GOVERNMENT DECENTRALIZATION UNDER CHANGING STATE CAPACITY:
EXPERIMENTAL EVIDENCE FROM PARAGUAY

Ernesto Dal Bó
Frederico Finan
Nicholas Y. Li
Laura Schechter

Government Decentralization Under Changing State Capacity: Experimental Evidence From Paraguay
Ernesto Dal Bó, Frederico Finan, Nicholas Y. Li, and Laura Schechter
NBER Working Paper No. 24879
August 2018, Revised July 2019
JEL No. D02,D04,D23,D61,D73,D78,D82,H11,H43,J45,O22,Q28

ABSTRACT

Standard models of hierarchy assume that agents and middle managers are better informed than principals about how to implement a particular task. We estimate the value of the informational advantage held by supervisors – middle managers – when ministerial leadership – the principal – introduced a new monitoring technology aimed at improving the performance of agricultural extension agents (AEAs) in rural Paraguay. Our approach employs a novel experimental design that elicited treatment-priority rankings from supervisors before randomization of treatment. We find that supervisors did have valuable information—they prioritized AEAs who would be more responsive to the monitoring treatment. We develop a model of monitoring under different allocation rules and roll-out scales (i.e., the share of AEAs to receive treatment). We semi-parametrically estimate marginal treatment effects (MTEs) to demonstrate that the value of information and the benefits to decentralizing treatment decisions depend crucially on the sophistication of the principal and on the scale of roll-out.

Ernesto Dal Bó
University of California, Berkeley
Haas School of Business
545 Student Services Building #1900
Berkeley, CA 94720-1900
and NBER
dalbo@haas.berkeley.edu

Frederico Finan
Department of Economics
University of California
508-1 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
ffinan@econ.berkeley.edu

Nicholas Y. Li
Department of Economics
University of California
530 Evans Hall, #3880
Berkeley, CA 94720
nicholasli@econ.berkeley.edu

Laura Schechter
Department of Agricultural and Applied Economics
University of Wisconsin, Madison
427 Lorch St.
334 Taylor Hall
Madison, WI
53706
lschechter@wisc.edu

1 Introduction

In standard models of decentralization, the principal delegates decision-making authority to agents in order to take advantage of their superior information (Aghion and Tirole, 1997; Dessein, 2002; Mookherjee, 2006). However, decentralization may be suboptimal for at least two reasons. First, agents may pursue their own objectives and offset any gain from superior information. Second, principals could acquire some of the information that agents hold and improve their own choices without relinquishing control.

Moreover, these two issues are complicated by the fact that the value of information necessarily depends on the scale of the program. For example, suppose that an organization plans to provide assistance to low-income families. If resources are sufficient to cover all households, then decentralizing the selection decisions has no value: the program can be rolled out from the center with universal coverage. Correspondingly, if resources are so meager that only a few families can be covered, then the aggregate gain to decentralization is necessarily small. The benefits to decentralization live in the middle ground where resorting to local branches to screen, prioritize, and ultimately ration access to households yields aggregate gains relative to using coarse information and proxy-means testing. Thus, organizations face the jointly related questions of the amount of resources to devote to new programs and how best to implement them.

How the state rolls out a new monitoring technology among its front-line providers—the subject of this paper—poses similar challenges. In 2014, the government of Paraguay decided to distribute GPS-enabled cell phones to agricultural extension agents (AEAs) and their supervisors to aid in monitoring. AEAs are tasked with visiting farmers scattered over large tracts of land and giving them access to various support services including timely information about prices and best farming practices. The central government suspected that AEAs were shirking due to the monitoring difficulties afflicting their supervisors. The authorities hypothesized that issuing GPS-enabled phones to AEAs could help mitigate the problem by making it possible to spatially track AEAs.

However, the government had limited resources, which made rolling-out phones to all AEAs difficult. This motivates several questions: 1) Is supervisor discretion to assign phones to AEAs valuable in light of their potentially superior information and potentially imperfect motives? 2) Can the informational advantage of supervisors be replicated with the more limited information of ministerial staff? 3) What should the scale of roll-out be, and should ministerial staff or supervisors allocate the phones to AEAs?

In the context of real life organizations, answering questions about implementation scale and mode

1

is difficult, even when experimental methods can be deployed. At best, standard RCT-based impact evaluations inform researchers and policymakers about which individuals most benefit from a program along typical demographic and economic characteristics. These indicators can guide organizations committed to a centralized implementation, but give little clue about the informational gains from a decentralized approach or the scale at which a change of implementation mode will lead to meaningfully different outcomes.

In this paper, we develop a model applied to a novel experimental design to overcome these difficulties. In particular, we utilize information collected on how supervisors prioritize AEAs for roll-out in our study of the impact of enhanced monitoring on AEA performance. We show that supervisors have valuable information about who to place under stronger monitoring. We then model supervisor choices and measure the value of their discretion at different levels of roll-out. Next, we predict the effects of the program under several centralized implementation schemes that vary the amount of information that principals may plausibly acquire. Finally, under each scenario, we compute the optimal roll-out scale and show that this scale varies with the amount of information available.

We begin by addressing the standard impact evaluation question: does the new monitoring technology reduce shirking? We find that randomly assigned cell phones have a sizable effect on AEA performance, increasing the share of farmers visited in the previous week by an average of 6 percentage points (pp). This represents a 22 percent increase over the AEAs in the control group. The cell phones also improve farmer satisfaction with their AEAs by 0.13 standard deviations. Because we do not find any impact of cell phones on AEAs who do not have supervisors, we interpret this effect as a reduction of shirking following increased monitoring, as opposed to a direct improvement in productivity due to the cell phones. Also consistent with our interpretation, we find that treated AEAs perceive more monitoring from their supervisors. Finally, we find no evidence that treated AEAs increased the number of visits at the cost of conducting shorter ones.

We then evaluate whether supervisors had superior information about which AEAs would be more responsive to treatment using a novel feature of our experimental design. Prior to randomization of treatment, we elicited the preferences of supervisors regarding which half of their AEAs should be selected to receive the phone. We then randomly assigned phones to AEAs independently of the supervisor opinion, generating a 2-by-2 treatment-by-selection matrix that is the basis of our empirical analysis.

We find that supervisor-chosen AEAs respond more to increased monitoring, entirely driving the average increase of 6 pp. Among these supervisor-chosen AEAs, treatment increased the likelihood

2

that a farmer was visited in the past week by 15.4 pp compared to a statistically insignificant decrease of 3.6 pp among those who were not selected. This finding corroborates the notion that going down the hierarchy from the top program officers to local supervisors could allow organizations to leverage superior, dispersed knowledge about how best to allocate resources.

While the research design divorces treatment from supervisor-selection, supervisors maintain authority over AEAs. If there are spillover or Hawthorne effects, the treatment effects we measure might not be solely due to monitoring. We rule out these threats by relying on another feature of our experimental design. For an auxiliary and random subset of AEAs, we implemented the supervisor preferences, temporarily allocating phones only to supervisor-selected AEAs.

Besides lending credibility to the elicitation of supervisor choices, this feature allows us to compare the performance of non-selected untreated AEAs depending on whether their selected coworkers were treated with phones. Likewise we can compare the performance of selected treated AEAs depending on whether their non-selected coworkers were treated with phones. We find no significant differences in an individual's performance depending on his coworkers' treatment, suggesting that spillovers from supervisors reassigning supervision effort are not important.

Finally, we compare impact in groups where supervisors had a meaningful selection to make—several AEAs to choose from—versus cases with as few as a single AEA where no supervisor selection was made. We again find no significant differences in the impact of the phone across these groups. This suggests Hawthorne effects, such as a supervisor shifting supervision effort to prove her choices had been right, are unlikely.

What underlies the informational advantage of supervisors? Only some of the AEA characteristics are observable to the principal or an econometrician. Having collected a rich dataset on the AEAs that includes information on both cognitive and non-cognitive traits, we develop a two-step estimation procedure in the spirit of a sample selection model to decompose the value of information into observable and unobservable AEA traits. We use this approach to compute a series of marginal treatment effects under various selection rules and roll-out scales. The approach we develop would allow program leadership to optimize the program's roll-out scale. The marginal treatment effects we estimate are critical inputs to such optimization and also for deciding whether to decentralize treatment assignment in the first place.

We find that both commonly observed demographic traits (e.g., gender) and harder-to-measure characteristics such as cognitive ability and personality type do a poor job of explaining supervisors' selection decisions. Among the few observable traits that predict an AEA being selected for treatment by a supervisor, the AEA's party affiliation is one of the most robust. Supervisors

are much less likely to place members of the incumbent party under additional monitoring. But whether this is evidence of preference bias is difficult to say, since members of the incumbent party also tend to perform relatively worse when treated. Any evidence of preference bias seems to be on the basis of age: older AEAs are significantly less likely to be selected, even though age is not predictive of response to treatment. When we allow the treatment effects to vary by a rich set of observable characteristics, the unobservable (to the econometrician) component of the supervisors' choices is among the characteristics that most robustly predicts the responsiveness of an AEA to the additional monitoring.

While our findings suggest that supervisors have valuable information, the convenience of decentralizing the treatment assignment to them depends on the counterfactual allocation regime of the principal. The latter regime depends on the information she has, the feasible allocation rules she can adopt, and the extent of available resources. We construct a number of counterfactual scenarios varying the information sophistication of the central authority for all possible roll-out scales in order to measure effect sizes of centralized versus decentralized assignment. In particular, we use our model estimates for the distribution of treatment effects to compare the improvement in farmer visits when the supervisor selects AEAs against four hypothetical allocation rules: 1) a totally uninformed principal who allocates randomly; 2) a minimally informed principal who targets AEAs who have to travel longer distances to visit farmers; 3) a more sophisticated principal who collects and analyzes baseline data on AEAs and targets predictably low productivity AEAs; and 4) the most sophisticated principal who pilots an experiment, learns about the connection between observable characteristics and response to treatment, and thereafter targets AEAs in descending order of predicted responsiveness to treatment.

We find that the value of supervisor information is substantial relative to a regime in which the principal simply allocates phones at random and that this difference in program impact is maximized at 53 percent roll-out. At this scale, the supervisor allocation increases the share of farmers visited by 6.9 pp versus only 3.4 pp under random assignment. A slightly more effective approach compared to random assignment would be to simply allocate the phones to the AEAs who have to travel the longest distance to attend to their farmers. This method generally outperforms random assignment (a 1.9 pp advantage at 50 percent coverage), but still underperforms relative to supervisor choice.

A more effective centralized policy identifies and treats workers who are expected to be the least productive according to observable characteristics. We operationalize this policy by estimating the relationship between AEA productivity and observable characteristics among the control AEAs without GPS phones. We then utilize this relationship to rank all AEAs based on their predicted

productivity. Governments that have the information and capacity to prioritize the AEAs with the lowest predicted productivity are able to perform at least as well as, and in some cases better than, the supervisors. Such a regime succeeds because the central government can make better use of the information at hand despite not having as much information as a supervisor. This in turn suggests that imperfect processing of information or bias prevents supervisors from being as effective as they could be.

The most effective but most information-demanding centralized policy we consider uses AEA observables to predict response to treatment rather than to predict baseline productivity. Under such a policy, the principal conducts a pilot experiment and uses the results to predict responsiveness among the remaining untreated AEAs based on their observable characteristics. Treating AEAs in descending order of predicted responsiveness, even when lacking information on unobservables known only to the supervisor, vastly outperforms assignment by the supervisors. A sophisticated principal would find it optimal to roll-out phones to 70 percent of AEAs, which is (weakly) lower than the optimal scale under all other approaches we consider, indicating that a sophisticated principal can attain a higher aggregate impact with a smaller scale intervention. The high performance of the last two centralized methods highlights that innovations in information and communication technologies as well as the introduction of experimental methods to inform policy can play a role in reducing information frictions and alter optimal organizational structure.

Our study speaks to several literatures. First and foremost, our study contributes to a large but mostly theoretical literature on why organizations decentralize decision-making authority.[1] Recently, some empirical progress has been made in understanding why private- sector firms decentralize. For instance, based on the insight by Aghion and Tirole (1997) that organizations are more likely to decentralize if the principal and agent have congruent preferences, Bloom et al. (2012) find that firms are more decentralized when located in regions that are judged by those in the headquarters location to contain more trustworthy people. The authors view trust as a proxy for congruent preferences.

Given the standard assumption that agents are better informed than the principal, access to costly information can also determine a firm's decision to decentralize. For example, Acemoglu et al. (2007) show using data on French and British firms in the 1990s that firms closer to the technological frontier, firms in more heterogeneous environments, and younger firms are more likely to choose decentralization—settings that presumably proxy for environments where learning is more difficult. Despite the progress that these and other studies have made, direct empirical evidence on

---

[1]Mookherjee (2006) provides an excellent review of the theory on decentralization.

the existence of superior information by agents is still lacking.

One exception is Duflo et al. (2018) who conducted a field experiment that increased the frequency of inspections of industrial plants in Gujarat, India. In the control group, plants were audited as usual at the discretion of the inspectors, whereas in the treatment group, the audits were conducted more frequently but at random. They found that despite the increased regulatory scrutiny, randomized audits did not significantly reduce pollution emissions. This is because the discretionary inspections targeted the plants with higher pollution signals. Because the largest penalties are reserved for extreme pollution violations, these are the firms most likely to be impacted by audits.

We complement Duflo et al. (2018) in important ways. Our experiment was designed to identify who the supervisors would target for monitoring without having to rely on strong functional form assumptions. As a result, we can experimentally identify the decentralized counterfactual to a centralized approach. Moreover, that counterfactual depends both on supervisors' informational advantage and on potential preference biases, which we allow for but are absent from the targeting rules in Duflo et al. (2018). Thus, we incorporate elements that are crucial to the evaluation of the relative merits of decentralization.

Similar to the public sector, private sector employers also need to monitor their employees. A paper by de Rochambeau (2017) discusses the roll-out of GPS tracking devices in a trucking company. She finds that managers choose to allocate the tracking device to drivers who perform less well at baseline and that these truckers benefit most from the device.

The problem of how best to deploy monitoring technology to workers is also similar to the problem of how best to target social programs. In this regard, our paper is most related to two studies. Alderman (2002) examines an Albanian social assistance program. He shows that even after controlling for the assets that local officials used to target beneficiaries, household consumption was still predictive of who received the program. The author interprets this as evidence that officials relied on their local information and discretion. Alatas et al. (2012) conducted a field experiment in Indonesia comparing proxy-means testing against community-based targeting of a social program meant to help the poor. They find that proxy-means testing does a better job at identifying the consumption-poor than community-based targeting. They argue that this difference is not due to elite capture or local information, but rather a difference in how local communities define poverty. Similar to difficulties in the context of social programs, banks could benefit from community knowledge to help them lend to the most entrepreneurial people. Hussam et al. (2017) find that community members have useful information on marginal returns and this information is useful above and beyond what a machine learning algorithm would predict from observables.

Our study also has parallels to the literature on the use of marginal treatment effects (MTEs) to construct policy-relevant counterfactuals (Heckman and Vytlacil, 2005). As in the MTE literature, we express the supervisor's problem of how to prioritize AEAs for treatment as a joint model of potential outcomes and selection as determined by a latent index crossing a threshold. In contrast with the standard MTE setup, our selection equation does not model an AEA's self-selection into treatment but rather the selection by a supervisor. Crucially, treatment is not contingent on being selected—only those AEAs who were randomized into treatment were in fact treated. Thus, when we compute the MTEs, we do not have to extrapolate to subgroups of "always-takers" and "never-takers" because we only have compliers by design. In this respect, our approach implements a variant of the selective trial designs proposed by Chassang et al. (2012). In that paper, the authors recast randomized control trials into a principal-agent problem and show theoretically how one can recover the MTEs necessary to forecast alternative policies and treatment assignments by eliciting subjects' willingness to pay for the treatment. Instead of eliciting our agents' willingness to pay for the treatment, we elicit the targeting preferences of the supervisor, who in our context is the relevant decision maker.

Finally, our study adds to a growing body of experimental evidence on the impact of new monitoring technologies for reducing shirking in the public sector. Similar to our setting, some of these studies involve weak or no explicit financial incentives. For example, Aker and Ksoll (2019) monitored teachers of adult education in Niger by calling both the teacher and the students to ask whether the class was held and how many students attended. They found that the calls led to fewer canceled classes and better student test scores. Callen et al. (2018) used a similar cell phone technology to monitor health facility inspectors and found that this increased the frequency of inspections, especially for those with 'better' personality traits.

Other studies have introduced new technologies for monitoring but have also overlaid financial incentives. For instance, Duflo et al. (2012) required teachers to take a picture of themselves with their students at the beginning and end of each school day using a camera with tamper-proof date and time functions, whereas Banerjee et al. (2008) asked nurses to time-stamp a register at the beginning, middle, and end of the day. Both studies found these treatments increased teacher and nurse attendance, but in both cases, the impact was found to be mostly due to concomitant financial incentives. Dhaliwal and Hanna (2017) found that fingerprint readers in health centers decreased absence even though financial incentives provided by the monitoring technology were rather weak. Banerjee et al. (2015) and Khan et al. (2016) look at on-the-job performance rather than attendance (among police and tax collectors respectively) and employ both monitoring and incentives. These papers do not give a definitive answer regarding whether most of the improvement in performance is

due to the monitoring or the incentives. The first paper suggests a significant impact of monitoring alone, while the second suggests an insignificant impact. We contribute to this literature by showing that a cell phone technology can be effective in reducing shirking for individuals such as agricultural extension agents whose job requires them to visit farmers who live out in rural areas, often quite far from the local agricultural ministry offices in town.

