Norwegian University of Life Sciences
School of Economics and Business

Philosophiae Doctor (PhD)
Thesis 2018:21

Incentivized forest conservation: spatial econometric and experimental evidence

Insentivisert skogvern: romlig økonometrisk og eksperimentell evidens

Amare Teklay Hailu

Incentivized forest conservation: spatial econometric and experimental evidence

Insentivisert skogvern: romlig økonometrisk og eksperimentell evidens

Philosophiae Doctor (PhD) Thesis

Amare Teklay Hailu

School of Economics and Business
Norwegian University of Life Sciences

Ås (2018)

Thesis number 2018:21
ISSN 1894-6402
ISBN 978-82-575-1501-0

Dedication

To my special other selves:
Rodas and *Yohana*

Acknowledgements

I try to look back at my journey from a little cow-boyhood in a remote village in Ethiopia all the way to where I am today, and I see a vast array of moments each of which was significant but for the important people whom I happened to meet then and there. If I were to acknowledge each of the family members, teachers, friends, acquaintances and even random people without even one of whom I feel the long journey would inevitably have been incomplete, I would have to write at least a book (perhaps with several volumes!).

Life as a PhD student is tough, and even much tougher when you are away from your kids who, it at times looks, know to ask only one question: 'when are you coming home?' My daughters, Rodas and Yohana paid the heaviest price at the time they would need me the most. Fortunately, they have such a strong and caring mother, my wife, Rahwa Telele, whom I owe the peace of mind I have had over the years.

One of the things that make me wonder if I am the luckiest person is working with my main supervisor, Prof. Arild Angelsen. Arild has been like a father (not just a father, indeed a special father!) to me during the four and half years I have worked with him. Admittedly, there were times—perhaps many of them, especially after my Mom passed away—when I had really low energy and lost all the stamina to turn to the next page, but every time I meet Arild, he uplifts me. Only if words were strong enough to express how grateful I am to Arild! I had known my co-supervisor, Prof. Stein Holden, before I came to Norway, and he has, ever since, been an inspiration for me. As soon as I arrived, he took me, along with colleagues, to his cabin in Tuddal, where I first stepped on the shining snow. During my stay at HH, I enjoyed Stein's courses and benefited a lot from his insightful ideas. It is an honor to have been Stein's student.

Without the generous funding by Norad through CIFOR's global comparative study, I would hardly be able to reach my goal of pursuing a PhD in one of the best countries in the world, Norway. I cannot thank CIFOR and its staff enough for the opportunity and continuous support and inspiration at different stages of my study. I thank Habtemariam, at CIFOR Addis for hosting me in 2015. I am grateful to the colleagues who I met and shared their ideas with me in Warsaw at the GLF, in Bogor at CIFOR's meeting and in different conferences. CIFOR has created a global family of professionals and scientists who are always curious to listen and open to share experience.

I have seen some of the best and the worst things of my life during the last four and half years. My heart was broken and my life devastated when I lost my Mother in February, 2015. I thought I would get used to the pain and start to see the bright side of life, but it only gets worse. I owe my Mom not only all my happiest moments of my childhood, but also my current academic status. Being indebted to her all my achievement and having seen the hardship she used to endure to raise and educate me, my one and only wish was to show her how her toil would pay off. Alas, she gave up on life and made my dream wither in vain too soon. *Mebat, my heart literally hurts; I miss you, and I miss who I used to be before you left!* It is with all this pain, regret, despair and, most often than not, a staggering nihilistic feeling that I somehow have reached here. And by my side have been my father, siblings, grandmas, my wife and many of my friends, upon whom I would like to bestow my most sincere gratitude.

I am particularly grateful to my colleague, friend and brother (from another mother), Abrha

Megos, who would read my face and make sure that I redirect the energy I would expend against my resentful life to more productive and meaningful causes. Abrha, I hope that we will find time to read and discuss all the life-changing books you nudged me to buy.

Dr. Hambulo Ngoma, my office-mate, colleague, friend and brother, has been my role model throughout my stay at 'the third floor' of Tårn. I will not mention what we did in Warsaw, Gothenburg, Bergen, Oslo and other places we have been to together. I miss all the long working hours, Freya's visit to my room and Mazuba's shy but curious looks. Your maturity, decency and discipline taught me lessons I would find nowhere else.

I also enjoyed the company of and learned a lot from many friends at HH: Øyvind (special thanks for the helpful comments and friendly advice—not to mention the skiing lessons!), Federico, Livingstone, Alam, Kevin, Viktorjia (SoDoC is awesome!), Muuz, Kidanemariam (calm and mature officemate after Hambulo left), Girmay, Selam, Desta (I will miss those lunch hours), Menasbo and Mesfin, among others, who made my stay more lively and the hard work demanded of me more natural. I found 'philosophy Friday' positively addictive, and I thank the organizers: Elena, Fredrik, Rani, and Samantha (who also deserves a million thanks for the invaluable comments).

I would like to extend my special words of gratitude to the wonderful staff members at HH. I am grateful to all, but I cannot help to mention particularly Berit, Hanne, Reidun, Silje and Kateryna who made money, resources and information flow smoothly. I have also acquired tremendous academic and life lessons from interacting, in class, at presentations, informal meetings or even Christmas parties, with Ragnar, Knut Einar, Frode and other professors. I cannot thank you all enough!

It was quite a privilege to have taken specialized PhD courses at Gothenburg University, and my stay was even more memorable but for the colleagues and friends I met there. I would like to thank the course organizers, participants and other friends, particularly Gunnar Köhlin, Elizabeth Foldi, Thomas Sterner, Tensay, Seid, Yonas and EfD staff in general.

Last but surely not least, I would like to acknowledge the welcoming gesture I received from the Environment and Climate Research Center (ECRC)—EfD center in Ethiopia—when I approached the center Director, Dr. Haileselassie Medhin, for a desk at his young office. It was a productive stay, and I feel indebted to him.

I know I have not mentioned all, nor have I mentioned every crucial support those I named have offered me, but deep down I acknowledge that and I am thankful to all! .

Amare Teklay Hailu
January 2018
Ås, Norway

Contents

List of papers

This thesis is based on the following four papers.

1. Drivers of deforestation in Indonesia: a spatial panel data analysis, (with Arild Angelsen and Arief Wijaya).

2. Performance payment and reference levels: a framed field experiment on forest conservation, (with Arild Angelsen).

3. Pay individuals or groups to conserve forests? Experimental evidence from Ethiopia, (with Arild Angelsen) *Submitted to Land Use Policy*.

4. Forest conservation and motivation crowding: a framed field experiment.

Summary of thesis

A key element in the implementation of an effective result-based mechanism for Reducing Emissions from Deforestation and forest Degradation (REDD+) is the reference level (RL). Setting RLs requires modeling and predicting deforestation trajectories for a business-as-usual (BAU) scenario. This thesis looks into two aspects of the design and implementation of a Payment for Environmental Service (PES) scheme for REDD+. First, we apply spatial econometric panel data analysis to explore the drivers of deforestation in Indonesian districts. Spatial models come in many forms, and we test and identify the most suitable spatial model, the Spatial Autoregressive (SAR) model. Incorporating a spatial lag of the dependent variable does not only help us measure neighborhood effects but also improves the accuracy of estimates of other predictor variables that drive deforestation. We found a strong inter-district dependence, which implies that there could be synergistic gains in the implementation of forest conservation policies. Deforestation is contagious, but conservation efforts may have positive leakage (spillover), much like the effect of vaccination on those not treated.

A second set of questions consists of the PES design issues: how different RLs affect conservation and implementation cost, who to pay, and whether any incentivized reductions will last beyond the project period. These questions are crucial for the efficacy of the PES, and engaging in large-scale implementation of a program without answering these questions entails high risk of failure. We created a stylized PES experiment for REDD+ in the field, where forest users from rural Ethiopia played Common Pool Resource (CPR) games. By setting different RLs, varying pay modality and revoking PES, we examined the effects of different treatments on forest conservation. The results of these Framed Field Experiments (FFEs) are presented in three papers of the thesis (see Chapters 3-5).

The paper exploring the effect of varying the RLs (REDD+ conditionality) is a novel contribution to the problem of setting RLs, which is a conundrum devoid of empirical justification owing to scant data from natural experimental settings. We found that setting the RL above historical average induces more conservation in general. However, the effect of the RLs is not the same for both pay types: an RL below historical average coupled with collective pay results in high reduction by triggering a sense of group achievement and possibly also activating pro-social preferences. A lower RL is also associated with higher cost efficiency in the form of a lower PES cost per tree saved.

The comparison between pay modalities (Chapter 4) shows that individual PES leads to relatively higher forest conservation. This is partly due to the obvious direct pecuniary incentive of individual PES and partly because of the uncertainty effect in collective PES; given group conditionality, individual forest users may undertake conservation but the aggregate effort may not still be sufficiently high to warrant payments.

Another conundrum of PES concerns how incentives affect intrinsic motivation to conserve forests. Though the empirical evidence is ambivalent, motivation crowding theory suggests that people may shift their frame and reasoning from altruistic, pro-social and pro-nature motives to pecuniary reward focus. This threatens the permanence of reduction realized if (or when) monetary incentives such as PES are removed. Despite the possibility to reverse or even escalate tree harvest following the withdrawal of a previously implemented PES scheme, I found, in the fourth paper (see Chapter 5), that PES might have crowding in effect on conservation, as forest users have reduced harvest, on average, after PES was removed.

This pattern holds both in individual and collective PES, but I also found suggestive results on heterogeneity of responses, for example, women are more likely to reverse their harvest when PES is removed.

Overall, spatial explicitness improves our ability to predict deforestation more accurately. This in turn aids the process and method of setting RLs, balancing the costs of 'too high' RLs and the disincentives of 'too low' RLs. This effect of setting RLs is not easily identified, but economic experiment is a cost efficient way of assessing the impacts *ex ante*. Generally, experimental evidence helps test how various designs of the PES interact with the local context.

Sammendrag

Et viktig element i gjennomføringen av en effektiv mekanisme for å redusere utslipp fra avskoging og skogforringelse (REDD +) er å sette referansebaner. Dette krever modellering og prediksjon av avskoging i et *business-as-usual* scenario. Denne avhandlingen ser på to aspekter ved utforming og implementering av en ordning med betaling for miljøtjenester (PES) for REDD +. Først anvendes romlige paneldata for å utforske drivkreftene bak avskoging i indonesiske distrikter. Det finnes ulike typer romlige modeller, og vi tester og velger vi den mest egnede, nemlig den romlige autoregressive (SAR) modellen. Vi fant at inkludering av lag av den avhengige variabelen ikke bare hjelper oss med å måle nabolagseffekter, men også forbedrer nøyaktigheten av estimater av andre strukturelle variabler som driver avskoging. Den sterke romlige avhengigheten innebærer at det kan være synergistiske gevinster i gjennomføringen av skogvernpolitikk. Avskoging er smittsomt, men vernetiltak kan ha positive bieffekter til andre områder, på samme måte som vaksinering har positive effekter også for de som ikke vaksineres.

