Journal of Economic Behavior and Organization 189 (2021) 111–131

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Internal versus top-down monitoring in community resource management: Experimental evidence from Ethiopia

Goytom Abraha Kahsay[a], Erwin Bulte[b,*]

[a] Department of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, DK-1958 Frederiksberg C, Denmark
[b] Development Economics Group, Wageningen University, Box 9101, 6700 HB Wageningen, the Netherlands

ARTICLE INFO

Article history:
Received 27 October 2020
Revised 3 June 2021
Accepted 15 June 2021

JEL Codes:
D02
D31
D78
O15
Q57

Keywords:
Participatory development
Internal monitoring
Top-down monitoring
Accountability
Community forestry

ABSTRACT

The evidence on the effectiveness of participatory development approaches in low-income countries is ambiguous. We randomly vary governance modalities to study elite capture in Ethiopian forest user groups and explore implications for livelihoods of group members. Top-down monitoring and punishment increases consumption and income, and decreases inequality. In contrast, internal monitoring has no effect on livelihoods. Additional heterogeneity analysis, based on observational data, reveals that while top-down monitoring works in groups where forest benefits are unimportant, internal monitoring improves economic outcomes in those groups where forest benefits are an important component of rural livelihoods. This suggests that participatory approaches work if targeted participants have strong incentives to voluntarily contribute effort.

1. Introduction

Efforts to engage local communities in development interventions and government reform programs are increasingly common in low-income countries. Community-driven development interventions have become popular, devolving priority-setting as well as financial and operational control over public good provision to local stakeholders. Increasingly, local beneficiaries are also made responsible for monitoring front-line service providers, or management of natural resources. These approaches are expected to overcome information and incentive constraints, and create a sense of ownership. An additional expected benefit is that participatory approaches may empower the poor––especially marginalized social groups such as women and youths. For that reason, participatory projects often involve efforts to build local capacity, reform institutions, and training of beneficiaries. However, the effectiveness and efficiency of participatory approaches remains controversial and ambiguous.[1]

* Corresponding author.
E-mail addresses: goytom@ifro.ku.dk (G.A. Kahsay), erwin.bulte@wur.nl (E. Bulte).

[1] For a long time, financial support was based on (optimistic) expectations rather than hard evidence (e.g., World Bank 2004; Mansuri and Rao 2003, 2012; and Casey 2018). In recent years, evidence from randomized controlled trials has come available to shed light on this question, and paints a mixed picture. Community-driven development interventions have more or less effectively delivered public goods to hard-to-reach communities, and sometimes

https://doi.org/10.1016/j.jebo.2021.06.030

We consider the interplay between two potentially complementary components of the agenda to push community participation in development: *community-based natural resource management* and *internal monitoring* of leaders or service providers by intended beneficiaries. Experimental evidence with respect to community engagement in the management of natural resources is virtually absent, and the available evidence based on observational data sends mixed signals. For example, while community-managed forests appear less prone to deforestation and degradation than centrally-managed forests (e.g., Ostrom and Nagendra 2006; Ribot et al., 2010; Lund et al. 2015), the livelihood impacts are unclear (e.g. Sikor and Nguyen 2007). The literature proposes several explanations why decentralization may fail to produce the expected benefits. In addition to free riding, (leadership) capacity constraints, and opportunity costs for villagers,[2] elite capture is consistently mentioned as an important cause. If local elites hijack the decentralization process, streams of benefits are diverted from the community to a small number of individuals––by corruption and graft, or the prioritization of goods and services preferred by elites (e.g., Bardhan 2002; Platteau and Gaspart 2003; Bardhan and Mookherjee 2006; Olken 2007; Casey et al., 2012).

Is there a role for local communities in curbing elite capture through internal monitoring? A recent experimental literature explores the issue of monitoring by local beneficiaries. While monitoring of service providers improves performance and reduces corruption in some contexts (e.g. Bjorkman and Svensson 2009; Bjorkman Nyqvist et al. 2017), it fails to generate such impacts elsewhere (e.g., Olken 2007; Banerjee et al., 2010; Raffler et al., 2019). It appears that various pre-conditions have to be satisfied for community-driven approaches to be effective. Perhaps alternative approaches to improve the governance of community-driven development projects are more effective. We probe this issue in the context of natural resource management in Ethiopia.

The main objectives of this paper are threefold: (i) to explore the effectiveness of internal monitoring relative to alternative governance modalities to improve leadership in the context of a specific participatory development intervention, (ii) to examine how these modalities affect income levels and inequality, and (iii) to analyze how the impact of governance interventions is mediated by baseline conditions, in particular dependence on the resource base for livelihoods. We use a randomized controlled trial to compare the performance of three approaches: top-down monitoring by government officials with possible punishment for under-performing leaders, top-down monitoring by government officials with rewards for the best leaders (instead of punishments for the worst ones), and internal monitoring by group members who can punish under-performing leaders. With the exception of Olken (2007), we are the first to include both internal and top-down monitoring interventions in a single RCT design.[3] We also include a treatment based on rewards for good leadership to study asymmetric responses to positive and negative incentives by group leaders. The theory of change is outlined in Section 3, below.

The specific case of participatory development that we consider is decentralization of forest user rights in Oromia region, Ethiopia. Forest User Groups (FUGs) are an important component of the livelihoods of many Ethiopians and millions of people elsewhere in low-income countries. They are responsible for managing 'forest blocks' for members' benefits (see below) and offer a unique opportunity to study community-based resource management. Observational and anecdotal evidence about capture of forest benefits by group leaders is widespread (Lund and Saito-Jensen 2013; Fritzen 2007). Elite capture causes an unequal distribution of benefits and reduces forest-based income of non-elite group members. To the best of our knowledge this is the first randomized controlled trial in the domain of participatory natural resource management.[4]

In our internal monitoring arm, we help groups to delegate an important part of the monitoring to specific individuals to attenuate collective action problems. We also provide groups with information and material inputs, and "empower" them by lowering the costs of replacing the group leader as a credible sanction. Forest benefits are rather important for group members, on average accounting for 20–25% of their income, so we expect group members to be motivated to contribute their effort. In other words, we consider a context where internal monitoring has a good chance to succeed. However, this is not what we find. While top-down monitoring and punishment increases average forest income and promotes equal sharing of forest benefits, internal monitoring on average fails to generate such outcomes. The same is true for rewarding good leadership. As argued by Casey (2018, p.148), this suggests that 'some functions, like coercive power to limit corruption, are better left to more central tiers of administration. However, while our analysis is based on a census of all forest user groups in the study area, we find that some of our regression models have lower ex post power than predicted. This means we cannot exclude the risk of a type II error for certain null results.

We also document an interesting asymmetry using additional heterogeneity analysis, based on observational data (so that identification of causal effects rests on additional assumptions). Internal monitoring *is* effective in the sub-sample of groups

have contributed to local economic development. However, by-and-large they failed to transform local institutions or empower the poor (see Casey 2018 for an overview). Refer to Beath et al. (2013) for evidence from Afghanistan, Casey et al. (2014) for evidence from Sierra Leone, Fearon et al. (2015) for evidence from Liberia, Heß et al. (2018a) and Heß et al., 2018b for evidence from the Gambia, and Humphreys et al. (2019) for evidence from the DRC.

[2] Refer to studies by Platteau (2004), Olken (2007), Grossman and Hanlon (2014), and Kosfeld and Rustagi (2015).

[3] Compared to Olken (2007), however, we analyze a case study where monitoring has a real "bite" and directly affects the risk of leader replacement. In Olken (2007), leaders who were caught misbehaving ran the risk of losing their re-election bid in the (possibly far) future. In our case, punishment is more direct and immediate, so potential impacts are expected to be greater.

[4] Despite a growing interest in using RCTs in the context of environmental and resource management, the application of the method has been limited. In the domain of tropical forest conservation, Jayachandran et al. (2017) study payments for ecosystem services to landowners in Uganda; Wilebore et al. (2019) study unconditional community transfers and deforestation in Sierra Leone; and Heß et al. (2018b) and Heß et al., 2018b study community-driven development and deforestation in Gambia. Pynegar et al. (2018) study payments for ecosystem services in the context of water delivery in Bolivia.

where forest benefits are important for the livelihoods of members––who at baseline exerted high levels of monitoring effort. The internal intervention leveraged existing monitoring effort, which is an important qualification of the earlier null result and helps to align the conflicting evidence on the effectiveness of community participation in the literature. Internal interventions work when people rely on forest resources – either because they developed a better understanding of the relevant monitoring technology or simply because they 'care enough.' In contrast, top-down monitoring and punishment was particularly effective in forest user groups that, at baseline, depended less on forest resources for their livelihoods and exerted less monitoring effort. Top-down and internal monitoring effort therefore appear to be substitutes, with aggregate monitoring effort subject to diminishing returns to scale.

This paper extends earlier work based on data collected among the same forest user groups. These earlier papers emphasize both the importance and malleability of local governance, and suggest that lack of accountability and transparency are key concerns at the group level. Kahsay and Bulte (2019) explore the association between trust within forest user groups, and formalisation of rules (speaking to the bi-directional relationship between culture and institutions). Groups with lower in-group trust are more likely to formalize forest extraction and monitoring rules. However, elites typically play a role in such formal monitoring committees as well, so their impact is limited. This cross-section study is based on the baseline data introduced below. Using the same baseline data, Kahsay and Medhin (2020) look into how past leader turnover affects group outcomes. Leader turnover increases quality of leaders, members' participation, forest stock and benefits, and decreases forest benefit inequality. These results suggest that if under-performing leaders can be replaced, then the result may be improved forest use and economic performance. Kahsay and Bulte (2020) use theory and these same observational baseline data as well as RCT data to look at the association between formalisation of monitoring and self-selection of leaders. It focuses on the impact of formalized monitoring on leader quality (proxied by education and experience) and leader effort. While monitoring disciplines leaders, it also crowds out high ability individuals from assuming leadership positions, especially when the opportunity cost of the leadership position is high. But, the discipline effect outweighs the self-selection effect, suggesting the importance of monitoring interventions in our context. The current paper looks at the first-order impacts of the monitoring interventions. It extends the analysis by looking at the impact of monitoring on (forest-based) income and consumption levels for group members and distributional outcomes. Moreover, it unpacks the monitoring treatment by distinguishing between three alternative monitoring modalities.

The rest of this paper is organized as follows. In Section 2 we discuss the literature on elite capture and inequality. In Section 3 we introduce the case study and outline the experiment including its main treatments. In Section 4 we present our experimental design. Section 5 introduces the data and outlines the identification strategy. Section 6 presents the results. The conclusion ensues.

2. Elite capture and participatory forest management

Participatory forest management (PFM) is a prominent example of the recent trend of devolving responsibilities to local communities. During the past decades, local communities in many low- and middle-income countries were granted user-rights to state-owned forest resources. About a third of the forests in low-income countries are now managed by local communities (Blackman et al., 2017). In return for respecting certain rules, local communities were made the custodians and managers of forest resources––allowed to extract timber and non-timber forest products for own consumption or sale. The idea was that establishing user-rights for local communities would attenuate principal-agent problems or incentive problems, and create a sense of "ownership". Communities were expected to monitor extraction and enforce sustainable management rules because it would be in their own interest to do so––increasing future flows of resource rents for within-group sharing. PFM therefore was expected to not only promote forest conservation, but also contribute to improvements of the livelihoods of rural poor.