2 Background

Agricultural extension services in Paraguay are centered around the Ministry of Agriculture based in Asunción. Below the central ministry are 19 Centros de Desarrollo Agropecuario (CDAs, which exist at the department level, similar to a state in the United States) and below the CDA level there are 182 Agencias Locales de Asistencia Técnica (ALATs, which are at the municipality level, similar to a county in the United States). The Paraguayan Ministry of Agriculture has close to 1000 agricultural extension agents working within ALATs spread across four main agencies. We work with the biggest of these agencies, Dirección de Extensión Agraria (DEAg).

The main job of extension agents is to help farmers access institutional services that will help them improve their production. The goal is to increase farmers' output directed both for own consumption as well as the market. Another goal is to increase farmers' participation in markets. The official thematic areas in which extension agents work are soil improvement, food security, product diversification, marketing, improving quality of life, and institutional strengthening. Extension agents come from a variety of backgrounds including agricultural sciences, veterinary sciences, nutrition, law, and teaching.

Much of what extension agents do resembles the role of middlemen, connecting farmers with cooperatives, private enterprises, and specialists. Extension occurs both one-on-one and in group meetings. Extension agents conduct farm visits in which agricultural problems are diagnosed and addressed. Group meetings are used to lead demonstrations or talk about technical topics. AEAs also organize farmer field trips. Each extension agent is assigned to work with approximately 80 producers. Extension agents do not usually offer free goods or services to farmers. Although the ALAT headquarters are in towns, most of their daily work involves driving out to rural areas to visit farmers.

Within every ALAT there is a supervisor who, in addition to working with his own farmers, must also monitor the other extension agents working in the ALAT. We will refer to individuals who

work purely as agricultural extension agents as 'AEAs.' By this definition, DEAg has over 200 AEAs working within the organization at any time.

In June 2014, the Ministry of Planning, in association with the Ministry of Agriculture, decided to provide AEAs with GPS-enabled cell phones. While all AEAs already owned their own personal cell phones, AEAs would be able to make work-related calls or send work-related messages for free on the government-provided phone. This initiative had several objectives. One was to improve coordination and communication between the AEAs and their supervisors. For example, it would give the AEA a mechanism to take a picture of a farmers' diseased crop, circulate it, and get advice for the farmer from a specialist. But crucially, it would allow the supervisors to see where AEAs were at all times, how long they spent in each place, and what they did there (since the AEA is supposed to document every meeting in which he participates). AEAs can submit reports and review reports they have already submitted through the phone. Supervisors can view reports submitted by all the AEAs they oversee.

In the terms of the hierarchical agency model we lay out in the next section, we view the ministerial leadership introducing the new technology as the principal, we will refer to the ALAT-level supervisors as "supervisors," and the AEAs as the "agents."

3 Model

Consider a hierarchy composed by a principal, a supervisor, and a continuum of agents with mass 1. The supervisor is responsible for monitoring the agents. In such a hierarchy there are two possible agency problems: that between the agent and his supervisor and that between the supervisor and the principal. We will focus mainly on the problem between agent and supervisor, and analyze how it changes when the agents are placed under a new monitoring technology. The question will be whether the principal can obtain better results by relying on supervisors in deciding how to deploy the technology.

Agents and monitoring Each agent caters to a mass 1 of farmers. Each visit by an agent i yields a benefit to the visited farmer. Agents receive a fixed wage w and choose a share $s_i \in [0, 1]$ of farmers to visit. The agent obtains an intrinsic motivation $z_i s_i$ from visiting a share s_i of farmers, but also incurs a cost $a_i s_i + b_i \frac{s_i^2}{2}$. The share s_i is a measure of agent effort, and the principal cares about it because it constitutes a measure of socially beneficial service provision. From now on, we will refer to s_i as effort and assume that it is noncontractible.

9

The supervisor operates a monitoring technology that with probability $q_i \in (0,1)$ reveals s_i. The principal then reprimands the agent in proportion to the amount by which his effort falls short, $1 - s_i$. The agent gets disutility from being reprimanded equal to $(1 - s_i)r_i$, with $r_i > 0$.[2] While monitoring allows the supervisor to obtain information about the agent's effort, it can potentially weaken intrinsic motivation. When monitored, the intrinsic motivation payoff of agent i becomes $(z_i - g_i)s_i$, which is potentially negative. It reflects the fact that agents may feel aggrieved to an extent $g_i \geq 0$ when under close supervision. In sum, wages and effort costs accrue to the agent regardless of supervision, while reprimand and intrinsic motivation payoffs accrue in relation to monitoring intensity q_i. Thus, agent i can be seen to maximize utility,

$$u_i(s_i) = w - a_i s_i - b_i \frac{s_i^2}{2} + q_i [(z_i - g_i) s_i - (1 - s_i) r_i] + (1 - q_i) z_i s_i,$$

or, collecting terms,

$$u_i(s_i) = \omega_i + \mu_i s_i - b_i \frac{s_i^2}{2} + q_i s_i \rho_i,$$

where $\omega_i \equiv w - r_i q_i, \mu_i \equiv z_i - a_i, \rho_i \equiv r_i - g_i$. Agent i chooses the share s_i of farmers to visit to maximize utility $u_i(s_i)$, and he does so after learning the level of monitoring intensity q_i he is under. Because $u_i(s_i)$ is concave, agent i's optimal effort is $s_i^*(q_i) = \max\left\{0, \frac{\mu_i + q_i \rho_i}{b_i}\right\}$. Since b_i only affects effort through ratios involving ρ_i and μ_i, parameters that can be scaled arbitrarily, we normalize $b_i = 1$, yielding,

$$s_i^*(q_i) = \max\{0, \min\{\mu_i + q_i \rho_i, 1\}\}. \tag{1}$$

The term μ_i – a proxy for net-of-cost intrinsic motivation – is individual-specific and for some agents potentially negative. Even more important for our purposes, the term ρ_i, which captures both the agent's distaste for being reprimanded (which raises effort) and his resentment at being monitored (which lowers effort) is also potentially negative for some agents. We will assume ρ_i to be drawn from a continuous distribution $F(\rho_i)$ over a support $[\rho_l, \rho_h]$, where $\rho_h > 0$ but ρ_l is potentially negative.

New technology and treatment effects We assume that q_i can take one of two levels $\{q_l, q_h\} \in$

---

[2]Alternatively, one may assume that the supervisor draws a farmer at random, and finds he has not been visited with probability $1 - s_i$, in the event of which she proceeds to reprimand the agent with a fixed intensity r_i. It is also possible to extend the model to make q_i a function of monitoring effort by the supervisor. The choice of monitoring effort remains unmodeled here, in order to stick with the simplest formulation that will deliver the results of interest. Such an extension could also involve an agency problem in the supervisor's choice of monitoring effort without affecting the essence of our results. The only tension between supervisor and principal that may arise in our simpler setting relates to the deployment of the monitoring technology to be described below.

$(0,1)$, with $q_h \equiv q_l + t_i \Delta q$, $\Delta q > 0$, where q_l denotes a status quo level of monitoring, and $t_i \in \{0,1\}$ reflects whether agent i is "treated" with a new monitoring technology.[3] Treating any given agent costs an amount c. In order to characterize treatment effects neatly and avoid awkward truncation issues, in what follows we will assume that $\mu_i + min\{q_h \rho_l, q_l \rho_l\} > 0$ and $\mu_i + q_h \rho_h < 1$, which guarantees interior solutions for s_i.

While μ_i and ρ_i both affect the level of effort, only ρ_i affects the response of effort to a change in monitoring technology. Thus, in the remainder of this section we will refer to different levels of ρ as agents' "types." Under increased monitoring, an agent of type ρ increases his effort by $\rho \Delta q \equiv T(\rho)$, which captures the treatment impact of the new technology for that type. Note that since ρ_l can be negative, $T(\rho)$ can be negative for some types. In addition, the definition of $T(\rho_i)$ implies that equilibrium agent effort s_i^* (weakly) increases in monitoring technology q_i for all agents with $\rho_i > 0$, and an improvement in monitoring technology (an increase in q_i) has a larger effect on the effort of agents with a higher type ρ_i. We use m to denote the roll-out scale of the intervention. Given a continuum of agents, the total treatment impact from targeting a random sample with measure m is $m \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho$, illustrated by the strictly increasing diagonal line in Figure 1, achieved at a total cost mc.[4]

Consider an intervention treating a random selection of agents, and a typical impact evaluation comparing performance between a treatment and a control group. What would that evaluation be testing? Inspection of the expression for the average treatment impact $\int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho$ yields the following:

Remark 1. *If $\rho_l \geq 0$, the total (and average) treatment impact is guaranteed to be positive. If $\rho_l < 0$, the total (and average) treatment impact is positive if and only if, given $\Delta q > 0$, the density $f(\cdot)$ places enough weight on positive types.*

Thus, when $\rho_l \geq 0$, a standard impact evaluation tests our assumption that better monitoring technology increases detection rates ($\Delta q > 0$); and when $\rho_l < 0$, a standard impact evaluation jointly tests that $\Delta q > 0$ and that $f(\cdot)$ places enough weight on positive types.

Centralization In this section we define centralization as the benchmark case in which the principal knows the distribution of types $F(\rho)$, but does not know the type of any specific agent.[5] Thus, she

---

[3]Here we assume treatment only affects the agent's problem by raising monitoring intensity, although it could in principle also affect μ_i via the agent's cost a_i. This is plausible as some technologies, like GPS phones, can be productivity-enhancing. However, as we will show later, in our context the data do not support that possibility.

[4]Our definition of the total treatment effect abstracts from spillover effects across agents. As we report below, we do not find empirical evidence of spillover effects. So we keep the theory consistent with the empirical measurement of marginal treatment effects, which will likewise abstract from spillovers.

[5]Under centralization, the principal makes all decisions based upon her own information. Under decentralization,

11

cannot determine whether any agent is more profitably treated than another. At roll-out scale m, the principal selects treated agents at random and attains an impact equal to $m \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho$ at cost mc. If the principal could choose m, she would choose roll-out scale of either zero or 100%. The principal will adopt, and do so at full scale, whenever $\int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho \geq c$ (breaking indifference in favor of adoption), i.e., whenever the average treatment effect of the new technology is larger than its marginal and average cost, as represented in Figure 1.

Decentralization Under decentralization, the supervisor selects which agents will receive treatment. The supervisor's criterion is a function $D^S(\rho|m) \in [0,1]$ indicating the probability that a type ρ is selected for treatment when a proportion m of agents must be treated. The total treatment impact under decentralization is $\int_{\rho_l}^{\rho_h} T(\rho) D^S(\rho|m) f(\rho) d\rho$. Supervisors may have superior information: they may know the distribution $F(\rho)$ and also obtain a signal about each agent's type. In the extreme, the supervisor observes individual types perfectly. If benevolent, a fully informed supervisor will select the highest set of types $\rho \in \left[\rho_m \equiv F^{-1}(1-m), \rho_h\right]$ that make for a measure m, and attain an impact equal to $\int_{F^{-1}(1-m)}^{\rho_h} T(\rho) f(\rho) d\rho$. The impact under decentralized supervisor assignment is larger than that from random assignment.[6]

Remark 2. *If supervisors know agents' types and assign treatment to maximize visits to farmers, the total (and average) treatment impact on agents selected by supervisors will be higher than on agents selected at random, and on agents who were not selected by supervisors.*

If $f(\rho)$ is well behaved (is atomless and has full support), and $\rho_l \geq 0$, then the impact attained by the supervisor is continuously increasing and concave in m as illustrated in Figure 1.[7] The total treatment impact attained by supervisors coincides with that under random assignment for $m = 0$ and $m = 1$. At $m = 0$ nobody is treated, so total treatment impact is zero. At 100% roll-out, everyone is treated, and there is no advantage of prioritizing high types for treatment. In both extremes, any potential informational advantage of the supervisor is worthless.

The typical concern with decentralization is that supervisors may pursue their own objectives, and select agents in a way that undoes, partly or completely, their informational advantage. An extremely biased supervisor who selected agents in order to minimize impact would select the lowest set of types $\rho \in \left[\rho_l, \rho_m \equiv F^{-1}(m)\right]$, and attain an impact equal to $\int_{\rho_l}^{F^{-1}(m)} T(\rho) f(\rho) d\rho$ (which,

---

the principal delegates decisions to supervisors, or, equivalently, supervisors submit information that mechanically drives the principal's decisions. Thus, we abstract from the interesting distinctions made by Dessein (2002) between delegation and strategic communication.

[6]The condition $m \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho < \int_{F^{-1}(1-m)}^{\rho_h} T(\rho) f(\rho) d\rho$ is equivalent to the condition $\int_{\rho_l}^{F^{-1}(1-m)} T(\rho) \frac{f(\rho)}{F[F^{-1}(1-m)]} d\rho < \int_{F^{-1}(1-m)}^{\rho_h} T(\rho) \frac{f(\rho)}{1-F[F^{-1}(1-m)]} d\rho$ which is clearly true from the fact that $T(\rho)$ is increasing –the left hand side is the average treatment effect among a lower set of types than the right hand side.

[7]If $\rho_l < 0$, then the total impact curve will have an interior maximum and be decreasing for m close enough to 1.

under the same distributional assumptions as before, would yield a convex impact curve in Figure 1 lying below the straight line corresponding to random assignment). Real-life situations will likely fall in between these extremes. Even partially informed supervisors will attain higher impact than random assignment if they are not too biased – a situation in which we say supervisors have a "net informational advantage." To determine if supervisors have such a net informational advantage, our experiment will compare the impact of treatment among agents that would be selected by a supervisor against agents that would not be.

Optimal roll-out, and the optimal value of decentralization

When information about agent types is available, there will be gains to choosing the roll-out scale judiciously. Consider the case of an informed and unbiased supervisor who has control over the scale of adoption. That supervisor will choose the lowest treated type ρ' to maximize $\int_{\rho'}^{\rho_h} T(\rho)f(\rho)d\rho - c(1 - F(\rho'))$, and that type satisfies $T(\rho^*) = c$. In words, the supervisor will choose to treat every agent down to a type ρ^* whose marginal treatment effect from the new technology equals the marginal cost, and whose type is weakly lower than that of a measure m^* of others, as illustrated in Figure 1. Thus, realizing the full value of decentralization depends on the ability to choose the roll-out scale.

Remark 3. *The value of decentralization varies with roll-out scale. In order to empirically determine the value of decentralization, it is necessary to estimate marginal treatment effects at multiple roll-out scales.*

There are at least three reasons why it may be important in practice to determine the optimal roll-out scale. First, if decentralization is the organizational status quo, implementing a new program in a decentralized way does not require costly organizational change, and the informational gains will be an added benefit. But in some cases centralized assignment may be the status quo, and a shift to decentralization would require investments in local organizational capacity. If the optimal roll-out scale is very close to 0 or 100%, the informational gain may be too small to justify investing in an organizational shift. Second, it may often be the case that limited resources make it infeasible to roll-out an intervention at its optimal scale, in which case the effective gains from decentralization may be lower than otherwise.

Third, there may be interventions that are not socially valuable when implemented by the principal, but which would be socially valuable if implemented at a scale close to what is optimal under decentralized choice by informed supervisors. To see this possibility, consider a costly intervention with marginal cost c', such that $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho < c'$. In this case, the average treatment effect is below marginal and average cost, and so both 100% adoption and random centralized adoption at

13

any scale would lead to a loss. But as represented in Figure 1, suppose also that treatment impact is larger than cost for a set of types $\rho \in [\rho', \rho_h]$, for which $(T(\rho) \geq c')$, with $T(\rho') = c'$, so that $\int_{\rho'}^{\rho_h} T(\rho)f(\rho)d\rho > c'(1 - F(\rho'))$, where $m' = 1 - F(\rho')$ is the measure of treated agents. In this situation, the principal would not gain by selecting any measure of agents to be treated at random, but would gain by delegating to the supervisor if the latter will focus treatment on types (weakly) larger than ρ'.

The broader implication is that impact evaluation should not abstract from the extent of roll-out and its implementation mode, since the implementation mode affects who gets treated. In other words, determining whether a technology is valuable – presumably the ultimate goal of an impact evaluation – requires assessing the total treatment impact under different scales of roll-out using the best assignment mechanism. The practical challenge for impact evaluation is that estimating treatment effects at all roll-out scales requires a very demanding experimental design. Most experiments, including ours, have one or at most two roll-out scales. In this paper we will develop a method to estimate the gains from decentralization at all levels of roll-out.

The comparison between decentralization and centralization-as-random-assignment is stark. Richer comparisons are possible. Once a treatment impact curve for all roll-out scales is estimated for decentralized assignment, we will consider less handicapped centralized approaches and ask how much information is necessary for the center to do as well as or even outperform decentralized choice. In an era of decreasing costs for information and communications technology (ICTs), central authorities can increasingly collect and process data and bridge the informational gap.

Summarizing, our empirical study will investigate three claims stemming from the remarks in this section. First, are there enough responsive types among AEAs such that the intervention at hand delivers a positive treatment impact on average? Second, do supervisors have valuable knowledge (net of potential preference biases) about which agents ought to be treated given partial roll-out? If so, we should observe treatment effects among those selected by supervisors to be larger than among agents who were not selected. Third, how does the scale of roll-out alter the impact of decentralized assignment, and can a realistically informed principal attain the same impact? Answering these questions will require a special experimental design, a method for ascertaining the marginal treatment effects on different types, and a characterization of centralization approaches utilizing varying levels of information.

4 Research Design

Our experiment was conducted on 180 local technical assistance agencies (ALATs, Agencia Local de Asistencia Técnica). On average, each ALAT consists of a supervisor and three agricultural extension agents (AEAs). Many ALATs have a single AEA. In those ALATs, the AEAs were randomized to receive a phone either in the first wave, the second wave, or not at all.

