Environment and Development Economics (2024), **29**, 154–177
doi:10.1017/S1355770X23000165

RESEARCH ARTICLE

Monitoring and management of common property resources: empirical evidence from forest user groups in Ethiopia

Goytom Abraha Kahsay,[1] and Erwin Bulte[2]*

[1]Department of Food and Resource Economics, University of Copenhagen, Frederiksberg, Denmark and [2]Development Economics Group, Wageningen University, Wageningen, The Netherlands
*Corresponding author. E-mail: erwin.bulte@wur.nl

(Submitted 31 January 2023; revised 5 September 2023; accepted 5 November 2023; first published online 11 January 2024)

Abstract

The presence of monitoring institutions affects quality and effort of leaders. We investigate the effect of intensified monitoring on the ability and effort of leaders for a sample of forest user groups in Ethiopia, and find experimental and non-experimental evidence of an important trade-off: monitoring increases leaders' effort but lowers their quality in terms of education and experience. This effort–ability trade-off only occurs in the presence of alternative income opportunities (affecting the opportunity cost of time) and only among a subsample of leaders with low prosocial motivation. For our context, we document that the net effect of monitoring on economic outcomes is positive.

Keywords: common property management; leadership ability; monitoring and auditing; participatory forest management; rent capture

JEL classification: D02; D71; Q57

1. Introduction

The performance of organizations depends to a large extent on the quality of its workers and leaders. How can organizations attract and retain the 'right' set of individuals, in terms of ability and motivation, and how can they be incentivized to provide sufficient effort? A small literature considers this issue, mainly focusing on the effects of remuneration (wages).[1] In this paper we explore another dimension of the challenge

[1]Theoretical and observational evidence suggests that while higher wages may attract more able candidates (e.g., Ferraz and Finan, 2009; Gagliarducci and Nannicini, 2013), and deter corruption (Wadho, 2016), there exists a risk that applicants are primarily motivated by money (Francois, 2000; Delfgaauw and Dur, 2007; Prendergast, 2007; Gugerty and Kremer, 2008). This would adversely affect performance if effort cannot be perfectly monitored and contracted. Experimental evidence on this issue is scarce and conflicting (Dal Bó *et al.*, 2013; Ashraf *et al.*, 2015; but also Deserranno, 2019).

of recruiting, selecting and incentivizing high-quality individuals – the intensity with which their activities are monitored.

We focus on the context of communal natural resource management in a low-income country, and analyze how variation in monitoring intensity affects the quality and effort of group leaders. Attracting the 'right' individual is not an issue of overcoming an adverse selection problem (we assume leader ability is known), but one of inducing high-quality individuals to apply for the position. While monitoring should incentivize leaders to supply more effort, it may discourage high-ability individuals from applying for the position, because this may lower their private returns and these individuals have the highest opportunity cost of time. The net effect on leadership performance is therefore ambiguous (Grossman and Hanlon, 2014). The main objective of this paper is to analyze how intensified monitoring affects the ability (quality) and effort of leaders, and the overall effect on payoffs for group members. As a secondary objective, we explore to what extent the effects of monitoring on leader ability and effort is mediated by the extent to which leaders display prosocial preferences as well as the presence or absence of alternative income opportunities.

As a motivating example we study the type and behavior of (locally-elected) forest user group (FUG) leaders in Ethiopia, a particular form of participatory development programs that has become a dominant model of international development assistance and devolution of government responsibilities. These programs emphasize bottom-up participation by local communities in priority setting and implementation to leverage local knowledge, and crowd-in local contributions through an increased sense of 'ownership' (e.g., Mansuri and Rao, 2003). However, due to concerns about free riding and elite capture, monitoring of local leaders has appeared as an important issue.

More specifically, the empirical analysis is based on data collected among 132 FUGs in Ethiopia, allowing us to probe the relationship between group-level monitoring and our outcomes of interest.[2] Group leaders are selected by the group, and do not receive a salary for their work (even if some other forms of compensation can be offered (see below)). For all groups we have access to proxies or signals of leader quality (related to education and experience) and leader effort (mainly related to time spent organizing and contributing to patrolling 'expeditions' to protect the group's forest from incursions by others). We have rich cross-section (baseline) data, enabling a detailed analysis based on observational data. We complement this analysis with a randomized controlled trial (RCT) where we exogenously vary monitoring intensity. Monitoring is varied by sending in government inspectors, or by encouraging and facilitating bottom-up monitoring by group members. The number of observations where leadership changes occurs is small in the RCT data, but the findings are consistent with the observational study, and we hope that, together, the two analyses tell a convincing story.

[2]The analysis is based on the same observational and experimental data as Kahsay and Bulte (2019, 2021) and Kahsay and Medhin (2020). This paper extends Kahsay and Bulte (2019), who use cross-section data to consider the bi-directional relationship between within-group trust and formalization of rules; Kahsay and Medhin (2020) who use the same cross-sectional data to investigate the association between leader turnover and forest management outcomes; and Kahsay and Bulte (2021) and Kahsay *et al.* (2023), who use a field experiment to probe the causal effects of alternative monitoring modalities on forest-based income and its distribution across group members and forest conservation. In the current paper we zoom in on the mechanism linking monitoring to income, or the governance implications of intensified monitoring – the impact on the quality and effort of group leaders.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Our main results are as follows. Using the observational data we find a significant and meaningful *positive* correlation between the presence of a monitoring committee at the group level and leader effort (as intended), and a *negative* correlation between the monitoring committee and leader ability. These correlations are consistent with experimental results. Our treatments increased the leader turnover rate, and our monitoring treatments crowd out high ability leaders but crowd in leader effort. Groups assigned to monitoring interventions have leaders with, on average, less education and business experience. We find that stronger prosocial preferences by candidate leaders (in the sense of being more likely to place the group's interests before their own) weaken both the negative effect of monitoring on ability and the positive effect on effort. An auxiliary analysis reveals that the overall effect of monitoring on forest-based income of group members is positive.

The main contribution of this paper is threefold. First, we are the first paper based on observational and experimental data to demonstrate that monitoring introduces a trade-off between ability and effort of leaders. Second, we are able to say something about the channel linking monitoring to outcomes: alternative opportunities to earn income and leaders' prosocial motivation. We demonstrate that the crowding-out effect of monitoring on ability occurs in locations where alternative income-earning opportunities are available. The explanation we provide is that monitoring reduces the scope for corruption, which reduces the private returns to becoming a leader for individuals prone to corrupt behavior. Such individuals will stop volunteering for leadership positions if alternative, more remunerative income sources are available. Finally, our finding that the incentive effect (effort) of monitoring is more important than the selection effect (ability) is surprising and presumably context specific. In our context, leaders spend on average more than 40 days per year patrolling the forest – a massive time commitment, for which education and business experience appear of secondary importance. It is important to explore to what extent this finding spills over to alternative settings, where formal qualifications of leaders matter more.

The paper speaks to the literature on management of common property resources (e.g., Velez *et al.*, 2010), and also to the literature on recruiting public service workers – the domain of personnel economics of the state (but it is an open question whether the external validity of our study of unremunerated leaders extends to the selection of paid professionals). It is also related to the literature on community-driven development interventions. Recent impact evaluations of community-driven development initiatives have produced disappointing results (e.g., Casey *et al.*, 2012; Humphreys *et al.*, 2019). Some evidence points to elite capture and bad local leadership as potential explanations; see section 3.[3] Performance can improve through monitoring – either top-down by designated inspectors (e.g., Olken, 2007) or bottom-up by community members (Björkman and Svensson, 2009; Björkman *et al.*, 2017). But it is an open question whether performance gains can be sustained if monitoring crowds out high-ability leaders.

The paper is organized as follows. In section 2, we sketch a simple theoretical framework that shows how outside income opportunities, organizational resources, and monitoring institutions interact to shape leaders' ability and effort. The theory is general and may be applied to many contexts, including communal resource management. In section 3, we describe the background and context of our case study, outline the leader's

[3]References include Platteau (2004), Beekman *et al.* (2014), Kosfeld and Rustagi (2015), and Voors *et al.* (2018).

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

role and responsibilities, and explain how leaders are selected. Section 4 presents the experiment and sampling strategy and also introduces our data. In section 5, we outline our identification strategy. In section 6 we present the main results and a robustness analysis. Section 7 concludes.