The role of communities in resource management has until now not been analyzed with aid of randomized trials. The available evidence based on alternative approaches suggests mixed experiences. While PFM has contributed to forest conservation, or slowed down forest degradation in many different contexts, it is not evident that the livelihoods of forest users have improved as a result. For instance, Jumbe and Angelsen (2006) and Ameha et al. (2014) study forest user groups in Malawi and Ethiopia respectively, and find both positive and negative impacts of PFM on forest income––depending on the location of the PFM program within each country. Persha and Meshack (2016) find no impact of devolution of forest rights on forest income, total income or assets in Tanzania. A recent review of the impact of devolution forest rights on poverty finds mixed impacts (Miller et al., 2021). Observational data from several countries suggests the unequal distribution of forest benefits and associated failure to lift the poorest group members out of poverty can often be explained by elite capture (Persha and Andersson 2014; Adhikari et al., 2004; Agrawal and Gupta 2005; García-López 2019; Sikor and Nguyen 2007).

Elite capture in community-driven development projects can take different forms, including rent grabbing by leaders and the "re-designing" of projects to serve the interests of specific community members (e.g., Bardhan 2002; Bardhan and Mookherjee 2006; Iversen et al., 2006; Fritzen 2007; Gugerty and Kremer 2008; Nagendra and Gopal 2011). The PFM literature describes many cases with clear "winners" and "losers" after the devolution of forest user-rights (Agrawal and Gupta 2005; Sikor and Nguyen 2007; Larson 2011; Ameha et al., 2014). Sometimes traditional users are excluded from accessing the forest altogether (Larson 2011; Ameha et al., 2014) and in some cases the poor are particularly excluded from participation (Agrawal and Gupta 2005). In other 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 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.

In sum, local elites tend to dominate the participatory forest management process. Current resource rents, and the promise of more rents in the future, may attract such elites (e.g., Gugerty and Kremer 2008, in the context of PFM in rural Ethiopia, see Gelo et al., 2013; Tesfay et al., 2015). One major lesson from earlier studies is that the local institutional capacity of FUGs needs to be strengthened (Lund and Saito-Jensen 2013; Persha and Andersson 2014). Can internal monitoring institutions effectively curb rent extraction? The evidence so far is mixed and suggests that, to be effective, interventions that aim to mobilizing community monitoring to discipline leader (or service providers) should satisfy four pre-conditions. First, communities should overcome collection action problems associated with monitoring (Barr et al., 2012). Second, beneficiaries should have access to information about relative performance, and about their own entitlements (Bjorkman Nyqvist et al. 2017; Banerjee et al., 2018). Third, local beneficiaries should be sufficiently empowered to hold local elites accountable, and have access to low-cost mechanisms to report or sanction under-performance without fearing retaliation in the future (Fiala and Premand 2018). Fourth, the issue at stake should be sufficiently important for local beneficiaries to motivate them to invest their time in it. It is an open question to what extent these conditions are satisfied in communities managing natural resources, including the ones studied in this paper. The interventions we introduce aim to help groups to meet the conditions (see below).

3. Context and theory of change

Our study focuses on the Adaba-Dodola PFM program, which was started in the mid-1990s in the Adaba and Dodola districts of West Arsi zone (Oromia regional state), located about 320 km south west of Addis Ababa. Under this program, so-called Forest User Groups (FUGs) were created. These groups of at most 30 members were given exclusive user-rights to blocks of on average 12 hectares per member (so 360 hectares in total). Membership in the groups is voluntary.[5]

The “typical FUG” has a General Assembly, consisting of all members, an Executive Committee of five members (chairperson, vice chairperson, secretary, cashier, and one member), and a few Technical Committees. FUG leadership is responsible for the equitable distribution of benefits and responsibilities among members, and should ensure that group rules are respected.[6] We provide additional information about the Adaba-Dodola PFM program in Appendix A.

Consistent with the evidence from other countries, the PFM program was fairly successful in terms of reducing deforestation and promoting sustainable resource use. However, the livelihood impacts are less favorable (Ameha et al., 2014) caused by lack of accountability in financial flows, and unfair distribution of benefits and responsibilities (Tesfay et al., 2015). Our own preparatory field visits to major PFM programs within Oromia regional state supported these concerns. We found evidence of group leaders colluding with local brokers to illegally sell forest products. Consistent with insights from the broader literature, we also learned that many group leaders (and key committee members) are representatives of the local elite. They have higher income and consumption levels, are more educated and wealthy, and often have a role as religious or clan leader (see Appendix Table A1). This raised concerns among OFWE officials about sustainability of the program. As a result, OFWE proposed to introduce a monitoring mechanism to increase the quality of FUG governance.

Several questions remained. Should the intervention focus on establishing and empowering internal monitoring committees, on external (i.e. top-down) monitoring by experts, or some sort of joint monitoring? Should the intervention aim to punish transgressions, or reward good behavior? The institutional landscape in Oromia regional state provided some suggestions about which approaches might work, all of which are consistent with theory. One FUG in Delo-Mena district (Bale zone) instituted an internal monitoring committee that successfully monitored group leaders. Leaders of FUGs in Jelo-Muktar (Hararge zone) appeared to respond to “special recognition” for performance by the Hararge OFWE branch office. We also observed variation in the intensity of top-down monitoring, and inspection of forest blocks by district OFWE branch offices. Especially in areas where NGO-funded projects have been phased out, inspection is infrequent.

Our theory of change is simple, based on the expectation that leaders respond to incentives. They self-select into and out of positions of authority, depending on relative rewards (Kahsay and Bulte 2020). They also choose how much effort to expend towards providing public goods (organizing the monitoring of the group’s forest block, enforcing group rules), and how to distribute the gains from forest product extraction. That is, monitoring affects the quality of leaders as well as the leader’s behavior (conditional on type), including the propensity to engage in the extraction of rents for private gain—corrupt behavior. Monitoring to punish bad leadership lowers the expected value of engaging in rent extraction and shirking, and should improve leader (and, hence) group performance. It is an empirical question whether top-down or internal monitor-

[5] Households residing inside the designated forest area were offered resettlement if they declined membership. Virtually none of the eligible households declined membership as they relied on the forest for subsistence and livestock grazing (and perhaps because of large transaction cost associated with resettlement) (Ameha et al. 2014). After formation, groups can only admit new members after existing members have passed away, or have permanently migrated out of the region.

[6] In an effort to keep leadership in check, about 30% of the groups have a so-called monitoring committee, charged with the responsibility to monitor the leader, and some 60% of the groups have an auditing committee, charged with the responsibility of “checking the books”. However, these committees typically fail to curtail elite capture as elite members frequently feature on these committees themselves and because of collective action problems associated with monitoring. Many groups heard nothing from their monitoring committee for years.

ing is more efficient. This depends on information asymmetries and the possibility to bribe monitors, but also on group empowerment, the group's ability to overcome social dilemmas, and incentives for voluntary monitoring by members.

Monitoring to reward good leadership should also improve leader performance by raising the expected value of effort provision and compliance (and possibly by incentivizing high-quality leaders to enter). Both material rewards and non-material rewards (signals of high performance) may induce the provision of public goods (e.g., Ashraf et al., 2014, Shikuku et al., 2018). We try to leverage both dimensions by combining the provision of a material gift and public recognition (see below), recognizing that these dimensions may interact (e.g., it has been argued that material rewards may crowd out intrinsic motivation or the signaling motive to engage in prosocial behavior). The effectiveness of positive rewards versus negative rewards is another empirical question.

4. Experimental design

We designed and implemented a randomized control trial to explore the impact of alternative governance modalities to curb elite capture, raise (forest) income levels for members, and promote an equitable distribution of forest benefits. The experiment includes all 132 FUGs in the Adaba-Dodola PFM program. FUGs were randomly assigned to one of four experimental arms: (i) control, (ii) internal monitoring, (iii) top-down monitoring and punishment, and (iv) top-down monitoring and leader reward.[7] All interventions were implemented by OFWE. Before randomization, we asked FUGs for their consent to participate. In Appendix Table A2, we provide the number of groups per experimental arm and numbers of group members. Details of the treatments are as follows.

4.1. Internal monitoring treatment

Representatives of selected FUGs were invited to a start-up meeting in which OFWE representatives introduced the program. In addition to group representatives, we also asked PFM experts as well as village and district officials to participate in the meeting. All invited FUGs participated in the meeting. Group representatives were asked to, after returning to their group, arrange a meeting and introduce the program to group members. Groups should set-up a new internal monitoring committee, consisting of three members and responsible for monitoring forest use and management. Members of this new monitoring committee should be elected through anonymous majority voting by all group members.

One month into the program, OFWE experts visited each FUG to verify that group members knew about the program and had, indeed, set up an internal monitoring committee. These experts further clarified the program, if necessary, and facilitated election of committee members for groups that had failed to do so already.[8] Next, members of the committee received training on how to monitor group leadership and performance, and forest management. As information collection strategies, OFWE recommended inspection of forest blocks, books and records, as well as organizing group discussions and holding interviews with anonymous group members and non-members. Key indicators of performance, as also used by the OFWE in its own evaluations, were shared and explained, and lists with these indicators were offered to committee representatives (a "scorecard"). In addition, committee members received stationery to aid report preparation.

The monitoring committee was expected to share the information with other group members during General Assembly meetings held approximately 6 months into the intervention. During these meetings, OFWE staff facilitated discussions and helped to "benchmark" group-specific outcomes – putting them in context by sharing outcomes of other groups. At least two-thirds of the group members had to be present during these meetings. Group members were invited to reach a decision with respect to FUG leadership, for which anonymous voting sessions were organized to minimize concerns about retaliation by frustrated leaders. Majority voting sessions led to one out of three possible outcomes: (i) verbal appreciation for leadership, (ii) group-level warning for the leader, or (iii) removal of the leader followed by election of a new one by anonymous majority voting.[9]

In sum, the treatment was a multi-pronged intervention involving assistance with selecting committee members, a training component on how and what to monitor, provision of inputs and information (including data about comparative performance) and an institutional innovation aimed at lowering the costs of holding leadership accountable. These elements

[7] While we did not stratify at the kebele level, the four experimental arms are fairly evenly distributed across the kebeles: control and top-down monitoring treatments are found in 11 kebeles and the reward and internal monitoring treatments are found in 12 kebeles.

[8] Newly-elected internal monitoring committees are different from any pre-existing monitoring committee present in some FUGs. First, the new committee is responsible for monitoring all matters related to forest use and management (as opposed to only scrutinizing the leader's behavior). This includes checking on the state of the forest, the behavior of group leaders and members, and overseeing group rules and enforcement, books and records, and elections in the group. Second, the new monitoring committee must present a report in the general assembly meeting (see below) which served as a basis for further discussions and decisions. There is no specific reporting requirement for existing monitoring committees, which not surprisingly implied that very little happened in many FUGs. Third, the new monitoring committee was given material inputs (e.g. stationary material) and monitoring guidelines to assist it in conducting its activities efficiently. More specifically, the monitoring guideline includes 42 indicators organized in 10 categories (forest condition, leadership committee, chairperson, monitoring committee, members, elections, books and records, rules, participation, punishments and sanctions). Almost all groups in the internal monitoring committee arm elected different members for the new committees than had been selected for the old ones. The selection rules for the pre-existing committees were opaque.