There are 48 larger ALATs that have at least 2 AEAs. We asked the supervisors of the latter group to indicate which half of her AEAs should receive the phones first given the program's objective to increase worker performance. We refer to these AEAs as "selected." AEAs were not told that their supervisor was asked to make such a decision, and were not told who was selected. These 48 ALATs were then randomly assigned into three groups according to how and when the agents would receive their phones.

The main group of ALATs is in cells *A*, *B*, *C*, and *D* in Figure 2. The ALATs in cells *B* and *D* (a quarter of the ALATs), serve as the treatment group. In these ALATs all AEAs, both selected and non-selected, received the GPS-enabled cell phone which increased monitoring in the first round. The ALATs in cells *A* and *C*, (half of the ALATs), serve as our control group as no AEAs received the phones in these groups. The average difference in performance between AEAs in cells *B* and *D* and AEAs in cells *A* and *C* estimates the average impact of treatment. And, the difference-in-differences computed as the performance by AEAs in cells $(B - A) - (D - C)$ estimates whether the impact on selected AEAs is larger than the impact on non-selected AEAs. This difference-in-differences allows us to determine whether supervisors had valuable information on how to direct treatment.

A third group of ALATs (cells *E* and *F*) received partial treatment. Only those AEAs who had been selected by their supervisors for treatment (cell *E*) received phones in the first wave. This design helped make the elicitation of supervisors' preferences credible and relevant. Eight months after the delivery of these phones, a second wave of phones were delivered to the non-selected AEAs in those ALATs, cell *F*. The difference in performance between the AEAs in cell *F* and in cell *C* provides a test of whether allocating phones to the selected AEAs can also affect the performance of non-selected AEAs in the same ALAT. This would be the case if the supervisors also responded to the treatment of those in cell *E* by changing the intensity with which they monitored the AEAs without the cell phone in cell *F*.

4.1 Taking the Theory to the Data

Recall that the performance of AEAs whenever interior is given by equation (1):

$$s_i^* = \mu_i + q_i \rho_i.$$

To operationalize this equation, recall that the level of monitoring for each AEA, q_i, is a function of the monitoring technology t_i, according to the expression $q_i = q_l + \Delta q t_i$ where t_i takes a value of 0 when AEAs do not get a cell phone and 1 when they do. Because our objective is to see AEAs respond to exogenous changes in monitoring q_i, we normalize $q_l = 0$ and rewrite the expected disutility of being reprimanded (net of monitoring grievance) as $q_i \rho_i = \beta_i t_i$, where $\beta_i = \Delta q \rho_i$.

The central goal of our approach is to model various selection criteria and estimate the marginal treatment effects under each criterion for varying levels of roll-out. A key element is to consider different degrees of observability of the individual parameters (μ_i, β_i), in an individual AEA's effort function in equation (1). In particular, we will map these parameters into a vector of fixed characteristics (X_i) and two independently random characteristics (ε_i, η_i), to write: $\mu_i(X_i, \varepsilon_i)$ and $\beta_i(X_i, \eta_i)$, so that,

$$s_i^* = \mu_i(X_i, \varepsilon_i) + \beta_i(X_i, \eta_i)t_i. \tag{2}$$

While the vector X_i may be observable to both the principal and supervisor, the elements (ε_i, η_i), may only be partially observed by the supervisor.

Average treatment effect   We can estimate the average treatment impact of the cell phone on effort by imposing some familiar (but mild) structure on individual parametric heterogeneity as follows: $\mu_i = \mu' X_i + \varepsilon_i$ and $\beta_i = \beta_0$. An individual AEA's effort function becomes

$$s_i^* = \mu' X_i + \beta_0 t_i + \varepsilon_i, \tag{3}$$

where s_i^* measures the share of farmers AEA i visited in the past week. The coefficient β_0 provides a causal estimate of the difference in performance between AEAs in both treated cells B and D relative to AEAs in the control cells A and C. Thus, we can test whether the intervention yields positive value (Remark 1 from Section 3) by contrasting the null hypothesis of $\beta_0 = 0$ against the alternative that $\beta_0 > 0$.

Given our research design, we cluster the standard errors at the ALAT level. Because we have relatively few clusters, the appendix reproduces our main tables with two sets of p-values: one

computed using a wild cluster score bootstrap procedure (Wu, 1986; Kline and Santos, 2012), and the other computed using Fisher's permutation-based randomization inference (RI) test as implemented by Heß (2017). In estimating equation (3), we can also include the single AEA ALATs, which were assigned phones at random. In this randomization, one-third of AEAs initially received a phone, with two-thirds serving as a control. When including these ALATs, the vector X_i contains an indicator for whether or not the ALAT has a single AEA.

Average treatment effect when supervisors choose   To test whether supervisors are able to select those AEAs whose effort would most increase when monitored, we can simply re-parameterize $\beta_i = \beta_0 + \beta_1 D_i^S$, where D_i^S is an indicator for whether AEA i was selected to receive a phone. Equation (2) then becomes

$$s_i^* = \mu' X_i + \beta_0 t_i + \beta_1 (D_i^S \times t_i) + \varepsilon_i, \tag{4}$$

where the indicator D_i^S is also included within the vector X_i. With this specification, we can compare the difference in performance between selected AEAs in the treatment and control groups (cells $B - A$) net of the difference in non-selected AEAs in the treatment and control group (cells $D - C$). Thus, we can test Remark 2 from Section 3, that supervisors have valuable information about which AEAs should be targeted, by contrasting the null hypothesis $\beta_1 = 0$ against the alternative $\beta_1 > 0$. Because treatment t_i is randomly assigned in cells A, B, C, and D unconditional on supervisor selection, we can estimate μ' and $\beta = (\beta_0, \beta_1)$ via ordinary least squares.

4.2   Estimating the Marginal Treatment Effects of the Program

A strictly positive value for β_1 in estimating equation (4) is a necessary condition for a decentralized approach to be preferred, but it is not a sufficient condition. Two other considerations are pertinent. First, is β_1 large enough to justify paying the cost d of decentralization? Second, what would the average treatment effect be at scales other than 50 percent? We asked supervisors to select half of their AEAs but this pilot implementation does not directly tell us what β_1 would be at different selection shares. In this section we develop a method for tracing out the impact for all possible roll-out scales under different implementation regimes that vary the degree of informational advantage associated with decentralization.

Marginal treatment effects under different selection models

In order to lay out the main intuitions surrounding the value of decentralization, our theory considered the stark contrast between a totally uninformed principal and a fully informed, benevolent supervisor. We will allow for intermediate cases in our empirical approach – the econometric operationalization of the theory will in fact extend it in two directions. First, we allow for supervisors to be less than fully benevolent. Second, we allow them to be less than perfectly informed about the responsiveness of AEAs to treatment. In addition, this framework will allow us to consider a principal who is partially informed.

Each organizational situation – decentralization or centralization under different informational capabilities of the principal – will be modeled as leading to the selection of AEAs according to a suitably defined latent index model.

The supervisor's perception of the value of treating a particular AEA is modeled as a function of observables X_i and unobservables u_i according to the function $\Gamma' X_i + u_i$. In what follows, we develop some structure to link this empirical object to the theory.

In the case of decentralization, supervisors select AEAs according to the value v_i they believe would be attained from treating AEA i,

$$v_i = \beta_i(X_i, \eta_i) + \psi_i(X_i, \zeta_i), \tag{5}$$

where $\beta_i(X_i, \eta_i)$ represents the heterogeneous effect of receiving the cell phone and ψ_i is a preference for treating AEA i that depends on X_i and an independent, idiosyncratic preference term ζ_i. A benevolent supervisor would only select AEAs based on an index $v_i = \beta_i(\cdot)$. Thus, the additional term ψ_i captures the potential non-benevolence of the supervisor. In addition, supervisors may not observe η_i perfectly but, instead, observe a signal $\theta_i = \eta_i + \xi_i$, where $\xi_i \sim F_\xi(\cdot)$ is a white noise (hence mean zero) term. As the variance of ξ_i goes to zero, the supervisor gets closer to being perfectly informed. Given the random element (ξ), the supervisor faces uncertainty. A risk neutral supervisor will assign monitoring technology to AEAs depending on the expected value $\mathbb{E}\{v_i | X_i, \theta_i, \zeta_i\}$. The expectation is taken over ξ, and conditional on ζ, since to the supervisor the former represents noise while the latter captures preferences.

Given a selection criterion (such as v_i), and a well-defined measure of diversity across AEAs as given by a joint distribution over (X_i, θ_i, ζ_i), the supervisor rank orders all AEAs according to the expected value $\mathbb{E}\{v_i | X_i, \theta_i, \zeta_i\}$, with minimum element $\underline{Ev}$ and maximum element $\overline{Ev}$. We assume

18

there is enough variation that the rank order is strictly monotonic. Therefore, any roll-out of scale m under a selection criterion based on v_i implies treating all AEAs who satisfy $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\} \geq c_p(m)$, where $c_p(m)$ is a putative cost (hence the subscript). This cost is putative in the sense that it is the marginal cost of treatment that the supervisor would have to perceive in order to decide to treat a share m of AEAs. Thus, $c_p(m)$ satisfies $\frac{dc_p}{dm} < 0$, $\lim_{m\to 0} c_p(m) = \overline{Ev}$, and $\lim_{m\to 1} c_p(m) = \underline{Ev}$. These conditions say that for the supervisor to want to treat more AEAs, the putative cost of treatment must be lower; for the supervisor to treat no AEAs, the putative marginal cost of treating a single AEA must exceed the benefit of treating the most valuable AEA; and that for the supervisor to treat all AEAs, the expected desirability of treating the least valuable AEA must cover the putative cost. When $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\} \geq c_p(m)$, the selection indicator denoted by $D_i^M(X_i, \theta_i, \zeta_i, c_p)$ takes the value 1, and 0 otherwise.

The fundamental difference between AEA characteristics X_i and η_i is that elements in the vector X_i are potentially observable by a sophisticated principal who can gather and analyze data. Elements in X_i could contain AEA-related demographic and psychometric data. The term η_i is fully unobservable to the principal, and can potentially be known only to a supervisor who establishes a more personal connection with the AEA. Thus, decentralization has two potential informational advantages: supervisors may (or may not) know and use data on X_i better than the principal, and they are the only ones who can potentially know something about η_i. To the extent that η_i enters the function $\beta_i(.)$, the supervisor will have an unassailable informational advantage over the principal.

To make further progress, we need to parameterize the dependence of $\mu_i(\cdot)$, $\beta_i(\cdot)$, and $\psi_i(\cdot)$ on X_i. We parameterize each of these linearly. Slightly abusing notation, and anticipating our assumption that η is mean zero, we can re-write equations (2) and (5) respectively as,

$$
\begin{aligned}
s_i^* &= \underbrace{(\mu'X_i + \varepsilon_i)}_{\mu_i(\cdot)} + \underbrace{(\beta'X_i + \eta_i)t_i}_{\beta_i(\cdot)} \\
&= \mu'X_i + (\beta'X_i) \times t_i + \varepsilon_i + \eta_i \times t_i.
\end{aligned} \tag{6}
$$

and

$$
v_i = \underbrace{(\beta'X_i + \eta_i)}_{\beta_i(\cdot)} + \underbrace{(\psi'X_i + \zeta_i)}_{\psi_i(\cdot)}. \tag{7}
$$

Marginal treatment impact under an uninformed principal An uninformed principal knows nothing about individual values of β_i, so she can only randomly select which AEAs should be placed under the new technology. Using equation (6), given a scale of roll-out m (the share of

AEAs to be treated), the total treatment effect on expected performance is

$$\int_{X_i} ((E_{\varepsilon,\eta}(s^*|t=1,X_i) - E_{\varepsilon,\eta}(s^*|t=0,X_i))m)d\Xi(X_i) = m\beta'\bar{X},$$

where Ξ is a cumulative distribution function describing variation in the vector X, which is unobservable to a fully uninformed principal. This equation says that if no AEAs are treated, the total gains are zero. If all AEAs are treated, the total gains are equal to the average treatment effect of the intervention. If a partial measure $m \in (0,1)$ is treated, the total gains are proportional to roll-out m, and the marginal impact of enhancing roll-out is always the average impact $\beta'\bar{X}$.

Marginal treatment impact under decentralization A supervisor observes each AEA's characteristics (X_i, θ_i, ζ_i), and selects AEAs to treat according to the value of the expected index $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\}$ as given by,

$$
\begin{aligned}
\mathbb{E}\{v_i|X_i,\theta_i,\zeta_i\} &= \underbrace{(\beta'X_i + \mathbb{E}\{\eta_i|X_i,\theta_i\})}_{\mathbb{E}\{\beta_i(\cdot)|\theta_i,X_i\}} + \underbrace{(\psi'X_i + \zeta_i)}_{\psi_i(\cdot)} \\
&= \underbrace{(\beta' + \psi')}_{\Gamma'}X_i + \underbrace{(\mathbb{E}\{\eta|X_i,\theta_i,\zeta_i\} + \zeta_i)}_{u_i}.
\end{aligned}
\tag{8}
$$

This equation is important for our linking the theory of AEA selection with the empirics. The AEA observables in X_i matter both because they affect response to treatment (through β', as in the theory), but also because supervisors may have biases (through ψ'). Unobservables in u_i may also reflect components that affect response to treatment (through η) and biases (through ζ).

A key hurdle is that we do not have a direct measure of $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\}$. We only observe the supervisor selection decision D_i^S. To recover Γ, we further assume that η_i, ξ_i, and ζ_i are mean zero, normally distributed random variables with variances σ_η^2, σ_ξ^2, and σ_ζ^2, respectively. Given all of these distributional assumptions, the variable u_i can be characterized as drawn from Φ, a cumulative Normal $(0, \sigma_u^2 = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\xi^2}\sigma_\eta^2 + \sigma_\zeta^2)$ (this stems from the fact that the supervisor is Bayesian and updates her expectation of η upon observing θ). Since a supervisor selects an AEA for treatment whenever $u > -\Gamma'X_i + c(m)$, the supervisor's selection criterion D_i^S takes the familiar form of a probit model:[8]

$$Pr\{D_i^S = 1|X_i\} = \Phi\left(\frac{1}{\sigma_u}(\Gamma'X_i - c_p(m))\right)$$

---

[8]In our estimation, c_p is not separately identified from the constant vector in X_i and thus we normalize it to zero. We revisit c_p in section 7.

Under our assumptions, standard arguments yield $\mathbb{E}\{\eta_i|u_i\} = \frac{\sigma_{\eta u}}{\sigma_u^2} u_i$, where $\sigma_{\eta u} = \left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\xi^2} \sigma_\eta^2 \right)$. Using iterated expectations, if $E(\eta|u) = \frac{\sigma_{\eta u}}{\sigma_u^2} u$, then $E(\eta|D, X_i, m) = \frac{\sigma_{\eta u}}{\sigma_u^2} E(u|D, X_i, m)$. And using a basic property of truncated normals, we know that,

$$E(u|D=1) = E\left(u|u > -\Gamma'X_i + c\right) = \sigma_u \frac{\phi\left(\frac{-\Gamma'X_i+c}{\sigma_u}\right)}{1 - \Phi\left(\frac{-\Gamma'X_i+c}{\sigma_u}\right)} \tag{9}$$

$$E(u|D=0) = E\left(u|u > -\Gamma'X_i + c\right) = -\sigma_u \frac{\phi\left(\frac{-\Gamma'X_i+c}{\sigma_u}\right)}{\Phi\left(\frac{-\Gamma'X_i+c}{\sigma_u}\right)}. \tag{10}$$

Plugging (9) and (10) into $E(\eta|D, X_i, m)$, it follows that $\mathbb{E}\{\eta_i|D_i^S, X_i, m\} = \frac{\sigma_{\eta u}}{\sigma_u} \frac{\phi(\frac{1}{\sigma_u}(\Gamma'X_i - c_p(m)))}{D_i^S - \Phi(\frac{1}{\sigma_u}(\Gamma'X_i - c_p(m)))} \equiv \frac{\sigma_{\eta u}}{\sigma_u} \lambda(D_i^S, X_i, m)$. This expression characterizes the expected value of the unobservable information supervisors have about an agent's responsiveness to treatment. This expression enters into the overall expected performance of an AEA conditional on being put into treatment. Taking conditional expectations in equation (6), we get [9]

$$\mathbb{E}\{s_i^*|X_i, D_i^S, t_i, m\} = \mu'X_i + (\beta'X_i) \times t_i + \mathbb{E}\{\varepsilon_i|X_i, D_i^S\} + \mathbb{E}\{\eta_i|X_i, D_i^S, m\} \times t_i$$

$$= \mu'X_i + (\beta'X_i) \times t_i + \frac{\sigma_{\eta u}}{\sigma_u} \lambda(D_i^S, X_i, m) \times t_i. \tag{11}$$

Note that ε_i and η_i are independent of X_i *by definition* and independent of t_i by way of the randomized experiment. Thus we can estimate equation (11) via a two-step procedure using OLS. The first step allows us to estimate the selection model that will yield $\lambda(D_i^S, X_i, m)$ and the second step yields estimates for the coefficients in equation (11). We show these estimates in Section 6.3.