2. Theoretical framework

We consider endogenous leadership selection in local community organizations in which leaders' ability, but not their effort, is observable. Our theoretical framework is based on Grossman and Hanlon (2014), but has two new features. First, we allow the level of prosocial motivation of the leader to vary. Second, reflecting the nature of our case study (elaborated below) in which natural resource abundance matters, we assume that the returns to providing effort by the leader varies with the abundance of the local resource (the standing stock, R).

The augmented Grossman-Hanlon model is as follows. Consider a group of N members who differ in terms of their innate ability, A_i, and can be ranked from low- to high-ability individuals: $A_i \in (0, \hat{A})$. Ability levels are public knowledge, or there are credible signals that all members can 'read' and that can be used to learn something about unobservable dimensions of quality (in the sense of Spence (1973)). Following Grossman and Hanlon (2014), we assume that one group member will be selected as the leader, and is henceforth responsible for producing a public good valued equally by all group members (but who does not receive a salary – leaders are 'volunteers'). Each group member has one unit of time. Ordinary group members allocate all their time to an income-earning activity, but leaders divide their time between the income-earning activity and public good provision.[4] The productivity in both activities depends on the leader's ability and effort. In the context of our case study, below, public good provision amounts to (overseeing the) management of a communal resource, which provides a flow of benefits for all group members.

Specifically, the value of public good P produced by leaders depends on their ability (A_l), the share of their effort (time) devoted to public good production (e_l), the resource stock (R), which is assumed to be exogeneous, and a random noise term with mean value of zero (η_p).[5] Assume complementarity between effort and ability in producing public goods ($\partial P_l/\partial A_l \partial e_l > 0$), and between leader's effort and the forest stock, as well as between the leader's ability and the forest stock in producing the public good ($\partial P_l/\partial A_l \partial R > 0$ and $\partial P_l/\partial e_l \partial R > 0$). The value of the local public good is given by $P = p(A_l, e_l, R, \eta_p)$.[6] The leader's private income depends on ability (A_l), the share of effort devoted to income generation ($1 - e_l$), as well as a random noise term (η_I), and is given by $I(A_l, 1 - e_l, \eta_p)$. We again assume complementarity in production ($\partial I_l/\partial A_l \partial e_l > 0$). Ordinary group member j earns a private income $I(A_j, 1, \eta_I)$, where η_I is a random noise term with mean value of zero. Assume the Inada conditions hold.

[4] Alternatively, and without loss, we could assume that ordinary members allocate a fixed amount of time to producing the public good. This situation is actually closer to the truth for the case study introduced below: all group members are supposed to contribute, but in addition to contributing himself, the leader also plans, manages and oversees the effort.

[5] In a more complex model one could make the forest stock endogenous, varying with the level of the public good provided by the leader. To keep the model tractable we treat the stock as a fixed endowment, or parameter.

[6] In the absence of a group leader, $e_l = 0$ and $p = \eta_p$.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

The leader also faces additional costs (or rewards), defined as $C(m, e_l)$. These costs depend on the level of monitoring by others, m, and effort. Monitoring enters the problem because effort levels, unlike ability, are not perfectly observed and cannot be inferred from P (because of the random error term η_p). Negative costs imply rewards, which may reflect (effort-dependent) compensation by group members, and graft or corruption. In our context, diversion of group resources is relevant, resulting in a flow of private benefits. The scope for engaging in corrupt behavior is a decreasing function of monitoring. Monitoring reduces opportunities for graft and corruption, and can also be costly for the leader because of direct costs associated with complying with certain behavioral imperatives or rules.[7] Assume the cost (reward) function is twice differentiable, increasing and convex in monitoring intensity m, and decreasing in effort allocated to public good provision e (i.e., leaders receive a lower penalty, or greater reward, if they work harder). We also assume that the marginal benefit of supplying effort is increasing in monitoring intensity: $\partial^2 C/\partial m \partial e < 0$.

In addition to behavior-specific costs and rewards there is a fixed candidacy cost (ϕ) associated with running for the leadership. This may include opportunity cost of time or social obligations that materialize as a consequence of efforts to mobilize sufficient support. In line with recent literature (e.g., Fedele and Naticchioni, 2016), payoffs of leaders depend on their prosocial motivation β. Leaders who intrinsically care about the well-being of their group members derive additional utility from supplying the public good as this makes others better off. Summarizing, the leader's payoff (Y_l) can be written as follows:

$$Y_l = \alpha I(A_l, 1 - e_l, \eta_I) + (1 - \alpha)(1 + \beta)P(A_l, e_l, R, \eta_p) - C(m, e_l) - \phi, \qquad (1)$$

where α is a group-level indicator of the availability of private income opportunities relative to the value of the public good. We assume all group members face the same market conditions for their non-forest activities, but payoffs per unit of time vary with individual ability.

The 'voters' side of the model is simple, even if the voter label is a bit of a misnomer in the context we study (see section 3). We abstract from social capital and prosocial preferences among voters and assume away the second-order social dilemma of who will invest in monitoring. We also assume a simple leader selection process, where voters support the leader who delivers the highest expected income.

Timing of the model is as follows. Group members first decide whether they want to be candidates. Then there is an election and, finally, the elected leader decides how much effort to exert. The model can be solved using backward induction: (i) potential leaders determine their optimal effort level e_l^*; (ii) group members decide whom to select; and (iii) the expected outcome of the election determines who runs for office (candidates compare the payoff from running and not running). See the online appendix for details. This simple model produces the following testable predictions:

1a. In the presence of outside income opportunities, increasing monitoring can cause high-ability leaders to self-select out of candidacy; but…
1b. … this effect is weaker for leaders with strong prosocial motivation.

[7]The group does not raise the leader's compensation (if any) in response to stricter monitoring. While potential rents by group leaders is not explicitly modelled for simplicity, the cost of the introduction of monitoring institution on leaders includes reduction in rents.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Increasing monitoring intensity reduces returns of leaders, but the impact of monitoring on the quality of group leaders varies with the opportunity cost of their time. In the presence of outside income opportunities, the subsample of candidates for whom 'running for office' is privately optimal is reduced. In particular, high-ability individuals with high opportunity costs will opt out of the race. The extensive margin shifts to an individual with lower ability and lower opportunity costs. However, this effect is attenuated by prosocial motivation. Prosocial individuals are less likely to be motivated by private profits when taking on a leadership role, and are less likely to display corrupt behavior.

2a. For any given leader, the effort level chosen is increasing in the level of monitoring and decreasing in outside income opportunities; but…
2b. … these effects are weaker for leaders with strong prosocial motivation.

Leaders have to determine how to divide their time between private activities and effort for the public good. As outside opportunities become more attractive, leaders re-allocate part of their time to private activities – raising the marginal benefit of time allocated to public good provision until the marginal benefit across activities is equalized. Since the benefits of increasing effort are greater for greater levels of monitoring, $\partial^2 C/\partial m \partial e < 0$, it follows directly that more monitoring will raise effort.[8] A stronger prosocial motivation increases effort levels but dampens the accentuating effect of intensified monitoring. Increasing the leader's valuation of the public good implies a re-allocation of labor towards the public good, raising the marginal value product of effort allocated to 'other activities'.

3. Increasing natural resource endowments induces higher-ability leaders to serve.

A greater resource stock raises the marginal value product of effort allocated to public good provision ($\partial^2 P/\partial e \partial R > 0$). This generates greater benefits for all group members, including the leader. Running for the leadership position will become optimal for a larger set of group members – especially more able ones as $\partial^2 P/\partial A \partial R > 0$.

3. Background and context

Ethiopian forests suffer from widespread degradation and deforestation – forest cover declined from 40 per cent at the end of the 19th century to 4 per cent in recent years (Dessie and Christiansson, 2008).[9] In response to this trend, the Ethiopian government initiated a large Participatory Forest Management (PFM) program – decentralizing forest management and giving local communities use rights over demarcated forest blocks. One of the earliest such programs was piloted in the mid-1990s in the Adaba and Dodola districts of West Arsi zone, Oromia regional state. In collaboration with

[8]We do not consider prosocial preferences of non-leader group members in our model, but it is possible that this may have an effect on monitoring. For instance, Nannicini *et al.* (2013) show theoretically and empirically that social capital improves political accountability (see also Rustagi *et al.*, 2010).