Et annet sett av spørsmål knytter seg til utformingen av PES: hvordan påvirker ulike referansebaner graden av vern og kostnadene, hvem skal motta betalingen, og vil eventuelle incentiverte reduksjoner vare utover prosjektperioden. Disse spørsmålene er avgjørende for effektiviteten av PES, og ethvert forsøk på å hoppe over dem og starte direkte implementering av et verneprogram gir høy risiko for fiasko. Vi utformet et stilisert PES eksperiment for REDD + i feltet, der brukere av skog på landsbygda i Etiopia deltok i et allmenningsspill om felleseide skogressurser. Vi testet effekten på skogvern av ulike versjoner av PES, i form av ulike referanseverdier, mottakere av betalingen, og tilbakekalling av PES. Resultatene av disse felteksperimentene presenteres i tre av artiklene i avhandlingen.

Artikkelen som undersøker effekten av å variere referansebanen gir et nytt bidrag til litteraturen om problemet med å sette referansebaner, et tema som er lite undersøkt på grunn av mangelfulle data fra naturlige eksperimenter. Vi fant at referansebaner over historisk gjennomsnitt generelt gir økt vern. Effekten av de ulike referansebanene er imidlertid ikke den samme for begge betalingsformer: en referansebane under historisk gjennomsnitt kombinert med gruppe-betaling gir høy reduksjon i bruken av skog ved å utløse en gruppedynamikk for å nå felles mål og ved å aktivere deltagernes fellesskaps preferanser. En lavere referansebane er også forbundet bedre kostnadseffektivitet i form av lavere kostnad per vernet tre.

Sammenligningen mellom betalingsformer (Chapter 4) viser at individuell PES betaling fører til høyere skogvern. Dette skyldes delvis det direkte økonomiske insentivene ved individuelle belønninger og dels på grunn av usikkerheten ved kollektiv betaling; selv om den enkelte bruker reduserer uttaket kan samlet uttak fortsatt være for høyt til å sikre utbetalinger.

Et annet problemstilling ved PES handler om hvordan insentiver påvirker egen indre motivasjon for å bevare skogene. Empiriske studier gir motstridende resultater. Motivasjonkrympende teorier sier at økonomiske insentiver kan føre til skift i beslutningsrammen og tankesettet, med mindre vekt på altruistiske, pro-sosiale og miljømessige motiver. Dette kan bety at bruken av skog reverseres eller til og med øker når monetære insentiver som PES fjernes. Den fjerde artikkelen viser at bruken av skog er redusert også etter at PES ble fjernet, sammenlignet med bruken før PES ble innført. Dette mønsteret holder både for individuell og kollektive betaling, men jeg finner også at responsen varierer mellom ulike brukergrupper, for eksempel er kvinner mer tilbøyelige til å reversere bruken når PES er fjernet.

Samlet sett viser avhandlingen at romlige modeller forbedrer prediksjonene av avskoging. Dette hjelper igjen prosessen og metodene for å sette referansebaner, og balansere kostnadene ved for "for høye" og dis-insentivene ved "for lave" referansebaner. Effekten av å sette ulike referansebaner er ikke lett å identifisere, men økonomiske eksperimenter er en kostnadseffektiv måte for *ex ante* analyser. Generelt bidrar eksperimentelle studier til å teste hvordan ulike utforminger av PES samhandler med den lokale konteksten.

Chapter 1: Introduction

Incentivized forest conservation: spatial econometric and experimental evidence

1 Introduction

1.1 Motivation

Tropical deforestation is a complex phenomenon, which contributes to global warming, loss of biodiversity and other ecological services, soil erosion and land degradation, and loss of forest-based income to rural dwellers. Studies show that between 6% and 17% of anthropogenic CO_2 emissions come from deforestation (Baccini et al., 2012), with ca. 10% being the commonly used estimate (IPCC, 2013). This worsens the effect of the biggest global externality of our era, climate change, not only by increasing global warming but also by eroding the resilience and adaptive capacity of the ecosystem and people living adjacent to the forests. The rate of forest depletion is the highest in the tropics, where its impact is also the most severe.

For about a decade, reducing emissions from tropical deforestation and forest degradation, fostering conservation, sustainable management of forests and enhancement of forest carbon stock (REDD+) has taken a center stage in the global climate change negotiations. REDD+ has been promoted as a relatively low-cost, quick and simple way to mitigate climate change (Angelsen, 2008). As a notion, REDD+ appears to have won support and captivated attention among scholars, activists, NGOs and politicians alike, but its practice has not yet seen promising stride forward (Angelsen et al., 2017). Lack of binding agreements and (therefore) reliable financial sources, for example, from a global carbon market, has hampered implementation. At national and local levels, REDD+ has been tied to other objectives such as poverty reduction, and different actors have conceived it differently (Angelsen et al., 2017). The divergence in the design and implementation of REDD+ necessitates questions about how various design issues influence its effectiveness.

REDD+ is a multilevel (international, national and local) suite of policies ranging from market-based incentive mechanism to land zoning (Angelsen, 2014). Corbera (2012) conceptualizes it as a giant international experiment in Payments for Environmental Services (PES). Designing REDD+ scheme in the form of PES has many challenges. PES involves a voluntary participation of buyers and sellers of a clearly defined Environmental Service (ES) and an agreed upon conditionality for performance measurement and payments (Wunder, 2005). When the PES system is for REDD+, additional challenges emerge. As Engel (2016) precisely put it, the devil in the design of an effective PES has been in the detail.

One of the key elements of the design stage of a typical PES contract in general and REDD+ mechanism in particular is setting a performance benchmark. The benchmark in REDD+, which is called forest reference (emission) level[1] (hereafter RL), is difficult to set, since it is

[1] Herold et al., (2012) note that reference emissions level (REL) is often used to refer to gross emissions from deforestation and forest degradation, and reference level (RL) refers to deforestation and

(based on) unknown or even unknowable future counterfactual values, i.e., the Business-as-Usual (BAU) scenario without any REDD+ policies. Future deforestation may be subject to social and economic shocks that are hardly predictable *a priori* and are thus fundamentally uncertain (Angelsen, 2008; Herold et al., 2012).

Estimating the BAU scenario and measuring the potential impact of different RLs are two different problems. The former has to do with modeling and understanding the BAU dynamics in the land use sector. There is huge literature on what factors drive deforestation (Angelsen and Kaimowitz, 1999; Busch and Ferretti-Gallon, 2017; Geist and Lambin, 2002), yet the results are inconclusive, if not ambiguous. The *ex post* impact of setting RLs at different levels and varying other terms of the REDD+ contract need to be analyzed at finer resolutions (local or agent level) in order to capture potential interactions between pre-existing behavior and the external policy parameters. This thesis touches on both sides of the challenges pertaining to RLs—how to set them and how they affect conservation behavior—by using spatial panel data econometrics and framed field experiments, respectively.

1.2 Objectives

The general objectives of this thesis are twofold. The first part is an attempt to understand the nature of deforestation, particularly its predictors. We explore the spatial interaction effects of deforestation at sub-national levels so that we can draw important lessons about potential inter-district *leakage* while implementing REDD+. In the second part, we design framed field experiments, which resemble a typical PES scheme for REDD+, to understand behavioral responses of forest users to variations in three different PES design issues (and ultimately: REDD+ policies). First, we took group level historical average forest extraction in a baseline game (i.e., without PES) and set one of three different RLs: at, above or below the historical average, to compare their effect on forest conservation and cost efficiency. These provide insights about varying the conditionality of PES and ensuring *additionality*. Second, we randomly assigned groups to two pay modalities, where the specific objective is to explore whether individual or group based pay leads to more forest conservation. We test for uncertainty effect of group pay. Finally, we look at motivation crowding and forest conservation vis-à-vis the external incentives and the *permanence* issue in REDD+, particularly when the reward is revoked, and whether there is heterogeneity in conservation crowding.

2 Climate change and tropical deforestation

The science of climate change is well established. Starting from the nineteenth century, scientists have identified the events that lead up to climate change, and more recently the possible consequences. Stern (2008) gives a brief outline of this science. Anthropogenic activities produce and emit greenhouse gases (GHG), which accumulate in the atmosphere. Carbon is the most important part of the equation—CO_2 is the main agent (accounts for three-quarters) of anthropogenic global warming. When the stock of GHGs exceeds the

forest degradation (REDD), enhancement of carbon stocks, sustainable forest management and forest conservation (i.e., the + after REDD).

assimilation capacity (through the carbon cycle) of the earth, the GHGs absorb heat and cause global warming. The climate changes due to global warming, i.e., the distribution of heat, wind and precipitation. Climate change is a global externality and poses the biggest threat to humanity and life on earth. Cognizant to this threat, there have, since the establishment of the UN Framework Convention of Climate Change (UNFCCC) in 1992, been negotiations to keep average global temperature increase below 2 (1.5) degrees.

As the sources of GHGs are generally the production and consumption activities of our economies, the solution to climate change should come from all sectors. The Intergovernmental Panel on Climate Change (IPCC, 2014) identified a number of key mitigation technologies and practices ranging from improving energy efficiency, changing lifestyles to reduced deforestation, afforestation, reforestation. Some of these measures have to be taken in sectors which emit GHG (e.g., energy, industry, transport) while others are policies that enhance the resilience of the ecosystem (e.g., forestry, agriculture) (IPCC, 2014).

While efforts in improving the efficiency of other sectors are crucial, forests also provide the biosphere with a natural shield against climate change, as they absorb CO_2. Saving tropical forests and enhancing their carbon storage capacity have gained significant attention since the inception of REDD+ at COP11 in Montreal (den Besten et al., 2014). This has central place in part of the Paris Agreement (UN, 2015). Tropical forests cover approximately 10% of the earth (Corlett, 2016), and extend mainly across South America, Africa and Asia. Despite their contribution to keep the climatic balance, tropical deforestation has been so high that it accounts for about 10% of global GHG emissions (IPCC, 2013). Besides its high mitigation potential, conserving forests is—at least in principle—considered easier, quicker and less costly way of climate change mitigation (Angelsen et al., 2012).

Reducing deforestation is easier said than done. Deforestation is a complex phenomenon that occurs in different places, at different scales and for different reasons. Globally, there is no consensus about the drivers of deforestation. Tropical deforestation occurs due to subsistence agriculture and shifting cultivation in Africa (Ickowitz, 2006), while large-scale commercial agriculture causes most deforestation in the Amazon (Simon and Garagorry, 2005). Prices of palm oil, domestic transmigration and decentralization policies increased deforestation in Indonesia since the 1980s (Tsujino et al., 2016). The drivers of deforestation that can be identified also depend on the scale of analysis. As a conceptual framework, studies distinguish between *immediate causes* and *underlying causes* of deforestation (Angelsen and Kaimowitz, 1999; Busch and Ferretti-Gallon, 2017; Geist and Lambin, 2002).

Even after decades of studies on drivers of deforestation, we do not know enough yet. This is partly because of lack of data, and in most part due to the spatio-temporal heterogeneity and dynamics of the problem. Locally, there is a tradeoff between personal benefits from deforestation (more agricultural land and hence more production) and public benefits from conservation (improved water quality, soil conservation). This represents a classical social dilemma and collective action problem. When carbon and the global effects of deforestation are added, this dilemma and conflict between individual and collective rationalities is enlarged further. The local farmer in a remote village in Ethiopia does not have sufficient private incentives to conserve forests, even if standing forests provide *some* benefits to him. To resolve this dilemma, the international climate change negotiations begot a mechanism whereby the Ethiopian farmer–and others–will be incentivized and compensated for any extra efforts they make to increase the sequestration and storage of carbon in forests.