Equation (11) is the crucial piece to estimate the marginal treatment impact of the intervention under different scenarios of decentralization and informational advantage. To see this, consider first the simplest case where neither the principal nor the supervisor can observe any AEA traits so the vector X_i is constant. The expected index on which the supervisor selects is $\mathbb{E}\{v_i|\theta_i, \zeta_i\} = \mathbb{E}\{\eta|\theta_i, \zeta_i\} + \zeta_i = u_i$. Given the 50 percent roll-out in the experiment, we know that under decentralization, the total treatment impact of 50 percent roll-out is $\beta_0 + \beta_1$ from OLS estimation of equation (4). In order to trace the marginal treatment impact at any other roll-out m, we only need

---

[9] We do not impose any restrictions on u_i and ε_i and so also include $\lambda(D_i, X_i, m)$ as a main effect without any interaction with t_i. This parameter (along with μ) is not of direct interest to us and is not required for identification, but may improve the efficiency of the other estimates.

to consult the value of u_i at the m percentile in the Normal distribution of u_i.[10] Thus, it is possible to trace the total treatment gain from following the supervisor's selection criterion for all m.

As the expression $\mathbb{E}\{\eta|\theta_i, \zeta_i\} + \zeta_i = u_i$ makes clear, we cannot tell whether a supervisor's selection is due to information on unobservables that affect true responsiveness to treatment (η) or due to unobservables that make the supervisor select an AEA for other reasons (ζ). But if $\beta_1 > 0$ we know the supervisor gets a precise enough signal on η, and places enough weight on it, so that even if she is biased in her choices, her selection yields higher treatment impact than selecting the AEAs at random.

In most situations, supervisors will know characteristics of their AEAs, and so the expected index $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\}$ on which supervisors select will indeed be a function of X_i. In this situation, each expansion of roll-out will imply extending treatment to new AEA types, where the type space as seen by the supervisor is some unidimensional path in a higher dimensional space of traits X_i and the supervisor-only observed u_i. The analyst does not observe u_i, but can form an expectation of it conditional on an AEA with traits X_i being selected. Knowing traits X_i and a conditional expectation on u_i for an AEA being selected at a given level of roll-out, equation (11) delivers the treatment impact. Thus, it is possible to derive the total treatment gain from following the supervisor's selection criterion for all m.

Further uses of the model: evaluating supervisors, and the potential for sophisticated centralization We have now described ways to obtain marginal treatment impacts at varying roll-outs for the cases of an uninformed centralized principal and an informed supervisor. But the selection model laid out in this section can be put to other uses. First, it is possible to evaluate the supervisors in a more complete way than simply checking whether they have an informational advantage over the principal. We can ask how much of their advantage is due to their knowledge of traits observable by the principal (X_i) versus things the principal cannot expect to learn (η). Moreover, the analyst can econometrically evaluate the extent to which supervisors make optimal use of observable data in X_i.

Second, with gains in the ability to gather and process data, a principal could learn some traits of her AEAs, captured by X_i. This opens up consideration to a new class of counterfactuals, with a natural one involving the marginal treatment impact for varying m for a decision maker that knows X_i but does not observe θ_i. Thus, we can ask whether a sophisticated centralized principal can emulate or surpass the performance of supervisors despite her informational disadvantage. We perform these exercises in Section 7.

---

[10]We do not recover separate values for σ_u and σ_η, since all parameters are scaled by σ_u in the probit regression.

Discussion We have presented a heterogeneous treatment effect model where supervisors have private information about the treatment effects. Equation (11) shares the same functional form as the "Heckit" selection model. However, in most settings where the Heckit is applied the treated are also the selected, $t_i = D_i^S$. In most settings like ours where selection D_i^S is assigned according to ε_i or η_i, the control function $\lambda(\cdot)$ must be included in estimation to account for the non-random censoring of potential outcomes, the raison d'etre for the literature on selection correction. However, control functions require credible instruments; without an instrumental variable that could be excluded from the outcome equation, identification relies on the parametric assumption governing u_i. Stated differently, if one assumed u_i was instead uniformly distributed and estimated the corresponding selection equation via OLS, $\lambda(\cdot)$ would be collinear with the vector X_i (Olsen, 1980) and thus would not be separately identified. Even in contexts where there are credible instruments that generate experimental variation, treatment effect heterogeneity among never-takers and always-takers are obtained by using selection models to extrapolate from instrument-implied local average treatment effects (Heckman and Vytlacil, 2005; Kline and Walters, 2019).

In our context, however, treatment t_i is independently and randomly assigned, and not equal to selection D_i^S. While supervisor preferences are elicited, they are not used to determine assignment in our main sample. This means that we neither have censored potential outcomes nor always-takers and never-takers. Instead, we have both a randomized experiment with full compliance and also information about supervisor preferences that were not implemented, allowing us to credibly estimate treatment effects along the full distribution of η_i, an exercise that requires no extrapolation. Because we observe treatment effects for non-selected AEAs (i.e., those with $D_i^S = 0$), even if misspecified, $\lambda(D_i^S, X_i)$ is just a transformation of D_i^S and X_i, and with inclusion of controls, its independent variation is driven primarily by D_i^S.[11]

5 Data

We collected two main sources of data. The first is a survey of AEAs. Each AEA and supervisor independently filled out answers on a paper questionnaire with survey enumerators available to answer any questions. The survey contains questions regarding the AEAs' demographics and work history, the digit span test measuring cognitive ability, and the Big-5 inventory (John et al.,

---

[11]Thus, if one wanted to stick with OLS in the first stage rather than a Probit, while continuing to assume a linear conditional expectation function $\mathbb{E}\{\eta_i|u_i\} \propto u_i$, the coefficient on $\lambda(\cdot)$ in the second stage would be numerically equivalent to estimating an OLS regression in one step with $\frac{D_i^S}{2}$ in place of $\lambda(\cdot)$. This is a result that follows immediately from the Frisch-Waugh theorem.

2008). We combine this five-dimensional inventory into two higher-order personality traits called stability and plasticity. Stability combines the traits neuroticism, agreeableness, and conscientiousness which have been found to predict earnings and job attainment, such as the tendency to remain emotionally stable and motivated and be organized and thorough. Plasticity, which aggregates extraversion and openness, is a measure of a person's gregariousness and openness to new experiences. These two meta-traits tend to account for much of the shared variance among the lower order dimensions (DeYoung, 2006).

The second source of data we have is two rounds of farmer phone surveys. One round was conducted after the first wave of phones was distributed and one round was conducted after the second wave of phones was distributed. We called farmers who were beneficiaries of the AEAs and asked questions about their interactions with the AEAs such as how often they saw the AEA and how satisfied they were with his work.

The timeline of events is as follows. In March of 2014, the ALAT-level supervisors chose which AEAs they would like to prioritize for receiving a phone with the objective of expanding effort in service to farmers. The first wave of phones was distributed to the AEAs between May and July, 2014. These phones were given to AEAs in cells *B* and *D* (all AEAs in the 100% coverage ALATs) and cell *E* (selected AEAs in the 50% coverage ALATs) as well as a randomly chosen third of AEAs in the small ALATs. Individuals from the central ministry traveled across the country to meet with the AEAs who were scheduled to receive phones, distribute the phones to them, and teach them how to use the phones. This process took over two months because it involved 19 meetings spread across the country.

After the first wave of phones was distributed, we conducted two types of data collection. From July through September, 2014 we conducted the first round of farmer phone surveys. Additionally, during September 2014, we conducted the survey of all AEAs as well as their supervisors. We treat AEA characteristics such as sex, age, years of education, and the personality indices as being fixed and not affected by the roll-out of the phones. On the other hand, we treat variables such as the AEAs' perceptions of whether their supervisors know where they are during the working week as potentially being affected by the roll-out of the phones. In the control group, AEAs in ALATs where no AEAs received phones, these responses should not be impacted by the roll-out of the phones.

After completing the first round of surveying, the second wave of phones was distributed in February and March, 2015. These phones were given to the AEAs in cell *F* (the remaining AEAs who were not selected in the 50% coverage ALATs) as well as another randomly chosen third of AEAs

in the small ALATs. We then conducted a second round of farmer phone surveys in April and May, 2015. The Ministry of Agriculture planned to distribute phones to all AEAs who had not yet received one before the end of 2015 but in the end did not do so.[12]

The ministry did not give any phones to AEAs who were not on our randomized list. There were a few cases in which phones broke down or sick AEAs were not able to pick up their phones. For this reason we look at intent-to-treat (ITT) estimates using our initial random assignment.

In early 2014, we were given full information, including job title, job location, and client names and phone numbers for 368 agricultural extension agents - 139 supervisors and 229 AEAs. In September 2014, we were able to interview 301 of these – 119 supervisors and 182 AEAs. We interviewed 79% of the AEAs in our original administrative data, 15% no longer worked for DEAg, and 6% were absent the day of the surveying.

The job description of an AEA involves working with 80 farmers. Thus, it is no surprise that the median AEA in our data listed the names of 80 farmers with whom he worked; the mean of the distribution is 75 with a standard deviation of 26. The median AEA in our data listed phone numbers for 78% of the farmers he served, while the mean share listed is 73%. These numbers vary very little for AEAs versus supervisors.

In total, we called 2,635 farmers in the first round and 2,642 in the second round for the 182 AEAs who responded to the AEA survey.[13] Of those, 68% led to completed surveys.[14] Conditional on completing the survey, 70% of farmers confirmed that the AEA that had provided their number worked with them and thus were asked more detailed questions about their interaction with that AEA.[15] This leads to 2,519 completed farmer phone surveys.

---

[12]The government was very happy with the phones, but their priorities shifted to focusing on the extreme poor. Thus phones were taken away from the AEAs of DEAg who work with the poor, and given to AEAs in a different organization which focused on the extreme poor.

[13]We conducted two rounds of farmer phone surveys, but we wanted to leave open the possibility of conducting three rounds. For AEAs and supervisors in multi-AEA ALATs who listed 75 or more farmer phone numbers, we randomly chose 75 farmers to call and then randomly divided them to call 25 farmers in each of three rounds. For those who listed fewer than 75 farmer phone numbers, we randomly divided their farmers into thirds to call in each of the three rounds. Similarly, for AEAs and supervisors in single-AEA ALATs who listed 24 or more farmer phone numbers, we randomly chose 24 farmers to call and then randomly divided them to call 8 farmers in each round. For those who listed fewer than 24 farmer phone numbers, we randomly divided their farmers into thirds to call in each of the tree rounds.

[14]In 18% of cases, we reached voicemail on all five tries, 7% of cases were wrong numbers, 4% were out-of service phone numbers, and 2% of farmers did not agree to complete the survey.

[15]We first asked the farmers to talk about any AEAs with whom they worked and did not offer up the name of the AEA we had on record for them. We only asked the farmer about the specific name we had on record if either the farmer worked with an AEA whose name he couldn't remember or if he did not list the name of the AEA we had on record on his own.

Table 1 presents sample means and a randomization check of the cellphone assignment for various AEA characteristics. The table shows differences between treated and control small ALATs (column 3) and 100% coverage and control large ALATs (column 6). Overall, the average AEA age is 38, and 71% of AEAs are male. The AEAs were able to recall an average of 5.3 digits in the memory digit span test, which is a commonly-used measure of cognitive ability.[16] AEAs are required to travel an average of 12 kilometers to visit a given farmer. Overall the results in Table 1 suggest that treatment, which was randomized at the ALAT level, is balanced.

We present some additional balance tables in the appendix. In Table A1, we show that balance holds when we compare the 100% coverage ALATs (i.e. cells *B* and *D*) and the 50% coverage ALATs (i.e. cells *E* and *F*) to the control group. In Tables A2 and A3, we also check for balance on a set of ALAT-level characteristics from the population and agricultural censuses, as well as the results of the 2013 presidential elections. In looking at 18 comparisons across the two tables, only one shows significant imbalance across treatment and control. Across a different dimension, Table A2 shows that large ALATs are located in larger and less rural districts than the small ALATs.

6 Results

In this section, we begin by estimating the impact of the cell phones on AEA performance. According to the model, under certain conditions (cell phones improve monitoring and there are sufficiently many AEAs who respond positively to it), the increase in monitoring induced by the phones should boost the effort levels of the AEAs and thus increase the number of farmers visited. Subsequently, we test whether the impact of cell phones was higher among the AEAs who were selected by the supervisors, which would be the case if supervisors had the ability and willingness to target the AEAs with highest responsiveness to treatment. Next we look for, but fail to find, evidence of spillovers and Hawthorne effects. Finally, we estimate heterogeneous treatment effects, which we use to evaluate impacts under various counterfactual scenarios with different scales of roll-out.

---

[16]For the digit span test, the enumerator read out loud a random number that the AEA was then required to recite back. The test began with a number that was two digits long and incrementally increased the number of digits until the AEA could no longer recall a number correctly on both of two chances.

6.1 Increased Monitoring and Performance

As we discussed in Section 2, the primary task of an AEA is to visit farmers. In columns (1) through (5) of Table 2, we estimate the impact of the phone on whether the farmer reported having been visited by the AEA in the last week. In columns (1) through (3), our estimation sample includes all AEAs in the small and large ALATs, excluding those randomized into the partial treatment ALATs (cells E and F). In column (1), we present the estimates without any additional controls. In column (2), we add a set of basic controls (e.g., age and gender), and in column (3), we further augment the specification to include controls measuring AEA type (e.g., Big 5 personality meta-traits and digit span). In column (4), we re-estimate the specification presented in column (3), excluding the small ALATs.

We find that the increase in monitoring leads AEAs to visit their farmers more often. They are approximately six percentage points more likely to have visited a given farmer in the past week, which is an increase of 22% over the control group. As expected given the random assignment, the estimated impact varies little with the different controls added and when we restrict estimation to the large ALATs, which will be our main sample moving forward. Appendix Table A4 shows both wild cluster bootstrap and randomization inference p-values which help account for the small number of ALATs under study and the coefficient on treatment status retains its significance. Overall, the demographic and personality-based controls have little predictive power.[17]

Supervisors are in charge of both supervising the AEAs in their ALAT as well as serving their own farmers. In column (5), we test the impact of the phone on supervisors' visits to their farmers. We find a small and insignificant impact (point estimate = -0.008; clustered standard error = 0.035). This suggests that the impact of the phone is related to the greater monitoring ability it gives supervisors and not due to productivity-enhancing functions of the phone (e.g., ease in communication), which would have the same effect on both supervisors and AEAs. As a further check, AEAs were asked whether they agreed with the statement that their supervisor usually knows where they are during the work week. In column (6), we see that having a phone significantly increased the extent to which AEAs agreed with this statement.

While the treatment led to more visits, this does not necessarily imply that the AEAs are exerting more effort. AEAs could be making more visits but making them shorter. In column (7), we do not find evidence support this idea. The point estimate is statistically insignificant and suggests that treated AEAs spend only one percent less time (approximately one minute) on each visit.

---

[17]In results not shown here we look at short and long-run impacts of the phones (using the first and second rounds of farmer phone surveys), and find that they are quite similar. The impact of the phones does not diminish over time.

In Appendix Table A5, we examine the effects of the treatment on other dimensions of AEA performance. We consider four additional measures: 1) how satisfied the farmer is with the AEA (1=not at all, 2=somewhat, 3=very); 2) an indicator for whether the farmer reported that the AEA conducted a helpful training session; 3) an indicator for whether the farmer did not report that the AEA was not helpful at all; 4) and the first principal component for the three measures, with higher values indicating better performance. In general, we find that the additional monitoring improved performance along these dimensions as well, although the estimates are measured with less precision. Based on our polychoric principal component measure, the treatment improved aggregate performance by 0.13 of a standard deviation (clustered standard error = 0.07). [18]

6.2 Do Supervisors Have Useful Information?

Recall that our model assumes AEAs differ in their responsiveness to enhanced monitoring and that supervisors know the AEAs' responsiveness. If supervisors wish to increase the number of farmers visited, then when tasked with the responsibility of assigning increased monitoring, they should target the AEAs for whom a larger increase in performance ought to be expected. Our research design allows us to test this directly.

Prior to the randomization, supervisors identified which half of their AEAs they believed should receive the phones. Given these selections, we test for the value of information using a simple difference-in-differences estimator for our sample of large ALATs. We compare the performance of AEAs who were selected and received the phone against those who were selected but did not receive the phone, net of the difference in performance between those who were not selected and received the phone against those who were not selected and did not receive the phone (i.e., $(B - A) - (D - C)$).

From Table 3, we see that the effects of the phones on performance are entirely driven by the effects on the AEAs selected to receive the phone prior to the randomization. These AEAs increased the share of farmers visited in the last week by approximately 15 percentage points. Compared to the selected AEAs in the control group, this effect represents a substantial increase of 58 percent. We present this result with bootstrapped clustered p-values in Appendix Table A7 and with randomization inference p-values in Appendix Table A8 and the significance remains. From column (4), we also see that selected AEAs in the control group are 4 percentage points less likely to have visited

---

[18]All four measures are significantly correlated with our preferred measure of productivity, AEA visits (see Appendix Table A6). Farmers who receive more visits from their AEAs are also more satisfied with their AEAs, more likely to think the AEA conducts helpful training sessions, and less likely to find the AEA useless.

their farmers relative to the non-selected. This potentially suggests that supervisors target the least productive AEAs, although this difference is not statistically significant. While our preferred specification focuses on cells *A*, *B*, *C*, and *D*, the 0% and 100% coverage ALATs, in Appendix Table A9, we present results incorporating the AEAs from cells *E* and *F* (the 50% coverage ALATs) and results are similar.

Spillover and Hawthorne Effects   The theory we laid out in Section 3 assumes the absence of spillovers across agents. An advantage of our experimental design is that we can test this assumption directly. Recall from Figure 2 that for the ALATs assigned to the 50% coverage treatment, only the selected AEAs (cell *E*) received phones in the first wave, while the non-selected (cell *F*) did not. Therefore we can test for the presence of spillover effects on the untreated by comparing the performance of the AEAs in the first survey round in cell *F* (i.e., non-selected AEAs who did not receive a cell phone but whose selected coworkers did) to the AEAs in cell *C* (i.e., the non-selected in the control group who, like their selected coworkers, did not receive a phone).