[9]Despite recent large-scale tree plantation efforts, Ethiopia still has a high rate of natural forest loss. For instance, the 2015 FAO forest resource assessment (FAO, 2015) show that Ethiopia lost about 17 per cent of its forest area between 1990 and 2015, and the 2020 assessment showed that the forest area lost is about 11 per cent between 1990 and 2020. Furthermore, the Global Forest Watch (2023) report that Ethiopia has lost 4.5 per cent of its humid primary forest and 3.9 per cent of its tree cover in the period between 2000 and 2022.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

the German Society for International Cooperation, so-called forest user groups or FUGs were established and given exclusive use rights.

FUGs have a maximum group size of 30 individuals, so villages have multiple FUGs (minimum 4 and maximum 20 FUGs per village). With the help of experts from Oromia Forest and Wildlife Enterprise (OFWE), FUGs drafted their own bylaws. This included details regarding the organizational structure, specification of the rights and responsibilities of group members and leaders, and establishment of specific committees (including the monitoring committee featuring in our empirical analysis). While OFWE provided sample bylaws, groups were free to specify their own rules. A 'typical FUG,' consists of a General Assembly (all members), an executive committee of five members (chairperson, vice chairperson, secretary, cashier, and one member), and perhaps some technical committees (forest development committee, forest rent determining committee, forest products marketing committee[10]).

The sample bylaws also propose that FUGs have a monitoring committee, and about 30 per cent of the FUGs in our sample indeed have such a committee – introduced when the group was established (see below). Monitoring committees are charged with the responsibility of monitoring the behavior of the group's chair (or leader). We will exploit the presence or absence of such monitoring committees in our analysis of observational data, below.

After drafting the group bylaws, leaders and committee members are selected by the group members in a participatory process. The details of this process deviate slightly from group to group, but the main elements are consistent across groups and reflect existing (informal) institutions. The selection process is not based on anonymous voting. Instead, as written by Deserranno *et al.* (2019: 240) who study leader selection in microfinance service groups in rural Uganda, '*groups are embedded in local power structures that can enter their governance and create a bias in favor of influential community members*'. During a meeting of the General Assembly, members engage in a public discussion about who are suitable candidates and whom should be selected for the various leadership positions. Any member may volunteer to take on such responsibilities, but not all members have an equal chance of being supported by their peers. Deserranno *et al.* (2019: 241) write '*when decision-making is public, less powerful members can be coerced or intimidated into supporting certain* [*candidates*], *tilting selection decisions towards outcomes that are less representative of the low-income group members' preferences*'. This is consistent with our understanding of how group leaders are selected in Ethiopia – if local elites believe that assuming a leadership position is privately profitable, they may 'muscle their way' into such positions (see also Gugerty and Kremer, 2008). Kahsay and Bulte (2021) document that leaders '*have higher income and consumption levels, are more educated and wealthy, and often have a role as religious or clan leader*'.

OFWE was concerned about elite capture and corruption in our study region. Kahsay and Bulte (2021) study elite capture and corruption by FUG leaders and discuss several approaches through which elites can 'grab' more than their fair share of forest rents: '*In* [*some*] *cases, the poor lack complementary assets to benefit fully from forest resources. For example, forage from forests is especially valuable for livestock-owning households – the elites. If forests are managed to maximize the flow of forage, poor community members may benefit little. There is also anecdotal evidence on cases of blatant abuse, where*

[10]While the sample bylaws include a forest products marketing committee, FUGs rarely organize joint marketing. Members typically utilize their individual quota and sell their harvested output individually on local markets.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

proceeds from the collective sale of forest products are appropriated or divided based on family and clan ties as well as bribes from (non) members to facilitate illegal extraction of forest resources.' (p. 114).

The group bylaws stipulate the role and responsibilities of group leaders. These responsibilities are varied, and include patrolling the forest and organizing and overseeing the patrolling of fellow group members. Illegal extraction of forest resources should be minimized, and individuals engaged in illegal extraction – whether they are group members or others – should be punished. In addition to patrolling, leaders play a role in legal extraction, perhaps negotiating sales with outside buyers of tradable forest resources (e.g., timber), and liaise with the government when necessary. The bulk of the time that group leaders allocate to group-based activities, however, goes to patrolling. As shown below, on average leaders spend some 40 days per year patrolling the forest. While being educated and having business experience are relevant characteristics for leaders, it is imperative that leaders are willing and able to allocate enough time (effort) to this key component of their job description. As mentioned, leaders do not receive financial compensation for their time. Instead, they may expect indirect financial and non-financial benefits, including benefits from the public good and legal forest harvesting; participation in workshops and trainings, which are sometimes organized by OFWE and involve per diems; free labor assistance from group members during agricultural activities; and non-pecuniary benefits in the form of status. In addition, there may be opportunities for covert diversion of group income or side-selling of forest assets.

This setting is ideal for testing the predictions of the augmented Grossman and Hanlon model. Leaders of the FUG are voluntary group members who are 'selected' by their peers, who know them well as they live in close proximity (e.g., Hussam *et al.*, 2022). It is reasonable to assume that leader ability is known, or at least signals of ability. Ability is relevant as leaders have to organize forest management, keep records, and negotiate with traders, the government and NGOs. On the other hand, leader effort is also important and likely imperfectly observable – it is at least partly private information. Group size is fixed, and there is no entry in response to rent grabbing opportunities.[11] Leadership is a part-time activity and leaders must decide how to allocate their time between public good production and income generation.

In our setting, the hypothesized link between monitoring and leadership quality is as follows. We hypothesize that more intensive monitoring will reduce the scope for engaging in 'cases of blatant abuse' – reducing the private profitability of taking on a leadership role for individuals who are prone to corrupt behavior (individuals with weak prosocial preferences). Some individuals therefore opt out of the race for leadership, and the quality of the remaining top candidate (weakly) goes down. This will be particularly true if relatively well-educated local elites opt out. Incentives to opt out will be strongest in settings where leaders have alternative income opportunities.

4. Experiment and data

This paper is based on observational data, and data from an RCT. We have access to two waves of data collection in Adaba and Dodola districts, involving all 132 FUGs from the

[11] Groups cannot admit new members, except when members pass away or have permanently migrated out of the *kebelle*. Entry and exit is very rare since group members are assumed to represent their respective households and thus a group member who dies or permanently migrates could easily be replaced by another household member.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

13 villages (*kebelles*) in the study area. Baseline household data and OFWE registration data were collected in the Spring of 2017 (see Kahsay and Bulte, 2019). We explore the correlation between the presence of monitoring committees and leader ability and effort using these baseline data (cross-section analysis). A randomized monitoring intervention started in May 2017, introduced below, and end-line data were collected in autumn of 2018 (see Kahsay and Bulte, 2021). We leverage exogenous variation in monitoring induced by the experiment, and estimate a diff-in-diff model combining baseline and end-line data. The experiment was based on informed consent of all FUG members and there was no attrition or non-compliance.

The RCT approach is superior from an identification perspective (due to exogenous variation in monitoring) but our experimental results are driven by a relatively small number of cases where leaders were replaced by new ones during the study period of 16 months due to the monitoring (leader turnover during the intervention increased from less than 3 per cent for the control group to more than 17 per cent for the treatment group). Instead, the results of the observational analysis are based on a cross-section of the RCT panel – using all 132 leaders in our sample. We, therefore, believe there is merit in presenting *both* the observational and RCT results. Together, these complementary approaches provide a rich picture of how monitoring affects leader ability and effort.

4.1 The experiment

We randomly assigned FUGs to experimental arms – three treatment arms (98 FUGs) and one control arm (34 FUGs). As described in detail in Kahsay and Bulte (2021), we had three distinct monitoring modalities: (i) 'bottom-up' by group members (33 FUGs), (ii) 'top-down' by the government (33 groups), and (iii) 'top-down' combined with a reward for good leadership (32 groups). Groups in the control arm rely on business-as-usual monitoring by group members, possibly organized in pre-existing monitoring committees. Approximately one-third of the groups had a monitoring committee at baseline. Random assignment of FUGs to treatment arms resulted in a balanced distribution, where all arms contain groups with and without pre-existing committees. Not all pre-existing committees were active, and some level of informal monitoring always takes place – also in groups without a committee. Conceptually, therefore, the interventions imply that the (average) intensity of monitoring in treatment arms is greater than in the control arm.