[9] If there was no majority for one of these three outcomes, the vote was repeated for the two most popular outcomes. This happened in 6% of the groups.

aim to make monitoring effort by group members more productive, and to address bottlenecks with respect to social action, information and empowerment in a low-cost, scalable manner. While the emphasis is on internal monitoring by group members it is evident that this is still a fairly heavy-handed intervention involving considerable top-down involvement by OFWE staff.

4.2. Top-down monitoring and punishment treatment

FUG representatives were invited to participate in a start-up meeting with OFWE experts and village and district officials where OFWE introduced the program. Participants were asked to, after returning to their group, arrange a meeting and introduce the program to group members. The main difference with the internal monitoring intervention was that OFWE experts were now charged with the responsibility for collecting data. OFWE focused on the same variables and indicators. In addition to collecting information on generic performance indicators, OFWE officials also verified whether leadership acted in accordance with the group's own bylaw. Examples of the former are checks whether the leadership keeps records of financial activities, or is able to present receipts. An example of the latter is to check whether punishments are according to the group's bylaw.

As in the internal monitoring, one month into the program, OFWE experts visited each FUG to verify whether group members knew and understood the program. Approximately 6 months later, OFWE presented a summary of its findings at a General Assembly meeting, facilitating group discussions to gain a better understanding of what happened and comparing group-level outcomes to those of other groups. Next, findings from the audit report, augmented with information collected during the General Assembly, were reported to OFWE management for decision-making. OFWE management subsequently chose one out of three possible outcomes: (i) verbal appreciation for the leader, (ii) a warning, or (iii) removal of the leader from its position. In the latter case, FUGs were instructed to elect a new leader by anonymous majority voting. All FUGs complied with the management decision.

4.3. Top-down monitoring and reward treatment

Like the top-down monitoring and punishment treatment, the reward treatment is based on scrutiny by OFWE officials. But instead of punishing (firing) under-performing leaders, this treatment is based on rewarding the best performing ones. OFWE experts collected data on forest use and management, and double-checked these findings during a General Assembly meeting. The experts reported their findings to OFWE management who subsequently selected the top-three performing leaders. A ceremony attended by leadership of all FUGs assigned to this treatment arm was organized to celebrate the winners (notably the chairperson from the winning groups), comprising the non-material reward.

There was also a material reward, as top-three performers also received a solar panel for private use with a market value of 60 USD (197 USD in PPP terms). Performance was based on the same criteria as applied in the top-down monitoring intervention, and supposedly in the internal monitoring intervention as well. Group leadership was informed about these criteria beforehand. While OFWE did not facilitate a General Assembly discussion about rewarding or punishing leaders by ordinary members, group members were of course free to independently organize their own general assembly meetings to follow up on any lessons learned. This will be explored below.

4.4. Timing

Baseline data was collected in March and April 2017, followed by the start-up meeting in May 2017. The first round of monitoring and follow-up explanation took place in July 2017, and additional information collection (by group members or OFWE officials) occurred during the rest of 2017. General Assembly meetings were held in December 2017-January 2018. End-line data for the analyses below were collected in the period September-December 2018, or some 16 months after starting the interventions.

The program is ongoing; internal committees are operational in the internal monitoring arm and OFWE scheduled another round of auditing in the top-down arm. However, an additional round of reward provision is not foreseen, even if FUGs in the reward arm were not explicitly informed about this.[10] If group leaders anticipate that there will be no future rewards, their incentive to behave in an inclusive and transparent manner is attenuated. Any impacts in this treatment arm during the end-line must then be due to group leaders "locked-in" in certain modes of behavior—for example because group members have updated their beliefs and expectations about flows of forest-resource benefits. Our reward treatment provides a *lower bound* of what a properly-timed incentive for good governance can achieve. However, we believe outcomes are of interest, if only because they involve top-down scrutiny (as in the top-down monitoring and punishment arm) without the threat of punishment.

[10] Importantly, we did not deceive the leaders and did nothing to make them believe there would be multiple rounds of rewards.

5. Data and identification

We use data from two sources: administrative data from OFWE and two waves of survey data. Administrative data includes pre-intervention estimates of the forest stock (measured as the number of potential crop trees (PCTs) and mature trees (MTs) per ha), group size and composition, year of establishment, and size and location of forest blocks.[11] Survey data were collected by interviewing FUG leaders and randomly selected members. Overall, 1222 members were interviewed at baseline, of which 1215 members were interviewed during the follow-up survey. The attrition rate is less than 1% and will be disregarded in what follows.[12] The survey data includes socio-demographic characteristics, welfare indicators, social capital and network indicators, and measures of forest use and management. We also collected data on punishment strategies used by FUG leadership in case of transgressions by members; grazing and encroachment, excessive extraction of forest resources, failure to contribute to patrolling. The most common punishments are fines, confiscation of resources, and reduced sharing in group benefits––used in equal proportions across treatment arms.

As welfare indicators, we measured income and consumption levels following the method used for the Ethiopia Rural Household Survey (ERHS). We presented members with a list of potential income sources (e.g. sale of agricultural products, non-farm income, remittances, forest income, etc.) and asked whether and, if so, how much, income they obtained per source. We also presented a list of consumption items (food and non-food items), and asked subjects to indicate how much of each consumption item they consumed during the past 12 months (own production or purchase). Through aggregation we obtain measures of total annual income and consumption per member. Forest benefits were measured by presenting FUG members with a list of commonly-extracted forest products (firewood, poles, tree bark, lianas and vines, bamboo, tree branches, logs, saw logs) and non-timber forest products such as honey, thatch grass, and wild vegetables. We asked about quantities extracted for each product (in local measures) and the purpose of extraction (sale or consumption). After generating a per-unit market price at the village level for each product type, we calculated a household-level measure of forest income. By aggregating across members, we generate measures of (average) consumption, income and forest benefits at the FUG-level. In a next and final step, we also use these household outcomes to compute a Gini index score at the group level for consumption, total income, and forest (extraction) benefits.

We used five-point Likert scales to construct indices for the subjective appreciation of leadership, participation and influence, and forest quality. For example, a measure of satisfaction with group leadership is obtained by computing the average of five questions about the appreciation of leader behavior. With respect to "participation and influence" we distinguish between own participation[13] and a subjective measure of other members' participation in the FUG: attending meetings, respecting rules, spending time in activities, engaging in discussions, and solidarity towards the group. Finally, we asked members about their perception of the forest condition, paying special attention to tree regeneration, charcoal burning sites, newly felled trees, encroachment and agricultural expansion. The questions on satisfaction and participation are in Appendix Table A3.

Appendix Table A4 summarizes our main dependent and explanatory variables, averaging baseline and end-line data. FUG-level variables are constructed based on the mean of individual-level variables. Table A5 demonstrates that random assignment to experimental arms achieved "balance" in terms of our explanatory variables at baseline (including the incidence of pre-existing auditing or monitoring committees) while there are few significant differences in the outcome variables.

To measure the treatment effect of our interventions we estimate difference-in-differences (DID) models. Our main dependent variables Y_{it} are group-level welfare indicators (average income, consumption, forest benefits) and within-FUG inequality measures, for group i at time t. Our main unit of analysis is the FUG which was also the unit of randomization, and which is the only unit of analysis at which our inequality measures make sense. However, we also explore the robustness of our findings by considering household-level welfare indicators. To facilitate comparison over time we deflate end-line data where appropriate using the Consumer Price Index (CPI) in Ethiopia. As explanatory variables we use a vector of time-varying and time-invariant FUG controls $\boldsymbol{X}_{\mathrm{i}}$, a vector of treatment dummies $\boldsymbol{T}_{\mathrm{j}}$ for treatment j, an end-line dummy (*Post*), and a vector of interaction terms ($\boldsymbol{T}_{\mathbf{j}}$x*Post*). We also include *kebele* fixed effects to capture any village-level factors, α_v, perhaps due to differences in geophysical variables or *kebele* governance. In sum, our estimation model reads as:

$$Y_{it} = \beta_0 + \beta_{1j}\sum_{j=1}^{3}\boldsymbol{T_j} + \beta_2 Post + \beta_{3j}\sum_{j=1}^{3}\boldsymbol{T_j} * \boldsymbol{Post} + \beta_4 \boldsymbol{X_{it}} + \alpha_v + \varepsilon_{it} \quad (1)$$

The term ε_{it} captures unobserved time-varying factors that induce heterogeneity in welfare outcomes. Our coefficients of interest are the β_3's. The coefficients β_1 capture potential pre-treatment differences in our outcome variables between ex-

[11] Forest stock data were collected before the start of the experiment, in 2004 for 117 FUGs and 2012 for 49 FUGs. OFWE is currently collecting a new wave of forest inventory data (counts of potential crop trees and mature trees as well as number of charcoal burning sites, trees planted, and trees regenerated), but this data collection effort is not yet completed.

[12] We have also estimated Lee bounds and obtain similar findings as the ones reported in section 6, below (not shown but available on request).

[13] This index captures the average response to questions about whether or not participants are free to participate in decision making; able to comment on performance of leaders; actively engage and make suggestions in meetings, and so on.

Table 1
Governance interventions and outcomes.

	Level outcomes			**Inequality outcomes**		
	Consumption (1)	*Income* (2)	*Forest benefit* (3)	*Consumption* (4)	*Income* (5)	*Forest benefit* (6)
Top-down	3127.713	3721.669	260.276	0.004	0.003	0.056
	(2002.074)	(2726.971)	(632.127)	(0.021)	(0.028)	(0.036)
Reward	1773.896	4857.783*	111.436	−0.009	−0.050*	−0.017
	(2122.174)	(2911.712)	(613.124)	(0.020)	(0.028)	(0.029)
Internal	2821.789	5558.448*	36.958	0.007	−0.020	−0.014
	(2103.919)	(2916.755)	(550.668)	(0.027)	(0.029)	(0.032)
Period	3404.904**	7133.237***	608.833	−0.076***	−0.128***	−0.095***
	(1721.104)	(2255.487)	(417.881)	(0.017)	(0.026)	(0.025)
Top-down x Period	6721.364**	9677.234**	5320.505***	−0.153***	−0.129***	−0.136***
	(3020.529)	(4109.497)	(1988.671)	(0.026)	(0.042)	(0.051)
Reward x Period	−1213.092	−6018.961	605.385	−0.005	0.013	−0.027
	(2700.023)	(3892.589)	(923.587)	(0.028)	(0.038)	(0.040)
Internal x Period	23.270	−3600.826	375.109	−0.044	0.025	0.025
	(2700.613)	(3838.917)	(702.071)	(0.032)	(0.038)	(0.047)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Model	DID	DID	DID	DID	DID	DID
Constant	1,039,960.87	1,100,638.31	936,964.7***	−1.465	18.039*	−19.802
	(665,455.785)	(1,309,871.628)	(226,087.349)	(8.620)	(10.216)	(12.550)
R^2	0.498	0.351	0.314	0.487	0.458	0.282
Mean of control-group 2018	22,875.955	17,391.792	3376.02	0.265	0.374	0.319
Observations	264	264	264	264	264	264

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include altitude, group size, year of establishment, share of female members, share of literate members, average age of members, and average land holding. Initial forest stock is not included as a control variable to increase the number of observations. However, our results do not change when we control for it (details available on request). The estimated coefficients for the treatments are significantly different from each other for each outcome variable except between internal monitoring and reward treatments (for income) (Wald test).

perimental arms, and β_2 captures effects due to general time-varying factors – affecting outcomes across all experimental arms in the same way.[14]

6. Results

The main results are summarized in Table 1. Our diff-in-diff estimates provide the causal effect of alternative governance interventions on our outcome variables: the level and distribution of consumption, forest benefits and overall income. Columns (1)-(3) present estimation results for level outcomes and columns (4)-(6) present results for inequality outcomes. In all the tables, *Top-down* refers to the top-down monitoring and punishment treatment, *Reward* refers to the top-down monitoring and reward treatment, and *Internal* refers to the (participatory) internal monitoring treatment by group members.