3 REDD+

The parties to the UNFCCC agreed in 2005 to consider the topic of Reducing Emissions from Deforestation (RED) in their agenda (Angelsen and McNeill, 2012; den Besten et al., 2014). This idea was then expanded in 2007 to include forest Degradation in developing countries and the conservation, sustainable management of forests and enhancement of forest carbon stocks, hence the term REDD+ (den Besten et al., 2014). Under the Clean Development Mechanism (CDM) of the Kyoto Protocol, conservation of tropical forests was not included (but tree planting in the form of afforestation and reforestation was). Initially, REDD+ aspired to generate funds from a global carbon market in a similar flexible offsetting mechanism as the CDM. These funds would then be used to incentivize forest users in developing countries to undertake more pro-environmental actions. Viewed this way, REDD+ is a Market Based Instrument (MBI).

However, the idea and practice of REDD+ have evolved significantly since its first inception (Angelsen et al., 2017; Angelsen and McNeill, 2012; den Besten et al., 2014). Globally, carbon markets did not emerge easily, and the attempt to reach a global climate change agreement has seen different gestures. Notable among the latter is the recent move (to withdraw from the Paris Agreement) by the US president Donald Trump. Lack of a binding international agreement with binding national emission caps have prevented the development of a global carbon market, making realizing (the initial idea of) REDD+ extremely difficult.

Against this backdrop, Angelsen (2016) notes three key changes that REDD+ has undergone over the last decade or so. First, it has evolved to include multiple objectives, in addition to carbon sequestration and storage. Second, the primary sources of funding are international donors instead of the initially envisioned global carbon markets. Even in the face of weak international cooperation, REDD+ has been practiced at different scales with the help of bilateral agreements (in particular, Norway providing large funds to Brazil, Indonesia and Guyana). Third, it has become a multilevel policy arena encompassing not only PES but also other conventional domestic policies such as command-and-control policies (e.g., protected areas) and traditional agricultural policies (Angelsen, 2016).

As a multi-objective and multilevel system, PES for REDD+ (or REDD+ as PES) (Pagiola, 2011) faces numerous challenges at different levels. Besides lack of global coordination, it is not clear how to create terms of contracts and institutional infrastructure, at local and national levels, which can reconcile the interests of various stakeholders. Essentially, the question is how to make the mechanism satisfy the incentive constraints of parties, i.e., the buyers of the ES want to ensure effectiveness at the least cost possible, while the sellers of the ES are at least as well off. Since REDD+ is taking the form of a giant global experiment on PES (Angelsen, 2014; Corbera, 2012), most of the issues of designing an effective REDD+ mechanism can be drawn from the PES design and best practices guide.

In this thesis, the focus is on REDD+ as a PES scheme, and to deal with some of the most important PES design issues. PES for REDD+ is an incentivized or result-based REDD+, where the environmental service is carbon sequestration and storage and incentives are conditional on the delivery or some close proxies that ensure the delivery, such as adoption of sustainable land use (Engel, 2016). The conditionality of REDD+ takes the form of an RL, which has to be high enough to induce compliance and low enough to ensure additionality.

4 PES design issues

PES is defined and conceptualized in various ways, and the literature that does so is huge. The most widely adopted and commonly cited definition comes from Wunder (2005), who characterizes PES as: (1) a voluntary transaction involving (2) a clearly defined environmental service, where (3) (at least one) buyer pays (4) (at least one) service provider (5) conditional on delivery, or input that ensures delivery, of the specified ES. Though this definition consists in five elements, there may be more aspects depending on how precise one wants to be. Without loss of the essence, the conceptual framework adopted here (Figure 1) considers four pillars of the PES mechanism.

The first pillar is the ES, which can be water quality, carbon sequestration, biodiversity or protection of soil erosion. Without a conceivable ES, there is no justification for the existence of a PES mechanism—ES is a necessary condition, but it is not sufficient, as (a) the ES may not be suitable for PES, or, (b) other considerations and the remaining pillars must be fulfilled.

The second pillar is the service provider (or seller). In the context of REDD+, sellers of the ES (carbon sequestration) are usually forest-dependent livelihoods in developing countries whose direct benefits from the ES are lower than from other land uses (Angelsen, 2014). These are decision-making agents having distinct objectives, constraints, characteristics, preferences and perceptions.

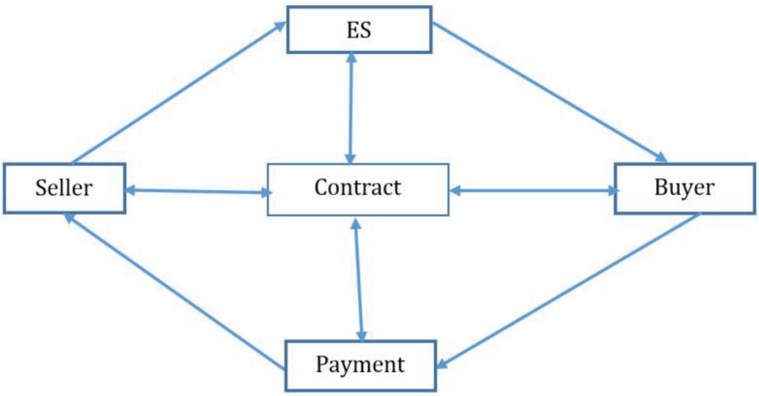

Figure 1: Conceptual framework and design issues of PES mechanism
Source: Adapted from Brenes et al. (2016) and Angelsen (2014)

Third, there should be buyers of the ES who benefit more from the ES and are willing and able to compensate the sellers at least for the cost of providing the service. Engel (2016) argues that the buyers could be direct beneficiaries of the ES or an intermediary body—the government, for instance—paying on behalf of the beneficiaries. These are, therefore, agents on the other (demand) side of the equation, and they have, likewise, their own distinct objectives, constraints, preferences and bargaining power.

A mirror image of the ES is the payment (or reward) that the two agents agree upon when entering the contract. The contract is, therefore, a platform, which brings the two agents together and enables them to bargain and agree on terms and conditions, exchange ES for payments and (if necessary) sanction one another. The contract is a key element of the PES mechanism for REDD+, but it exists only if the four pillars exist. The contract specifies the level of ES to be provided and the reward (amount, modality, duration etc.). These details in turn affect how both parties will make their strategic decisions, including the seller to comply and the buyer to monitor.

The conundrum in the design of an effective PES system for REDD+ stems from the multidimensional nature of this system as well as standard moral hazard and asymmetric information issues. When the objective of the PES program is as big as contributing to climate change mitigation, issues like *additionality*, *leakage* and *permanence* are crucial to the buyer (Angelsen et al., 2013; Engel, 2016). Both the buyer and the seller of the ES face varying sources and degrees of uncertainty. The former suffers from information asymmetry and may face moral hazard problem. The incentive constraints of the agents must be taken into account. Meanwhile, the type of the ES, its measurement, dynamics, value and cost of provision on the one hand, and the amount, mode (cash or in-kind) and type (individual vs. collective) of payment have to, at least in principle, be specified. From the buyer's perspective, the objective is effectiveness and (cost) efficiency, which depend on a number of crucial design issues and considerations (see Engel (2016) for a comprehensive list and discussion). This section highlights four issues, which stand out in the context of REDD+ and to which the papers in this thesis contribute: *leakage*, *additionality*, *permanence*, and *pay type*. The first three were discussed as the key challenges of REDD+ from its early days.

4.1 Leakage

Leakage is also referred to in the debate as "displaced emissions", i.e., an increase in emissions outside the project or program boundaries due to the reduced emissions inside those boundaries. In a general sense, leakage can be positive—the PES scheme produces more benefit than stipulated in the contract, or negative—economic activities that the PES intends to restrict migrate to places where there is no PES and at least part of the ES gained through the program is reversed elsewhere (IPCC, 2014). One way of minimizing this effect is to understand the prospects for leakage prior to creating a PES contract. When targeting potential sellers of the ES, the buyer can explore the possible ways leakage could occur.

Two categories of leakage (in its negative sense) have been identified. First, *primary leakage* occurs when the activity that used to reduce the ES shifts to locations beyond the boundaries of the project area and ends up offsetting the gains in ES realized by the PES program. Second, *secondary leakage* may occur as a result of changes in market signals (such as increased prices for agricultural commodities due to shortage of supply caused by the efforts to reduce deforestation) (Atmadja and Verchot, 2012; Aukland et al., 2003). Targeting and delimiting the spatial scope of the PES should, therefore, be part of the exploration phase of the design process (Sattler and Matzdorf, 2013). Paper I (Chapter 2) deals with the spatial interaction effects of deforestation at district level in Indonesia, and thus sheds light on the potential for positive leakage.

4.2 Additionality

Additionality simply refers to the emission reductions being beyond what would have occurred if the project or activity had not been undertaken. In other words, a climate policy—to be additional—should compensate users only for emissions below their BAU levels. The question of additionality is, therefore, closely linked to the question of setting the RL.

The UNFCCC, under Decision CP.16/1/Add.1/par.71, demands developing countries, which voluntarily want to participate in REDD+, to develop a national action plan and RLs along with transparent monitoring system (Sandker et al., 2014; UNFCCC, 2010). A stepwise approach (Herold et al., 2012) is being used for two main practical reasons: lack of data at all scales of analyses and low implementation capabilities of developing countries. Using spatially explicit socioeconomic models may improve the accuracy of predictions (Angelsen et al., 2013). The spatial interaction effects may also indicate what types of domestic policies can harness the synergies in working together with neighboring administrative units. Paper I (see Chapter 2) intends to aid this process by accounting for the spatial interaction effects of deforestation among neighboring districts in Indonesia.

Many of the REDD+ countries that have submitted their RLs to the UNFCCC are taking historical average emission levels as their baseline[2]. Emissions from forest loss may show different trends over time, depending on which stage of the forest transition (FT) a country is (Angelsen and Rudel, 2013; Culas, 2007). High historical forest loss which would be followed by low forest loss even without any policy will lead to ‘payment for nothing’ when historical average is used as a crediting baseline (Engel, 2016). In principle, REDD+ design should ensure additionality by identifying and saving the tree that would inevitably be cut down and the land that would actually be converted to other uses than forest.

At the project level, in order to agree on a certain level of ES, both parties have to determine the future counterfactual emission levels in a business-as-usual (BAU) scenario. They use these to set a benchmark acceptable to both. This benchmark (the RL) is an indispensable element of the contract and it sets the conditionality, identifies outcomes, which qualify for payment, and determines the amount of payment (Angelsen, 2008). The underlying idea of PES is that changing behavior of resource users by compensating them for changing their land use patterns will make additional ES, which would otherwise not be, available (Börner et al., 2017). This presupposes the existence of such a potential, which the program can tap into. A good REDD+ design requires setting RLs at the point beyond which the resource user will find it too costly to reduce emission without external incentive to cover the opportunity cost (Angelsen, 2008; Engel, 2016; Wunder, 2005). This ensures additionality in that each money unit spent fetches new units of ES.

In practice, a precise BAU cannot be established due to the high uncertainty, and there are costs of both setting the RL too high or too low compared to the “true” BAU. In addition, there might be political or popular pressure to setting the RL higher than the BAU. These issues are explored further in Paper II (Chapter 3 of this thesis), which experimentally explores the effects on effectiveness and costs efficiency of setting the RL higher or lower than the historical average (the most common formula used in RL-setting).