For the bottom-up arm, OFWE experts visited selected FUGs and engaged in a series of activities in the presence of all group members. They assisted in the creation of new monitoring units (not to be confused with pre-existing committees, analyzed in the observational study), oversaw an anonymous voting process where unit leaders were selected, trained these unit members to monitor group leadership and performance, and provided them with material inputs to facilitate the monitoring task (e.g., pencils and score cards with key indicators). OFWE recommended inspection of forest blocks, books and records, and organization of group discussions and interviews with anonymous group members and non-members. After six months, OFWE organized and moderated group-level meetings where units presented their findings to the general assembly, which were subsequently 'benchmarked' by OFWE experts. After this, group members were encouraged to express appreciation for their leaders, to warn them, or in severe cases to punish them (dismiss them).

The two top-down monitoring arms were based on regular inspections by OFWE officials. They evaluated leadership performance based on the same criteria as offered to

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

the group members in the bottom-up arm. Some 6–8 months later, findings were shared and discussed with group members during a general assembly meeting, an event during which additional information was collected. Next, a final report was prepared and shared with regional OFWE headquarters. While the monitoring process was identical for these two arms, the implications of monitoring were not. In one arm, OFWE management could warn or punish the worst performing leaders (by dismissing them), and in the other arm OFWE management provided the top performing group leaders with a material reward (a solar panel with a market value of USD60). Leaders dismissed both in the bottom-up and top-down monitoring treatments continued as group members.

The choice of treatment interventions was partly motivated by OFWE's own observations regarding elite capture and governance challenges as the main factors undermining the sustainability of the PFM program. The three treatments imply intensified scrutiny for the group leader. For the analysis, we combine the three treatments into one large treatment group (98 FUGs) for two reasons. First, given the low intervention-induced leader turnover rate (17 per cent for the treatment groups), combining the treatments increases statistical power.[12] Both treatment and control groups have on average the same duration of leaders at baseline (5.2 years *vs* 6.2 years, respectively, $p = 0.12$). Second, the theory does not distinguish between alternative monitoring modalities, although the intensity of monitoring (and, hence, the impact) may vary across treatments. The number of FUGs, the total population and our sample of respondents for each experimental arm are presented in online appendix table A1.

Changes in the rate of leader turnover are an important outcome of the experiment. Groups whose leader was dismissed selected new leaders, and members have the option to self-select into candidacy for this leadership position. Since group size and composition remained the same in all groups, and members (the 'pool of possible candidates') in treatment and control groups have the same average education, this experiment enables investigation of the impact of monitoring on (self-) selection of high ability candidates into or out of the leadership position.[13]

4.2 Baseline data

We interviewed all FUG leaders and a random sample of one-third of the group members – 1,222 respondents in total. We also use OFWE administrative data, including the year of group establishment, group size and composition, the location of forest blocks, and measures of resource abundance (i.e., counts of PCTs/ha and MTs/ha).

As dependent variables we collect information on leader effort and ability. Leader effort is measured as the number of days in a year spent patrolling the forest and overseeing the patrolling of others. As mentioned, active engagement of leaders in patrolling is a key part of their responsibility and a necessary ingredient for group performance. It sets an example for other group members to engage in patrolling, helps leaders to get information on forest extraction and patrolling behavior of members, and provides

[12]The results do not change, qualitatively, if we drop observations from the 'top-down monitoring and reward' arm, and only consider the impact of top-down and bottom-up monitoring followed by possible punishment.

[13]Ideally we would compare the distribution of individual characteristics among (a) the potential pool of candidates in each group (all group members); (b) the actual pool of candidates, or those who run for office in each group; and finally, (c) the leader ultimately elected. While this comparison would provide a complete picture of the effects of the intervention on selection into leadership, we do not have data on the actual pool of candidates (group b).

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

them with an opportunity to meet other members and discuss group issues. On average, leaders work 42 days per year to provide this public good.

We collected two types of measures of leader ability, although it is probably better to think of them as signals of leader ability.[14] First, we collected information on education levels (as a categorical indicator, from zero reflecting 'never attended school' to six reflecting 'university diploma or degree') and years of schooling. Leaders' schooling ranges from 0 to 13 years, and on average leaders have 4 years of schooling. Second, to gauge business experience we asked leaders whether they earned any income during the past 12 months from running a business or engaging in trade.[15] Some 12 per cent of the group leaders answered in the affirmative.

Another dependent variable in some models is forest-based income. We measured forest benefits by presenting FUG members with a list of commonly extracted forest products (e.g., firewood, poles, tree bark, lianas and vines, bamboo, tree branches, logs). We asked whether they extracted the forest product, quantities extracted (in local measures), and the purpose of extraction (sale or consumption). We generated a per unit market price at the village level for each of the product types, and calculated a household-level measure of forest income.[16]

For the analysis based on cross-section baseline data, our key explanatory variable is the presence or absence of a monitoring committee (called *Monitoring*). We inspected the local bylaws, and also asked members whether their group has a separate 'committee that monitors the activities of leaders'. Enumerators were clearly instructed to explain the role and function of these committees (as described in section 3) to members. We find that 32 per cent of the groups have a monitoring committee at baseline.

For the cross-section analysis we consider several moderating variables. First, our theory predicts that the impact of monitoring varies with the prosocial motivation of leaders, which we proxied by asking the following question: 'How would you rate your own financial contribution to organize community activities or overcome collective problems?'[17] Leaders could respond on a scale from zero 'Not at all' to four 'A great deal'. We categorize leaders as pro-socially motivated if they scored a two, three or four (the results are robust with respect to alternative thresholds). Observe that this survey question is self-reported and not incentivized. This question may leave out time contributions of leaders. As a robustness analysis we used an alternative question: 'suppose that a community project does not directly benefit you or any other member of your household, but

[14]Our education and experience variables are perhaps best thought of as signals of quality – not as measures of leader quality. Signals are credible if the cost of sending a signal of high quality is relatively low for high-quality individuals (Spence, 1973). For intelligent individuals, educational achievements are relatively easy to obtain, and for business-savvy individuals it is not so risky to start a small enterprise. If we assume that (i) group members care about intelligence and 'business-savviness', and that (ii) obtaining schooling and running a business are easily-observable signals of these characteristics, then our quality variables are correlated with quality as demanded by group members and our story goes through.

[15]Most business activities in the study area include cattle fattening and dairy business, retail business, and commercial crop farming.

[16]We have also re-estimated these models based on an aggregate market price (the average of the various local market prices) and we found that the results are unaffected (not shown).

[17]In rural Ethiopia, members commonly contribute money (and time) to informal local institutions, which is used to purchase inputs for the construction of village roads, schools, or religious institutions, or may be used to help families affected by idiosyncratic shocks. This measure of prosocial motivation is not significantly correlated with our proxy for wealth.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

has benefits for many others in your village/community, would you be willing to contribute time or money to the project?' While this question does not ask about actual prosocial behavior of leaders as in the first question, such measures of willingness to contribute are often used as a proxy for prosocial behavior. Using the willingness to contribute instead of the financial contribution does not change our results (results not reported but available on request).

Other moderating variables are outside income opportunities (called *OIO*) and the value of the standing forest stock (*VT*). Following Grossman and Hanlon (2014), we measure the former as the share of group members earning an income from casual and regular employment, running their own business, or involved in trading. The average share of group members earning an income from off-farm employment or business is 18 per cent. The resource value was measured using the number of PCTs/ha (Rustagi *et al.*, 2010). This data was collected in 2004 for only a subsample of the FUGs, and we use this as a measure of the initial forest stock. Forest blocks have on average 35 PCTs/ha, but the value of standing timber varies with distance to infrastructure and (local) markets and it may be unprofitable to harvest and sell the PCTs in remote areas. To capture variation in the stumpage value of trees across groups we multiply the number of PCTs/ha with the inverse of distance to the nearest market (measured in hours of walking time). This new variable is called *VT* in what follows. We, alternatively, generate the *VT* variable as (i) the product of average village price and number of PCTs/ha, and (ii) the product of average village price and number of PCTs/ha divided by the distance to the nearest market. See the distribution of these key dependent and moderating variables in online appendix figure A1. We present both models *with* the resource value variable to see how outcomes compare to the theory, and *without* the variable to benefit from the full sample of FUG data.

We also collected information on a vector of socio-economic and demographic respondent characteristics (including gender, literacy, household size, wealth, and so on), and on group-level variables. FUG members in our study area are dominated by one ethnic group (the Oromo) and one religion (Islam). However, since members belong to different clans we can generate a clan fractionalization index to measure socio-cultural heterogeneity (Poteete and Ostrom, 2004). The value of this index ranges from 0 to 0.84 in our data. We also measured wealth inequality (income, land, livestock and assets) using a Gini index, and collected geographic variables (altitude, distance to market, and distance to asphalt road) as well as group characteristics such as age of the group, group size and share of female members.