The main potential source of such spillovers is a reallocation of supervision effort from the treated to the untreated when only a fraction of a supervisor's AEAs are treated with phones. In that case, not only would the non-selected workers without phones whose co-workers had phones perform better than those whose co-workers did not, but those selected workers with phones whose non-selected co-workers also had phones would perform better than those whose non-selected co-workers did not have phones. That is, AEAs in cell *B* would perform better than those in cell *E* in the first round. We test for this second possibility as well.

In column (1) of Table 4, we add two new variables to the specification presented in column (4) in Table 3. The first new variable, labeled 'Spillover *(Non-selected)*,' is an indicator equal to one for the AEAs in cell *F* in survey round one, i.e., those who were not selected but whose selected co-workers received a phone, and zero otherwise. If there are spillovers resulting from supervisors reassigning supervision effort from the treated in *E* to the untreated in *F*, this coefficient ought to be positive. The second new variable, labeled 'Spillover *(Selected)*,' is an indicator equal to one for the AEAs in cell *B* in both rounds and cell *E* in survey round 2, i.e., for those who were selected and treated, and have treated co-workers. The coefficient on this variable measures the additional effect of having treated coworkers on the selected who were treated.

Neither the non-selected untreated, nor the selected treated, perform any differently whether or not their coworkers are treated with phones. The spillover on the non-selected from having co-workers with phones is small at -0.0137 and not statistically significant. This finding suggests the absence

of spillovers, either directly in the form of peer effects, or through effort substitution by supervisors. Meanwhile, the spillover on the selected from having co-workers with phones is of larger magnitude at -0.0534 but this is not statistically significant. The overall effect of treatment for a selected AEA whose non-selected co-workers do not have phones is 16 percentage points. This is consistent with our interpretation that supervisors can supervise all AEAs with phones regardless of what share of their AEAs have phones, and that supervisors possess useful information about their AEAs.

An alternative story, akin to a Hawthorne effect, is that supervisors may have increased their supervision effort on those treated AEAs whom they had selected. They may have done this to make sure their selected AEAs performed better under the new monitoring technology, sending a signal that their selection had been judicious. In this case, our findings would not necessarily reflect supervisors' informational advantage but rather their additional effort.

There are four reasons we find this explanation unlikely. First, an ancillary test points to the absence of Hawthorne effects. We exploit the fact that some ALATs are small (a single AEA), and thus supervisors were not asked to identify which half of the AEAs ought to be treated. If Hawthorne effects are present, the average treatment effect in multi-AEA ALATs should be stronger than in the single-AEA ALATs where no selection was made, as the signaling motive is absent in the latter. We estimate the effects of the phone distinguishing between single and multi-AEA ALATs. As we see in column (2) of Table 4, we cannot reject the hypothesis that the treatment effects are similar across small and big ALATs. If anything, the point estimate of the interaction term is positive, suggesting that the treatment effect might be slightly larger in the single-AEA ALATs. This test assumes that the propensity to have been "selected" is balanced across small and large ALATs. In column (3) we test whether our point estimates change when we control for our AEA characteristics, which would be the case if AEAs are systematically different across the two types of ALATs. Based on our rich set of observables, we do not find any evidence of AEAs sorting by ALAT size.

The data speak against potential Hawthorne effects in a second way. If Hawthorne effects were present, then we might expect them to be more pronounced in the 50% coverage treatment arm (where supervisors' selection rule was enforced) relative to the 100% coverage treatment. Alternatively, one could conjecture that the Hawthorne effects would be less pronounced in the 50% coverage treatment arm if, for instance, supervisors in the 100% coverage treatment felt slighted. But as we noted from the results presented in column (1) of Table 4, the effects for the selected AEAs across the two treatment arms are similar in magnitude in round 1, rejecting this possibility.

Third, the structure of incentives in the field makes Hawthorne effects unlikely. Supervisors had little incentive to prove to the Ministry that they could select the most responsive AEAs. They

had no reason to think that the Ministry would or even could evaluate their selection process, let alone reward or punish supervisors for the selections made. Those beliefs would have been correct since the Ministry never performed, nor planned to perform, such an evaluation. Given the lack of incentives, it seems unlikely that supervisors would exert costly effort to make sure the selected AEAs who were treated performed better.

Fourth, if Hawthorne effects were driving our results, then selected AEAs should have larger treatment effects in both observable and unobservable dimensions. For example, we will see later that being married significantly predicts being selected but is negatively associated with the treatment effects. For such a result to be consistent with Hawthorne effects, the supervisor would have to perceive that they were being judged primarily on dimensions that could not be predicted from a simple probit model and respond in kind. We believe such a manipulation to be unlikely in our context. We explore both predictors of selection and treatment effect heterogeneity in the next section.

In sum, we find strong evidence that the phones increase AEA effort and that supervisors possess useful information regarding which AEAs' performance will improve most after receiving a phone. This indicates that in the absence of enough phones to treat all AEAs, it may be valuable to decentralize the assignment of phones to supervisors. This begs the question of what characteristics the supervisors used to create their prioritized list and the extent to which supervisors used information analysts could hope to obtain. The next subsection answers these questions.

6.3   Heterogeneous Treatment Effects

In Table 5, we present estimates from a Probit regression, in which the dependent variable is an indicator for whether the AEA was selected by the local supervisor.[19] Based on standard observable characteristics, we find that supervisors tended to select AEAs who were younger, married, and had to travel further distances to visit their farmers. In terms of their personality traits, supervisors were more likely to select AEAs with lower levels of the Big-5 Stability meta-trait. Individuals with higher stability scores may be more likely to stay motivated and have better relationships with their supervisors without extra monitoring technologies.

Interestingly, we also find that supervisors of the large ALATs, all but one of whom are registered

---

[19]Appendix Table A10 shows the same regression with bootstrapped *p*-values. Our preferred specification uses the estimation sample from Table 3, including only the 0% and 100% coverage ALATs. In Appendix Table A11, we present similar results incorporating the AEAs from the 50% coverage ALATs.

with the incumbent political party, are significantly less likely to place AEAs registered with the incumbent party under increased monitoring. This either suggests that supervisors are acting non-benevolently or, as we will subsequently test, that party affiliation serves as a marker for those who are less likely to respond to treatment.

Despite the richness of our data, our ability to predict the choices of the supervisors is fairly low: the highest pseudo R^2 is only 18.5%. This opens the possibility that supervisors are also selecting AEAs based on unobservable but productive characteristics (η) or unobservable and idiosyncratic characteristics (ψ), features that are not captured by demographic traits or indicators of cognitive and non-cognitive ability. Ultimately, the only way to determine whether there could be an advantage to decentralization is to rely on our experimental design, and ask whether supervisors select AEAs who will be more responsive to treatment.

In Table 6, we present a series of second stage estimates based on Equation 11.[20] In column (1), we present a specification without any additional controls or interaction terms, whereas in columns (2) and (3) we include additional controls along with their interactions with the treatment indicator. For columns (2) and (3), the first stage regressions correspond to the ones presented in Table 5.

The key finding in Table 6 concerns the inverse Mills ratio and particularly its interaction with treatment. As we explained in Section 4, the inverse Mills ratio captures the expected unobservable traits that recommended an AEA for selection by the supervisor. Because no controls were included in column (1), the coefficient on the inverse Mills ratio interacted with treatment in that column replicates the findings from Table 3 that supervisors are selecting individuals with higher treatment effects.

When we allow the effects of the treatment to vary by the characteristics that we found were predictive of the likelihood of selection (columns (2) and (3)), we find that the inverse Mills ratio is still highly predictive of responsiveness to treatment, direct evidence that the unobservable reasons supervisors are selecting AEAs are, on balance, productive. In addition to the unobservable traits, the treatment effect also varies by the cognitive ability of the AEAs; those who performed worse on the digit span test exerted more additional effort in response to the treatment. Moreover, in column (2) it appears that members of the incumbent party respond less to the treatment. But, once we account for the differential effects of cognitive and non-cognitive ability in column (3), this effect loses significance. We also do not find any differential effects by age, even though older AEAs were much less likely to get selected. This suggests that some favoritism towards older AEAs may

---

[20]Appendix Table A12 shows the same regression with bootstrapped p-values. See Appendix Table A13 for similar results including the 50% coverage ALATs.

have influenced the supervisors' targeting.

The results so far raise several questions. What is the basis for the supervisors' informational advantage? At a given cost of decentralization, does the informational advantage justify the cost? To answer these questions, one needs to know two elements. First, what is the scale of roll-out anticipated? Under decentralization, the anticipated scale of roll-out should be whatever is optimal, and this motivates the need to identify that optimal level. Second, how much information does the central authority have? In the next section, we apply the framework introduced in Section 4 to provide answers to these questions.

7   Counterfactuals

In this section, we exploit our heterogeneous treatment effects model to compute counterfactual treatment effects under alternative selection rules. This allows us to assess the benefits of decentralization relative to centralization under different informational assumptions.

The first step is to define a counterfactual aggregate benefit under an arbitrary selection rule D_i^{CF} as:

$$\Delta Y^{CF} = \int \mathbb{E}\{\beta | D_i^{CF} = 1, X_i\} Pr\{D_i^{CF} = 1 | X_i\} dF_X \tag{12}$$

In keeping with the rest of our notation, we write our arbitrary selection rule as a threshold problem, $D_i^{CF}(X_i, u_i) = 1[\tilde{\Gamma}'X_i + \tilde{u}_i \geq c_p]$; because we have not made any distributional assumptions about $\tilde{u}_i$, this does not impose additional assumptions. Note that the assumed cost c_p is not directly observable, and the threshold problem is not a unique representation of the selection rule—any monotonic transformation of the latent index and c_p will yield the same choices. However, we are not trying to directly obtain either of these objects. We only predict the consequences $Pr\{D_i^{CF} = 1 | X_i, c_p\}$ and $\mathbb{E}\{\beta | D_i^{CF} = 1, X_i\}$, which map into the scale of roll-out m and the aggregate counterfactual impact ΔY^{CF}.

One example of a selection rule is the one implicitly applied by supervisors, $D_i^S(X_i, u_i)$, which anchors our portrait of what can be achieved under decentralization. Note that from our estimation of Equation 11, we have recovered $\mathbb{E}\{\beta_i(\cdot) | X_i, u_i\}$, and under distributional assumptions, the selection rule under decentralization, $D_i^S(X_i, u_i)$. Given this, we can use Equation 12 to trace out the expected treatment effects of the cell phones under decentralization for any given threshold c_p or, by extension, any scale of roll-out m. But, we can impose any other selection rule capturing different

counterfactual scenarios corresponding to different forms of centralized assignment and trace out the expected treatment effects for all roll-out levels in each scenario.

Uninformed Principal A natural, if extreme, benchmark is that of a principal who does not have any information about how best to target roll-out. In this situation, the selection rule is random allocation. At a roll-out level m, a fraction m of all AEAs receives a cell phone, and the expected total treatment effect is $m\%$ of the average treatment effect (considering, as in the theory section, a large number of AEAs who can then be approximated by a continuum).

The dotted line in Figure 3 plots this counterfactual selection rule at various roll-out levels. For example, if the principal decided to allocate the phones to everyone, then the expected aggregate treatment of the program would be 6.4 percentage points, which corresponds to the average treatment effect in column (3) in Table 6. If instead she decided to treat only half of the AEAs, then we would expect an aggregate treatment effect of only 3.2 percentage points. Thus, it is easy to see that with a random selection rule, we get a set of counterfactuals that traces a straight line from zero to the average treatment effect. In Table 7, we present our estimated treatment effects at different roll-out levels for the various allocation rules we consider. The number displayed in bold represents the largest treatment effect under a given allocation rule.

Supervisor We contrast the aggregate benefits from the random allocation rule with the benefits from the supervisor's selection rule. The supervisor's selection rule is given by $Pr\{D_i^S = 1 | X_i\} = \Phi(\frac{1}{\sigma_u}(\Gamma'X_i - c_p))$ and the expected aggregate treatment effect is $\Delta\mathbb{E}\{s_i^* | X_i, D_i^S, T_i\} = \beta'X_i + \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i)$. The solid line in Figure 3 depicts this counterfactual. Note that, by construction, the curve must cross three points: the origin (no impact with no roll-out), 0.064 at 100% roll-out, and 0.070 at 53.8% roll-out which corresponds to the share of AEAs that received the phones under the actual research design.

At each level of roll-out, the difference in impact between the supervisor counterfactual and the random allocation rule measures the benefits of decentralization under the assumption that the principal possesses no information about the AEAs. The figure shows that the difference in benefits between random allocation and supervisor selection is maximized at a roll-out scale of 53% where the additional treatment effect is over 3.5 percentage points. The optimal scale of roll-out under decentralization is not 53%, however, but at 77% where the total treatment effect is 7.7 percentage points. The total treatment effect starts to decline at a roll-out scale of 77% as we begin to assign the treatment to individuals for whom the treatment effect is negative. The existence of individuals

34

for whom the treatment effect is negative is consistent not only with our model (as ρ is allowed to be negative), but with the findings in de Rochambeau (2017), who showed that the introduction of a new monitoring device for truck drivers in Liberia lowered the productivity of the intrinsically motivated.

What underlies the informational advantage of the supervisor? One way to tackle this question is to ask how much of the supervisor's advantage is predicated on the use of information on observables X_i versus information on unobservables η_i. The dot-dash line in Figure 3 traces out the counterfactual treatment effect under the assumption that the supervisor does not use his signal of η. In other words, the dot-dash line tells us what the treatment effects would be under a supervisor who cannot use information on unobservables. In this case, the selection rule and expected treatments are only computed based on the observable (to the econometrician) traits, setting $\lambda = 0$. The dot-dash curve is much closer to the one under random assignment. This suggests that, in our setting, most of the supervisor's informational advantage is driven by access to information that is likely hard to collect for a centralized authority lacking personal contact with the AEAs.

Giving Centralization A Chance: Counterfactual Treatment Effects With A Partially Informed Principal

Minimally informed principal: Assignment based on distance traveled   Thus far, we have assumed that the principal does not have any prior information about how AEAs will respond to the program which, though extreme, may not be a wholly unreasonable approximation to the situation facing the leadership of government programs when state capacity is low. This does not suggest however that adopting a sensible heuristic might not outperform a random assignment mechanism, which would of course affect the centralization versus decentralization calculus. One such heuristic might be to simply allocate the phones to the AEAs who have to travel the farthest in order to visit their farmers. This requires some information on the work environment of AEAs, and it constitutes the case we associate with a minimally informed principal. This counterfactual is displayed in Figure 4 with a dashed light gray line. We find that this method generally outperforms random assignment (a 2.0 p.p. advantage at 50 percent coverage), but it cannot beat the supervisor at any roll-out level.

Significantly informed principal: Assignment based on predicted baseline performance   Next we consider a partially-informed principal with the capacity to gather information on AEA characteristics map them onto baseline productivity. To this end, we run a simple prediction model

35

in which, among the AEAs in the control ALATs, we regress the share of farmers visited on our set of basic and cognitive controls (see Appendix Table A14 for the estimation results). Using the estimated coefficients, we can then compute an AEA's expected productivity based on his or her observable traits. Given this information, a sensible centralized policy would be to assign cell phones starting with the AEAs who had the lowest predicted productivity, and as roll-out increases, expand coverage to AEAs with higher productivity. As we see in Figure 4, under this approach centralization would dominate decentralization at virtually all levels of roll-out. It is worth noting that the data requirements to estimate our performance-prediction model are not trivial and often beyond the capacity of government programs in several developing countries.

A sophisticated principal: Experimentation and assignment based on response to treatment
The approach in the previous section does not exhaust the possibilities open to a central authority with the capacity to gather and analyze data. The key shortcoming of that approach is that baseline performance may not be a strong predictor of responsiveness to treatment. While baseline performance reflects individual heterogeneity in the cost of effort, response to treatment depends on other cost drivers such as the disutility from receiving a reprimand.

To overcome this mismatch, a sophisticated principal can conduct a pilot experiment at a low roll-out level and establish a map between AEA observable characteristics and response to treatment. Using that, the principal can construct an assignment rule $D_i^{CF}(X_i)$ allocating phones to AEAs predicted to have the highest response to treatment and working downwards to treat progressively less responsive AEAs. Note that we are privileging principals in the "sophisticated" case because they would need to have digit span and Big-5 measures for all of their workers, which may be as much of a data constraint as running the pilot RCT.

As shown in Figure 4, this approach outperforms all others by a wide margin. The largest gap relative to the decentralized supervisor-choice approach is above 1.7 percentage points and occurs at a roll-out level of 38.4%. A sophisticated principal would be more interested in setting roll-out at its optimal scale: the maximum total treatment effect for an 'experimenting principal' is 9.0 percentage points and is achieved at a roll-out of 70%.

Note that relative to "blind centralization," which treats everyone and attains a total treatment effect of 6 percentage points, this arrangement saves on almost a third of the phones and attains almost 1.5 times the total treatment impact. Relative to the decentralized supervisor choice, the sophisticated centralized approach distributes roughly 10% fewer phones and attains roughly 1.3 additional percentage points in total treatment effect.

8 Conclusions

One of the primary benefits of decentralization is the possibility of exploiting information that is dispersed throughout the organization. Nevertheless, many organizations choose not to decentralize. Individuals in organizations pursue their own objectives, and even if there were an information gain net of biased preferences, the value of decentralized information often depends on the scale of activities, and on how costly it would be for the principal to acquire information directly. Despite the fact that the informational advantage of agents and middle managers is a maintained assumption in much of principal-agent theory, evidence of the presence and extent of that advantage has been scarce. We also have little evidence on the effects of decentralization in a context in which the scale of implementation affects average treatment effects.