[2] http://redd.unfccc.int/submissions.html?topic=6 accessed 10.10.2017

4.3 Payment

As one of the pillars of any PES system, payment constitutes choosing the level of payment (price of the ES which may be based on cost of provision), the pay mode (cash or in-kind) and pay modality (individual or collective). Focusing on the last element, deciding whom to pay should begin with a unit of control that makes for an easier and effective contract (Engel, 2016). Some environmental services are by their very nature best suited for group-based contracts and thus group pay may be more suitable. Unfortunately, the classic free riding problem might undermine performance in group-based pay if there is no strong social capital within the group or individuals do not have strong intrinsic motivation and pro-social preferences. The empirical literature on the performance differential between individual and group pay is inconclusive (Midler et al., 2015; Narloch et al., 2012). Paper III (Chapter 4 of this thesis) contributes to this debate by exploring both direct pecuniary and potential uncertainty effects of being in a group pay with group level conditionality but unobserved individual behavior.

The differences between individual and collective payment goes beyond simple payoff calculations. There is substantial evidence in behavioural economics that the resource users are not (only) selfish profit-maximizing *homo economicus* as assumed in neoclassical models. The effects of rewarding individuals vs. rewarding groups may result in different performance depending on the strength of social preference and trust (Cardenas et al., 2000; Narloch et al., 2012). Moreover, the institutional setup may in itself activate or strengthen social preferences, and we find evidence consistent with this in Paper II (see Chapter 3).

4.4 Permanence

Another critical issue of PES is protecting an ES produced, delivered and paid for at one point in time from being reversed after termination of the payment. According to Engel (2016), permanence may be viewed as a question of avoiding leakage in time (i.e., making the distinction between temporal and spatial leaskage, with the latter being what is normally referred to as leakage). In the context of REDD+, the carbon storage services must be permanent, as the value of just postponing emissions is limited. Paying the ES providers forever can both economically and politically not be an option (Dutschke and Angelsen, 2008).

One way of addressing this problem in developing countries is to combine PES schemes with anti-poverty interventions to help transform and permanently shift livelihoods from natural resource dependence to other sustainable means. Once the PES scheme stops paying the ES providers on this and other pretexts, it is not self-evident whether people will fall back to old habits or stick to their newly acquired sustainable behaviors. There is increasing evidence and convincing warning that revoking PES, which may have shifted people's frame of thinking towards money instead of moral concerns, may, under certain circumstances, exacerbate resource depletion (Kaczan et al., 2016; Rode et al., 2015). Paper IV (Chapter 5) presents results from a field experiment along this design.

Table 1 relates these big PES design issues to the contributions made by the papers in this thesis.

Table 1: Contribution of the papers in this thesis to four big design issues of PES systems.

Papers	PES issues	Explanation and findings
Paper I	Leakage	The converse of the contagiousness of deforestation is that districts may reap synergistic gains from conservation efforts. Leakage is not a problem.
Paper II	Additionality	The conditionality (RL) of PES determines the degree of additionality. By setting different RLs, this paper compares their effectiveness and cost-efficiency. In general, high RLs are costly.
Paper III	Pay modality	Individual vs. group pay is an important issue that affects the effect of the PES on the motivation of participants. Individual pay performs better, with some caveats.
Paper IV	Permanence	Does the change in behavior due to PES last beyond the program period? The paper suggests 'yes'.

5 Methods

The issues in the design of PES for REDD+ make a tandem with logical sequence where deliverables in one stage serve as inputs in subsequent stages. Before implementing REDD+, the main question is how to design a contract that ensures compliance of the seller and least cost to the buyer of the environmental service (i.e., carbon sequestration). In the context of deforestation, how much reduction to expect (or how much deforestation to use as a benchmark) is an integral part of the terms of the contract. This necessitates that deforestation in the future with no policy, be forecast. Paper I revisits the old question of deforestation drivers and employs spatial econometrics analysis, to explore the predictors of and spatial interactions in deforestation in Indonesian districts.

Even when a contract has been designed, i.e., reference level set and performance based marginal incentives determined, whether we reach our target—of inducing reductions in emissions that could have significant contribution to climate change—will depend on the compliance of forest users. The *ex post* outcome is hard to foresee beforehand without conducting an appropriate experiment, if not a pilot. To circumvent this and as a second vantage point, we create framed field experiment (FFE) to mimic the REDD+ project life cycle and evaluate the impact of three dimensions along which the incentive system could vary and differently affect the propensity to conserve forests.

These approaches are necessitated by the nature of the problems we want to address. To the extent that these different phases form a feedback loop, the two general methods, i.e., the spatial panel and experimental approaches, are complementary approaches.

5.1 Spatial econometrics analysis

Many spatial econometrics studies allude to the first law of geography, due to Tobler (1970, p.4): "everything is related to everything else, but close things are more related". The nature of this closeness could be in the form of spatial proximity and neighborhood or social networks. In the context of deforestation at subnational jurisdictions, the relevant spatial interaction will be among spatially closely situated units, such as districts which share a border.

The need for spatial econometric modeling of deforestation stems from the two main merits of doing so. First, a conventional non-spatial model of deforestation drivers that omits spatial lags of relevant variables on the right hand side is likely to suffer from omitted variable bias (Kostov, 2013). Second, when the objective of modeling deforestation is prediction, estimating the parameters which capture the interaction effects is important in and of itself (Elhorst, 2014).

Studies show that drivers of deforestation may differ across regions, among subnational units or over time (Brun et al., 2015). When this is the case and if the main intention is predicting deforestation, using spatial lags of deforestation may be crucial. Predicting deforestation by using all possible drivers in all possible spatial scales requires much more data than is required in the spatial models.

5.1.1 Spatial interactions

The nature of the neighborhood effects that can be explored using spatial models is not limited to spatial units situated next to one another, but also social and other networks involving a web of nodes connected in some way. Suppose there are N geographical units. Let Y be an outcome of interest, which is potentially determined by a number of factors specific to the units or to a particular point in time or both. One way of modeling the relationship between Y and a host of other factors, say X, will be to put an equation with Y_i on the left and the X_i's on the right. Simply stated, the outcome variable of unit i is a function the characteristics of the same unit i.

The choice between spatial and non-spatial models depends on what X consists of. If X includes some or all of the X's and/or the outcome variable of neighboring units to i, then we call it a spatial model. The justification for this is that spatial units interact with their neighbors, and they do so in systematic ways, which, if we do not take them into account, may create bias in the estimates of the parameters of the X_i's.

Elhorst (2014) distinguishes among three types of spatial interactions among neighboring units. These are summarized in Table 1. While fitting a spatial econometric model to a given data, one may consider any combination of these interaction effects. For example, if the spatial interaction effects occur through both the endogenous and the exogenous variables, the Spatial Durbin Model (SDM) must be used. The other variants and corresponding statistical tests to distinguish them are discussed in Chapter 2, where we used the Spatial Autoregressive (SAR) model.

Table 2: Three types of spatial interactions.

Nomenclature	Spatial dependence
Spatial Autoregressive	Dependent variable Y of unit i $\leftrightarrow$ Dependent variable Y of unit j
Spatial lag X model	Independent variable X of unit i $\leftrightarrow$ Independent variable X of unit j
Spatial error model	Error term ϵ of unit i$\leftrightarrow$ Error term ϵ of unit j

Note: the $\leftrightarrow$ symbol indicates the interaction effects go both ways. If districts i and j are neighbors, then i's characteristics affect j and vice versa.

Source: Adapted from Elhorst (2014).

5.1.2 Spatial weighting matrix (W)

The spatial weighting matrix (W) represents the underlying assumption about the nature of the spatial dependence among neighboring units. It generates weighted averages of the lagged variables or error term such that closer units get more weight. As Kostov (2013) puts it, W has two main functions. First, it shows the geographical location of each spatial units relative to other units in the sample (Elhorst, 2014). Based on this, units that are linked could be identified. This is usually shown by the elements of $W = w_{ij}$ where i and j are two spatial units and $w_{ij} = 1$ if they are linked, and $w_{ij} = 0$ otherwise. Second, W determines the strength of the interaction between the neighboring units. There have been inconsistencies between the theoretical guidance and empirical practice in the choice of W (Getis, 2009). The former alerts scholars to take utmost care when choosing W, because using the wrong W may make inexistent spatial dependence look significant, or existing spatial interaction appear insignificant or existing effect inaccurately estimated (Kostov, 2013).

The empirical practice has been confined to only a few common types of W: either contiguity based matrices—with varying degrees of contiguity, i.e., taking just neighbors or neighbors of neighbors etc.—or distance based weights where the effect decays as distance increases (Elhorst, 2014). In any case, there is no rule of thumb as to how to pick a particular type of W. The researcher has to speculate about the most plausible nature of interaction based on predetermined criteria (Getis, 2009). The spatial econometric analysis in Chapter 2 is based on a first-order contiguity matrix, but distance based weights were also generated to make robustness checks.

5.1.3 Study area: Indonesia

Tropical forests in Indonesia—the third most populous country in Asia—are rich in biodiversity, but they are threatened by activities within and outside of the forest sector, such as logging and commercial agriculture respectively (Angelsen et al., 2013). Indonesia, being one of the most active REDD+ countries, has submitted its RL to the UNFCCC and also has a large number of REDD+ pilot projects promoted by NGOs and bilateral agreements between Indonesia and donors like Norway.

According to Tsujino et al. (2016), different factors drove deforestation in Indonesia at different times. For instance, export-oriented log production and cultivation of rice, magnified by global demands and growing population were the major causes of deforestation in the 1970s and 1980s. These gave way to the disproportionate global demand for Indonesian timber and oil palm, which, in turn, led to illegal or non-sustainable timber harvest and the expansion of permanent crop cultivation areas starting from the mid-1990s.

Until promising forestry policies were promoted after 2011, measures like the decentralization of the authority of the Ministry of Forestry (MoFor), following economic and political instability, are believed to have accelerated deforestation and forest degradation in Indonesia. Tsujino et al. (2016) noted that this enabled districts to issue small logging parcel leases, which resulted in the virtually uncontrolled harvest of remaining accessible lowland forests (Curran et al., 2004). Many districts saw forests as an easy source of financial revenue to be exploited rather than managed (Wollenberg et al., 2009).

Recent forest cover data by FAO forest statistics indicate that the forest area of Indonesia has declined from 118.5 Mha in 1990 to 91.0 Mha in 2015 (FAO, 2015). In addition to these temporal dynamics of the magnitude and sources of deforestation, there are regional differences worth capturing. For example, annual forest cover changes from 2006 to 2012 in Kalimantan(-0.78%) and Sumatra (-0.53%) were lower than in Sulawesi (0.79%), Java and Bali (0.73%), or in the other islands (0.29%).

Commercial agriculture such as palm oil plantation is an important driver of deforestation in Indonesia. In its reference level submission, Indonesia set its future emissions trajectory from deforestation at its historical average (UNFCCC, 2016b). Given the complexity of the dynamics in the forest sector, its interaction with other sectors and the multitude of other potential factors affecting land use patterns at different scales, future deforestation may not just be its own historical average. In the first paper of this thesis (see Chapter 2), we take advantage of the availability of relatively rich data to study spatial dynamics and spatial interaction effects of deforestation in Indonesian districts.

The rationale behind the use of a spatial econometric analysis is that these dynamics may be contagious in the sense that districts sharing borders with high deforestation districts may experience similar pressure. This may occur because either the same overarching drivers are at work in clusters of geographical units, or the outcomes in one district compel people in the other to do the same. There may also be idiosyncratic factors, which cause significant correlation among unobserved driving forces of deforestation.