4.3 End-line (experimental) data

In 2018 we revisited the FUGs and collected additional data on forest benefits and leadership. We collected data from 1,215 households (out of the 1,222 interviewed in the baseline period). In addition, we interviewed the new leaders who replaced the old leaders who were fired during the experiment. Leader education and effort was measured as at baseline, enabling the estimation of a diff-in-diff model. On average, achieved education levels decreased during the study period. Unfortunately, business experience of the group leader was not measured consistently across the two survey waves. While at baseline we proxied business experience by whether or not leaders earned any income from household-owned business or trade-related business, we directly asked leaders about their business experience in the end-line survey. All variables, at baseline and end-line, are summarized in online appendix table A2.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

5. Econometric method and identification strategy

To identify the effect of monitoring on leaders' ability and effort, we estimate OLS and diff-in-diff models. Consider the former first. Using the baseline data, we regress measures of ability and effort of the leader of group i, Y_i, on a dummy variable indicating the presence of a monitoring committee, $Monitoring_i$, a measure of Outside Income Opportunities (OIO_i), the interaction between these terms, and additional controls, X_i:[18]

$$Y_i = \beta_0 + \beta_1 Monitoring_i + \beta_2 OIO_i + \beta_3 Monitoring_i \times OIO_i + \beta_4 X_i + \alpha_z + \varepsilon_i. \quad (2)$$

In (2), α_z captures *kebelle* (village) fixed effects to capture time invariant factors like geophysical conditions and local governance (one *kebelle* hosts multiple FUGs), and ε is a random error term. We are mainly interested in coefficient β_3 that captures the conditional association between ability or quality and the presence of a monitoring committee (when outside income opportunities are available). In additional models we include (interaction terms involving) our measure of the value of the baseline forest stock. This enables exploration of whether high-ability leaders are attracted by the prospect of greater resource wealth.

To interpret the association captured by β_3 as the 'causal effect' of monitoring institutions (in the presence of outside income opportunities) on leader ability or effort requires additional assumptions. Specifically, we need to assume that the presence of the monitoring committee is an exogenous variable in (2). However, it is possible that monitoring institutions are established *in response* to (expectations about) weak leadership, or that certain factors for which we cannot control drive both leadership and monitoring. In other words, the presence of monitoring institutions may be endogenous in models explaining dimensions of leadership.

This concern is mitigated, but not eliminated, by the following considerations. First, we have access to a rich set of controls, attenuating concerns of omitted variables. Second, as mentioned, much of the variation at the group level is caused by expectations and beliefs of the implementing agent of OFWE. Local OFWE officers worked with villagers to modify the sample bylaws in accordance with local needs. While this process could be informed by local conditions, our data suggest this is not the case.[19] Third, our institutional measures are quite persistent. Personal communication with villagers and OFWE representatives revealed that groups that have a monitoring committee established the committees at the beginning of the FUG-creating process – when local bylaws were drafted. The election of the first cohort of leaders happened afterwards, and since

[18] In models explaining variation in leader effort we estimate specification (2), and use as dependent variable the number of days spent by the leader patrolling the forest, e_l. We follow the specification of Grossman and Hanlon (2014) and include ability, the interaction between ability and outside income opportunities, and the interaction between monitoring and ability in vector X.

[19] Our focus-group discussions with FUGs suggest that variation in groups' monitoring institutions is mainly due to variation in the recommendations of local OFWE advisors. This is also confirmed by previous and current OFWE experts and leaders in Adaba and Dodola districts, including personal communication with Dr. Aklilu Ameha, who was an advisor and a team leader at the beginning of the Adaba-Dodola PFM program (2000–2007). Unfortunately, we do not have information linking specific advisors to FUGs, so we cannot use the identity of the advisor as an instrumental variable to predict the existence of Monitoring Committees. Online appendix table A3 documents that the presence of committees is not associated with explanatory variables. This is generally the same for the sample of groups with less pro-socially motivated leaders and sample of groups with more pro-socially motivated leaders, as can be seen from the second and third columns respectively of online appendix table A3.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

these days nearly all groups have changed leadership (typically more than once) (Kahsay and Medhin, 2020). As a robustness analysis, we consider the subsample of FUGs with their leaders elected during the past six years – with the institutional setting firmly in place. Finally, we examine whether our results are sensitive to omitted variable bias by the coefficient stability approach (Altonji *et al.*, 2005; González and Miguel, 2015; Oster, 2019).

Notwithstanding these considerations, we understand that not all concerns about endogeneity of the monitoring variable can be eliminated with our cross section data. We therefore also combine our baseline and end-line measures of ability and effort, and the random assignment to the intensified monitoring treatment, and estimate the following ability model:

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 OIO_i + \beta_3 Year + \beta_{4j} OIO_i \times T_i + \beta_5 OIO_i \times Year + \beta_6 T_i \times Year + \beta_7 T_i \times OIO_i \times Year + \beta_8 X_{it} + \alpha_v + \varepsilon_{it}. \quad (3)$$

In (3), T_i is the treatment dummy, OIO_i is the measure of outside income opportunity, and *Year* is a year dummy (1 if end-line period, 0 otherwise). Our coefficient of interest is β_7, associated with the triple interaction of outside income opportunities, the treatment, and the end-line dummy. In the diff-in-diff model, the monitoring committee variable measured at baseline enters as one of the controls. For the effort model we estimate a simple diff-in-diff model in which we interact the treatment and end-line dummies (and include outside income opportunities as a control variable).

Finally, to probe the net welfare effects of monitoring institutions we estimate a model explaining variation in average forest-based income, *F*, across FUGs. Since monitoring is expected to increase effort but lower ability, the net effect on welfare is ambiguous. We estimate a simple OLS model with a dummy variable for the presence (or absence) of a monitoring committee as the main explanatory variable which enables us to probe the net effect.

$$F_i = \beta_0 + \beta_1 Monitoring_i + \beta_2 \boldsymbol{X}_i + \alpha_z + \varepsilon_i. \quad (4)$$

Again, X_i is a vector of controls and α_Z are *kebelle* fixed effects. We also estimate a diff-in-diff model with assignment to the monitoring treatment, T_i, as the main explanatory variable (interacted with the end-line dummy):

$$F_i = \beta_0 + \beta_2 T_i + \beta_2 Year + \beta_3 T_i \times Year + \beta_2 \boldsymbol{X}_i + \alpha_z + \varepsilon_i. \quad (5)$$

We estimate the OLS and diff-in-diff models using two different samples. First, we include many of the key control variables also used by Grossman and Hanlon (2014), and use the largest sub-sample for which these variables are available. Including controls is especially important for the OLS analysis, but also helps to increase precision of the diff-in-diff estimations.[20] Second, we re-estimate all models for larger samples without controls (but including *kebelle* fixed effects). The results and statistical significance of the two samples are very similar, and the same is true for models based on a restricted vector of controls (selected using LASSO). To economize on space we only report the regression

[20]For a few new leaders who were initially *not* part of the panel, we collected the following variables: age, education, years of schooling, forest patrolling time, and business experience. We did not collect data on other characteristics such as income, livestock holding and household size (included as controls in baseline regressions).

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

results of the former sample.[21] Third, we estimate models where we imputed missing values by the mean of the variable, and included dummy variables to indicate imputation. Fourth, we estimate a series of ANCOVA models. Finally, we estimated models with the three monitoring interventions entering separately, rather than pooled into one monitoring dummy. In all models, we cluster standard errors at the *kebelle* level. However, results without clustered standard errors (not reported here, but available upon request) remain the same.

6. Results

6.1 Results based on observational data

Our first set of results based on the observational (baseline) data is presented in table 1, explaining variation in leader ability. Columns (1)–(2) use the categorical education indicator as a measure for leader's ability and include *kebelle* fixed effects, and in column (2) we include FUG-level controls. Columns (3)–(6) present results using years of schooling and business experience as alternative proxies for leader's ability and include village fixed effects (again, without and with controls). Adding controls reduces the number of observations due to missing values for some variables. For comparability, we re-estimated columns (1), (3) and (5) on the sample used for columns (2), (4) and (6) and results remain the same (not reported, but available on request).