In this paper, we establish that middle managers in the government hierarchy do have information which can improve the targeting of an intervention. The context of our study is an initiative by the federal government in Paraguay to introduce a new monitoring device that enables supervisors in rural areas to track their agricultural extension agents. We show that supervisors have superior information relative to central authorities about which agents should be placed under the new monitoring. In addition, we show how the aggregate value of that information varies depending on the share of agents that can be covered with the new technology. We develop an approach to trace out the total treatment effect of the intervention at all levels of roll-out, and compare decentralized assignment against centralized criteria predicated on different data requirements that the central authorities could meet.

Our experimental design randomly assigned monitoring devices across AEAs and independently elicited the preferences of their supervisors as to which AEAs should be prioritized for monitoring. Crucially, in the main sample, treatment assignment was independent of supervisor recommendations. The AEAs selected by supervisors are far more responsive to treatment, establishing the fact that supervisors have valuable knowledge. The informational advantage of supervisors is tied to information other than observables that analysts might reasonably collect.

We argue theoretically that the value of the information advantage varies with the scale of anticipated roll-out for the new technology. We then write an empirical version of our theory with two objectives. One objective is to decompose the information advantage of supervisors into the piece stemming from AEA characteristics observable to an analyst, and the piece stemming from characteristics only observable to supervisors. The other goal is to estimate the full schedule of marginal treatment effects as roll-out scale is expanded from 0 to 100 percent. We do this for the selection

37

rule that supervisors are seen to have used as well as several other counterfactual assignment rules.

Our counterfactual assignment rules approximate what principals with varying levels of information might achieve when targeting AEAs for treatment in a centralized fashion. In our setting, impacts resulting from treatment decided upon by a minimally informed principal are not as high as those attained under decentralization. However, a reasonably well-informed principal can approach the level of impact of decentralization. And in the best case scenario for centralization, a principal who can conduct a pilot RCT to obtain predictors of individual response to treatment can outperform supervisor choices. Such a principal would substantially reduce the roll-out scale and still attain larger aggregate gains in AEA performance.

Overall our findings suggest that, as information and communication technologies continue to improve the capabilities of organizations, the benefits from harnessing decentralized information will decrease. Although studies have shown that innovation in information technologies can lead to more decentralization (Bresnahan et al., 2002; Bloom et al., 2014), our findings suggest the opposite may occur, particularly if these technologies primarily serve to reduce the information gap between principals and agents (or middle managers such as supervisors).

Of course, the value of supervisor information is specific to the task and context, which may raise concerns of external validity. But while our findings may not be generalizable, our method is, as it can be easily exported to other settings. Our approach is designed for settings in which spillovers across treatment units are minimal. Thus, it would be interesting to extend our framework to incorporate the potential effects of spillovers in the calculus to decentralize. We view this as a potential avenue for future research.

References

Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J., and Zilibotti, F. (2007). Technology, information, and the decentralization of the firm. *Quarterly Journal of Economics*, 122(4):1759–1799.

Aghion, P. and Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy*, 105(1):1–29.

Aker, J. C. and Ksoll, C. (2019). Call me educated: Evidence from a mobile phone experiment in Niger. *Economics of Education Review*. Forthcoming.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *American Economic Review*, 102(4):1206–40.

Alderman, H. (2002). Do local officials know something we don't? Decentralization of targeted transfers in Albania. *Journal of Public Economics*, 83(3):375 – 404.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., and Singh, N. (2015). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. Unpublished Working Paper.

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

Bloom, N., Garicano, L., Sadun, R., and Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12):2859–2885.

Bloom, N., Sadun, R., and Van Reenen, J. (2012). The organization of firms across countries. *Quarterly Journal of Economics*, 127(4):1663–1705.

Bresnahan, T., Brynjolfsson, E., and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quarterly Journal of Economics*, 117(1):339–376.

Callen, M., Gulzar, S., Hasanain, A., Khan, M. Y., and Rezaee, A. (2018). Data and policy decisions: Experimental evidence from Pakistan. Unpublished Working Paper.

Chassang, S., Padró I Miquel, G., and Snowberg, E. (2012). Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, 102(4):1279–1309.

de Rochambeau, G. (2017). Monitoring and intrinsic motivation: Evidence from Liberia's trucking firms. Unpublished Manuscript.

Dessein, W. (2002). Authority and communication in organizations. *Review of Economic Studies*, 69(4):811–838.

DeYoung, C. G. (2006). Higher-order factors of the big five in a multi-informant sample. *Journal of Personality and Social Psychology*, 91(6):1138–1151.

Dhaliwal, I. and Hanna, R. (2017). The devil is in the details: The successes and limitations of bureaucratic reform in India. *Journal of Development Economics*, 124:1–21.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2018). The value of regulatory discretion: Estimates from environmental inspections in India. *Econometrica*, 86(6):2123–2160.

Duflo, E., Hanna, R., and Ryan, S. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *Stata Journal*, 17(3):630–651.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738.

Hussam, R., Rigol, N., and Roth, B. (2017). Targeting high ability entrepreneurs using community information: Mechanism design in the field. Unpublished Manuscript.

John, O. P., Naumann, L. P., and Soto, C. J. (2008). *Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues*. Guilford Press, New York, NY.

Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *Quarterly Journal of Economics*, 131(1):219–271.

Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1(1):23–41.

Kline, P. and Walters, C. R. (2019). On Heckits, LATE, and numerical equivalence. *Econometrica*, 87(2):677–696.

Mookherjee, D. (2006). Decentralization, hierarchies, and incentives: A mechanism design perspective. *Journal of Economic Literature*, 44(2):367–390.

Olsen, R. J. (1980). A least squares correction for selectivity bias. *Econometrica*, 48(7):1815–1820.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14(4):1261–1295.
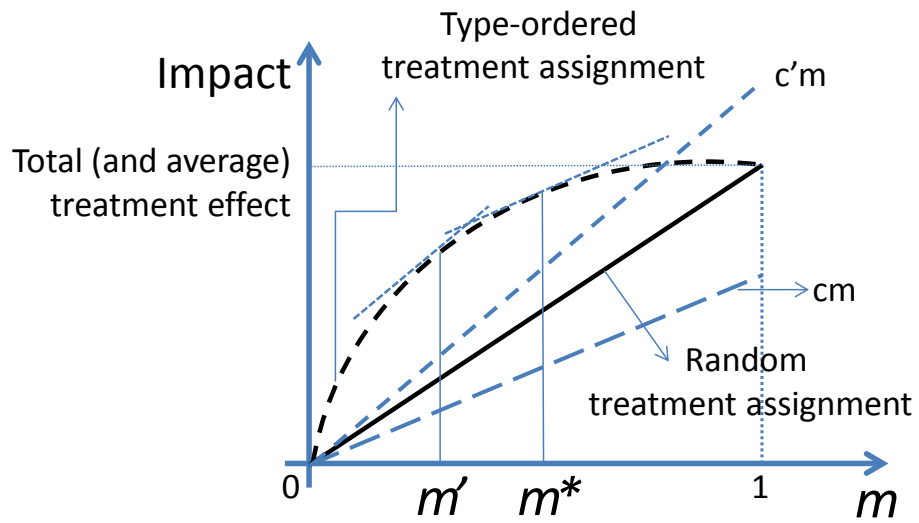
The *y*-axis shows the total treatment effect at different scales of roll-out *m* on the *x*-axis. Treating any agent costs c or c'.

Figure 1: Optimal Roll-Out and the Value of Information

	ALATs		
	Control Group	100% Coverage	50% Coverage
Selected AEA	A	B	E
Non-selected AEA	C	D	F

The columns correspond to the large ALATs, and the rows correspond to AEAs. The control group consists of cells *A* and *C*, in which none of the AEAs received a cell phone. For the ALATs in the column labeled 100% Coverage (cells *B* and *D*), every AEA received a cell phone independently of whether or not they had been selected by their supervisor. For the ALATs in the column labeled 50% Coverage (cells *E* and *F*), the selected AEAs (cell *E*) received cell phones in the first wave while the non-selected (cell *F*) received them in the second wave.

Figure 2: Experimental Design in Large ALATs
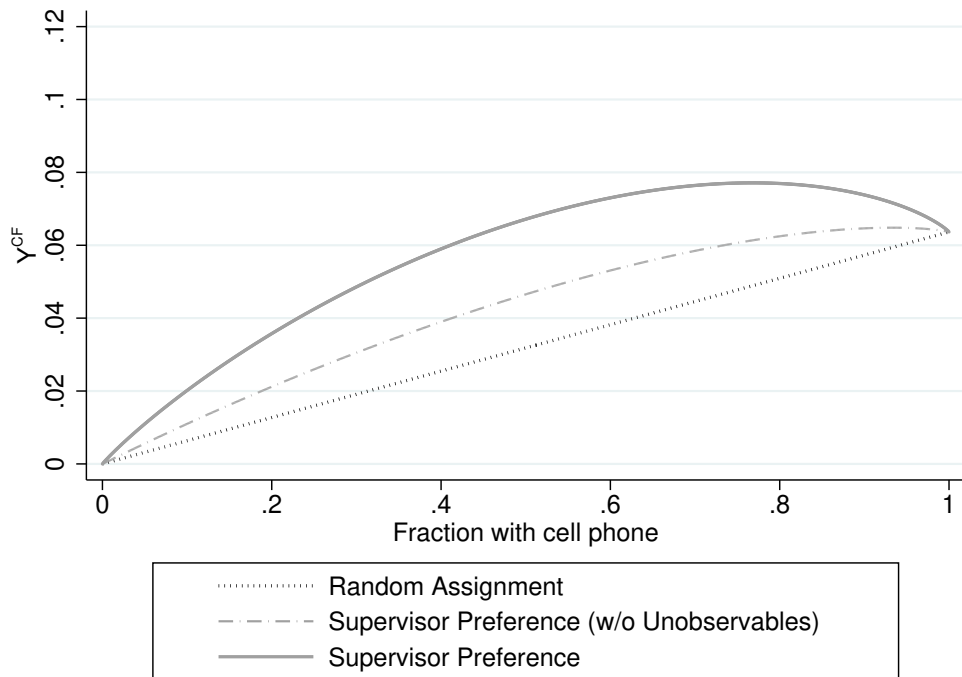
The *y*-axis shows the total treatment effect at different scales of roll-out under different assignment rules. Under random assignment, treatment is assigned randomly. Under supervisor preference without unobservables, supervisors make the assignment decision based purely on observable AEA characteristics. Under supervisor preference, supervisors make the assignment decision based on all information they have.

Figure 3: Supervisor versus Random Assignment
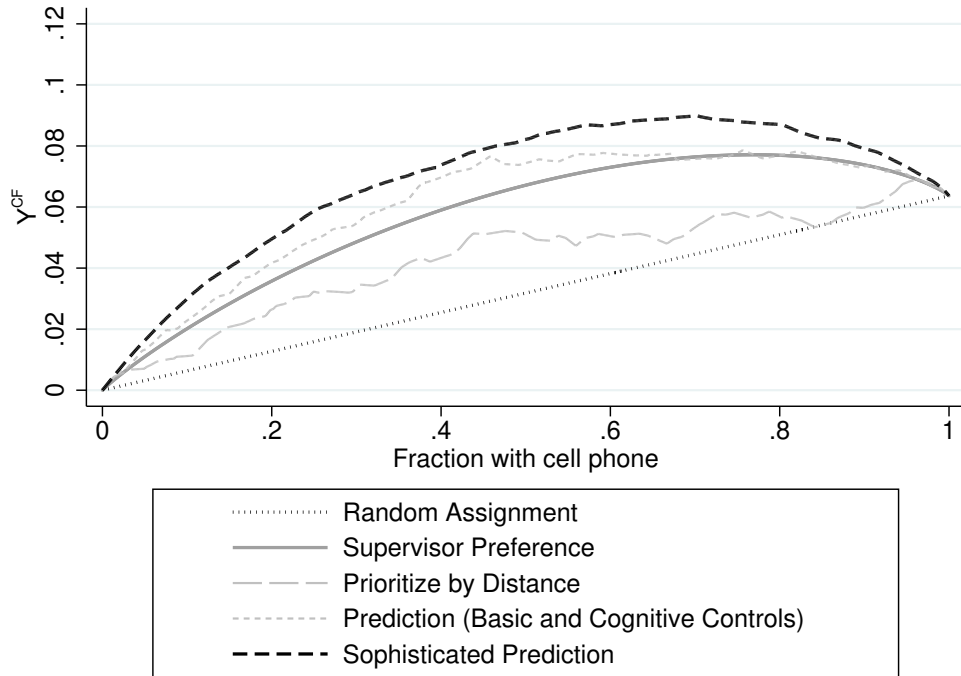
The *y*-axis shows the total treatment effect at different scales of roll-out under different assignment rules. Under random assignment, treatment is assigned randomly. Under supervisor preference, supervisors make the assignment decision based on all the information they have. Under prioritize by distance, treatment is assigned first to those AEAs whose beneficiaries live further from the local ALAT office. Under prediction (basic and cognitive controls), the observable characteristics of the control group are used to predict baseline performance and AEAs who are predicted to be the worst performers in the baseline are prioritized. Under sophisticated prediction, the principal runs a pilot experiment at low roll-out to establish a map between treatment response and observables and then prioritizes AEAs predicted to have the highest treatment response.

Figure 4: Supervisor versus Alternative Allocation Rules

Table 1: Covariate Balance Across AEAs and ALATs, AEA Characteristics

	Pooled	Small ALATs		Large ALATs		
	All (Wgt.)	Ctrl (Unwgt.)	Diff (T-C)	Ctrl (Wgt.)	Ctrl (Unwgt.)	Diff (T-C)
Number of AEAs				4.647	3.091	-0.727
				[2.862]	[2.245]	(0.722)
AEA Averages						
Male	0.706	0.618	0.097	0.750	0.724	-0.065
	[0.358]	[0.485]	(0.154)	[0.275]	[0.355]	(0.136)
Age	38.412	36.412	1.398	37.838	37.934	3.271
	[8.720]	[10.642]	(3.756)	[7.295]	[7.650]	(2.992)
Married	0.412	0.265	0.092	0.441	0.396	0.081
	[0.337]	[0.437]	(0.150)	[0.286]	[0.337]	(0.119)
Average Distance	12.280	10.701	3.376	12.678	13.802	-2.151
	[7.302]	[8.673]	(3.361)	[6.328]	[8.377]	(2.702)
Incumbent Party	0.574	0.588	0.055	0.559	0.614	-0.114
	[0.390]	[0.476]	(0.156)	[0.359]	[0.378]	(0.143)
Digit Span	5.346	5.382	0.189	5.191	5.312	-0.016
	[0.790]	[0.928]	(0.330)	[0.652]	[0.958]	(0.333)
Big 5 — Stability	0.017	-0.053	0.375	-0.057	-0.086	0.233
	[0.758]	[1.097]	(0.323)	[0.682]	[0.828]	(0.282)
Big 5 — Plasticity	-0.062	-0.390	0.620*	-0.140	-0.241	0.352
	[0.758]	[0.923]	(0.308)	[0.679]	[0.950]	(0.313)
Num ALATs (Ctrl)	39	17		22		
Num ALATs (Treat)	32	21		11		
Num AEAs (Ctrl)	88	20		68		
Num AEAs (Treat)	48	22		26		

"Treatment" for small ALATs refers to ALATs that received phones in either round 1 or 2. "Treatment" for large ALATs refers to the 100% coverage ALATs where all AEAs received a phone in round 1. Columns 1 and 4 report AEA-average AEA characteristics (weighting ALAT averages by the number of AEAs in the ALAT). Columns 2 and 5 report ALAT-average AEA characteristics. Unweighted and weighted standard deviations are in square brackets. Columns 3 and 6 report mean differences in AEA-average characteristics between treatment and control. Standard errors from a two-sample t-test are in parentheses. Appendix Table A1 reports balance tests using randomization inference (RI), and also adds in the 50% coverage large ALATs.

Table 2: Average Effects of Receiving a Cell Phone on Productivity

	Farmer was visited in the last week					Supervisor knows	Log length of meeting (mins)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.0656**	0.0633**	0.0626**	0.0569*	-0.00753	0.185*	-0.00959
	(0.0307)	(0.0262)	(0.0264)	(0.0305)	(0.0351)	(0.103)	(0.0589)
Selected	0.0111	0.0150	0.0173	0.0109		-0.0799	-0.0335
	(0.0327)	(0.0271)	(0.0289)	(0.0291)		(0.138)	(0.0365)
Male		0.0391	0.0387	0.0317	-0.0293	0.301**	-0.0467
		(0.0272)	(0.0295)	(0.0329)	(0.0459)	(0.147)	(0.0420)
Age		0.00258**	0.00280**	0.00276*	-0.000143	-0.00806	0.00437**
		(0.00125)	(0.00138)	(0.00152)	(0.00172)	(0.00824)	(0.00168)
Married		-0.00414	-0.00294	0.0198	-0.0409	0.0136	-0.0532
		(0.0299)	(0.0296)	(0.0312)	(0.0340)	(0.141)	(0.0424)
Distance to Farmers (log)		0.0386	0.0365	0.0316	-0.0693**	-0.139	-0.0187
		(0.0268)	(0.0275)	(0.0339)	(0.0280)	(0.119)	(0.0469)
Incumbent Party		-0.0216	-0.0206	-0.0397	0.0582	0.254*	0.00210
		(0.0244)	(0.0249)	(0.0248)	(0.0574)	(0.151)	(0.0524)
Digit Span			0.00987	0.0125	0.0303	0.0509	0.0125
			(0.0117)	(0.0129)	(0.0185)	(0.0516)	(0.0181)
Big 5 — Stability			0.0118	0.0106	0.0351*	0.141***	0.0175
			(0.0149)	(0.0160)	(0.0199)	(0.0497)	(0.0245)
Big 5 — Plasticity			-0.0107	-0.0111	-0.0235	0.111*	-0.0177
			(0.0105)	(0.0102)	(0.0169)	(0.0637)	(0.0173)
Servicer	AEA	AEA	AEA	AEA	Supervisor	AEA	AEA
Mean of Control Dep. Var	0.271	0.271	0.271	0.274	0.308	4.593	4.414
R^2	0.005	0.012	0.013	0.012	0.032	0.199	0.012
Number of Phone Surveys	1842	1842	1842	1584	1173		1819
Number of AEAs	136	136	136	94	107	126	136
Number of ALATs	71	71	71	33	107	65	71
Includes Small ALATs	✓	✓	✓		✓	✓	✓

Outcomes in columns other than (6) are from farmer phone survey; these regressions include indicators for small ALATs and survey round. Outcome in column (6) is from AEA survey; outcome is an index from (1) "Strongly Disagree" to (5) "Strongly Agree;" this regression includes indicator for small ALATs. Big 5 stability and plasticity have mean zero and unit variance. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Appendix Table A4 reports results using wild cluster bootstrap and randomization inference (RI).