5.1.4 Spatial panel data

Lack of data is one of the major challenges in studies of deforestation focusing on finer geographical units. Various remote sensing data sources are available at global level, but preparing those for subnational level econometric analyses is not easy, in part due to the lack of corresponding socioeconomic data. The data for the spatial panel econometric analysis in this study come from different sources. We obtained land cover and changes in land cover data from the Ministry of Forestry (MoFor) of Indonesia.

For the explanatory variables, the study relies partly on Indonesia's Database for Policy and Economic Research (DAPOER)[3] from the World Bank. Additional socioeconomic variables, e.g., population size, were also obtained from Indonesia's Central Bureau of Statistics (BPS).

[3] http://databank.worldbank.org/data/reports.aspx?source=1266

Climate variables—precipitation and temperature—were extracted from CRU TS3.21 available at the University Corporation for Atmospheric Research website [4].

5.2 Experiments

Economists have used experiments to elicit the response of an outcome of interest to an exogenously determined treatment while controlling the environment and thus keeping the effect of confounding factors minimal. Lab experiments, where subjects are selected mostly from undergraduate university students and asked to make economic decisions in a laboratory setting, were popular before economists moved the lab to the field (Anderies et al., 2011; Ostrom et al., 1994). Though experiments are useful in testing theories and identifying impact, many economists have reservations regarding their usefulness vis-à-vis generalizability and external validity (Camerer, 2011). Peter Bohm is believed to have conducted pioneering experiments in the field in the 1970s, and by many considered the father of experimental economics (Dufwenberg and Harrison, 2008). Since then, there has been an increasing interest in and growing practice of field experiments with more relevance added to the subject pool, the commodity and even the information set.

Experiments can be used to evaluate the potential impact of a policy *ex ante*, and as such save money that could have been squandered by scaling up a policy that would not work. The literature on experimental economics is rich, and there are different kinds of experiments, such as lab experiments, artefactual field experiments, framed field experiments (FFE) and natural experiments (Harrison and List, 2004). We used FFE to tackle questions raised in three of the papers (II-IV) in this thesis (see Chapters 3-5).

5.2.1 Framed field experiment (FFE)

FFEs differ from lab and artefactual experiments in the type of subject pool they take, the nature of the commodity they consider, and the place of the experimental sessions. FFEs are conducted in the place where subjects make actual economic decisions in their daily lives. The commodity is framed such that it is made as close to reality as possible. When the subjects are recruited from the population of actual decision makers, they bring their real-world experience with and expectations about other participants (Harrison and List, 2004; Levitt and List, 2009). These are important factors, which enhance the representativeness of the overall experimental design to the real world decision problem being studied as well as the validity of the conclusions thus drawn.

FFEs in resource and environmental economics are common and useful. Since Ostrom et al. (1994), there is increasing momentum and rigor in the use of field experiments in the study of common pool resources. Among the earliest FFEs on natural resource management issues is Cardenas et al. (2000) from rural Colombia, who showed that regulation crowds out other-regarding behavior. Later examples of relevant FFEs include: Rodriguez-Sickert et al. (2008)—also from Colombia—who found that externally imposed 'institutions influence social preferences'. The results from an FFE conducted by Vollan (2008) in Namibia and South Africa also corroborate the fact that external regulations may undermine good

[4] https://climatedataguide.ucar.edu/climate-data/cru-ts321-gridded-precipitation-and-other-meteorological-variables-1901

practices of the local communities. More recently, studies by Narloch et al. (2012) as well as Handberg and Angelsen (2015) have explored the effects of varying pay modality (individual vs. collective in the former) and general conservation policies (i.e., command-and-control, PES and community forest management in the latter). Drawing on the FFE that Handberg and Angelsen (2015) conducted in Tanzania, we took the framing a step further and presented the subjects with real tree branches.

5.2.2 Study area: Ethiopia

In its global forest resources assessment, FAO (2010) estimated the total forest cover in Ethiopia to be around 13 million ha (11.4% of total land area). With its population estimated over a 100 million and its fast growing economy, it is expected that the business-as-usual scenario will put a lot of pressure on Ethiopia's existing forests. The government has shown its commitment to embark on a green and climate resilient development path (FDRE, 2011), as part of which Ethiopia has become a member of the UN-REDD countries since 2011 and has submitted a readiness preparation proposal (R-PP) (Gonzalo et al., 2017). It has also established some institutional structure in the form of a national REDD+ secretariat that coordinates efforts to a full-scale participation (Bekele et al., 2015), and a specialized Environment and Climate Research Center (ECRC)—with the support of Environment for Development (EfD) initiative—which supports the ongoing efforts to realize green growth.

However, limitations in implementation capacity along with limited knowledge on how incentivized forest conservation policies work might leave many caveats in the actual implementation of and level of success in these policies. Modeling deforestation in Ethiopia at subnational levels is difficult as data is hardly available at this scale. The present studies focus on the behavior of smallholder and natural resource-dependent farmers who use forest products for different purposes. A study such as ours is highly relevant for Ethiopia, as it is one of the REDD+ countries which have submitted their reference level to the UNFCCC (UNFCCC, 2016a) and also has become an important REDD+ partner country for Norway.

The major sources of deforestation and forest degradation in Ethiopia are linked to cropping, fuelwood, charcoal, logging, and livestock. By participating in REDD+, Ethiopia aspires to, and can potentially, reduce emissions and preserve biodiversity (co-benefits). The challenges include, but are not limited to, high natural resource dependence, high and increasing population pressure, poverty and low capacity for implementation. At a finer resolution, culture and existing social and behavioral aspects may also be important factors, which, together with the economic incentives, affect behavior towards forest conservation.

Ethiopia's ambitious plan of achieving and sustaining green growth needs to be backed by research. Three papers in this thesis are directly relevant for Ethiopia while the paper on spatial panel analysis from Indonesia may still guide future attempts to use spatially explicit models of deforestation in Ethiopia's administrative zones and districts.

5.2.3 Design and framing

The experimental papers of this thesis use data from a FFE conducted between February and June 2016. We undertook the experiments in forest-rich villages in Northern Ethiopia.

The experiments were designed with three objectives in mind. First, we compare the effect on conservation and on cost of setting different forest reference levels above, equal to or below the historical average forest harvest. Second, we explore performance (forest conservation) under two PES modalities: individual and group pay. Third, we assess and characterize the response of participants to the termination of a PES program.

The three objectives required different treatments. We used a two by three design to allocate observations in each cell. The experiment was done in two or three stages—each lasting for five rounds—depending on whether the subsample was used to assess the third objective. The first, pre-treatment stage is common to all, and sought to observe baseline levels of tree harvest in a typical common pool resource (CPR) game. The purpose of this stage is to get historical data to set the RL in the subsequent stage, and as a reference to compare the impact of the treatments (within group design). In the second stage, subjects were informed about the RL and the PES incentives. Only the subsample, which were given their historical averages as RLs in the second stage played the game for one more stage (five more rounds); this time with the PES removed and the payoff structure thus identical to the first stage.

The treatments and distribution of subsamples are summarized in Table 3.

Table 3: Assignment of subjects into different treatment groups and number of rounds in different sessions.

Group	FRL	Pay	Sessions	Participants	No. of rounds		
					Pre-PES	PES	Post-PES
1	Above	Individual	8	64	5	5	-
2		Group	8	64	5	5	-
3	Historical	Individual	11	88	5	5	5
4		Group	11	88	5	5	5
5	Below	Individual	8	64	5	5	-
6		Group	8	64	5	5	-

5.2.4 External validity

Taking the lab to the field and—as a novel feature—framing the commodity to represent real tree branches increase the external validity of the results. External validity is one of the challenges related to the use of experiments in economics (Lusk et al., 2006). FFEs are by design meant to have more external validity than the other forms of experiments or artefactual experiments. We used tree branches to make the product real, and selected *kebelles* (peasant associations) where there is forest within a walking distance from the *kebelle* center to help the participants relate the task to the decisions they make in their daily lives. We also selected different districts from the National Regional State of Tigray to represent communities with different socioeconomic backgrounds and agro-ecological endowments.

People are generally aware that cutting trees is illegal, but they also admit that they cannot live without forest products. The upfront interview questions about sales of fuelwood and charcoal demonstrated some reluctance to reveal the true forest use. Instead, we used average weekly visits to collect some forest product as a proxy for actual use and found stronger correlation than Harrison and List (2004) did in a comparable FFE in Tanzania. An interesting anecdote is that the participants thought they were given a training on forest conservation, and they made remarks that it showed them how they were destroying their forests.

In some villages, people asked if cutting some branches of a standing tree would be counted, but the working definition was that cutting a tree means cutting it from the its roots. When we investigate further, we realized that people would usually have a purpose in mind (e.g., making a yoke) before they decide to cut a tree, and they knew from experience that some could be fulfilled with selective cutting from the branches of a big tree. Tailoring the working definition to the daily use patterns in the community would increase external validity, which may require a design on forest degradation.

6 Main findings

6.1 Is there spatial spillover effect in deforestation among Indonesian districts? (Paper I)

This paper revisits the question of predictors of deforestation in Indonesia, with a focus on the potential benefits of spatial explicitness. The overarching motivation stems from the fact that deforestation has several adverse effects such as global warming, land degradation and soil erosion, loss of biodiversity and ecosystem services, and other indirect socio-economic imbalances (Portela and Rademacher, 2001). Many factors have been identified as drivers of tropical deforestation. Income (or lack thereof) and population density are among the most common ones (Boubacar, 2012), but it is also ascribed to structural factors, macroeconomic dynamics and specific incentive (price) changes (Hargrave and Kis-Katos, 2012; Wheeler et al., 2013).

Our revisit to the old problem is further justified by the facts that most existing studies use non-spatial models and might thus suffer from lack of accuracy and rigor, and second, that the focus on prediction (for RL setting) may justify new approaches. Paper I (Chapter 2 of this thesis) develops a spatially explicit model of deforestation in 190 Indonesian districts, and shows that the spatial explicitness aids to both estimate the spatial spillover effect of deforestation and obtain more accurate estimates for other explanatory variables. We also included forest cover and forest cover squared to test for forest transition (FT) hypothesis.

The panel data captures dynamics over time with an overall spirit along Hargrave and Kis-Katos (2012) and Boubacar (2012), while focusing on subnational scale and aiming to contribute to the design of a PES mechanism for REDD+. Part of the effort to fit a spatial explicit model is making a choice between non-spatial and spatial models, and we found that the latter fits the data at hand better. Among the variants of the spatial panel econometric models, we undertook a series of statistical tests and found that the SAR model is superior. The spatial interaction effect occurs through the lagged values of the dependent variable, suggesting that social planners ought to not only tailor their policies to district specific characteristics but also consider dynamics in neighboring districts. REDD+ programs at subnational levels can benefit from harnessing the synergy in conservation efforts, which spills over to neighboring districts. This is a good news in light of the problem of leakage in emissions from deforestation and forest degradation, as conservation efforts in one district can be expected to also reduce deforestation in neighboring districts.

6.2 How should RLs be set to increase forest conservation at lower cost? (Paper II)

Paper II (Chapter 3) explores how various forms of reference levels (RL) might lead to different levels of forest conservation and associated costs of doing business and why. To this end, we developed and implemented a framed field experiment (FFE) that portrays a realistic PES scheme for REDD+. We observed forest extraction in the first (baseline) stage, and then set three RLs in the second stage of the experiment: historical (pre-treatment) average, above historical average, and below historical average. This is, to the best of our knowledge, the first paper to report on experimental tests of the impacts of varying RLs on forest conservation.