The main results are in line with prediction (1a) of our theoretical model. In the presence of outside income opportunities, there exists a negative association between monitoring institutions and leader's ability for all proxies of ability, both in the parsimonious specifications and elaborate models with fixed effects and controls. The estimated coefficients are also economically meaningful. For instance, keeping *OIO* at 30 per cent, groups with a monitoring committee have leaders that score up to 1.69 lower in terms of the categorical education indicator; have 1.27 fewer years of schooling; and 39.18 percentage points less business experience.

According to prediction (1b) of the theoretical model, leaders with strong prosocial motivation should be less sensitive to the attractions of outside income opportunities because they derive utility from supplying the public good. This prediction is supported by the data. Column (7) presents results for leaders with 'relatively low prosocial motivation,' and column (8) presents results for the complementary subsample of prosocial leaders. In both models, we explain variation in ability as measured by the categorical education variable. The results in earlier columns are driven by leaders with relatively low prosocial motivation, who appear to respond strongly to incentives associated with the opportunity cost of their time. The relevant coefficient for the group of prosocial leaders is not statistically different from zero and has the opposite sign.

Column (9) presents estimation results for the subsample of leaders who have been in office for a relatively short period – years after monitoring committees were established (and therefore less likely to suffer from endogeneity concerns). We restrict the sample to leaders who have been in power for less than 6 years. Our estimation results confirm that monitoring committees are associated with lower leader ability in the presence of outside income opportunities.

[21] The regression results for the sample without controls or the sample based on LASSO are available on request. For only one specific model do we find that the regression results of the larger sample without controls loses its statistical significance at conventional levels. We explicitly mention this case, below.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Table 1. Monitoring and leader ability (observational data)

	Education	Education	Years of schooling	Years of schooling	Business experience	Business experience	Education	Education	Education
							Less pro-socially motivated	More pro-socially motivated	Duration in power <6 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Monitoring	0.244	0.918	0.696	2.569	−0.009	−0.007	0.286	−0.552	5.022
	(0.290)	(0.607)	(0.882)	(1.410)	(0.029)	(0.123)	(0.496)	(1.442)	(0.872)
OIO	1.074	2.081	1.271	3.982	0.606	0.592	0.890	−3.092	4.097
	(0.992)	(1.663)	(3.035)	(3.787)	(0.200)	(0.507)	(2.599)	(5.367)	(3.001)
Monitoring × *OIO*	−3.284	−5.633	−7.874	−12.585	−0.395	−1.306	−5.295	0.248	−21.901
	(1.217)	(2.124)	(3.522)	(5.515)	(0.468)	(0.619)	(2.350)	(5.984)	(9.518)
Controls	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	2.431	251.392	2.989	751.061	−0.032	67.374	18.199	1,836.349	253.248
	(0.142)	(140.174)	(0.482)	(460.415)	(0.025)	(46.102)	(176.655)	(2,885.483)	(315.452)
R^2	0.208	0.497	0.155	0.489	0.284	0.493	0.668	0.903	0.865
Observation	125	99	125	99	130	99	59	39	49

Notes: Clustered (at village level) standard errors in parentheses. Included explanatory variables: age of the leader, income of the leader, livestock holding of the leader, clan of the leader, household size of the leader, altitude, year of establishment, group size, share of female members, average age of members, share of members who can read and write, average household size of members, share of trusting members, share of members who have non-farm income (employment and business), income heterogeneity, land heterogeneity, clan fractionalization index, average distance to market, average distance to asphalt road, value of standing timber stock, and number of potential crop trees per ha. The coefficient of *Monitoring* × *OIO* for less pro-socially motivated leaders is statistically different from more pro-socially motivated ones (Wald test).

Table 2. Monitoring, resource rents, and leader ability (observational data)

	Education	Years of schooling	Business experience
Monitoring	1.415 (0.386)	3.901 (1.136)	0.065 (0.193)
Monitoring × *VT*	−0.024 (0.021)	−0.049 (0.075)	−0.002 (0.011)
VT	0.019 (0.013)	0.086 (0.036)	0.003 (0.004)
OIO	2.709 (1.823)	6.476 (5.064)	1.318 (0.956)
VT × *OIO*	0.008 (0.072)	0.048 (0.192)	−0.037 (0.029)
Monitoring × *OIO*	−5.570 (2.052)	−12.853 (4.974)	−1.336 (0.882)
Monitoring × *OIO* × *VT*	−0.001 (0.000)	−0.002 (0.001)	0.000 (0.000)
Controls	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes
Constant	187.447 (162.407)	451.374 (492.210)	38.361 (44.017)
R^2	0.498	0.489	0.453
Observations	101	101	101

Notes: Clustered (at village level) standard errors in parentheses. Included explanatory variables: age of the leader, income of the leader, livestock holding of the leader, clan of the leader, household size of the leader, altitude, year of establishment, group size, share of female members, average age of members, share of members who can read and write, average household size of members, share of trusting members, share of members who have non-farm income (employment and business), income heterogeneity, land heterogeneity, clan fractionalization index, value of standing timber stock, and average distance to market.

In table 2 we re-estimate the first three models, but now also include our measure of initial resource wealth, measured in 2004, so before the current crop of leaders started. This analysis is based on a much smaller subsample of FUGs for which VT data are available (101 out of 132 FUGs). Interpretation of this table is problematic due to potential endogeneity problems: resource wealth is likely affected by leader ability, and leader ability may be persistent over time (bad leadership may bequeath bad leadership). With that important caveat in mind, we highlight a number of suggestive observations. First, the main results of table 1 remain largely unaffected. Specifically, the interaction between the presence of a monitoring committee and outside income opportunities has a negative sign in the ability models. We find that the net effect of *VT* is positive, albeit not always significant, which provides weak support for prediction (3) of our theoretical model that greater resource wealth is associated with higher ability members volunteering to become leaders. Re-estimating our model with the alternative measures of *VT* does not affect our conclusions (see online appendix table A5).

Table 3 summarizes the main coefficients for the effort models. The theory predicts (prediction (2a)) a direct effect of monitoring on leader effort (i.e., not moderated by outside income opportunities). Indeed, the interaction term is not significant as can be seen from columns (1) and (2). Consistent with the prediction of the model, leader effort is positively associated with the presence of a monitoring committee. Groups with

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Table 3. Monitoring and leader effort (observational data)

	Full sample of leaders (1)	Full sample of leaders (2)	Less pro-socially motivated (3)	More pro-socially motivated (4)
Monitoring	15.991	18.714	28.427	−46.019
	(6.999)	(7.273)	(15.521)	(109.672)
OIO	−7.962	1.152	22.477	−35.995
	(24.427)	(20.035)	(30.937)	(166.939)
Monitoring × *OIO*	18.974	3.013		
	(17.104)	(19.539)		
Controls	No	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Constant	40.541	6, 465.799	7, 110.735	−42, 912.275
	(3.158)	(1, 359.212)	(3, 157.401)	(42, 903.112)
R^2	0.329	0.647	0.797	0.944
Observations	124	98	59	38

Notes: Clustered (at village level) standard errors in parentheses. Included explanatory variables: age of the leader, education of the leader, income of the leader, livestock holding of the leader, clan of the leader, household size of the leader, altitude, year of establishment, group size, share of female members, average age of members, share of members who can read and write, average household size of members, share of trusting members, share of members who have non-farm income (employment and business), income heterogeneity, land heterogeneity, clan fractionalization index, average distance to market, average distance to asphalt road, value of standing timber stock, and number of potential crop trees per ha.

a monitoring committee have leaders who spent 18.71 more days patrolling the forest and overseeing patrolling by others. Effort is negatively associated with outside income opportunities, but this coefficient is far from significant. As was the case with the results for ability, above, the impact of monitoring appears to be driven by leaders with low prosocial motivation (columns (3)–(4)) consistent with prediction (1b) of our theoretical model. The large negative estimated coefficient, albeit insignificant, for prosocial motivated leaders is consistent with the idea that monitoring may crowd out effort of some prosocial leaders.[22]

In the online appendix we use the coefficient stability approach to demonstrate that the correlations between monitoring and leader ability (effort) are not sensitive to omitted variable bias. While this may attenuate endogeneity concerns, we realize that such concerns are not eliminated. Therefore, we now turn to the results of our experimental analysis.

6.2 Experimental evidence

In online appendix table A2, we present balance tests by treatment status on baseline characteristics. These tests demonstrate that random assignment to experimental arms created groups that are comparable in terms of observables (even if the groups are not perfectly balanced). More specifically, groups in the treatment and control groups have similar prosocial preferences of both leader and members as well as similar effort levels

[22]This is in line with the literature on crowding out effects of extrinsic incentives and external control (Bowles and Polania-Reyes, 2012).