Table 3: Do Supervisors Have an Informational Advantage?

	Farmer was visited in the last week			
	(1)	(2)	(3)	(4)
Treated	0.0607*	-0.0233	-0.0397	-0.0361
	(0.0341)	(0.0433)	(0.0347)	(0.0377)
Treated × *Selected*		0.142**	0.161***	0.154***
		(0.0582)	(0.0509)	(0.0503)
Selected	0.0113	-0.0332	-0.0445	-0.0409
	(0.0328)	(0.0363)	(0.0274)	(0.0282)
R^2	0.00	0.01	0.02	0.02
Number of Phone Surveys	1584	1584	1584	1584
Number of AEAs	94	94	94	94
Number of ALATs	33	33	33	33
Includes Basic Controls			✓	✓
Includes Cognitive Controls				✓

Sample is 100% coverage and control large ALATs. Regressions include survey round indicator. Basic controls are sex, age, marital status, average distance to farmers, and incumbent party registration. Cognitive controls are digit span, Big 5 stability, and Big 5 plasticity. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results using the wild bootstrap in Appendix Table A7, using randomization inference (RI) in Appendix Table A8, and adding in the 50% coverage large ALATs in Appendix Table A9.

Table 4: Testing for Spillover and Hawthorne Effects

| | Farmer was visited in the last week | | |
	(1)	(2)	(3)
Treated	-0.0241	0.0607*	0.0580*
	(0.0365)	(0.0339)	(0.0293)
Treated × *Selected*	0.161**		
	(0.0762)		
Treated × Small		0.0331	0.0302
		(0.0793)	(0.0738)
Spillover (*Non-selected*)	-0.0137		
	(0.0490)		
Spillover (*Selected*)	-0.0534		
	(0.0605)		
Selected	-0.0465	0.0110	0.0171
	(0.0377)	(0.0326)	(0.0289)
Small		-0.0205	-0.0112
		(0.0467)	(0.0455)
R^2	0.01	0.01	0.01
Number of Phone Surveys	2228	1842	1842
Number of AEAs	134	136	136
Number of ALATs	44	71	71
Includes Basic Controls	✓		✓
Includes Cognitive Controls	✓		✓

Sample in column (1) is all large ALATs. Sample in columns (2) and (3) is 100% coverage and control large ALATs and all small ALATs. Outcome is from farmer phone survey. Spillover (*Non-selected*) indicates AEAs who were non-selected and not treated but whose selected peers were treated. This is cell *F* (non-selected, 50% coverage) in round 1. Spillover (*Selected*) indicates AEAs who were selected and treated and whose non-selected peers were treated. This is cell *B* (selected, 100% coverage) in both rounds and cell *E* (selected, 50% coverage) in round 2. Basic controls are sex, age, marital status, average distance to farmers, and incumbent party registration. Cognitive controls are digit span, Big 5 stability, and Big 5 plasticity. Regressions include survey round indicator. Cluster robust standard errors reported in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: First Stage Probit Regressions

	AEA was selected	
	(1)	(2)
Male	-0.519	-0.511
	(0.349)	(0.395)
Age	-0.0361**	-0.0373**
	(0.0181)	(0.0170)
Married	0.863*	0.791*
	(0.459)	(0.441)
Average Distance	0.331*	0.318
	(0.173)	(0.197)
Incumbent Party	-0.808**	-0.840**
	(0.359)	(0.367)
Digit Span		-0.111
		(0.142)
Big 5 — Stability		-0.234*
		(0.124)
Big 5 — Plasticity		0.122
		(0.154)
Pseudo R^2	0.159	0.185
Number of AEAs	94	94
Number of ALATs	33	33

Sample is 100% coverage and control large ALATs. Coefficients from probit of indicator for the AEA being selected on AEA characteristics. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results using the wild bootstrap in Appendix Table A10, and adding in the 50% coverage large ALATs in Appendix Table A11.

Table 6: Treatment Effect Heterogeneity on Observable and Unobservable Characteristics

	Farmer was visited in the last week		
	(1)	(2)	(3)
Main Effects:			
Inverse Mills Ratio	-0.0207	-0.0188	-0.0158
	(0.0225)	(0.0158)	(0.0158)
Average Treatment Effect	0.0616*	0.0585**	0.0639**
	(0.0346)	(0.0255)	(0.0280)
Interactions with Treatment:			
Inverse Mills	0.0880**	0.0640***	0.0636**
	(0.0362)	(0.0217)	(0.0250)
Male		-0.0208	-0.00389
		(0.0508)	(0.0664)
Age		-0.00334	-0.00524
		(0.00311)	(0.00350)
Married		-0.0839	-0.0941*
		(0.0605)	(0.0521)
Average Distance to Farmers (log)		0.101	0.104
		(0.0675)	(0.0774)
Incumbent Party		-0.0812**	-0.0695
		(0.0384)	(0.0477)
Digit Span			-0.0638**
			(0.0274)
Big5 — Stability			-0.0379
			(0.0488)
Big5 — Plasticity			0.0349
			(0.0253)
R^2	0.009	0.023	0.028
p-value for Observable Interactions		0.031	0.001
Number of Phone Surveys	1584	1584	1584
Number of AEAs	94	94	94
Number of ALATs	33	33	33
Basic Controls		✓	✓
Cognitive Controls			✓

Sample is 100% coverage and control large ALATs. Dependent variable for all regressions is whether the AEA visited the farmer in the week prior to the phone survey. The inverse Mills ratio is the generalized residual—the expected value of the error term—from the probit selection regression in the same column of Table 5. Regressions include survey round indicator. Main effects of basic and cognitive controls omitted for space. The *p*-value in the bottom row for observable interactions is from the joint Wald test for coefficients on treatment interacted with all included controls. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results using the wild bootstrap in Appendix Table A12, and adding in the 50% coverage large ALATs in Appendix Table A13.

Table 7: Treatment Effects by Roll-Out Levels and Allocation Rules

Rollout	Random	Supervisor	Distance	Prediction	Sophisticated
0.25	0.016	0.042	0.032	0.049	0.059
0.50	0.032	0.067	0.051	0.074	0.082
0.53	0.034	0.069	0.049	0.075	0.084
0.70	0.045	0.076	0.054	0.076	**0.090**
0.75	0.048	0.077	0.057	0.079	0.088
0.76	0.048	0.077	0.057	**0.079**	0.088
0.77	0.049	**0.077**	0.057	0.077	0.088
0.97	0.062	0.069	**0.069**	0.069	0.071
1.00	**0.064**	0.064	0.064	0.064	0.064

Treatment effects for the various allocation rules as estimated at the different roll-out scales labeled in the first column. Bold numbers represent the largest treatment effect under a given allocation rule. The allocation rules are random - treatment is assigned randomly; supervisor - supervisors make the assignment decision based on all the information they have; distance - treatment is assigned first to those AEAs whose beneficiaries live further from the local ALAT office; prediction - the observable characteristics of the control group are used to predict baseline performance and AEAs who are predicted to be the worst performers in the baseline are prioritized; and sophisticated - the principal runs a pilot experiment at low roll-out to establish a map between treatment response and observables and then prioritizes AEAs predicted to have the highest treatment response.

Table A1: Extended Covariate Balance, AEA Characteristics

	Small ALATs	Large ALATs		
	Diff (T-C)	Diff (T-C)	Diff (T100-C)	Diff (T50-C)
Number of AEAs		-0.091	-0.727	0.545
		(0.708)	(0.861)	(0.861)
		⟨0.898⟩	⟨0.403⟩	⟨0.530⟩
		⟦0.946⟧	⟦0.449⟧	⟦0.549⟧
AEA Averages				
Male	0.097	-0.179	-0.065	-0.294
	(0.154)	(0.110)	(0.133)	(0.133)
	⟨0.535⟩	⟨0.112⟩	⟨0.630⟩	⟨0.034⟩**
	⟦0.579⟧	⟦0.107⟧	⟦0.640⟧	⟦0.029⟧**
Age	1.398	3.148	3.271	3.026
	(3.756)	(2.643)	(3.276)	(3.276)
	⟨0.712⟩	⟨0.240⟩	⟨0.324⟩	⟨0.361⟩
	⟦0.738⟧	⟦0.266⟧	⟦0.314⟧	⟦0.347⟧
Married	0.092	0.103	0.081	0.125
	(0.150)	(0.099)	(0.123)	(0.123)
	⟨0.542⟩	⟨0.306⟩	⟨0.514⟩	⟨0.316⟩
	⟦0.603⟧	⟦0.304⟧	⟦0.536⟧	⟦0.283⟧
Average Distance	3.376	-0.619	-2.151	0.913
	(3.361)	(2.342)	(2.873)	(2.873)
	⟨0.322⟩	⟨0.793⟩	⟨0.458⟩	⟨0.752⟩
	⟦0.295⟧	⟦0.785⟧	⟦0.473⟧	⟦0.732⟧
Incumbent Party	0.055	0.028	-0.114	0.169
	(0.156)	(0.118)	(0.141)	(0.141)
	⟨0.728⟩	⟨0.816⟩	⟨0.422⟩	⟨0.236⟩
	⟦0.725⟧	⟦0.828⟧	⟦0.439⟧	⟦0.228⟧
Digit Span	0.189	-0.162	-0.016	-0.307
	(0.330)	(0.245)	(0.301)	(0.301)
	⟨0.571⟩	⟨0.512⟩	⟨0.957⟩	⟨0.313⟩
	⟦0.599⟧	⟦0.512⟧	⟦0.958⟧	⟦0.301⟧
Big 5 — Stability	0.375	0.180	0.233	0.127
	(0.323)	(0.215)	(0.266)	(0.266)
	⟨0.253⟩	⟨0.406⟩	⟨0.386⟩	⟨0.634⟩
	⟦0.264⟧	⟦0.405⟧	⟦0.384⟧	⟦0.637⟧
Big 5 — Plasticity	0.620	0.426	0.352	0.501
	(0.308)	(0.244)	(0.302)	(0.302)
	⟨0.052⟩*	⟨0.088⟩*	⟨0.250⟩	⟨0.105⟩
	⟦0.049⟧**	⟦0.073⟧*	⟦0.264⟧	⟦0.115⟧
Num ALATs	38	44		

Column 1 reports differences in small ALAT-average characteristics between treatment and control ALATs as in column 3 of Table 1. Column 2 reports differences in large ALAT-average characteristics between the pooled 100% and 50% coverage ALATs and the control. Columns 3 and 4 compare the 100% coverage ALATs and 50% coverage ALATs to the control respectively. Column 3 is as in column 6 of Table 1. Conventional *p*-values in angled brackets, and *p*-values from randomization inference (RI) in double square brackets.   52

Table A2: Covariate Balance Across AEAs and ALATs, ALAT Characteristics

	Pooled	Small ALATs		Large ALATs		
	All (Wgt.)	Ctrl (Unwgt.)	Diff (T-C)	Ctrl (Wgt.)	Ctrl (Unwgt.)	Diff (T-C)
Number of Rural Hhds	3413.191	2297.824	320.129	3688.779	3218.182	878.636
	[2316.480]	[1977.709]	(643.835)	[2026.628]	[1695.731]	(833.460)
Share of Rural Hhds	0.712	0.809	-0.094	0.705	0.708	-0.039
	[0.188]	[0.104]	(0.057)	[0.200]	[0.217]	(0.077)
Average Hhd Size	4.753	4.705	0.008	4.758	4.699	0.143
	[0.358]	[0.335]	(0.137)	[0.315]	[0.348]	(0.132)
Land Per Farm (Hectares)	42.121	41.346	7.630	38.328	43.403	0.676
	[48.311]	[43.495]	(18.661)	[38.732]	[42.478]	(17.560)
Cropland Per Farm (Hectares)	7.280	12.578	-3.228	5.069	6.880	2.305
	[11.746]	[21.459]	(5.610)	[5.719]	[8.155]	(4.034)
Share of Farmers Working With DEAg	0.087	0.082	-0.020	0.087	0.096	0.014
	[0.089]	[0.109]	(0.032)	[0.064]	[0.072]	(0.036)
Corn Yield (metric tons) Per Hectare	2.008	2.142	0.160	1.804	1.967	0.264
	[1.116]	[1.490]	(0.497)	[0.682]	[0.898]	(0.393)
Share of Farms With Running Water	0.476	0.466	0.035	0.478	0.476	-0.051
	[0.209]	[0.247]	(0.084)	[0.200]	[0.210]	(0.075)
Colorado (Winner) Vote Share	0.452	0.455	-0.007	0.452	0.472	-0.023
	[0.083]	[0.093]	(0.029)	[0.073]	[0.083]	(0.034)
Num ALATS (Ctrl)	39	17		22		
Num ALATS (Treat)	32	21		11		
Num AEAs (Ctrl)	88	20		68		
Num AEAs (Treat)	48	22		26		

"Treatment" for small ALATs refers to ALATs that received phones in either round 1 or 2. "Treatment" for large ALATs refers to the 100% coverage ALATs where all AEAs received a phone in round 1. Columns 1 and 4 report AEA-average ALAT characteristics (weighting ALAT averages by the number of AEAs in the ALAT). Columns 2 and 5 report ALAT-average ALAT characteristics. Unweighted and weighted standard deviations are in square brackets. Columns 3 and 6 report mean differences in AEA-average characteristics between treatment and control. Standard errors from a two-sample t-test are in parentheses. Appendix Table A3 reports balance tests using randomization inference (RI), and also adds in the 50% coverage large ALATs. The first three variables come from the 2002 census, the next five come from the 2008 agricultural census, and the final variable comes from the 2013 presidential elections.

Table A3: Extended Covariate Balance, ALAT Characteristics

	Small ALATs	Large ALATs		
	Diff (T-C)	Diff (T-C)	Diff (T100-C)	Diff (T50-C)
Number of Rural Hhds	320.129	620.955	878.636	363.273
	(643.835)	(679.528)	(839.447)	(839.447)
	⟨0.622⟩	⟨0.366⟩	⟨0.301⟩	⟨0.667⟩
	⟦0.617⟧	⟦0.356⟧	⟦0.304⟧	⟦0.671⟧
Share of Rural Hhds	-0.094	-0.047	-0.039	-0.054
	(0.057)	(0.060)	(0.074)	(0.074)
	⟨0.106⟩	⟨0.443⟩	⟨0.601⟩	⟨0.474⟩
	⟦0.118⟧	⟦0.444⟧	⟦0.604⟧	⟦0.447⟧
Average Hhd Size	0.008	0.071	0.143	-0.001
	(0.137)	(0.112)	(0.138)	(0.138)
	⟨0.953⟩	⟨0.530⟩	⟨0.305⟩	⟨0.995⟩
	⟦0.948⟧	⟦0.529⟧	⟦0.339⟧	⟦0.995⟧
Land Per Farm (Hectares)	7.630	-3.916	0.676	-8.508
	(18.661)	(12.726)	(15.727)	(15.727)
	⟨0.685⟩	⟨0.760⟩	⟨0.966⟩	⟨0.591⟩
	⟦0.717⟧	⟦0.783⟧	⟦0.968⟧	⟦0.623⟧
Cropland Per Farm (Hectares)	-3.228	0.989	2.305	-0.326
	(5.610)	(2.965)	(3.658)	(3.658)
	⟨0.569⟩	⟨0.740⟩	⟨0.532⟩	⟨0.929⟩
	⟦0.573⟧	⟦0.751⟧	⟦0.573⟧	⟦0.942⟧
Share of Farmers Working With DEAg	-0.020	0.017	0.014	0.020
	(0.032)	(0.033)	(0.041)	(0.041)
	⟨0.550⟩	⟨0.605⟩	⟨0.731⟩	⟨0.623⟩
	⟦0.548⟧	⟦0.620⟧	⟦0.721⟧	⟦0.624⟧
Corn Yield (metric tons) Per Hectare	0.160	0.368	0.264	0.472
	(0.497)	(0.311)	(0.385)	(0.385)
	⟨0.749⟩	⟨0.244⟩	⟨0.496⟩	⟨0.227⟩
	⟦0.754⟧	⟦0.232⟧	⟦0.498⟧	⟦0.243⟧
Share of Farms With Running Water	0.035	-0.031	-0.051	-0.011
	(0.084)	(0.063)	(0.077)	(0.077)
	⟨0.678⟩	⟨0.625⟩	⟨0.518⟩	⟨0.887⟩
	⟦0.676⟧	⟦0.605⟧	⟦0.485⟧	⟦0.870⟧
Colorado (Winner) Vote Share	-0.007	-0.020	-0.023	-0.018
	(0.029)	(0.025)	(0.031)	(0.031)
	⟨0.804⟩	⟨0.422⟩	⟨0.467⟩	⟨0.568⟩
	⟦0.813⟧	⟦0.394⟧	⟦0.478⟧	⟦0.578⟧
Num ALATs	38	44		

Column 1 reports differences in small ALAT-average characteristics between treatment and control ALATs as in column 3 of Table A2. Column 2 reports differences in large ALAT-average characteristics between the pooled 100% and 50% coverage ALATs and the control. Columns 3 and 4 compare the 100% coverage ALATs and 50% coverage ALATs to the control respectively. Column 3 is as in column 6 of Table A2. Conventional p-values in angled brackets, and p-values from randomization inference (RI) in double square brackets. The first three variables come from the 2002 census, the next five come from the 2008 agricultural census, and the final variable comes from the 2013 presidential elections.