This paper raises and attempts to tackle three specific questions: how do RLs affect forest conservation? Does this effect differ between pay modalities? Which type of RL gives the highest cost efficiency, i.e., largest reduction (avoided harvest) per Ethiopian Birr paid to the participants? Given the different RL treatments, the effects might be framed in terms of a positive incentive effect and a negative anchoring effect of higher RL. The results were nuanced: with individual pay, a higher RL increases conservation, while this is not the case with group pay – possibly due to the limited pecuniary incentives provided in group pay combined with the anchoring effect. When the RL was set below the pre-treatment (historical) average, we found no effect under individual pay, but a large and significant conservation effect under collective pay. We propose that this is due to the RL being interpreted as an 'aspirational target' for the group, which then collectively aimed to achieve it by reducing their harvest. And possibly also through the activation of pro-social preferences under group payment. The combination of group dynamics and the anchoring and normative effects outweighed the weak incentive effect of a low RL. In other words, our results give some support to claims of the PES design affecting to the extent social or environmental preferences are being expressed.

In terms of cost efficiency, our experiment demonstrated the high costs to the PES scheme sponsor of a high RL. This is simply due to the fact that the starting point for payment is higher (with some non-additional payments) as well as a generally weak net incentive effect of higher RL. This result points to a general problem in real-world PES design and implementation.

6.3 Should REDD+ pay individuals or groups to conserve forests? (Paper III)

Using data from the framed field experiment (FFE) we conducted in Northern Ethiopia, Paper III (Chapter 4) explores performance differential in individual and collective pay in terms of limiting forest use in a common pool forest resource setting. A total of 432 randomly selected smallholder farmers, who make forest use decisions in their daily lives, participated in the experiment. Besides taking the lab to the field, we framed the commodity to real dry tree branches to add relevance to the task (Harrison and List, 2004), and we found significant correlation between actual forest use and harvest in the experiment.

The exploration of an uncertainty effect with group pay in FFEs is a novel aspect of this paper. Payments are made in the experiment only if the harvest (individual or group average) is below a predefined benchmark, which resembles experiments of public goods provision with a threshold (Cadsby and Maynes, 1999; Dickinson, 1998), and mirrors a typical PES for REDD+ scheme. A forest reference level (RL) is set and used as a benchmark for program evaluation and financial transfers (Busch et al., 2009). We introduce three different RLs, creating different levels of uncertainty, as participants do not know whether the group will reach the RL and be eligible for payment.

We found that individual pay leads to more forest conservation. There are several explanations for this. Although forest users in Ethiopia are not pure *homo economicus* agents, the direct pecuniary incentives in individual pay provides a strong motivation. Besides lowering the marginal benefit to the individual of saving a tree, group pay opens up the possibility of free riding, and introduces uncertainty as to whether the aggregate group harvest will be sufficiently low to qualify for payment. Lowering own harvest for the common good might be squandered. Finally, group level pay demands that you trust the authorities to distribute payments in a fair way to all participants. Nevertheless, the difference is perhaps not as large as expected. Group PES does not change the participants' dominant strategy, which is still maximum harvest after the introduction of the treatment. Yet there is a substantial reduction, indicating significant pro-social preferences.

6.4 Does revoking PES crowd out motivation to conservation? (Paper IV)

The literature on motivation crowding theory has, so far, inconclusive results: there are reports of both motivation crowding out effect of economic instruments (e.g., Cardenas et al. (2000) and Kerr et al. (2012)), and there is crowding in effect (e.g., Rodriguez-Sickert et al. (2008) and Narloch et al. (2012)). Rode et al. (2015) also reported, in a comprehensive review of the literature, that many studies found no significant effect, i.e., neither crowding out nor crowding in. The study of motivation crowding is applied in diverse areas of business and nature conservation, and the findings are similarly inconclusive. Paper IV (Chapter 5) contributes to the debate on whether PES programs lead to motivation crowding in or crowding out. In a stylized PES scheme that resembles a program for Reducing Deforestation and forest Degradation (REDD+), we observed forest harvest in a baseline stage, introduced PES and finally removed it. The objective is to statistically test if there is evidence for motivation crowding out, or if the PES entices participants to consider lower harvest the norm, and this effect persists beyond the project period. We also varied the pay modalities (individual vs. collective) to compare if the effect might differ based on pay modality, and other sources of heterogeneity.

We found strong evidence that the PES increased motivation to conserve. Though average tree harvest went up in stage 3 relative to stage 2, it remained below the baseline case on average. This pattern is similar across pay modalities, but we observe a stronger tendency of crowding out among female participants. Paying the poor for environmental services could indeed create more environmental awareness and hence induce a lasting pro-nature behavior. As these conclusions are always context dependent, this provides another strong evidence on one side of the contentious issues of motivation crowding from the introduction (and removal) of monetary incentives.

7 Contributions, limitations and future research

7.1 Contribution

The questions addressed in this thesis belong to two broad strands of the literature. In the drivers' analysis, we draw on several spatial and non-spatial econometric models and seek to add some evidence as to how (if any) spatial explicitness could enhance our ability to predict deforestation. This contributes to literature on setting RLs for incentive-based conservation schemes, particularly to REDD+ initiative at the UNFCCC. Detecting spatial interaction effects is informative of the nature of potential leakage in REDD+. The second set of questions is about the behavioral aspects of the implementation of incentivized conservation policies. Both parts contribute to the tandem of activities—which go from understanding deforestation *ex ante* to the *ex post* evaluation of performance—involved in the design and implementation of REDD+.

The novelty of this thesis lies mainly in the experimental design of a stylized REDD+ scheme with RLs as treatments. While the results of each paper in the thesis are of significance in and of themselves, scholars and practitioners may benefit from following this approach and employing a combination of methods in accordance with the essence of the research questions they pose and scale of analysis they thus identify.

7.2 Limitations

This thesis embarked on addressing a globally significant and complex problem. One of the main challenges we faced was lack of readily available data. That pushed us towards questions, though as important as those we deferred, whose answers could be found from available data or from relatively manageable field experiments. This 'meta-limitation' cuased admittedly that some important research questions had to give way to those we have raised and addressed in this thesis. At a finer resolution, the questions we have posed in these papers are again admittedly likely to be subject to other sources of limitations. Using secondary data collected for a different purpose from what we aim at has limited our ability to include as many explanatory factors as theory dictates. When using district level aggregate data for various indicators, we risked accepting as much aggregation errors.

In the experimental data, we took utmost care of the factors we suspected would confound the behavior of participants and followed a strict and thoroughly developed research protocol. Yet, it is not realistic to claim that there were no noise. For example, low level of understanding, experimental fatigue and demand effects are potential problems while, in some cases, participants also looked strongly influenced in their daily interactions with dominant political principles. They sometimes went to the extent of mentioning the government in many of the post-experiment survey questions despite efforts to assure them that the survey had nothing to do with government programs in the study area and that their anonymity would be guaranteed. It is possible that some participants chose to harvest less for fear that their villages be viewed as less environmentally caring than other villages.

7.3 Further research

An important insight we observed from the field experiments is that participants had many practical questions about the RL as well as about the operational definition of deforestation. In the experimental literature focusing on natural resource management, involving participants in setting up institutions results in significant difference in performance. One way of doing so would be to let participants suggest and vote on alternative ways of setting the benchmark.For example, Gatiso and Vollan (2016) found that participatory procedures such as electing leaders and selecting rules endogenously are more effective than imposing them exogenously.

It is possible that the scale of the experiments will oversimplify the complex interaction between livelihoods and forests. I believe that large scale randomized control trials could be useful. Besides, using random RLs, as opposed to those based on historical averages in our case, may give additional insight. Replicating the experiments in similar settings is also another important future research agenda, especially for the effect of varying RLs as there is so far no, to the best of my knowledge, other experimental study on this.

8 Overall conclusion

Apart from the need to conserve biodiversity, protect soil erosion and land degradation, forests have now been viewed as the natural shield protecting planet earth from overheating. They do so by capturing and storing carbon both in their above- and below-ground biomass. To the extent that they can rescue humanity from potential catastrophe from climate change, they demand that we prepay their service by saving them first. How do we save our forests and even bring those we destroyed back?

Logically, any effort to reduce forest depletion should begin from understanding the causes of the problem. This in turn depends on the specific time and location one chooses to study as the drivers of deforestation vary across continents, regions, countries and even districts as well as over time. A more subtle, albeit important variant of this challenge also arises in the definition of causes of deforestation: whether one refers to *immediate causes* or *underlying drivers* points to different potential sets of factors. Yet another challenge is that studies of deforestation—what it is caused by and how to reduce it—are constrained by lack of available data in the right format at the right temporal span and spatial scale.

Despite these constraints and challenges, there have been numerous and significant advances in both the theoretical underpinnings and empirical evidence on the drivers of deforestation, how to predict it and what policy measures may help mitigate it. The most frequently mentioned drivers of deforestation include population pressure, poverty, both subsistence and commercial agricultural land expansion, fuelwood, urbanization, forest fires and so on. Among the mitigation efforts, REDD+ has gained utmost popularity among scholars, governments, NGOs and other institutions alike.

We draw the following conclusions and policy implications from both the spatial econometric analysis and experimental evidence.

1. **Deforestation is contagious, meaning leakage is not a (big) problem.** Deforestation spreads like a fire across borders of districts. The spatial spillover effects

included in our model may carry important information about the dynamics of deforestation in neighboring spatial units. This has an important implication for the implementation of sustainable forest conservation mechanisms as districts can harness the inter-district conservation synergy . That is, if deforestation in neighboring districts exacerbates the problems in a given district, the converse is that slowing it down (conservation) in one district will slow it down in nearby districts.

2. **Forest users have pro-social and/or pro-nature preferences.** Generally, the pre-treatment harvest rates were on average not more than 50% of the maximum allowable amount, which was also the Nash Equilibrium solution for each participant making their decisions simultaneously. This result, not unique to this study, is a significant lesson and stands in opposition to a view that only economic motives drive the actions and behavior of the poor.

3. **Effects of higher/lower reference level depends on the pay modality.** Alternative approaches for setting RLs have been suggested in different studies and in the UNFCCC decisions. In terms of performance, setting RLs above historical average leads to more conservation in general. When we control for the pay modality, however, the story changes. Setting RL below historical average leads to more conservation in group PES as it triggers a sense of group achievement to reach the benchmark, and possibly also activate pro-social preferences. An RL above historical average is more effective in individual PES where incentives are more direct, and individuals are certain to earn for any amount of reduction they choose below the benchmark.

4. **Generous (high) reference levels are costly.** Setting high reference levels lead to more conservation in general, but has a cost in terms of higher overall payment to the participants. The latter effect dominates, thus the costs efficiency (tree saved divided by total PES transfers) declines with higher reference levels.

5. **Individual pay yields more conservation.** Regarding the pay modality, results show that paying individuals rather than groups and then sharing, leads to more conservation. We argue that group pay both reduces the individual incentives to reduce forest use, and that there is an additional an uncertainty effect arising from the fact that the group target (reference level) might not be reached.

6. **No general crowding out of intrinsic motivation.** We observed baseline data without PES, introduced PES and eventually removed it to compare conservation patterns between the baseline and the post-PES. Results show that there is a general crowding in effect both during and after the PES program.