Multiple insights can be obtained from Table 1. The results in the top three rows confirm that randomization worked rather well: there are few significant correlations between our outcome measures at baseline and treatment assignment (the β_1's in Eq. (1)).[15] Next, the post-intervention dummy (β_2) enters significantly with a positive sign for income and consumption outcomes and a negative sign for the inequality outcomes. Hence, during the study period, most "level outcomes" in the control group improved (assuming the CPI adequately captures local inflation dynamics), and became more equal. A simple before-after comparison that fails to control for this trend would produce biased estimates of the impact of governance interventions.

Our main result is that there are large differences in the effectiveness of our treatments, as reflected in the coefficients associated with the product of the treatment and period dummies (β_{3j}). The top-down monitoring and punishment intervention has significant effects on the level and distribution of income, consumption, and forest benefits. Top-down monitoring and punishment increases consumption and overall income by 0.20 and 0.22 standard deviations. The impact on forest benefits is even larger (0.34 standard deviations), which is not surprising given that the flow of forest benefits is

[14] Random assignment to treatment implies we need not be concerned about the parallel trends assumption that is at the basis of diff-in-diff models. Nevertheless, it would be comforting to note that, prior to the intervention, the FUGs in different experimental arms were on similar trends. While we do not have access to multiple waves of pre-intervention data for our key variables to test the parallel trends assumption underlying the diff-in-diff model, as highlighted above, we do have access to forest stock data for a subsample of our FUGs that date back to 2004 and 2012. As expected, our data indicate that trends in forest stocks were similar for all the FUGs in our sample (details in Appendix Table A6).

[15] FUGs in the internal monitoring and reward arms were better off at baseline in terms of income, which is presumably caused by the modest number of clusters we assign to treatment––while we use the universe of FUGs in our study area, the number of groups is limited to only 132. To control for pre-existing differences we use the diff-in-diff estimator, and estimate ANCOVA models in the robustness analysis below.

more directly tied to the intervention.[16] We estimate that the household gain in forest-based income is about USD 15–20 per month, which seems plausible given the value of the timber and non-timber forest resources that are extracted. For example, the market value of one 'bukra' of honey equals USD 22 and of one horse load of poles equals USD 5. The average price of a log equals USD 12. Gains of these magnitude are not surprising in light of the anecdotal evidence we heard about side-selling of large amounts of forest products by leaders to commercial traders, often at night. If additional scrutiny diverts this stream of benefits towards ordinary members rather than FUG leadership, then these effects are plausible.

Top-down monitoring and punishment also decreases inequality: income and forest benefit Gini coefficients go down by 0.18 standard deviations, and the consumption Gini coefficient by 0.25 standard deviations. The resulting distributions are very equal, suggesting that top-down monitoring (nearly) eliminates elite capture in forest user groups. These outcomes were not obtained at the expense of forest degradation as OFWE officials continuously monitored over-harvesting.

We do not document significant impacts of the internal monitoring treatment––neither for the level variables, nor for the inequality variables. The relevant coefficients are much smaller than the top-down monitoring coefficients. Similarly, the relevant coefficients of the top-down monitoring and reward treatment are small and statistically insignificant.

6.1. Robustness

With two time periods and both village fixed effects and a time dummy, our diff-in-diff estimator puts a lot of structure on the analysis. To allow for more flexibility we have also estimated a series of ANCOVA models (endline cross-section and baseline controls). These results are reported in Appendix Table A7 and are qualitatively similar as the diff-in-diff results in Table 1.

We also re-estimated the models in columns (1–3) of Table 1, using individual member-level data. While such models make no sense for the inequality results – based on group-level variables –variation in income levels can be analyzed both at the individual level and group level. Models based on aggregate data discard a substantial amount of the variation in our income level data. We now explain this variation with the treatment dummy, and group-level as well as household-level controls (and cluster standard errors at the group level). Results are reported in Appendix Table A8 and A9, for household-level models based on group-level and household-level fixed effects. Again, we obtain results that are qualitatively very similar to the ones reported in Table 1.

Next we consider the issue of spillovers. Multiple treatments occurred in all kebeles, which introduces the risk of information spillovers. Group members of control FUGs may be aware of OFWE's efforts to improve FUG governance in other groups. However, this did not emerge as an issue during end-line surveys. Specifically, only 22% of the respondents from the control group indicated that they heard about the program, and 11% indicated that this caused them to discuss governance in their group. In none of these groups, new measures based on monitoring, rewarding or punishing leaders were introduced. We also probed this issue using simple regression analysis, and re-estimated the models in Table 1 for the sub-sample of 34 control group FUGs while adding variables capturing the number of FUGs in their kebele assigned to the top-down or internal monitoring intervention. In these models we obviously did not include kebele fixed effects. 11 out of 12 new variables did not enter significantly (at 10%) (see Appendix Table A10), further attenuating concerns about spillovers.

Finally, we briefly discuss possible bias due to the fact that the analysis is based on self-reported income and consumption data. This raises the question to what extent our findings for top-down monitoring are caused by an experimenter demand effect. We are not overly concerned about demand effects because survey-based data were collected independently from OFWE, and did not enter OFWE records. Moreover, if top-down scrutiny affects attitudes towards the experimenter, then we would expect treatment effects for both the top-down monitoring and punishment regime as well as the reward regime. The finding that treatment effects only materialize for the top-down monitoring and punishment intervention, and not the top-down monitoring and reward regime, suggest that monitoring *per se* does not affect self-reporting.

6.2. Statistical power

Our study is based on a census of the FUGs in our study area. While we could not increase the sample size, concerns about statistical power remain and should be addressed. Appendix Table A11 reports the results of a statistical power analysis. We report both the results of an *ex ante* power analysis, based on baseline statistics and intra-cluster correlations, and the results of an *ex post* power analysis, based on the estimated standard errors of the main coefficients. Both analyses are based on two-sided tests, a power level of 0.8, and a significance level of 0.05. The minimum detectable effect size (MDE) of these power analyses should be compared to a meaningful threshold value. Ideally, we would base such a threshold on a comparison of the intervention's benefits to the per-group intervention cost, assuming that the government of Ethiopia is particularly interested in upscaling cost-effective interventions. However, detailed treatment costs per group are not available. Instead, we first use a threshold value equal to 25% of the value of the dependent variable of the control group, but also explore lower threshold values (20 and 10%).

[16] In absolute terms, the increase in total income is larger than the increase in forest benefits. The reason is that income from other sources likely goes up as the flow of forest benefits increases. For example, improved access to forage facilitates livestock fattening and milk production, and forest income facilitates investment in agricultural inputs, increasing agricultural productivity.

Table 2
Governance interventions and leadership.

	Leaders responsiveness		Members' participation		Forest condition
	Chairperson	*Executive committee*	*Participation and influence*	*Members behavior*	
	(1)	(2)	(3)	(4)	(5)
Top-down	0.239*	0.316**	0.288***	0.297***	0.417***
	(0.129)	(0.125)	(0.078)	(0.113)	(0.152)
Reward	0.012	0.077	0.107	0.072	0.130
	(0.139)	(0.135)	(0.083)	(0.124)	(0.157)
Internal	0.010	0.044	0.102	0.115	0.115
	(0.131)	(0.128)	(0.069)	(0.109)	(0.134)
Controls	Yes	Yes	Yes	Yes	Yes
Village FE	Yes	Yes	Yes	Yes	Yes
Model	OLS	OLS	OLS	OLS	OLS
Constant	37.247	17.678	27.399	36.454	9.576
	(70.130)	(67.289)	(48.741)	(74.953)	(73.704)
R^2	0.451	0.463	0.548	0.499	0.480
Mean of control-group 2018	3.487	3.452	3.963	3.64	3.106
Observations	132	132	132	132	132

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include altitude, group size, year of establishment, share of female members, share of literate members, average age of members, and average land holding. The estimated coefficients for the treatments are significantly different from each other for each outcome variable except between internal monitoring and reward treatments (Wald test).

While a 25% increase in income seems an ambitious threshold, we believe that impacts of this magnitude are consistent with the local context dominated by a large number of low-income members who stand to gain considerably from a more equitable distribution of forest-based income.[17] The threshold is also consistent with earlier "impacts" documented in studies that evaluate the effects of top-down and internal monitoring. For instance, Olken (2007) found a 30% decrease in missing expenditure as compared to mean of the control group as a result of top-down monitoring, and Bjorkman and Svensson (2009) found a 20% increase in outpatient services and 33% decrease in under-5 child mortality.

The ex-ante power analysis produces MDEs that are consistently smaller than the threshold, suggesting we are sufficiently powered to pick up meaningful treatment effects. However, the outcomes of the ex post power analysis are less favorable as the standard errors are larger than expected. The results of the ex post analysis in Table A11 are based on regression models with individual-level fixed effects, but we obtain qualitatively similar results when using the simple difference-in-differences specification. While the consumption model is always sufficiently powered (power = 0.8, $\alpha = 0.05$), the power level of the income and forest benefits models is lower for some (but not all) interventions under our assumed threshold value of 25%. The consumption model remains sufficiently powered with the 20% of the threshold value, but not with 10%. This implies a certain risk that we fail to reject the null of no effect when an effect actually does exist, which is a relevant caveat when interpreting the results.[18]

6.3. Why does top-down monitoring and punishment work?

What is the mechanism linking top-down monitoring and punishment to elevated income levels and reduced inequality? This part of the analysis does not rely on experimental variation, so identification relies on observational data and additional assumptions. With this caveat in mind, let us turn to the evidence. Table 2 explores how the three interventions affect end-line measures of (i) group member satisfaction with the leader and executive committee, (ii) member participation and behavior in group activities, and (iii) measures of members' own monitoring effort and appreciation of the forest condition. These variables were only collected at end-line, so we estimate OLS models. For all variables, we find a positive treatment effect of top-down monitoring and punishment. Columns (1–2) indicate that top-down monitoring results in greater appreciation for group leadership, and columns (3–4) reveal that group members are more likely to participate in group activities and respect group rules better.

Where does the enhanced appreciation for leadership come from? In theory, external monitoring can work through (at least) two channels :[19] (i) it can *discipline incumbent* leaders, incentivizing them to behave more inclusively and transparently, and (ii) it can stimulate the *replacement* of bad leaders by good ones. As is evident from Table 3, our data indeed provide evidence of accelerated leader turnover. While only one leader was replaced in the control group, no less than 9 leaders were replaced by OFWE management.