Table A4: Average Effects of Receiving a Cell Phone on Productivity, Wild Bootstrap and Randomization Inference (RI)

	Farmer was visited in the last week					Supervisor knows	Log length of meeting (mins)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.066	0.063	0.063	0.057	-0.008	0.185	-0.010
	$\langle 0.036 \rangle$**	$\langle 0.018 \rangle$**	$\langle 0.020 \rangle$**	$\langle 0.071 \rangle$*	$\langle 0.831 \rangle$	$\langle 0.076 \rangle$*	$\langle 0.871 \rangle$
	$\{0.028\}$**	$\{0.020\}$**	$\{0.023\}$**	$\{0.089\}$*	$\{0.843\}$	$\{0.066\}$*	$\{0.883\}$
	$[\![0.015]\!]$**	$[\![0.012]\!]$**	$[\![0.015]\!]$**	$[\![0.048]\!]$**	$[\![0.866]\!]$	$[\![0.174]\!]$	$[\![0.880]\!]$
Servicer	AEA	AEA	AEA	AEA	Supervisor	AEA	AEA
Mean of Control Dep. Var	0.271	0.271	0.271	0.274	0.308	4.593	4.414
R^2	0.005	0.012	0.013	0.012	0.032	0.199	0.012
Number of Phone Surveys	1842	1842	1842	1584	1173		1819
Number of AEAs	136	136	136	94	107	126	136
Number of ALATs	71	71	71	33	107	65	71
Includes Small ALATs	✓	✓	✓		✓	✓	✓
Includes Basic Controls		✓	✓	✓	✓	✓	✓
Includes Cognitive Controls			✓	✓	✓	✓	✓

Outcomes in columns other than (6) are from farmer phone survey; these regressions include indicators for small ALATs and survey round. Outcome in column (6) is from AEA survey; outcome is an index from (1) "Strongly Disagree" to (5) "Strongly Agree;" this regression includes indicator for small ALATs. Column (2) includes basic controls, and columns (3)-(7) include basic and cognitive controls. Basic controls are sex, age, marital status, average distance to farmers, and incumbent party registration. Cognitive controls are digit span, Big 5 stability, and Big 5 plasticity. p-values from cluster robust standard errors in angled brackets (as in Table 2), from a wild cluster bootstrap with 100,000 bootstrap samples in curly braces, and from randomization inference (RI) with 1,000 permutations in double square brackets. Significance stars correspond to the adjacent p-values, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Average Effects of Receiving a Cell Phone on Other Measures of Performance

	Satisfied	Received Training	Useful	PCA
	(1)	(2)	(3)	(4)
Treated	0.0747	0.0511*	0.0335	0.131*
	(0.0469)	(0.0299)	(0.0216)	(0.0708)
Selected	-0.0727	-0.0149	-0.0172	-0.0728
	(0.0512)	(0.0233)	(0.0210)	(0.0649)
Male	0.0601	-0.00163	0.0111	0.0450
	(0.0569)	(0.0297)	(0.0247)	(0.0805)
Age	0.00294	0.00247	0.00291**	0.00739*
	(0.00211)	(0.00160)	(0.00132)	(0.00378)
Married	-0.0627*	-0.0224	-0.0244	-0.0854
	(0.0365)	(0.0295)	(0.0212)	(0.0659)
Distance to Farmers (log)	0.00454	-0.0270	-0.0244	-0.0522
	(0.0533)	(0.0275)	(0.0227)	(0.0734)
Incumbent Party	-0.153***	-0.0380	-0.0316	-0.159**
	(0.0511)	(0.0240)	(0.0246)	(0.0670)
Digit Span	0.0267*	0.00161	0.00621	0.0236
	(0.0152)	(0.0106)	(0.0102)	(0.0271)
Big 5 — Stability	-0.00742	0.000479	0.0101	0.00735
	(0.0187)	(0.0118)	(0.00763)	(0.0276)
Big 5 — Plasticity	-0.0344	-0.0135	-0.0121	-0.0482
	(0.0237)	(0.0118)	(0.0115)	(0.0341)
Servicer	AEA	AEA	AEA	AEA
Mean of Control Dep. Var	2.542	0.765	0.817	-0.042
R^2	0.025	0.017	0.025	0.017
Number of Phone Surveys	1838	1841	1841	1838
Number of AEAs	136	136	136	136
Number of ALATs	71	71	71	71
Includes Small ALATs	✓	✓	✓	✓

Sample is 100% coverage and control large ALATs and all small ALATs. Outcome in column (4) is the first principle component from a polychoric principal component analysis (PCA) of the outcome variables in the first three columns. Regressions include indicators for small ALATs and survey round. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Correlation Matrix of Performance Measures

	Visited	Satisfied	Recieved Training	Useful	PCA
Visited	1				
Satisfied	0.174***	1			
Received Training	0.147***	0.397***	1		
Useful	0.145***	0.368***	0.722***	1	
PCA	0.189***	0.704***	0.878***	0.860***	1

Correlation matrix for the 1248 control group observations from Table A5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Do Supervisors Have an Informational Advantage? Wild Bootstrap Inference

	Farmer was visited in the last week			
	(1)	(2)	(3)	(4)
Treated	0.061	-0.023	-0.040	-0.036
	⟨0.085⟩*	⟨0.595⟩	⟨0.261⟩	⟨0.345⟩
	{0.073}*	{0.611}	{0.280}	{0.390}
Treated × *Selected*		0.142	0.161	0.154
		⟨0.021⟩**	⟨0.003⟩***	⟨0.004⟩***
		{0.036}**	{0.015}**	{0.027}**
Selected	0.011	-0.033	-0.044	-0.041
	⟨0.734⟩	⟨0.366⟩	⟨0.114⟩	⟨0.156⟩
	{0.756}	{0.481}	{0.188}	{0.273}
R^2	0.00	0.01	0.02	0.02
Number of Phone Surveys	1584	1584	1584	1584
Number of AEAs	94	94	94	94
Number of ALATs	33	33	33	33
Includes Basic Controls			✓	✓
Includes Cognitive Controls				✓

Sample is 100% coverage and control large ALATs. Regressions include survey round indicator. Basic controls are sex, age, marital status, average distance to farmers, and incumbent party registration. Cognitive controls are digit span, Big 5 stability, and Big 5 plasticity. p-values from cluster robust standard errors in angled brackets (as in Table 3), and from a wild cluster bootstrap with 100,000 bootstrap samples in curly braces. Significance stars correspond to the adjacent p-values, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Do Supervisors Have an Informational Advantage? Randomization Inference (RI)

	Farmer was visited in the last week			
	(1)	(2)	(3)	(4)
Treated	0.061	-0.023	-0.040	-0.036
	$\langle 0.085 \rangle^{*}$	$\langle 0.595 \rangle$	$\langle 0.261 \rangle$	$\langle 0.345 \rangle$
	$[\![0.028]\!]^{**}$	$[\![0.634]\!]$	$[\![0.321]\!]$	$[\![0.398]\!]$
Treated $\times$ *Selected*		0.142	0.161	0.154
		$\langle 0.021 \rangle^{**}$	$\langle 0.003 \rangle^{***}$	$\langle 0.004 \rangle^{***}$
		$[\![0.011]\!]^{**}$	$[\![0.001]\!]^{***}$	$[\![0.010]\!]^{**}$
R^2	0.00	0.01	0.02	0.02
Number of Phone Surveys	1584	1584	1584	1584
Number of AEAs	94	94	94	94
Number of ALATs	33	33	33	33
Includes Basic Controls			✓	✓
Includes Cognitive Controls				✓

Sample is 100% coverage and control large ALATs. Regressions include survey round indicator. Basic controls are sex, age, marital status, average distance to farmers, and incumbent party registration. Cognitive controls are digit span, Big 5 stability, and Big 5 plasticity. *p*-values from cluster robust standard errors in angled brackets (as in Table 3), and from randomization inference (RI) with 1,000 permutations in double square brackets. Significance stars correspond to the adjacent *p*-values, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Do Supervisors Have an Informational Advantage? Including 50% Coverage ALATs

	Farmer was visited in the last week			
	(1)	(2)	(3)	(4)
Treated	0.0503**	-0.00818	-0.0196	-0.0175
	(0.0249)	(0.0336)	(0.0319)	(0.0331)
Treated × *Selected*		0.101*	0.113**	0.114**
		(0.0517)	(0.0495)	(0.0494)
Selected	0.0122	-0.0304	-0.0429	-0.0422
	(0.0253)	(0.0305)	(0.0307)	(0.0319)
R^2	0.00	0.01	0.01	0.01
Number of Phone Surveys	2228	2228	2228	2228
Number of AEAs	134	134	134	134
Number of ALATs	44	44	44	44
Includes Basic Controls			✓	✓
Includes Cognitive Controls				✓

Regressions analogous to Table 3 but sample includes all large ALATs (100% coverage, 50% coverage, and control). Regressions include survey round indicator. Basic controls are sex, age, marital status, average distance to farmers, and incumbent party registration. Cognitive controls are digit span, Big 5 stability, and Big 5 plasticity. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: First Stage Probit Regressions, Wild Bootstrap Inference

	AEA was selected	
	(1)	(2)
Male	-0.519	-0.511
	$\langle 0.137 \rangle$	$\langle 0.196 \rangle$
	$\{0.163\}$	$\{0.258\}$
Age	-0.036	-0.037
	$\langle 0.046 \rangle^{**}$	$\langle 0.028 \rangle^{**}$
	$\{0.029\}^{**}$	$\{0.028\}^{**}$
Married	0.863	0.791
	$\langle 0.060 \rangle^{*}$	$\langle 0.073 \rangle^{*}$
	$\{0.046\}^{**}$	$\{0.066\}^{*}$
Average Distance	0.331	0.318
	$\langle 0.056 \rangle^{*}$	$\langle 0.106 \rangle$
	$\{0.118\}$	$\{0.112\}$
Incumbent Party	-0.808	-0.840
	$\langle 0.024 \rangle^{**}$	$\langle 0.022 \rangle^{**}$
	$\{0.044\}^{**}$	$\{0.043\}^{**}$
Digit Span		-0.111
		$\langle 0.434 \rangle$
		$\{0.472\}$
Big 5 — Stability		-0.234
		$\langle 0.059 \rangle^{*}$
		$\{0.084\}^{*}$
Big 5 — Plasticity		0.122
		$\langle 0.429 \rangle$
		$\{0.431\}$
Pseudo R^2	0.159	0.185
Number of AEAs	94	94
Number of ALATs	33	33

Sample is 100% coverage and control large ALATs. Coefficients from probit of indicator for the AEA being selected on AEA characteristics. p-values from cluster robust standard errors in angled brackets (as in Table 5), and from a wild cluster bootstrap with 100,000 bootstrap samples in curly braces. Significance stars correspond to the adjacent p-values, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: First Stage Probit Regressions, Including 50% Coverage ALATs

	AEA was selected	
	(1)	(2)
Male	0.107	0.117
	(0.275)	(0.309)
Age	-0.0101	-0.00513
	(0.0148)	(0.0157)
Married	0.851***	0.771**
	(0.306)	(0.317)
Average Distance	0.133	0.233
	(0.167)	(0.185)
Incumbent Party	-0.667**	-0.744**
	(0.319)	(0.317)
Digit Span		0.0726
		(0.118)
Big 5 — Stability		-0.270**
		(0.118)
Big 5 — Plasticity		0.111
		(0.129)
Pseudo R^2	0.105	0.135
Number of AEAs	134	134
Number of ALATs	44	44

Regressions analogous to Table 5 but sample includes all large ALATs (100% coverage, 50% coverage, and control). Coefficients from probit of indicator for the AEA being selected on AEA characteristics. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Treatment Effect Heterogeneity on Observable and Unobservable Characteristics, Wild Bootstrap Inference

	Farmer was visited in the last week		
	(1)	(2)	(3)
Main Effects:			
Inverse Mills Ratio	-0.021	-0.019	-0.016
	⟨0.366⟩	⟨0.245⟩	⟨0.325⟩
Average Treatment Effect	0.062	0.058	0.064
	⟨0.085⟩*	⟨0.029⟩**	⟨0.029⟩**
	{0.093}*	{0.067}*	{0.065}*
Interactions with Treatment:			
Inverse Mills	0.088	0.064	0.064
	⟨0.021⟩**	⟨0.006⟩***	⟨0.016⟩**
	{0.036}**	{0.017}**	{0.035}**
Male		-0.021	-0.004
		⟨0.686⟩	⟨0.954⟩
		{0.717}	{0.962}
Age		-0.003	-0.005
		⟨0.292⟩	⟨0.144⟩
		{0.329}	{0.210}
Married		-0.084	-0.094
		⟨0.175⟩	⟨0.081⟩*
		{0.241}	{0.134}
Average Distance to Farmers (log)		0.101	0.104
		⟨0.143⟩	⟨0.187⟩
		{0.251}	{0.361}
Incumbent Party		-0.081	-0.069
		⟨0.042⟩**	⟨0.155⟩
		{0.054}*	{0.149}
Digit Span			-0.064
			⟨0.026⟩**
			{0.036}**
Big5 — Stability			-0.038
			⟨0.442⟩
			{0.562}
Big5 — Plasticity			0.035
			⟨0.177⟩
			{0.315}
R^2	0.009	0.023	0.028
p-value for Observable Interactions (Conventional)		0.031	0.001
p-value for Observable Interactions (Wild)		0.207	0.146
Number of Phone Surveys	1584	1584	1584
Number of AEAs	94	94	94
Number of ALATs	33	33	33
Basic Controls		✓	✓
Cognitive Controls			✓

Sample is 100% coverage and control large ALATs. Dependent variable is whether the AEA visited the farmer in the week prior to the phone survey. The inverse Mills ratio is the generalized residual—the expected value of the error term—from the probit selection regression in the same column of Table A10. Regressions include survey round indicator. Main effects of basic and cognitive controls omitted for space. The *p*-values for observable interactions are from the joint Wald test for coefficients on treatment interacted with all included controls. *p*-values from cluster robust standard errors in angled brackets (as in Table 6), and from a wild cluster bootstrap with 100,000 bootstrap samples in curly braces. Significance stars correspond to the adjacent *p*-values, * $p <$

Table A13: Treatment Effect Heterogeneity on Observable and Unobservable Characteristics, Including 50% Coverage ALATs

	Farmer was visited in the last week		
	(1)	(2)	(3)
Main Effects:			
Inverse Mills Ratio	-0.0190	-0.0223	-0.0240*
	(0.0191)	(0.0136)	(0.0135)
Average Treatment Effect	0.0477*	0.0432*	0.0471**
	(0.0244)	(0.0231)	(0.0224)
Interactions with Treatment:			
Inverse Mills	0.0633*	0.0738***	0.0808***
	(0.0324)	(0.0232)	(0.0224)
Male		-0.118**	-0.102**
		(0.0447)	(0.0459)
Age		0.00100	0.000332
		(0.00251)	(0.00278)
Married		-0.0762	-0.0747
		(0.0480)	(0.0492)
Average Distance to Farmers (log)		-0.0144	-0.0342
		(0.0550)	(0.0600)
Incumbent Party		-0.122***	-0.120***
		(0.0379)	(0.0433)
Digit Span			-0.0400*
			(0.0222)
Big5 — Stability			0.00882
			(0.0227)
Big5 — Plasticity			0.00617
			(0.0169)
R^2	0.007	0.018	0.022
p-value for Observable Interactions		0.002	0.002
Number of Phone Surveys	2228	2228	2228
Number of AEAs	134	134	134
Number of ALATs	44	44	44
Basic Controls		✓	✓
Cognitive Controls			✓

Regressions analogous to Table 6 but sample includes all large ALATs (100% coverage, 50% coverage, and control). Dependent variable for all regressions is whether the AEA visited the farmer in the week prior to the phone survey. The inverse Mills ratio is the generalized residual—the expected value of the error term—from the probit selection regression in the same column of Table A11. Regressions include survey round indicator. Main effects of basic and cognitive controls omitted for space. The *p*-value in the bottom row for observable interactions is from the joint Wald test for coefficients on treatment interacted with all included controls. Cluster robust standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Predictors of Productivity in the Control Group

	Farmer visited last week
Male	0.0398
	(0.0405)
Age	0.00201
	(0.00245)
Married	0.0700*
	(0.0371)
Average Distance to Farmers (log)	0.0188
	(0.0337)
Incumbent Party	0.0139
	(0.0307)
Digit Span	0.0317**
	(0.0127)
Big5 — Stability	0.00776
	(0.0160)
Big5 — Plasticity	-0.00979
	(0.00962)
R^2	0.018
p-value for Model	0.000
Number of Phone Surveys	1091
Number of AEAs	68
Number of ALATs	22

Sample is AEAs in large control ALATs. Dependent variable is whether the AEA visited the farmer in the week prior to the phone survey. Regressions include survey round indicator. The p-value for the model reports the F-test for the null that all coefficients equal zero. Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.