References

Anderies, J. M., Janssen, M. A., Bousquet, F., Cardenas, J.-C., Castillo, D., Lopez, M.-C., Tobias, R., Vollan, B., and Wutich, A. (2011). The challenge of understanding decisions in experimental studies of common pool resource governance. *Ecological Economics*, 70(9):1571–1579.

Angelsen, A. (2008). REDD models and baselines. *International Forestry Review*, 10(3):465–475.

Angelsen, A. (2014). The economics of REDD+. *Handbook in forest economics, London and New York: Routledge*.

Angelsen, A. (2016). REDD+ as result-based aid: General lessons and bilateral agreements of norway. *Review of Development Economics*.

Angelsen, A., Ainembabazi, J., Bauch, S., Herold, M., Verchot, L., Hänsel, G., Schueler, V., Toop, G., Gilbert, A., and Eisbrenner, K. (2013). Testing methodologies for REDD+: Deforestation drivers, costs and reference levels. Report, Ecofys.

Angelsen, A., Brockhaus, M., Duchelle, A. E., Larson, A., Martius, C., Sunderlin, W. D., Verchot, L., Wong, G., and Wunder, S. (2017). Learning from REDD+: a response to fletcher et al. *Conservation Biology*, 31(3):718–720.

Angelsen, A., Brockhaus, M., Sunderlin, W. D., and Verchot, L. V. (2012). *Analysing REDD+: Challenges and choices*. CIFOR.

Angelsen, A. and Kaimowitz, D. (1999). Rethinking the causes of deforestation: lessons from economic models. *World Bank Res Obs*, 14(1):73–98.

Angelsen, A. and McNeill, D. (2012). *The evolution of REDD+*. CIFOR.

Angelsen, A. and Rudel, T. K. (2013). Designing and implementing effective redd+ policies: A forest transition approach. *Review of Environmental Economics and Policy*, 7(1):91–113.

Atmadja, S. and Verchot, L. (2012). A review of the state of research, policies and strategies in addressing leakage from reducing emissions from deforestation and forest degradation (REDD+). *Mitigation and Adaptation Strategies for Global Change*, 17(3):311–336.

Aukland, L., Costa, P. M., and Brown, S. (2003). A conceptual framework and its application for addressing leakage: the case of avoided deforestation. *Climate Policy*, 3(2):123–136.

Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., and Friedl, M. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2(3):182–185.

Bekele, M., Tesfaye, Y., Mohammed, Z., Zewdie, S., Tebikew, Y., Brockhaus, M., and Kassa, H. (2015). *The context of REDD+ in Ethiopia: Drivers, agents and institutions*, volume 127. CIFOR.

Boubacar, I. (2012). Neighboring effects of deforestation: a spatial econometric approach. *Environmental Economics*, 3(3).

Brenes, E., Alpizar, F., Madrigal-Ballestero, R., and Kim, J. (2016). Bridging the policy and investment gap for payment for ecosystem services: learning from the costa rican experience and roads ahead. Report, The Global Green Growth Institute.

Börner, J., Baylis, K., Corbera, E., Ezzine-de Blas, D., Honey-Rosés, J., Persson, U. M., and Wunder, S. (2017). The effectiveness of payments for environmental services. *World Development*, 96:359–374.

Brun, C., Cook, A. R., Lee, J. S. H., Wich, S. A., Koh, L. P., and Carrasco, L. R. (2015). Analysis of deforestation and protected area effectiveness in indonesia: A comparison of bayesian spatial models. *Global environmental change*, 31:285–295.

Busch, J. and Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? a meta-analysis. *Review of Environmental Economics and Policy*, 11(1):3–23.

Busch, J., Strassburg, B., Cattaneo, A., Lubowski, R., Bruner, A., Rice, R., Creed, A., Ashton, R., and Boltz, F. (2009). Comparing climate and cost impacts of reference levels for reducing emissions from deforestation. *Environmental Research Letters*, 4(4).

Cadsby, C. B. and Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics*, 71(1):53–73.

Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to levitt and list.

Cardenas, J. C., Stranlund, J., and Willis, C. (2000). Local environmental control and institutional crowding-out. *World Development*, 28(10):1719–1733.

Corbera, E. (2012). Problematizing REDD+ as an experiment in payments for ecosystem services. *Current Opinion in Environmental Sustainability*, 4(6):612–619.

Corlett, R. T. (2016). The impacts of droughts in tropical forests. *Trends in plant science*, 21(7):584–593.

Culas, R. J. (2007). Deforestation and the environmental kuznets curve: An institutional perspective. *Ecological Economics*, 61(2-3):429–437.

Curran, L. M., Trigg, S. N., McDonald, A. K., Astiani, D., Hardiono, Y., Siregar, P., Caniago, I., and Kasischke, E. (2004). Lowland forest loss in protected areas of indonesian borneo. *Science*, 303(5660):1000–1003.

den Besten, J. W., Arts, B., and Verkooijen, P. (2014). The evolution of REDD+: An analysis of discursive-institutional dynamics. *Environmental Science & Policy*, 35:40–48.

Dickinson, D. L. (1998). The voluntary contributions mechanism with uncertain group payoffs. *Journal of economic behavior & organization*, 35(4):517–533.

Dufwenberg, M. and Harrison, G. W. (2008). Peter bohm: Father of field experiments. *Experimental Economics*, 11(3):213–220.

Dutschke, M. and Angelsen, A. (2008). *How do we ensure permanence and assign liability?* Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Springer.

Engel, S. (2016). The devil in the detail: A practical guide on designing payments for environmental services. *International Review of Environmental and Resource Economics*, 9(1-2):131–177.

FAO (2010). Global forest resources assessment 2010 - country report: Ethiopia.

FAO (2015). Global forest resources assessment 2015: Country report indonesia. Report, FAO.

FDRE (2011). Ethiopia's climate-resilient green economy, green economy strategy. Report, Federal Democratic Republic of Ethiopia.

Gatiso, T. T. and Vollan, B. (2016). Democracy and cooperation in commons management: experimental evidence of representative and direct democracy from community forests in ethiopia. *Environment and Development Economics*, 22(2):1–23.

Geist, H. J. and Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. *BioScience*, 52(2):143–150.

Getis, A. (2009). Spatial weights matrices. *Geographical Analysis*, 41(4):404–410.

Gonzalo, J., Zewdie, S., Tenkir, E., and Moges, Y. (2017). *REDD+ and Carbon Markets: The Ethiopian Process*, pages 151–183. Springer International Publishing, Cham.

Handberg, Ø. N. and Angelsen, A. (2015). Experimental tests of tropical forest conservation measures. *Journal of Economic Behavior & Organization*, 118:346–359.

Hargrave, J. and Kis-Katos, K. (2012). Economic causes of deforestation in the brazilian amazon: A panel data analysis for the 2000s. *Environmental and Resource Economics*, 54(4):471–494.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4).

Herold, M., Angelsen, A., Verchot, L. V., Wijaya, A., and Ainembabazi, J. H. (2012). A stepwise framework for developing REDD+ reference levels. *Analysing REDD+: Challenges and choices*, pages 279–300.

Ickowitz, A. (2006). Shifting cultivation and deforestation in tropical africa: critical reflections.

Development and Change, 37(3):599–626.

IPCC (2013). Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Report.

IPCC (2014). Climate change 2014: Mitigation of climate change. contribution of working group iii to the fifth assessment report of the intergovernmental panel on climate change. Report.

Kaczan, D. J., Swallow, B. M., and Adamowicz, W. L. (2016). Forest conservation policy and motivational crowding: Experimental evidence from tanzania. *Ecological Economics*.

Kerr, J., Vardhan, M., and Jindal, R. (2012). Prosocial behavior and incentives: Evidence from field experiments in rural mexico and tanzania. *Ecological Economics*, 73:220–227.

Kostov, P. (2013). Spatial weighting matrix selection in spatial lag econometric model. *Econometrics*, 1(1):20–30.

Levitt, S. D. and List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1):1–18.

Lusk, J. L., Pruitt, J. R., and Norwood, B. (2006). External validity of a framed field experiment. *Economics Letters*, 93(2):285–290.

Midler, E., Pascual, U., Drucker, A. G., Narloch, U., and Soto, J. L. (2015). Unraveling the effects of payments for ecosystem services on motivations for collective action. *Ecological Economics*, 120:394–405.

Narloch, U., Pascual, U., and Drucker, A. G. (2012). Collective action dynamics under external rewards: Experimental insights from andean farming communities. *World Development*, 40(10):2096–2107.

Ostrom, E., Gardner, R., and Walker, J. (1994). Rules, games, and common-pool resources.

Pagiola, S. (2011). Using PES to implement REDD.

Portela, R. and Rademacher, I. (2001). A dynamic model of patterns of deforestation and their effect on the ability of the brazilian amazonia to provide ecosystem services. *Ecological Modelling*, 143(1–2):115–146.

Rode, J., Gómez-Baggethun, E., and Krause, T. (2015). Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*, 117:270–282.

Rodriguez-Sickert, C., Guzmán, R. A., and Cárdenas, J. C. (2008). Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior & Organization*, 67(1):215–227.

Sandker, M., Lee, D., Crete, P., and Sanz-Sanchez, M. (2014). Emerging approaches to forest reference emission levels and/or forest reference levels for REDD++. *UN-REDD Programme, FAO, UNDP, UNEP. FAO, Rome*, page 54.

Sattler, C. and Matzdorf, B. (2013). Pes in a nutshell: From definitions and origins to pes in practice—approaches, design process and innovative aspects. *Ecosystem services*, 6:2–11.

Stern, N. (2008). The economics of climate change. *American Economic Review*, 98(2):1–37.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, pages 234–240.

Tsujino, R., Yumoto, T., Kitamura, S., Djamaluddin, I., and Darnaedi, D. (2016). History of forest loss and degradation in indonesia. *Land Use Policy*, 57:335–347.

UNFCCC (2010). Framework convention on climate change: report of the conference of the parties on its 16th session, held in cancún from 29 november to 10 december. Report, Technical report, FCCC/CP/2010/7/Add. 1.

UNFCCC (2016a). Ethiopia's forest reference level submission to the unfccc.

UNFCCC (2016b). National forest reference emissions level for REDD+ in the context of decision 1/cp.16 paragraph 70 UNFCCC.

Vollan, B. (2008). Socio-ecological explanations for crowding-out effects from economic field experiments in southern africa. *Ecological Economics*, 67(4):560–573.

Wheeler, D., Hammer, D., Kraft, R., Dasgupta, S., and Blankespoor, B. (2013). Economic dynamics and forest clearing: A spatial econometric analysis for indonesia. *Ecological Economics*, 85:85–96.

Wollenberg, E., Moeliono, M., and Limberg, G. (2009). Between state and society: decentralization in indonesia. *The Decentralization of Forest Governance: Politics, Economics and the Fight for Control of Forests in Indonesian Borneo*, pages 3–24.

Wunder, S. (2005). Payments for environmental services: some nuts and bolts. Technical report.

Appendices

Intructions for field experiment

Good morning! [Introduce oneself and the enumerator] First of all, thank you for your cooperation. This is a study about the use and management of forest resources. We will have an entertaining time, and we kindly request your attention and participation.

When you pay full attention and make better decision, you will earn money according to your performance. We will explain this later. We will then have a brief interview in the end. Enjoy your stay with us!