[17] In addition, there may be indirect effects, for example through relaxing liquidity constraints that help to raise agricultural productivity (some farmers used extra income from forest-based resources to purchase agricultural inputs such as improved seeds and fertilizers, or engage in new economic activities like bee production.

[18] Low power also reduces the likelihood that a statistically significant result reflects a true effect.

[19] One alternative mechanism may be signaling the social value by involving an outside organization like OFWE. However, this begs the question why similar results do not emerge for the other treatments where OFWE also plays an important role.

Table 3
Leader turnover for different experimental arms.

Treatments	Number of forest user groups	Leader turnover (%)
Control group	34	2.94
Top-down	33	27.27
Reward	32	6.25
Internal	33	18.18

Table 4a
Selection vs discipline effects.

Monitoring	Observations		Mean values			
	Old leaders	*New leaders*	*Old leaders*	*New leaders*	*differences*	*p-values*
Top down monitoring						
Satisfaction with chairperson	24	9	4.043	3.166	.878	.001
Satisfaction with executive committee	24	9	4.053	3.215	.838	.001
Members' participation and influence	24	9	4.292	3.912	.38	.018
Internal monitoring						
Satisfaction with chairperson	27	6	3.538	3.343	.195	.511
Satisfaction with executive committee	27	6	3.51	3.305	.203	.468
Members' participation and influence	27	6	4.096	3.936	.161	.351

Table 4b
Assessment of leaders by group members: propensity score matching.

	Top-down	Control	Difference	T-stat
	(1)	(2)	(3)	(4)
Unmatched sample	4.042	3.539	0.503	3.34
Matched sample	4.123	3.458	0.664	3.10

In light of the small number of groups in the top-down monitoring treatment it is hazardous to explore the relative importance of *selection* effects vis-a-vis the *disciplining* (or incentive) effect of top-down monitoring. Nevertheless, as a first and tentative step in this direction, we compare group members' appreciation for remaining and newly-elected leaders. Results are reported in the top panel of Table 4a, and indicate that old leaders are more appreciated than new ones. This is consistent with theory and empirical work proposing that stricter enforcement will attract new leaders that are, on average, of "lower quality" in terms of their opportunity cost than pre-existing ones (Grossman and Hanlon 2014; Kahsay and Bulte 2020). The reason is that intensive monitoring and enforcement erodes the benefits (and raises the costs) from being in power.[20] In terms of observables, we find that new leaders are less educated and have less business experience than their predecessors. More importantly, we also find that the number of displaced leaders is statistically similar for the top-down and internal monitoring regimes (9 versus 6 leaders, respectively). Taken together, the data suggests the selection channel is unlikely to fully explain the difference in outcomes we observe.

However, there is some empirical support for the alternative assumption that top-down monitoring and punishment disciplines incumbent leaders. Table 4b summarizes the outcomes of a propensity score matching model. Remaining leaders of treatment groups are appreciated higher than "similar peers" from the control arm. This suggests that top-down monitoring and punishment affected their behavior.

6.4. Why does internal monitoring not work?

On average, the internal monitoring treatment does not improve livelihood outcomes for group members. This is consistent with Olken (2007) and Raffler et al. (2019), and raises the question why internal accountability does not seem to work. Three of the earlier reasons why internal monitoring may fail seem less relevant in the context that we study. First, OFWE officials helped groups to overcome social dilemmas and assisted in the formation of the monitoring committee. Second, since OFWE officials participated in General Assembly meetings, and placed the findings of the committee in context, group members could be informed about the relative performance of their leaders. Third, lack of empowerment seems unimportant because voting was anonymous. It was possible to indicate one's disapproval without having to fear future retaliation by the elite. Indeed, no less than six groups actually dismissed their leaders (Table 3), so internal punishment appears a credible course of action. It also seems reasonable to assume that committee members feel empowered and 'legitimate' because of the training and (material) support they received from OFWE.[21]

[20] For recent evidence on how to select prosocial and productive individuals into public office, refer to Ashraf et al. (2020). Another explanation for our finding may be that new leaders need to gain experience and learn-by-doing on the job, so that appreciation for their work will likely increase over time as their performance improves.

[21] Nevertheless, we also believe that our analysis does not produce the final word on empowerment and internal monitoring, We have performed a heterogeneity analysis based on two proxies of "group empowerment"––a measure of *leader turnover* (based on the assumption that leader's power increases

Table 5
Governance interventions, consumption and forest dependence (consumption outcome).

	Median cut-off		30% cut-off	
	Forest-dependent groups	*Non-forest-dependent groups*	*Forest-dependent groups*	*Non-forest dependent groups*
Top-down	5024.763	1880.204	3902.006	2269.843
	(3095.334)	(2997.874)	(5732.825)	(2297.546)
Reward	1262.175	2386.055	406.651	3133.026
	(3210.209)	(2707.243)	(3466.631)	(2565.612)
Internal	−1721.664	5816.556*	−3940.142	4772.308*
	(2203.582)	(3490.750)	(3258.573)	(2729.632)
Period	3582.454	3419.772	600.862	4519.974**
	(2208.873)	(3106.847)	(2454.846)	(2120.217)
Top-down × Period	2477.03	10,767.40**	3154.566	6796.627**
	(3959.39)	(4482.18)	(5515.701)	(3405.202)
Reward × Period	1592.04	−4790.50	5886.114	−4565.307
	(3727.61)	(3841.34)	(4219.343)	(3172.852)
Internal × Period	5440.73*	−4919.19	8716.248**	−3943.692
	(2985.94)	(4454.10)	(3409.250)	(3508.258)
Controls	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Model	DID	DID	DID	DID
Constant	1,642,155.03	872,428.34	−5,729,594.16	1,352,020.80*
	(990,462.74)	(861,924.28)	(3,970,132.47)	(713,481.04)
R^2	0.554	0.565	0.718	0.465
Mean of control-group 2018	23,687.35	28,483.28	19,996.59	28,970.45
Observations	132	132	70	194

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include altitude, group size, year of establishment, share of female members, share of literate members, average age of members, and average land holding. The estimated coefficients for the treatments are significantly different from each other for each outcome variable except between internal monitoring and reward treatments for non-forest dependent groups and between top-down monitoring and reward treatments for forest-dependent groups (Wald test).

This leaves us with the final potential reason for failure of internal monitoring. Community members may be "*uninterested*" to hold their leaders accountable, perhaps because doing so would be an uneconomic usage of their time. Group members may not care enough to monitor their leaders and fellow members if they have alternative sources of income, which, at the margin, are more worthy of their time and attention. Indeed, our data are consistent with such a "lack of interest" perspective. From columns (1–2) of Table 2, observe that internal monitoring does not increase appreciation for leadership––it did not motivate group leaders to make their management more inclusive or transparent. This would be a logical outcome if elected committee members allocated little effort to monitoring their leaders or the standing forest.

A key responsibility of FUG leadership is to be engaged in forest monitoring and to organize forest monitoring by other members—protecting it from incursions by non-members and theft by members. Forest monitoring is therefore a useful proxy for effort. Self-reported levels of monitoring by committee members in the internal monitoring arm are statistically identical to effort provided by ordinary members in their own group, or effort provided by members from the control group. On average, ordinary group members spend 109 days per year on monitoring forest use and illegal extraction. Members of the internal monitoring committee, charged with special responsibilities in this domain, allocate only 99 days to monitoring.[22] The difference is not statistically significant ($p = 0.48$) so, on average, committee members are not motivated by their mandate.

To probe this issue further we distinguish between two types of FUGs: groups that, at baseline, rely on forest-based resources as an important part of their income, and groups that are less reliant on forest-based resources. In Table 5 we summarize the results of the basic model explaining variation in consumption, after splitting the sample based on the share of forest income in total income at baseline.[23] We re-iterate that this part of the analysis is not based on experimental variation in forest dependence, but on observational data, so identification of causal effects requires additional assumptions.

length of tenure, and length of tenure increases power), and the *share of female group members* (which for cultural reasons are extremely unlikely to challenge group leadership). These data enable investigation of whether the impact of interventions varies with empowerment. Both empowerment data are observational, and so causal interpretation rests on additional assumptions. We split the sample based on median for the share of female members and we use a binary indicator for the leader turnover (whether groups changed their leaders previously, or not). We re-estimate the diff-in-diff models using forest benefits as the dependent variable. Results are summarized in Appendix Table A12. As expected, top-down monitoring and punishment works best in under-empowered groups—groups with many women or with leaders who have been "in office" for a long time. In these groups, members fail to hold their leaders accountable and rely on outside enforcement for an equitable sharing of benefits. Surprisingly we do not find that the internal monitoring intervention works better in more empowered groups. The impact does not depend on the share of female members at baseline, and the internal monitoring intervention performs worse in groups with new leadership. The relationship between group power and the outcome of governance interventions is complex, and deserves more attention.

[22] Often, FUG members monitor in groups of 3-4. The variable "monitoring days" does not necessary reflect full days allocated to monitoring, but rather the number of days that members are to some extent involved in monitoring the forest.

[23] We view this sub-set of results as plausible, yet tentative, as they are not based on exogenous variation in the strength of the incentive for monitoring. However, there are no significant differences in leadership quality and inequality of benefit distributions among more and less forest-dependent FUGs. Our

In columns (1–2) we split the sample at the median, and in columns (3–4) we only include FUGs in the set of forest-dependent groups if forest resources constitute at least 30% of the average member's income. The latter cut-off point is simply to demonstrate a higher dependence on forest resources and only 35 FUGs qualify as highly-dependent on forest resources thus defined. As expected, the two types of FUGs did not monitor equally intensively at baseline––in forest-dependent groups members allocated 10 days per year more to monitoring.[24]

The results for both sample-splits are comparable and suggest an interesting asymmetry in the impact of our interventions. Top-down monitoring "works" in groups with low forest-based income––presumably groups suffering from bad leadership (coefficients are different from each other, $p<0.01$). Large gains eventuate when an outside agency intervenes in a context where group members have "given up" or are unaware of the potential gains from enhanced scrutiny. In contrast, top-down monitoring does not (further) improve outcomes for groups who already obtain a large share of their income from the forest and who are heavily involved in monitoring themselves. These findings are consistent with the interpretation that internal and top-down monitoring are *substitutes*, and that the value-added of OFWE monitoring is limited when groups themselves monitor their leaders and forest.

Outcomes are reversed for the internal monitoring treatment. Internal monitoring "works" in groups where members depend on the forest––where members have a direct and strong incentive to invest in monitoring in order to protect a pillar of their livelihoods, and have acquired ample experience doing so. Such groups have an incentive to invest in monitoring, and presumably have a better understanding of relevant monitoring technologies. Regardless of the exact motivation, internal monitoring is successful when it builds on a foundation of an already-engaged group and it fails when forest-based income is relatively unimportant. Observe that the treatment effect is larger for the smaller sub-group of highly-dependent FUGs, consistent with the idea that group members are more motivated to engage in monitoring if the stakes are higher. We speculate that the OFWE intervention is akin to a technical improvement for FUGs—empowering monitoring members and raising their productivity while monitoring. The intervention leverages effort levels, and has impact when group members are motivated to supply monitoring effort.

