district-watch level. Watches are defined by 8-hour shifts throughout the day. Officers can be matched to watches, and outcomes are available at the watch-level. This makes it possible to estimate manager-rank fixed effects that are disaggregated over time. It also allows us to relax our sorting assumption to some degree, since manager fixed effects are time-varying.

Event Studies

To understand the effect on productivity of better managers along the hierarchy, one can leverage manager switches across districts, as quasi-experimental variation in the assignment of managers. Using the estimated manager-rank fixed effects, we can estimate the effect of switching from a high- to low-quality manager (and vice versa) on district performance using event studies.

The event studies would compare districts experiencing a change in manager to those not experiencing a change in manager. This would allow us to place an estimate on the degree of bias from temporal sorting of our manager-rank fixed effects for reported property crime (Section 3.4.3). It would also allow us to explore mechanisms through which managers are effective at their job, in their position within the hierarchy. For example, we could decompose the manager's effect by whether other managers they are working with along the hierarchy are of low or high ability. This would help us understand the extent to which a manager's effectiveness depends on the composition of the chain of command.

Bibliography

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). "High Wage Workers and High Wage Firms". In: *Econometrica* 67.2, pp. 251–333.

Adda, Jérôme, Brendon McConnell, and Imran Rasul (2014). "Crime and the depenalization of cannabis possession: Evidence from a policing experiment". In: *Journal of Political Economy* 122.5, pp. 1130–1202.

Afridi, Farzana and Vegard Iversen (2014). "Social Audits and MGNREGA Delivery: Lessons from Andhra Pradesh". In: *India Policy Forum* 10.1, pp. 297–341.

Andrews, M. J. et al. (2008). "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.3, pp. 673–697.

Arrow, Kenneth J. (1962). "The economic implications of learning by doing". In: *Review of Economic Studies* 29.3, pp. 155–173.

Avis, Eric, Claudio Ferraz, and Frederico Finan (2018). "Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians". In: *Journal of Political Economy* 126.5, pp. 1912–1964.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti (2009). "Active and passive waste in government spending: Evidence from a policy experiment". In: *American Economic Review* 99.4, pp. 1278–1308.

Banerjee, Abhijit et al. (2019). *The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India.*

Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster (2008). "Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system". In: *Journal of the European Economic Association* 6.2-3, pp. 487–500.

Banerjee, Abhijit V. et al. (2020). "E-Governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India". In: *American Economic Journal: Applied Economics (forthcoming).*

Bergemann, Dirk and Stephen Morris (2019). “Information Design: A Unified Perspective”. In: *Journal of Economic Literature* 57.1, pp. 44–95.

Bertrand, M. and A. Schoar (2003). “Managing with Style: The Effect of Managers on Firm Policies”. In: *The Quarterly Journal of Economics* 118.4, pp. 1169–1208.

Bloom, Nicholas et al. (2015). “Does Management Matter in schools?” In: *The Economic Journal* 125.584, pp. 647–674.

Bobonis, Gustavo J. and Frederico Finan (2009). “Neighborhood peer effects in secondary school enrollment decisions”. In: *Review of Economics and Statistics* 91.4, pp. 695–716.

Bobonis, Gustavo J., Luis R.Cámara Fuertes, and Rainer Schwabe (2016). “Monitoring corruptible politicians”. In: *American Economic Review* 106.8, pp. 2371–2405.

Bonhomme, Stephane (2021). “Teams: Heterogeneity, Sorting, and Complementarity”. In: *SSRN Electronic Journal*.

Carrillo, Paul, Dina Pomeranz, and Monica Singha (2017). “Dodging the taxman: Firm misreporting and limits to tax enforcement”. In: *American Economic Journal: Applied Economics* 9.2, pp. 144–164.

Casaburi, Lorenzo and Ugo Troiano (2016). “Ghost-house busters: The electoral response to a large anti-tax evasion program”. In: *Quarterly Journal of Economics* 131.1, pp. 273–314.

Centre for Sustainable Employment (2019). *State of Working India*. Tech. rep. Azim Premji University.

Chaisemartin, Clement de and Xavier D’Haultfœuille (2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects”. In: *American Economic Review (forthcoming)*.