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Table 4. Monitoring, ability and effort (experimental data)

	Ability			Effort
	Education (1)	Years of schooling (2)	Business experience (3)	Days patrolling (4)
Monitoring intervention	−0.794 (0.314)	−2.156 (0.949)	−0.196 (0.136)	−11.980 (7.228)
OIO	−1.718 (1.124)	−4.192 (3.070)	1.433 (0.490)	12.198 (30.735)
Monitoring intervention × *OIO*	1.226 (1.209)	2.443 (3.438)	−0.917 (0.541)	
Year	−1.205 (0.359)	−2.633 (0.803)		34.642 (14.737)
Monitoring intervention × *Year*	0.869 (0.406)	2.803 (0.892)		39.681 (18.024)
OIO × *Year*	3.932 (1.615)	9.272 (3.940)		
Monitoring intervention × *OIO* × *Year*	−3.421 (1.883)	−7.265 (4.459)		
Controls	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Model	DID	DID	OLS	DID
Constant	165.489 (151.401)	278.346 (385.720)	83.568 (48.000)	−4, 099.485 (6, 052.683)
R^2	0.318	0.279	0.372	0.374
Observations	211	211	107	210

Notes: Clustered (at forest user group level for DID and village level for OLS) standard errors in parentheses. Included explanatory variables: age of the leader, education of the leader, income of the leader, livestock holding of the leader, clan of the leader, household size of the leader, altitude, year of establishment, group size, share of female members, average age of members, share of members who can read and write, average household size of members, share of trusting members, share of members who have non-farm income (employment and business), income heterogeneity, land heterogeneity, clan fractionalization index, average distance to market, average distance to asphalt road, value of standing timber stock, and number of potential crop trees per ha.

(guarding the forest) of group members, outside income opportunities and pre-existing monitoring committee at baseline. Difference-in-differences estimation results for education as a proxy of leader quality are summarized in columns (1)–(2) of table 4. Our coefficient of interest is the triple interaction between the monitoring intervention, outside income opportunities and the indicator for the end line period. Both coefficients are negative and statistically significant.[23] For example, keeping OIO at 30 per cent, groups with a monitoring committee have leaders with 1.53 less years of schooling. These are economically meaningful estimates.

[23] An alternative explanation for the leader turnover in our treatment group is that the groups' periodic election of leaders coincides with the treatment period. However, this is not the case for two reasons. First, many groups do not have a regular and fixed election timeline. Second, randomization implies that the time elapsed since the leader started is not correlated with treatment status.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Column (3) is based on the OLS estimator as our data for business experience are not comparable across baseline and end-line, as mentioned above. The estimated coefficient for the interaction between monitoring and outside income opportunity is negative and significant. Again keeping OIO at 30 per cent, this implies a 27.5 percentage point reduction in business experience of leaders. This provides further support for the claim that intensified monitoring adversely affects the quality of the group leader – a negative selection effect.

Column (4) summarizes the diff-in-diff results for our effort variable, based on the interaction between monitoring and the end-line period. The estimated coefficient supports the earlier results obtained using the baseline data. The estimated coefficient in column (4) suggests that groups with a monitoring committee spend 39.68 more days patrolling the forest and overseeing patrolling by others, which is quite a large effect. As predicted by the theory, intensified monitoring increases leaders' effort.[24]

6.3 Monitoring and forest income

Taken together, results in tables 1–4 reveal a potential trade-off associated with implementing monitoring. While intensified monitoring invites more effort by the leader, as intended, it also lowers leader ability in the presence of outside income opportunities. If ability and effort of the leader are both inputs in the production of public goods, then the effect of monitoring on benefits for group members is theoretically ambiguous.[25]

We probe the net effect in table 5. Columns (1)–(2) summarize the results of the OLS model in equation (4) based on the baseline data, with and without controls respectively. Columns (3)–(4) report the coefficient of the interaction between the treatment dummy and the end-line dummy (the diff-in-diff model in equation (5) based on baseline and end-line data). The dependent variable in all columns is average forest-based income.[26] We find a positive correlation between the presence of a monitoring committee and forest income (column (2)) at baseline, and a positive causal effect of the monitoring treatment on forest income (columns (3)–(4)) at endline. Columns (2) and (4) estimate that members of FUGs with monitoring institutions have, respectively, a forest-based income that is 1463 Ethiopian Birr (ETB) (or 0.3 standard deviations) and 1311 ETB (0.17 standard deviations) higher than members from other groups.

As mentioned, the job description of group leaders involves both brains and brawn. Leaders should keep records, organize group activities, and negotiate. But the great majority of their time goes into physically-demanding activities such as contributing to

[24]In online appendix table A6 we report ANCOVA results and these are largely consistent with the diff-in-diff estimates. In online appendix table A7, we report regression results of models with the three monitoring interventions entering separately. Again, signs of the estimated coefficients are largely consistent with the results presented above, but statistical power is obviously lower. The coefficient for top-down monitoring with punishment is significant only for the business experience model while the coefficients for top-down monitoring with reward and bottom-up monitoring are statistically significant for the education models. Furthermore, both top-down monitoring treatments significantly increase leaders' effort.

[25]Kahsay and Bulte (2021) explore how monitoring affects additional outcomes of interest, distinguishing between the impact of top-down and bottom-up monitoring regimes introduced in separate experimental arms. They demonstrate that top-down monitoring not only raises forest income and reduces within-group inequality, it also promotes forest conservation. Outcomes for bottom-up monitoring have the same sign, but are not statistically significant.

[26]We deflated both the baseline and end-line forest income by the corresponding consumer price index (CPI) in Ethiopia in order to facilitate comparison across both periods.

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

Table 5. Monitoring and forest income

	(1)	(2)	(3)	(4)
Monitoring	119.703 (765.000)	1,462.477 (792.085)	1,238.535 (573.585)	1,311.112 (670.439)
Controls	No	Yes	No	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Model	OLS	OLS	DID	DID
Constant	1,591.331 (344.250)	999,810.623 (303,228.327)	1,092.029 (216.983)	892,900.118 (222,980.367)
R^2	0.339	0.637	0.035	0.240
Observations	132	99	264	214

Notes: Clustered (at village level for OLS and at forest user group level for DID) standard errors in parentheses. Included explanatory variables in column (1): age of the leader, income of the leader, livestock holding of the leader, clan of the leader, household size of the leader, altitude, year of establishment, group size, share of female members, average age of members, share of members who can read and write, average household size of members, share of trusting members, share of members who have non-farm income (employment and business), income heterogeneity, land heterogeneity, clan fractionalization index, average distance to market, average distance to asphalt road, value of standing timber stock, and number of potential crop trees per ha. Column (2) regression does not include income of the leader, livestock holding of the leader, clan of the leader and household size of the leader, since these variables were not collected from some new leaders.

patrolling the forest. For these activities, education and business experience are arguably less relevant. An obvious implication is that our result, that the net effect of the trade-off between increased leader effort and reduced quality is 'positive', may not extend to other contexts with a more 'brainy' job description. It is also an open question whether the reduced form effect of monitoring remains positive in the long run. It is possible that quality-reducing impacts of monitoring will become manifest more gradually over time. More research over prolonged time horizons is necessary to address these issues.

7. Conclusion and discussion

A growing literature investigates the impact of incentives and monitoring on ability and effort of public sector employees. This literature points to inherent trade-offs. Financial incentives to attract high-ability individuals may decrease effort by attracting individuals with lower public service motivation. Related to this, we find that monitoring institutions that aim to increase effort by the leader may do so at the expense of lowering leader ability – an argument based on opportunity costs of high quality leaders. In this paper, we use unique micro-level data among forest user group institutions in Ethiopia and investigate the effect of monitoring on two dimensions of leadership. These findings also speak to important concerns about governance and the quality of leadership in the context of common property resource management.

Our analysis is based on cross-section variation in observational data, and on exogenous variation created in an RCT. Together these analyses tell a consistent and coherent story. More intensive monitoring increases leaders' effort, and on average also lowers the ability of leaders, but only if leaders have low prosocial preferences and if there are other activities for leaders to undertake – groups with monitoring institutions are led by individuals with less education and less experience in business in the presence of other income opportunities. High-quality leaders behaving corruptly are dismissed and

https://doi.org/10.1017/S1355770X23000165 Published online by Cambridge University Press

replaced with new volunteers. However, individuals with low prosocial preferences now expect that the returns to becoming a leader are lower than before, and may therefore choose not to run. As a result of some candidates pulling out, the quality of the remaining top candidate (weakly) goes down. Instead, individuals with high prosocial motivation are not discouraged from 'running for office' by additional scrutiny.