Indeed, we find that, compared to peers from other groups, committee members in forest-dependent FUGs spend an extra 49 days monitoring ($p<0.05$). This extra input translates into extra output. As discussed above, 18% of the groups in the internal monitoring arm replaced their leaders. All these replacements occurred in the sub-sample of forest-dependent FUGs, so that 40% of the forest-dependent FUGs in the internal treatment replaced their leaders.[25]

6.5. Why does the top-down monitoring and reward intervention not work?

The reward treatment mimics the main components of the top-down monitoring and punishment intervention, except that the OFWE management no longer punishes under-performing leaders but instead provides (in-kind) rewards for the best performing leaders. Why does the "carrot" not seem to invite the same response as the "stick"? Leaving the issue of low power of our income and forest benefits models aside, several factors might matter. First and foremost, we believe this may be due to the 10 months gap between the timing of the reward and the follow-up data collection. Without any promises regarding a second-round reward, the incentive effect of the reward may have faded. Second, there may be important asymmetries between rewards and punishments. For example, rewards may crowd out leaders' intrinsic motivation. Third, design issues may play a role. Our reward may negatively affect cooperation among executive committee members because it is intended for the chairperson only. Alternatively, while each under-performing leader could be punished by OFWE representatives, the reward system was based on a tournament with only three winners (which arguably dilutes the incentive to put in extra effort). Unfortunately, we do not have access to additional data to probe these explanations further. We therefore believe that additional research on "top-down monitoring with carrots" is warranted before writing off this treatment as a potential remedy to local governance challenges—it is possible that alternative designs would have generated better results, even in the same context.

7. Discussion and conclusions

The idea of involving local communities in development initiatives, decision-making and monitoring of public service delivery has gained a lot of traction in international development circles. National governments in low-income countries also increasingly devolve part of their responsibilities to local levels, for example in the domain of natural resource management. The participatory approach to development is predicated on the idea that grassroots-level engagement of community members helps to overcome information and incentive constraints, hopefully contributing to efficient and effective service delivery or the creation and fair distribution of wealth. However, the empirical evidence regarding the success of community-driven approaches is inconclusive.

Our study focuses on the devolution of forest user rights to groups of villagers in Ethiopia, and the governance arrangements to improve group-level economic outcomes. While the introduction of forest user groups has contributed to

measure of (lagged) forest dependence is potentially endogenous, and additional research into how the effectiveness of monitoring is driven by incentives is a priority for future research.

[24] This is true both in groups assigned to top-down and internal monitoring, and differences are significant at the $p<0.1$ level in both treatment arms.

[25] Additional analysis (not shown) reveals that monitoring committees in forest-dependent FUGs send away similar types of leaders as the OFWE officials (in terms of observables).

conservation and sustainable management of forests, the impacts on group members' livelihoods is typically considered disappointing. Anecdotal evidence suggests that elite capture explains why many groups fail to live up to their potential. We use an RCT with 132 forest user groups in the Adaba-Dodola participatory forest management program to evaluate the impact of three interventions aimed to improve group performance, especially the impact on the generation and distribution of (forest-based) income. The three interventions are internal monitoring by group members, top-down monitoring and punishment by officials, and top-down monitoring and rewards for high-quality leadership. Implementation was done by the Oromia Forest and Wildlife Enterprise – the government organization actually responsible for forest management in our study region.

Our main findings reconcile conflicting evidence in the literature. Our result that top-down monitoring and punishment has, on average, large positive impacts on (forest-based) income for group members, matches results by Olken (2007), even if these were obtained in another country (Indonesia) and for another sector (infrastructure). Income and consumption levels increase by 0.2 standard deviations, and income is distributed more equally among group members. We find evidence supporting an incentive effect of top-down monitoring for incumbent leaders. Overall appreciation by group members for existing leaders in groups that are scrutinized by outside experts is very high. The threat of top-down monitoring has a real bite in our context; more than a quarter of the leaders in this treatment arm were replaced because they were underperforming according to the experts.

In contrast, the other governance interventions produced, on average, disappointing outcomes (but keep in mind the lower power of some of our tests). The reward treatment had no significant impact, perhaps because of the timing of our reward. More importantly in light of on-going discussions about the merits of engaging local communities, we also find that internal monitoring fails to raise forest-based income, on average. Our internal monitoring treatment contained several components identified in the literature as important (e.g. Barr et al., 2012, Bjorkman Nyqvist et al. 2017, Fiala and Premand 2018), including efforts to empower and inform group members, and address collective action problems. Nevertheless, the average impact was very small––fitting in a small emerging literature warning against too much enthusiasm about community-driven development. Many group members appear to lack incentives to engage in costly monitoring.

Interestingly, however, and possibly building a bridge to work of Bjorkman and Svensson (2009), internal monitoring works seems to better in the subsample of forest user groups heavily relying on the forest for their livelihoods, as compared to villages where forest-based income is a relatively small complement to agricultural income. Correlations in an analysis of observational data suggest that if members see the benefit of costly monitoring, and allocate effort to it, then assistance and technical improvements in internal monitoring can have a large impact on leadership and group performance. Similar to our finding that the effectiveness of the internal monitoring intervention seems to vary with local conditions, we also find that the effectiveness of top-down monitoring and punishment appears context-specific. If group members actively monitor themselves, then additional top-down monitoring and punishment seems to have little impact. Additional experimental research probing these suggestions is welcome.

Acknowledgements

We thank the editor, two anonymous reviewers, seminar participants in Bergen (NHH), Copenhagen, Lund, Geneva, Leuven and Wageningen for helpful comments and suggestions. Remaining errors are our own. 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

Appendix A. The Adaba-Dodola PFM program

The Ethiopian government, supported by international organizations, initiated a large PFM program towards the end of the previous century. Prior to the introduction of PFM, forests were under government control, which in practice meant that extraction took place under conditions resembling "open access". The PFM program decentralized forest management, and gave local communities exclusive user-rights over clearly demarcated forest blocks. In collaboration with the German Society for International Cooperation (GIZ), 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).

With the help of experts from the Oromia Forest and Wildlife Enterprise (OFWE), FUGs drafted their own bylaws. These include details regarding the organizational structure, and the rights and responsibilities of group members and leaders. OFWE provided a sample bylaw, which subsequently became the basis for most group bylaws, but groups were allowed to further specify their own rules. After agreeing on the ground rules, FUGs enter into contract with Oromia Rural Land and Natural Resources Administration Authority (ORLNRAA) to specify contracts spelling out rights and duties of the group, and the conditions for contract termination. FUGs are allowed to use forest products for own consumption and sale, and maintain existing farm plots inside the forest. In return, they should manage the forest in a sustainable manner, restrict further settlement and agricultural expansion, and pay an annual rent. The rental agreement is valid for an indefinite period, provided rents are paid and timber extraction does not exceed the allowable cut by more than 10%.

Currently, about 50,000 hectares of forest are managed under the PFM program in Adaba and Dodola districts by 132 FUGs (3600 members in 13 villages or kebeles), each having a clearly demarcated forest block, A kebele is a collection of

hamlets, with an average population size and land area of 7600 persons and 67 km2, respectively, in ours sample. The Adaba-Dodola OFWE branch office is responsible for the PFM program in Adaba and Dodola districts, verifying whether FUGs' actions are in accordance with the contract they entered with ORLNRAA. Among others, OFWE monitors forest use and management patterns, intervenes when conflicts arise within or between FUGs, estimates and collects forest rent, conducts forest inventories, and provides trainings.

From a conservation perspective, the program has been rather successful (Ameha et al., 2014, 2016) despite substantial heterogeneity across FUGs. Deforestation has nearly stopped, from a pre-program annual deforestation rate of 3% (Tadesse 1999). Forest inventories conducted by OFWE show that, on average, the number of potential crop trees and mature trees had increased by, respectively, 0.5 and 2 trees per ha, in 2012 compared to stocks in 2004. Conservation impacts vary with the share of conditional cooperators in the group (Rustagi et al., 2010) and the quality of group leadership (Kosfeld and Rustagi 2015).

The impact of PFM on income of local beneficiaries is less evidently favorable. While the great majority of FUG bylaws stipulates that forest benefits should be divided equally among members, there is little evidence of broad welfare gains for members. Consistent with the evidence from other countries, discussed above, Ameha et al. (2014) found no welfare gains for group members as compared to non-members in Dodola district, and even a decrease in welfare in another district. Gelo and Koch (2014) find an increase in average income for FUG members, but benefits are concentrated among rich members. This raised concerns among OFWE officials about sustainability of the program. In 2014, OFWE commissioned a study to learn more about the performance of FUGs as it had plans to upscale the PFM program throughout Oromia regional state. The study identified, as key problems, a lack of transparency and inadequate participation of members in group decision-making, a lack of accountability in financial flows, and unfair distribution of benefits and responsibilities (Tesfay et al., 2015).

Appendix B. Additional tables

Tabe A1
Descriptive statistics at baseline: leaders and members.

	members	leader
Age	48.212	45.248
Years of schooling	4.117	5.808
Household size	9.084	10.433
Literacy	.466	.754
Income	32,642.06	47,312.68
Land holding	9.058	10.995
Asset holding	1928.239	8786.281
Livestock	21.505	25.944
Consumption	40,210.5	60,307.17

Table A2
Sample size per experimental arm.

Treatments	Number of forest user groups	Total members	Members sampled
Control group	34	933	320
Top-down	33	897	297
Reward	32	866	296
Internal	33	902	302
TOTAL	132	3598	1215

Table A3
Governance questions.

Satisfaction with executive committee	
1	The Executive committee makes sure group rules are respected, to punish violators, and encourage members to spend enough time in group activities
2	They are fair, impartial, and care for the wellbeing of the group
3	They invite members to meetings and listen to members' comments
4	They are transparent and responsible regarding group activities such as finance, forest management, etc.
5	They consult members on important issues
Satisfaction with chairperson	
1	The leader makes sure group rules are respected, to punish violators, and encourage members to spend enough time in group activities
2	The leader is fair, impartial, and cares for the group wellbeing
3	The leader invites members to meetings and listens to members' comments
4	The leader is transparent and responsible regarding group activities such as finance, forest management, etc.
5	The leader consults members on important issues
Members participation and influence	
1	I am free to participate in FUG decision making
2	I present complaints to FUG leadership when things are not going well
3	I actively take part in FUG meetings
4	I make suggestions to FUG leadership
5	My suggestions and complaints influence WAJIB leadership decisions
Members' behavior	
1	Most members attend meetings
2	Most members respect group rules
3	Most members spend enough time in group activities
4	Most members actively engage in discussions
5	There is solidarity among members
Forest condition	
1	The past 6 months there has been an increase in the regeneration of trees
2	The past 6 months there has been an increase in forest area under protection management (Area closure)
3	The past 6 months there has been a decrease in the number of charcoal making spots
4	The past 6 months there has been a decrease in the number of a newly felled trees
5	The past 6 months there has been a decrease in encroachment and agricultural expansion
Monitoring committee	
1	The monitoring committee is elected appropriately
2	The monitoring committee consults members on important issues
3	The monitoring committee is fair, impartial, and cares for the group wellbeing
4	The monitoring committee is capable enough to monitor the performance of leaders

Note: Groups members were asked the extent to which they agree with the above statements (1=Strongly disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree).