Chetty, Raj (2009). “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods”. In: *Annual Review of Economics* 1.1, pp. 451–488.

Dal Bó, Ernesto et al. (2021). “Information Technology and Government Decentralization: Experimental Evidence From Paraguay”. In: *Econometrica* 89.2, pp. 677–701.

Dieye, Rokhaya, Habiba Djebbari, and Felipe Barrera-Osorio (2014). “Accounting for Peer Effects in Treatment Response”. In: *AMSE Working Papers*.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan (2012). “Incentives work: Getting teachers to come to school”. In: *American Economic Review* 102.4, pp. 1241–1278.

Duflo, Esther et al. (2013). “Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india”. In: *Quarterly Journal of Economics* 128.4, pp. 1499–1545.

Eeckhout, Jan, Nicola Persico, and Petra E. Todd (2010). “A theory of optimal random crackdowns”. In: *American Economic Review* 100.3, pp. 1104–1135.

Evans, William N. and Emily G. Owens (2007). “COPS and crime”. In: *Journal of Public Economics* 91.1-2, pp. 181–201.

Fenizia, Alessandra (2020). “Managers and Productivity in the Public Sector”. Working Paper.

Ferraz, Claudio and Frederico Finan (2008). “Exposing corrupt politicians: The effects of Brazil’s publicly released audits on electoral outcomes”. In: *Quarterly Journal of Economics* 123.2, pp. 703–745.

— (2011). “Electoral accountability and corruption: Evidence from the audits of local governments”. In: *American Economic Review* 101.4, pp. 1274–1311.

Finan, F, B A Olken, and R Pande (2017). “Chapter 6 - The Personnel Economics of the Developing State”. In: *Handbook of Economic Field Experiments*. Ed. by Abhijit Vinayak Banerjee and Esther Duflo. Vol. 2. Handbook of Economic Field Experiments. North-Holland, pp. 467–514.

Garicano, L. (2000). “Hierarchies and the organization of knowledge in production”. In: *Journal of Political Economy* 108.5, pp. 874–904.

Garicano, Luis and Timothy Van Zandt (2012). "Hierarchies and the division of labor". In: *The Handbook of Organizational Economics*. Ed. by Robert Gibbons and John Roberts. Princeton University Press, pp. 604–654.

Gerardino, Maria Paula, Stephan Litschig, and Dina Pomeranz (2017). "Can Audits Backfire? Evidence from Public Procurement in Chile". Working Paper.

Hart, Oliver and John Moore (2005). "On the design of hierarchies: Coordination versus specialization". In: *Journal of Political Economy* 113.4, pp. 675–702.

Holmstrom, Bengt and Paul Milgrom (1991). "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design". In: *Journal of Law, Economics, and Organization* 7.Special Issue, pp. 24–52.

Imbert, Clément and John Papp (2011). "Estimating Leakages in India's Employment Guarantee Using Household Survey Data". In: *The Battle for Employment Guarantee*. Oxford University Press.

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

Kamenica, Emir and Matthew Gentzkow (2011). "Bayesian Persuasion". In: *American Economic Review* 101.6, pp. 2590–2615.

Kelling, George L. and James Q. Wilson (1982). "Broken Windows". In: *The Atlantic*.

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

Lalive, Rafael and M. Alejandra Cattaneo (2009). "Social interactions and schooling decisions". In: *Review of Economics and Statistics* 91.3, pp. 457–477.

Lazear, Edward P. (2006). "Speeding, terrorism, and teaching to the test". In: *Quarterly Journal of Economics* 121.3, pp. 1029–1061.

Levitt, Steven D (1997). "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime". In: *The American Economic Review* 87.3, pp. 270–290.

Lichand, Guilherme and Gustavo Fernandes (2019). "The Dark Side of the Contract: Do Government Audits Reduce Corruption in the Presence of Displacement by Vendors?" Working Paper.

Londoño-Vélez, Juliana and Javier Ávila-Mahecha (2020). "Can Wealth Taxation Work in Developing Countries? Quasi-Experimental Evidence from Colombia". Working Paper.