One important remark concerning the finding that monitoring increases leader effort is as follows: our reduced form impact estimates cannot distinguish between two competing explanations. Our theoretical model predicts a direct effect of monitoring on effort because non-complying leaders incur a cost. However, there may also be an indirect effect. Monitoring will cause (self-) selection of leaders with pro-social preferences in office – leaders supplying greater levels of effort, conditional on monitoring intensity. Future work could usefully probe the mechanism linking monitoring to effort more closely, preferably in a larger sample.

Given the opposing effects of monitoring on ability and effort, the net effect on welfare is ambiguous. In our case, we find that the overall effect of monitoring is positive, at least in the short run. The discipline effect of monitoring appears to dominate the selection effect, suggesting it is reasonable to recommend the establishment of monitoring committees to groups of users organized in the joint management of valuable natural resources. It is an open question to what extent this insight spills over to alternative settings, where leader quality is more important than the time allocated to physically-demanding activities. It is also an open question whether this result remains valid in the long run – perhaps the effect of increased effort matters immediately while the effects of reduced leader quality materialize with a delay. We hope this paper inspires future work on the topic from other contexts, including ones with remunerated leadership positions, more 'brainy' job descriptions, and longer time horizons.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1355770X23000165.

Acknowledgements. We thank two anonymous reviewers and seminar participants in Bergen (NHH), Copenhagen, Lund, Essen, Geneva, Leuven and Wageningen for helpful comments and suggestions. We also thank Oromia Forest and Wildlife Enterprise and Environment and Climate Research Center (ECRC) at the Ethiopian Development and Research Institute (EDRI) for their kind support. We thank the Danish Research Council for financial support. This project was supported by the Danish Council for Independent Research – social sciences – under grant DFF 6109-00296. We also thank Rufford Small Grants for their financial support to implement the project. AEARCT identification number: AEARCTR-0006582.

Competing interest. The authors declare none.

References

Altonji J, Elder T and Taber C (2005) Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. *Journal of Political Economy* **113**, 151–184.

Ashraf N, Bandiera O and Lee SS (2015) Do-gooders and go-getters: career incentives, selection, and performance in public service delivery. *Harvard Business School*, Working Paper, 1–60.

Beekman G, Bulte E and Nillesen E (2014) Corruption, investments and contributions to public goods: experimental evidence from rural Liberia. *Journal of Public Economics* **115**, 37–47.

Björkman M and Svensson J (2009) Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics* **124**, 735–769.

Björkman NM, de Walque D and Svensson J (2017) Experimental evidence on the long-run impact of community-based monitoring. *American Economic Journal: Applied Economics* **9**, 33–69.

Bowles S and Polania-Reyes S (2012) Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature* **50**, 368–425.

Casey K, Glennerster R and Miguel E (2012) Reshaping institutions: evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics* **127**, 1755–1812.
Dal Bó E, Finan F and Rossi MA (2013) Strengthening state capabilities: the role of financial incentives in the call to public service. *The Quarterly Journal of Economics* **128**, 1169–1218.
Delfgaauw J and Dur R (2007) Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization* **62**, 605–624.
Deserranno E (2019) Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. *American Economic Journal: Applied Economics* **11**, 277–317.
Deserranno E, Stryjan M and Sulaiman M (2019) Leader selection and service delivery in community groups: experimental evidence from Uganda. *American Economic Journal: Applied Economics* **11**, 240–267.
Dessie G and Christiansson C (2008) Forest decline and its causes in the south-central rift valley of Ethiopia: human impact over a one hundred year perspective. *AMBIO: A Journal of the Human Environment* **37**, 263–271.
FAO (Food and Agriculture Organization) (2015) *Global Forest Resources Assessment 2015*. Rome, Italy: Food and Agriculture Organization of the United Nations.
Fedele A and Naticchioni P (2016) Moonlighting politicians: motivation matters!. *German Economic Review* **17**, 127–156.
Ferraz C and Finan F (2009) Motivating politicians: the impacts of monetary incentives on quality and performance. National Bureau of Economic Research No. w14906.
Francois P (2000) 'Public service motivation' as an argument for government provision. *Journal of Public Economics* **78**, 275–299.
Gagliarducci S and Nannicini T (2013) Do better paid politicians perform better? Disentangling incentives from selection. *Journal of the European Economic Association* **11**, 369–398.
Global Forest Watch (2023) Ethiopia. Available at https://www.globalforestwatch.org/dashboards/country/ETH/?category=summary&treeLossPct=eyJoaWdobGlnaHRlZCI6ZmFsc2V9
González F and Miguel E (2015) War and local collective action in Sierra Leone: a comment on use of coefficient stability approaches. *Journal of Public Economics* **128**, 30–33.
Grossman G and Hanlon WW (2014) Do better monitoring institutions increase leadership quality in community organizations? Evidence from Uganda. *American Journal of Political Science* **58**, 669–686.
Gugerty MK and Kremer M (2008) Outside funding and the dynamics of participation in community associations. *American Journal of Political Science* **52**, 585–602.
Humphreys M, de la Sierra RS and Van der Windt P (2019) Exporting democratic practices: evidence from a village governance intervention in Eastern Congo. *Journal of Development Economics* **140**, 279–301.
Hussam R, Rigol N and Roth B (2022) Targeting high-ability entrepreneurs using community information: mechanism design in the field. *American Economic Review* **112**, 861–898.
Kahsay GA and Bulte E (2019) Trust, regulation and participatory forest management: micro-level evidence on forest governance from Ethiopia. *World Development* **120**, 118–132.
Kahsay GA and Bulte E (2021) Bottom-up versus top-down monitoring in community resource management: experimental evidene from Ethiopia. *Journal of Economic Behavior & Organization* **189**, 111–131.
Kahsay GA and Medhin H (2020) Leader turnover and forest management outcomes: microlevel evidence from Ethiopia. *World Development* **127**, 104765.
Kahsay GA, Bulte E, Alpizar F, Medhin H and Hansen LG (2023) Leadership accountability in community-based forest management: experimental evidence in support of governmental oversight. Wageningen University Working Paper.
Kosfeld M and Rustagi D (2015) Leader punishment and cooperation in groups: experimental field evidence from commons management in Ethiopia. *American Economic Review* **105**, 747–783.
Mansuri G and Rao V (2003) *Evaluating Community Driven Development: A Review of the Evidence*. Washington, DC: World Bank Development Research Group.
Nannicini T, Stella A, Tabellini G and Troiano U (2013) Social capital and political accountability. *American Economic Journal: Economic Policy* **5**, 222–250.
Olken BA (2007) Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy* **115**, 200–249.

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* **37**, 187–204.

Platteau JP (2004) Monitoring elite capture in community-driven development. *Development and Change* **35**, 223–246.

Poteete A and Ostrom E (2004) Heterogeneity, group size and collective action: the role of institutions in forest management. *Development and Change* **35**, 435–461.

Prendergast C (2007) The motivation and bias of bureaucrats. *American Economic Review* **97**, 180–196.

Rustagi D, Engel S and Kosfeld M (2010) Conditional cooperation and costly monitoring explain success in forest commons management. *Science* **330**, 961–965.

Spence M (1973) Job market signalling. *The Quarterly Journal of Economics* **87**, 355–374.

Velez MA, Murphy JJ and Stranlund JK (2010) Centralized and decentralized management of local common pool resources in the developing world: experimental evidence from fishing communities in Colombia. *Economic Inquiry* **48**, 254–265.

Voors M, Turley T, Bulte E, Kontoleon A and List JA (2018) Chief for a day: elite capture and management performance in a field experiment in Sierra Leone. *Management Science* **64**, 5855–5876.

Wadho W (2016) Corruption, tax evasion and the role of wage incentives with endogenous monitoring technology. *Economic Inquiry* **54**, 391–407.

Cite this article: Kahsay GA, Bulte E (2024). Monitoring and management of common property resources: empirical evidence from forest user groups in Ethiopia. *Environment and Development Economics* **29**, 154–177. https://doi.org/10.1017/S1355770X23000165