Table A4
Descriptive statistics.

	N	Mean	St.Dev	min	max
Welfare indicators					
Consumption	264	28,214.67	11,433.2	9095.34	75,728.93
Income	264	24,241.78	14,293.03	3695.892	102,000
Extraction	264	4643.446	5131.262	443.667	53,148.38
Inequality indicators					
Consumption	264	0.208	0.105	0.012	0.608
Income	264	0.288	0.141	0.011	0.742
Extraction	264	0.273	0.156	0.015	0.836
Members satisfaction with leadership and participation					
Executive committee	132	3.57	.593	2.17	4.75
Chairperson	132	3.577	.62	1.875	4.702
Members' participation and influence	132	4.057	.375	2.929	4.823
Group characteristics					
Altitude	264	2.227	.659	1	3
Year of establishment	264	2005.629	3.456	1999	2011
Group size	264	27.273	4.445	11	30
Share of female members	264	.205	.117	0	.727
Average age	264	48.245	5.906	33.667	61.125
Average education	264	.495	.212	0	1
Average land holding	264	8.036	5.461	.814	59.627
Forest stock					
Number of potential crop trees ha	234	45.326	29.622	1.1	147.63
Number of mature trees per ha	234	42.989	40.281	2.12	423

Note: The welfare indicators are deflated both the baseline and follow-up periods by the corresponding CPI.

Table A5
Baseline balance of covariates across treatments.

	Control group	Top-down		Reward		Internal	
	Mean1	Mean	p value	Mean	p value	Mean	p value
Consumption	22,875.955	29,509.145	.011	26,244.792	.178	25,349.239	.302
Income	17,391.791	23,363.633	.032	22,880.787	.061	21,135.45	.22
Forest benefit	3376.019	3896.663	.425	3584.743	.716	3291.899	.871
Consumption inequality	.265	.26	.819	.241	.226	.264	.961
Income inequality	.374	.384	.756	.334	.165	.346	.365
Forest benefit inequality	.319	.389	.059	.314	.822	.302	.578
Altitude	2.264	2.182	.628	2.219	.779	2.243	.888
Year of establishment	2005.677	2005.515	.842	2005.625	.955	2005.697	.981
Group size	27.5	27.182	.767	27.063	.696	27.334	.879
Share of female members	.208	.193	.575	.212	.903	.208	.994
Age	48.572	48.813	.873	46.977	.321	47.325	.35
Education	.462	.528	.193	.542	.147	.458	.941
Land holding	9.574	10.511	.654	9.063	.806	7.498	.277
Monitoring com	.383	.303	.502	.375	.952	.212	.132
Auditing com	.676	.666	.933	.656	.865	.455	.069
Potential crop trees	45.064	39.657	.443	50.047	.538	46.524	.854

Table A6
Parallel trend assumption.

	Potential crop trees (1)	Mature trees (2)
Top-down	−4.202	−0.101
	(5.946)	(4.946)
Reward	4.059	1.824
	(6.022)	(5.023)
Internal	0.211	16.479
	(6.362)	(14.499)
Period	3.450	10.887
	(9.388)	(7.311)
Top-down x Period	13.170	−0.395
	(13.359)	(7.737)
Reward x Period	−3.861	−10.380
	(13.464)	(6.653)
Internal x Period	−9.584	−15.913
	(11.422)	(18.691)
Controls	Yes	Yes
Village fixed effects	Yes	Yes
Model	DID	DID
Constant	11,234.423***	9154.197***
	(2406.725)	(1634.818)
R^2	0.526	0.227
Observations	166	166

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. OLS estimates explaining variation in forest health, based on forest inventories in 2004 and 2012.

Table A7
Governance interventions and outcomes, ANCOVA regressions.

	Level outcomes			Inequality outcomes		
	Consumption	*Income*	*Forest benefit*	*Consumption*	*Income*	*Forest benefit*
	(1)	(2)	(3)	(4)	(5)	(6)
Top-down	10,852.211***	15,400.568***	5444.409**	−0.095***	−0.070**	−0.031
	(1628.529)	(3455.797)	(2272.437)	(0.023)	(0.024)	(0.025)
Reward	1014.842	1374.362	964.280	−0.011	−0.023	−0.042
	(2658.567)	(3320.873)	(932.841)	(0.018)	(0.022)	(0.033)
Internal	4292.377	5460.396*	679.993	−0.045**	0.010	0.022
	(2629.980)	(2930.425)	(755.745)	(0.019)	(0.030)	(0.029)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Constant	705,201.695	2,078,951.950*	1,045,515.5***	1.772	26.509***	−28.812***
	(510,357.105)	(1,110,033.566)	(155,183.443)	(2.132)	(4.388)	(7.255)
R^2	0.584	0.529	0.342	0.266	0.220	0.138

(*continued on next page*)

Table A7 (*continued*)

	Level outcomes			Inequality outcomes		
	Consumption	*Income*	*Forest benefit*	*Consumption*	*Income*	*Forest benefit*
Mean of control-group 2018	22,875.955	17,391.792	3376.02	0.265	0.374	0.319
Observations	132	132	132	116	124	125

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include altitude, group size, year of establishment, share of female members, share of literate members, average age of members, and average land holding. Initial forest stock is not included as a control variable to increase the number of observations. However, our results do not change when we control for it (details available on request).

Table A8
Governance interventions and outcomes, individual member level estimates (group-level fixed effects).

	Level outcomes		
	Consumption	*Income*	*Forest benefit*
	(1)	(2)	(3)
Top-down	3743.679**	−19,692.643***	−2809.801**
	(1770.554)	(2685.978)	(1102.909)
Reward	10,470.297***	1081.751	6643.419***
	(1496.683)	(3141.991)	(422.829)
Internal	−1582.749	−24,433.756***	−984.234***
	(1406.890)	(2036.165)	(346.570)
Period	4412.003***	8195.718***	703.550*
	(1614.512)	(2019.836)	(386.747)
Top-down x Period	6306.697**	8930.736**	5170.115**
	(3091.692)	(3889.048)	(2078.329)
Reward x Period	−1868.265	−6580.957	485.248
	(2571.898)	(4040.062)	(715.017)
Internal x Period	1026.346	−2973.312	220.843
	(2668.991)	(2814.546)	(671.782)
Controls	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes
Model	DID	DID	DID
Constant	6946.214***	5070.564	1312.771**
	(2385.719)	(4492.297)	(630.036)
R^2	0.350	0.269	0.104
Mean of control-group 2018	22,875.955	17,391.792	3376.02
Observations	2264	2264	2264

Notes: Clustered (at group level) standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include age, education, household size, marital status and land holding.

Table A9
Governance interventions and outcomes, individual member level estimates (household level fixed effects).

	Level outcomes		
	Consumption	*Income*	*Forest benefit*
	(1)	(2)	(3)
Period	3713.391***	7172.044***	573.078*
	(964.230)	(1571.171)	(334.191)
Top-down x Period	6400.985***	9745.169***	5975.682**
	(2104.592)	(2924.997)	(2446.149)
Reward x Period	−2632.347	−7383.802*	325.196
	(1533.911)	(2031.257)	(674.396)
Internal x Period	1535.494	−1561.789	554.219
	(1665.054)	(3287.836)	(488.316)
Controls	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes
Constant	17,574.216***	8431.521	7286.562***
	(5630.040)	(9857.952)	(1603.278)
R2	0.128	0.087	0.041
Mean of control group 2018	22,875.955	17,391.792	3376.02
Observations	2264	2264	2264

Notes: Clustered (at group level) standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include age, education, household size, marital status and land holding.

Table A10
Spillover effects, ANCOVA regressions.

	Level outcomes			Inequality outcomes		
	Consumption	Income	Forest Benefit	Consumption	Income	Forest benefit
	(1)	(2)	(3)	(4)	(5)	(6)
Number of groups in top-down monitoring	1052.627	127.288	316.066	−0.006	−0.011*	0.017
	(635.225)	(760.018)	(198.150)	(0.008)	(0.006)	(0.010)
Number of groups in internal monitoring	−1395.113	−777.507	−178.149	−0.001	−0.013	−0.005
	(1150.10)	(1639.41)	(200.42)	(0.01)	(0.01)	(0.01)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village fixed effects	No	No	No	No	No	No
Constant	437,606.03	1,857,841.32*	105,369.34	−8.02	39.58***	−14.31
	(1,162,036.6)	(1,062,050.0)	(223,158.6)	(9.74)	(7.86)	(12.59)
R^2	0.338	0.311	0.425	0.361	0.429	0.387
Mean of control group 2018	22,875.955	17,391.792	3376.02	0.265	0.374	0.319
Observations	34	34	34	34	34	34

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include altitude, group size, year of establishment, share of female members, share of literate members, average age of members, and average land holding.

Table A11
Power calculation.

	Level outcomes			Inequality outcomes		
	Consumption	Income	Forest benefits	Consumption	Income	Forest benefits
Regression-based standard errors (SE)						
Top-down	1628.53	3455.80	2272.44	0.023	0.024	0.025
Reward	2658.57	3320.87	932.84	0.018	0.022	0.033
Internal	2629.98	2930.43	755.75	0.019	0.030	0.029
Sample mean	30,460.09	27,332.23	5751.14	0.151	0.207	0.207
Intra-cluster correlation (ICC) based on baseline data						
Top-down	0.44	0.45	0.48			
Reward	0.43	0.44	0.46			
Internal	0.42	0.42	0.45			
Ex-ante power calculation: Minimum Detectable Effect (MDE), power of 0.80 and significance level of 0.05						
Top-down (MDE)	5287	5570	1285	0.039	0.059	0.078
Reward (MDE)	4941	5842	1124	0.038	0.058	0.053
Internal (MDE)	4722	6024	1014	0.051	0.062	0.063
Ex-post power calculation: Minimum Detectable Effect (MDE)=2.8*SE						
Top-down (MDE)	5893	8190	6849	0.064	0.067	0.070
Reward (MDE)	4295	5688	1888	0.050	0.062	0.092
Internal (MDE)	4662	9206	1367	0.053	0.084	0.081
Critical effect = 0.25*sample mean control group	**7615**	**6833**	**1438**	**0.038**	**0.052**	**0.052**
Critical effect = 0.20*sample mean control group	6092	5466	1150	0.030	0.041	0.041
Critical effect = 0.10*sample mean control group	3046	2733	575	0.015	0.021	0.021

Note: Ex-ante power was calculated using STATA power command.

Table A12
Governance interventions, forest benefit and group empowerment.