Lum, Cynthia and Daniel S. Nagin (2017). "Reinventing American policing". In: *Crime and Justice* 46.1, pp. 339–393.

Manski, Charles F. (1993). "Identification of endogenous social effects: The reflection problem". In: *Review of Economic Studies* 60.3, pp. 531–542.

Mas, Alexandre (2006). "Pay, Reference Points, and Police Performance*". In: *Quarterly Journal of Economics* 121.3, pp. 783–821.

Mastrobuoni, Giovanni (2020). "Crime is Terribly Revealing: Information Technology and Police Productivity". In: *The Review of Economic Studies* 87.6, pp. 2727–2753.

McCrary, Justin (2002). "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment". In: *American Economic Review* 92.4, pp. 1236–1243.

Mello, Steven (2019). "More COPS, less crime". In: *Journal of Public Economics* 172, pp. 174–200.

Milgrom, Paul and John Roberts (1992). *Economics, Organizations, and Management.* Englewood Cliffs, New Jersey: Prentice Hall.

Milgrom, Paul and Chris Shannon (1994). "Monotone Comparative Statics". In: *Econometrica* 62.1, pp. 157–180.

Mummolo, Jonathan (2018). "Modern Police Tactics, Police-Citizen Interactions and the Prospects for Reform". In: *The Journal of Politics* 80.1.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar (2016). "Building state capacity: Evidence from biometric smartcards in India". In: *American Economic Review* 106.10, pp. 2895–2929.

Narayanan, Rajendran, Sakina Dhorajiwala, and Rajesh Golani (2019). "Analysis of Payment Delays and Delay Compensation in MGNREGA: Findings Across Ten States for Financial Year 2016–2017". In: *Indian Journal of Labour Economics* 62.1, pp. 113–133.

Niehaus, Paul and Sandip Sukhtankar (2013). "Corruption dynamics: The golden goose effect". In: *American Economic Journal: Economic Policy* 5.4, pp. 230–269.

Olken, Benjamin A. (2006). "Corruption and the costs of redistribution: Micro evidence from Indonesia". In: *Journal of Public Economics* 90.4-5, pp. 853–870.

— (2007). "Monitoring corruption: Evidence from a field experiment in Indonesia". In: *Journal of Political Economy* 115.2, pp. 200–249.

Olken, Benjamin A and Rohini Pande (2012). "Corruption in Developing Countries". In: *Annual Review of Economics* 4.1, pp. 479–509.

Radner, Roy, Radner, and Roy (1992). "Hierarchy: The Economics of Management". In: *Journal of Economic Literature* 30.3, pp. 1382–415.

Rasul, Imran and Daniel Rogger (2018). "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service". In: *The Economic Journal* 128.608, pp. 413–446.

Rim, Nayoung et al. (2021). "The Black-White Recognition Gap in Award Nominations". Working Paper.

Rivera, Roman and Bocar Ba (2019). "The effect of police oversight on crime and allegations of misconduct: Evidence from Chicago". Working Paper.

Rogger, Daniel and Ravi Somani (2018). "Hierarchy and Information". Working Paper.

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

Shleifer, Andrei and Robert W Vishny (1993). "Corruption". In: *The Quarterly Journal of Economics* 108.3, pp. 599–617.

Sukhtankar, Sandip (2017). "India's National Rural Employment Guarantee Scheme: What Do We Really Know about the World's Largest Workfare Program?" In: *Brookings-NCAER India Policy Forum.* Vol. 13, pp. 231–86.

Svensson, Jakob (2005). "Eight questions about corruption". In: *Journal of Economic Perspectives* 19.3, pp. 19–42.

Syverson, Chad (2011). “What determines productivity?” In: *Journal of Economic Literature* 49.2, pp. 326–365.

Tsai, Thomas C. et al. (2015). “Hospital Board And Management Practices Are Strongly Related To Hospital Performance On Clinical Quality Metrics”. In: *Health Affairs* 34.8, pp. 1304–1311.

Vazquez-Bare, Gonzalo (2017). “Identification and Estimation of Spillover Effects in Randomized Experiments”. Working Paper.

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