	Leader turnover		*Share of female members*	
	Yes	No	High	Low
	(1)	(2)	(3)	(4)
Top-down	−4924.69**	125.57	408.09	733.21
	(2231.32)	(759.46)	(1284.93)	(807.62)
Reward	−3251.53	59.59	−73.31	811.71
	(2878.84)	(627.12)	(997.72)	(768.07)
Internal	−5265.94*	−107.50	−466.57	−244.03
	(2728.10)	(635.07)	(1100.11)	(661.88)
Period	3047.78***	538.84	121.24	1099.44**
	(271.33)	(415.23)	(814.94)	(529.29)
Top-down × Period	−831.26	6889.32***	7984.34*	2579.92*
	(1198.59)	(2610.24)	(4094.84)	(1390.23)
Reward × Period	−2651.51***	675.43	1212.48	−153.27
	(672.54)	(965.53)	(1338.62)	(1285.60)
Internal × Period	−1939.14**	547.81	977.25	−340.56
	(743.05)	(807.38)	(1236.67)	(743.44)

(*continued on next page*)

Table A12 (*continued*)

	Leader turnover		*Share of female members*	
	Yes	No	High	Low
Controls	Yes	Yes	Yes	Yes
Village fixed effects	Yes	Yes	Yes	Yes
Constant	−2,960,741**	1,001,245***	991,487	991,353***
	(1,289,888)	(248,230)	(1,428,948)	(272,029)
R^2	0.770	0.345	0.301	0.527
Observations[¤]	36	228	130	134

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included control variables include altitude, group size, year of establishment, share of female members, share of literate members, average age of members, and average land holding. The median split of the sample did not generate equal value given the limited number of possible values in share of female members and thus we added those exactly at the median to the 'Low' group.

References

Adhikari, B., Di Falco, S., Lovett, J.C., 2004. Household characteristics and forest dependency: evidence from common property forest management in Nepal. Ecol. Econ. 48, 245–257.

Agrawal, A., Gupta, K., 2005. Decentralization and participation: the governance of common pool resources in Nepal's Terai. World Dev 33 (7), 1101–1114.

Ameha, A., Nielsen, O., Larsen, H., 2014. Impacts of access and benefit sharing on livelihoods and forest: case of participatory forest management in Ethiopia. Ecol. Econ. 97, 162–171.

Ashraf, N., Bandiera, O., Jack, K., 2014. No margin, no mission? A field experiment on incetives for public service delivery. J Public Econ 120, 1–17.

Ashraf, N., Bandiera, O., Davenport, E., Lee, S., 2020. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. Am. Econ. Rev. 110, 1355–1394.

Banerjee, A.V., Duflo, E., Glennerster, R., Khemani, S., 2010. Pitfalls of participatory programs: evidence from a randomized evaluation in education in India. American Economic Journal: Economic Policy 2, 1–30.

Banerjee, A., Hanna, R., Kyle, J., Olken, B., Sumarto, S., 2018. Tangible information and citizen empowerment: identification cards and food subsidy programs in Indonesia. J. Polit. Econ. 126, 451–491.

Bardhan, P., 2002. Decentralization of governance and development. J. Econ. Perspect. 16, 185–205.

Bardhan, P., Mookherjee, D., 2006. Decentralisation and accountability in infrastructure delivery in developing countries. The Economic Journal 116 (508), 101–127.

Barr, A., Mugisha, F., Serneels, P., Zeitlin, A., 2012. Information and Collective Action in Community-Monitoring of schools: Field and Lab Experimental Evidence from Uganda. University of Nottingham Working Paper.

Beath, A., F. Christia, and R. Enikolopov, 2013. *Randomized impact evaluation of Afghanistan's national solidarity programme.*Working Paper

Björkman, M., Svensson, J., 2009. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. Q J Econ 124 (2), 735–769.

Björkman Nyqvist, M., de Walque, D., Svensson, J., 2017. Experimental evidence on the long-run impact of community-based monitoring. American Economic Journal: Applied Economics 9 (1), 33–69.

Blackman, A., Corral, L., Lima, E.S., Asner, G.P., 2017. Titling indigenous communities protects forests in the Peruvian Amazon. Proc. Natl. Acad. Sci. 114 (16), 4123–4128.

Casey, K., Glennerster, R., Miguel, E., 2012. Reshaping institutions: evidence on aid impacts using a preanalysis plan. Q J Econ 127 (4), 1755–1812.

Casey, K., 2018. Radical decentralization: does community-driven development work? Annu Rev Econom 10, 139–163.

Fearon, J.D., Humphreys, M., Weinstein, J.M., 2015. How does development assistance affect collective action capacity? Results from a field experiment in post-conflict Liberia. American Political Science Review 109 (3), 450–469.

Fiala, N. and P. Premand, 2018. *Social accountability and service delivery: experimental evidence from Uganda*. World Bank Policy Research Working Paper No. 8449.

Fritzen, S.A., 2007. Can the design of community-driven development reduce the risk of elite capture? Evidence from Indonesia. World Dev 35 (8), 1359–1375.

Gelo, D., Koch, S., 2014. The impact of common property right forestry: evidence from Ethiopian villages. World Dev 64, 395–406.

Gelo, D., Koch, S., Muchapondwa, E., 2013. Do the poor benefit from devolution policies? Evidences from quantile treatment effect evaluation of joint forest management. Economic Research Southern Africa (ERSA) Working Paper 400.

García-López, G.A., 2019. Rethinking elite persistence in neoliberalism: foresters and techno-bureaucratic logics in Mexico's community forestry. World Dev 120, 169–181.

Grossman, G., Hanlon, W., 2014. Do better monitoring institutions increase leadership quality in community organizations? Evidence from Uganda. Am J Pol Sci 58 (3), 669–686.

Gugerty, M.K., Kremer, M., 2008. Outside funding and the dynamics of participation in community associations. Am J Pol Sci 52 (3), 585–602.

Heß, S.H., Jaimovich, D., Schündeln, M., 2018a. Development Projects and Economic Networks: Lessons From Rural Gambia. Goethe University Frankfurt Working Paper.

Heß, S.H., Jaimovich, D., Schündeln, M., 2018b. Community-driven deforestation? Experimental evidence from a Rural Development Program in West African Drylands. Goethe University Frankfurt Working Paper.

Humphreys, M., de la Sierra, R., Van der Windt, P., 2019. Exporting democratic practices: evidence from a village governance intervention in Eastern Congo. J Dev Econ 140, 279–301.

Iversen, V., Chhetry, B., Francis, P., Gurung, M., Kafle, G., Pain, A., Seeley, J., 2006. High value forests, hidden economies and elite capture: evidence from forest user groups in Nepal's Terai. Ecol. Econ. 58 (1), 93–107.

Jayachandran, S., De Laat, J., Lambin, E., Stanton, C., R. Audy, R., Thomas, N., 2017. Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. Science 357 (6348), 267–273.

Jumbe, C.B.L., Angelsen, A., 2006. Do the poor benefit from devolution policies? Evidence from Malawi's forest co-management program. Land Econ 82 (4), 562–581.

Kahsay, G.A., Bulte, E., 2019. Trust, regulation and participatory forest management: micro-level evidence on forest governance from Ethiopia. World Dev 120, 118–132.

Kahsay, G.A., Bulte, E., 2020. Monitoring and Management of Common Property resources: Observational and Experimental Evidence from Forest User Groups in Ethiopia. University of Copenhagen Working Paper.

Kahsay, G.A., Medhin, H., 2020a. Leader turnover and forest management outcomes: micro-level evidence from Ethiopia. World Dev 127, 104765.
Kosfeld, M., Rustagi, D., 2015. Leader punishment and cooperation in groups: experimental field evidence from commons management in Ethiopia. Am. Econ. Rev. 105 (2), 747–783.
Larson, A.M., 2011. Forest tenure reform in the age of climate change: lessons for REDD+. Global Environ. Change 21 (2), 540–549.
Lund, J.F., Saito-Jensen, M., 2013. Revisiting the issue of elite capture of participatory initiatives. World Dev 46, 104–112.
Lund, J.F., et al., 2015. Mixed method approaches to evaluate conservation impact: evidence from decentralized forest management in Tanzania. Environ Conserv 42 (2), 162–170.
Mansuri, G. and V. Rao, 2003. *Evaluating community-based and community-driven development: a critical review of the evidence*. Washington DC: World Bank Policy Research Working Paper, 3209.
Mansuri, G., Rao, V., 2012. Localizing development: Does participation work?. The World Bank, Washington DC.
Miller, D.C., Rana, P., Nakamura, K., Irwin, S., Cheng, S.H., Ahlroth, S., Perge, E., 2021. A global review of the impact of forest property rights interventions on poverty. Global Environ. Change 66, 102218.
Nagendra, H., Gopal, D., 2011. Tree diversity, distribution, history and change in urban parks: studies in Bangalore, India. Urban Ecosyst 14 (2), 211–223.
Olken, B., 2007. Monitoring corruption: evidence from a field experiment in Indonesia. J. Polit. Econ. 115, 200–249.
Ostrom, E., Nagendra, H., 2006. Insights on linking forests, trees, and people from the air, on the ground, and in the laboratory. Proc. Natl. Acad. Sci. 103 (51), 19224–19231.
Persha, L., Andersson, K., 2014. Elite capture risk and mitigation in decentralized forest governance regimes. Global Environ. Change 24, 265–276.
Persha, L., & Meshack, C. (2016). A triple win. Te impact of Tanzania's Joint Forest Management programme on livelihoods, governance and forests, 3ie Impact Evaluation Report, 34.
Platteau, J.P., 2004. Monitoring elite capture in community-driven development. Dev Change 35, 223–246.
Platteau, J.P., Gaspart, F., 2003. The risk of resource misappropriation in community-driven development. World Dev 31 (10), 1687–1703.
Pynegar, E.L., Jones, J.P.G., Gibbons, J.M., et al., 2018. (2018) The effectiveness of payments for ecosystem services at delivering improvements in water quality: lessons for experiments at the landscape scale. PeerJ Life and Environment (10) 1–29.
Raffler, P., Posner, D., Parkinson, D., 2019. The Weakness of Internal accountability: Experimental evidence from the Ugandan health Sector. Harvard University Working Paper.
Ribot, J., Lund, J., Treue, T., 2010. Democratic decentralization in sub-Saharan Africa: its contribution to forest management, livelihoods, and enfranchisement. Environ Conserv 37 (1), 35–44.
Rustagi, D., Engel, S., Kosfeld, M., 2010. Conditional cooperation and costly monitoring explain success in forest commons management. Science 330, 961–965.
Shikuku, K., Pieters, J., Bulte, E., Laderach, P., 2018. Incentives and the diffusion of agricultural knowledge: experimental evidence from Northern Uganda. Am J Agric Econ 101, 1164–1180.
Sikor, T., Nguyen, T., 2007. Why may forest devolution not benefit the rural poor? Forest entitlements in Vietnam's central highlands. World Dev 35 (11), 2010–2025.
Tadesse, T., 1999. A New Look Towards the Conservation of the Adaba-Dodola forest [Unpublished]. IFMP, Dodola (Ethiopia), p. 13.
Tesfay, Y., Bekele, M., Tefera, F., Kebede, H., Kassa, H., 2015. Enhancing the Role of Forestry in Ethiopia: Strategy for Scaling Up Effective Forest Management Practices in Oromia with Emphasis On Participatory forest Management. CIFOR Ethiopia Office, Addis Ababa.
Wilebore, B., Voors, M., Bulte, E., Coomes, D., A. Kontoleon, A., 2019. Unconditional transfers and tropical forest conservation: evidence from a randomized control trial in Sierra Leone. Am J Agric Econ In Press.
World Bank, 2004. Making Services Work For Poor People. World Bank, Washington, DC World Development Report.