We would like to remind you that all the information you give us and the decisions you make in the following activities will be used only for scientific research and will be kept confidential. When you have questions, please raise your hands and ask. We will answer all your questions so that you fully understand the activities.

Now let's see the activities. Suppose this group represents people living as a community. Also suppose that there is a common forest resource nearby. [Pointing to the tree branches] Let us say this is the forest and there are 60 trees in the forest that everyone could access. In each round of the activities we are going to do, a member (any one of you) will decide how many trees to harvest without knowing what other members will do. It is possible to decide not to harvest any tree. The maximum number of trees one could cut in one round is five. When we say one round, you can think of it as one month or one season in real life. When one round of the activity ends, [mention the name of the enumerator] will add the total harvest in that round of the group and disclose it to the group. Then we start another round with the same 60 trees in the forest.

To indicate your decision, you will use [showing the decision form] this form. You will choose how many trees to cut and indicate it by putting an X mark in this [show them] column. Example, if I choose not to harvest, I mark it here [point to the right cell]; 3 here, 5 here and so on. Any questions? [Answer all questions]

In real life, you cut trees to use them as firewood, to make agricultural tools, and to build houses. In this activity, each member will get one Birr for each tree he/she decides to cut as direct benefit. On the other hand, you may know that standing forests also give members of the community and its surroundings indirect benefits. For instance, forests attract rain, regulate above and underground water, serve as sources of other non-timber forest products such as incense, honey etc. To represent these indirect benefits, we will give the group two Birr for each standing tree, which the group members will share equally.

We would like to remind you once again that everything you do is anonymous, so only you know how much you harvest. Any questions? [Answer all questions]

Let us try a practice round! This is just for learning so you will not earn anything from this round. [Complete a full round. Answer any further questions] Thank you. Now let's start the real experiment. [For all groups] Anything you will earn from now on will be noted and paid to you in actual money at the end of the experiment.

[At the end of the fifth round, calculate the RL and]

[For groups 1, 3 and 5,] According to the performance of this group in the last five rounds, the average harvest was [average amount] per person per round. Now the government and other organizations, which care about forest protection, want to encourage you to reduce your harvest.

[For group 1] If you could reduce your individual harvest below [average amount + 1], we calculate the difference between [average amount + 1] and your individual harvest.

[For groups 3] If you could reduce your individual harvest below [average amount], we calculate the difference between [average amount] and your individual harvest.

[For groups 5] If you could reduce your individual harvest below [average amount - 1], we calculate the difference between [average amount - 1] and your individual harvest.

Then you will get a compensation of 0.75 Birr per tree for the difference, in addition to the benefits we mentioned earlier.

[For groups 2, 4 and 6,] According to the performance of this group in the last five rounds, the average harvest was [average amount] per person per round. Now the government and other organizations, which care about forest protection, want to encourage you to reduce your harvest.

[For groups 2] If you could reduce, *as a group*, the average harvest below [average amount + 1], we calculate the difference between [average amount + 1] and the group average harvest in each round.

[For groups 4] If you could reduce, *as a group*, the average harvest below [average amount], we calculate the difference between [average amount] and the group average harvest in each round.

[For groups 6] If you could reduce, *as a group*, the average harvest below [average amount - 1], we calculate the difference between [average amount - 1] and the group average harvest in each round.

Then the group will get a compensation of 0.75 Birr per tree for the difference, in addition to the benefits we mentioned earlier. The group members will share this compensation equally.

[For groups 3 and 4, at the end of the 10_{th} round] Now we will have similar activities but this time, you will not get the additional incentive you got in the last five rounds. The government thinks that you have seen how important conservation is and it does not have enough resources to continue to compensate you for your reduction. You will make similar decisions and indicate them using the same form. The benefits you get are one Birr per tree you decide to cut (direct benefit) and 2 Birr for the group per standing tree (indirect benefit). The group shares it equally. Let's start! [Play five rounds without PES]

Table 4: Decision form handed over to participants along with a pencil.

በዝሒ/Amount	ውሳነ/Decision
5 	
4 	
3 	
2 	
1 	
0	

Table 5: Post-experiment questionnaire

Basic information		
Participant no.		
Age		
Gender (1=female, 0=male)		
Education (1=illiterate, 2=read and write, 3=one to five, 4=six to eight, 5=high school and above)		
Position in the kebelle		
Environmental awareness		
Have you ever heard about climate change? (1=yes, 0=no)		
Do you think forests can help reduce the effect of climate change? (1=yes, 0=no)		
Which one is more valuable for you? For the community? (1=standing trees, 0=cut trees)		Individual______ Community______
Would you rather get direct benefit from trees you cut today than wait for shared benefits in the future? (1=yes, 0=no)		
Do you use irrigation to cultivate in the dry season? (1=yes, 0=no)		
About forest use		
1	Do you use forest products? (1=yes, 0=no)	
2	What forest products do you collect? (1=fire wood, 2=incense, 3=honey, 4=charcoal)	
3	How many times per week do you go to the forest to collect forest products? (In the dry season and the rainy season)	dry:____ rainy:_____
4	Have you sold any forest products during the last month? (1=yes, 0=no)	
5	How much forest products do you use compared to other families in the village? (1=less, 2=about the same, 3=more)	
6	Have you sold any forest product last month? (1=yes, 0=no)	
7	How important is the forest to you? (1= not important, 2=important, 3=essential)	
8	Do you consider the impact on others in the village of your actions when you harvest forest products? (1=yes, 0=no)	
About forest conservation and wealth		
8	What are your major sources of income? (Agriculture, Gold mining, own business, livestock, wage, forest products) [*you may select more than one*]	
9	What do you think is the most effective measure the government can do to decrease deforestation and forest degradation?	

10	If payments for conserving forests are to be introduced, would you prefer the payments to be made to the community or directly to the individuals? (1=community, 0=individual)	
11	How many *timads* of land does your family own?	
12	How many *timads* of land do you have?	
13	How much livestock do you have? (number of cattle, goats, donkeys, sheep, camels)	c:____ g:____ d:____ s:____ cml:_____
14	What type of material is (most of) your house's roof? (1=thatch; 2=wood and earth; 3=iron or other metal sheets; 4=tiles; 9=other, specify)	
15	Please mention the major income shortfalls or unexpectedly large expenditures during the past 12 months that your household faced.	
Trust and social interdependence		
16	Do you trust your neighbors? (1=yes, 2= somehow 3= no)	
17	Do you get support from your neighbors in times of shock? (1=yes, 2= sometimes 3= no)	
18	Considering everything, how happy were you in the last 12 months? (1=very unhappy 2=unhappy 3=average 4=Happy 5=very happy)	
19	Suppose there is a dry tree in your neighborhood, and you notice that it is dry before your neighbors do. What would you do? 1= I would cut it before my neighbors do. 2= I would tell my neighbors and discuss how we should share it. 3= I would leave it so that others could use it.	
About the experiment		
20	Did the tree branches we used in the experiment represent the forest in your neighborhood? (1=yes, 2=no)	
21	Did the RL given after the fifth round reflect your harvesting pattern? (1=yes, 2=no)	
22	Did you participate together with any close friends or family in the experiment? (1=yes, 0=no) If yes, how many?	
24	Did you have any particular harvest strategy in the experiment? Why/why not?	
25	Was your decision pattern close to your actual behavior in real life? (1=yes, 0=no)	

Chapter 2:
Drivers of deforestation

Drivers of deforestation in Indonesia: a spatial panel data analysis *

Amare Teklay Hailu[1], Arild Angelsen[1,2], and Arief Wijaya[3]

[1]School of Economics and Business, Norwegian University of Life Sciences (NMBU), PO Box 5003, 1432 Ås, Norway
[2]Center for Internatioanl Forestry Research, Bogor, Indonesia
[3]World Resources Institute, Washington D.C., USA.

Emails: amare.teklay.hailu@nmbu.no; arild.angelsen@nmbu.no; Arief.Wijaya@wri.org

Abstract
This paper revisits the old problem of drivers of tropical deforestation, with new data and state-of-the-art spatial econometrics methods. We use a balanced panel data from 190 Indonesian districts for every three years between 2003 and 2012. We apply spatial panel data econometric analysis to model the dynamics and interaction effects of the weighted average of neighbors' deforestation. Results show that its spatial lag as well as other socio-economic factors are important predictors of deforestation, including population density, share of agricultural GDP, poverty gap, and literacy rate. In contrast, aggregate GDP, access to electricity, and climate variables do not have significant predictive power. When comparing non-spatial and spatial models, the latter fits our data better, making coefficients of the exogenous variables in the model more accurate and estimation of spatial interaction effects possible. The positive spatial interaction of deforestation suggests that deforestation is contagious, and that policies to contain deforestation in one district can help reduce it in neighboring districts. Thus, rather than spatial leakage we observe conservation synergies across district borders.

Key words: Forest, REDD+, spatial econometrics, PES, forest transition
JEL codes: Q23, Q57, C33

*We would like to acknowledge funding from the Norwegian Agency for Development Cooperation [agreement no. GLO-3945 QZA 13/0545; grant on Learning from REDD+ an enhanced Global Comparative Analysis] through the Center for International Forestry Research (CIFOR).

1 Introduction

Tropical deforestation has several adverse effects such as global warming, land degradation and soil erosion, loss of biodiversity and ecosystem services, and other indirect socio-economic imbalances (Portela and Rademacher, 2001). It accounts for 10-15% of global greenhouse gas emissions (IPCC, 2013). Since Allen and Barnes (1985) seminal article more than 30 years ago, hundreds of studies have attempted to identify the causes/drivers or predictors of tropical deforestation at various scales.

Despite the large number of studies, three reasons motivate our revisit to the old problem. First, many of the existing studies suffer from lack of accuracy and rigor as they rely on less sophisticated models with strong assumptions. In particular, we use spatial panel data econometric analysis to take account of the spatial nature of the deforestation process (and the data). Second, even studies that explicitly account for the spatial dependence have scant data. We have balanced panel data spanning 2003 to 2012 at district level. Third, new ways of mitigating the old problem—in particular Payments for Environmental Services (PES)—have gained popularity, out of which additional issues have emerged. Particularly, targeting and prioritizing deforestation hotspot areas for policy measures such as incentive based REDD+ (Reducing Emissions from Deforestation and forest Degradation) require modeling the dynamics of deforestation in a way that enables its better prediction. Further, to establish a PES contract one needs to define the counterfactual trajectories of deforestation or reference levels. Thus, accurately predicting deforestation is a critical element in the design and implementation of performance-based REDD+.

For these reasons, we focus on prediction rather than causality, i.e., on identifying good predictors rather than the underlying causes. We step back and look into the nature of deforestation in Indonesia vis-à-vis structural variables, i.e., demographic metrics (in particular population density), economic activities (GDP, share of agricultural GDP), etc., as well as spatial lags of deforestation. Specifically, we explore three questions: (i) Is there spatial autocorrelation, i.e., is there a difference between non-spatial and spatial models? (ii) What is the nature of the spatial dependence, if any? (iii) Is there a forest transition pattern to be observed, i.e., a bell-shaped relationship between deforestation and forest cover?

We posit a spatially explicit model of deforestation in 190 Indonesian districts. Spatial explicitness aids to not only estimate the spillover effect of deforestation among neighboring spatial units but also improve the accuracy and consistency of the estimates for other explanatory variables in the econometric model. We use panel data to capture dynamics over time, in the spirit of Hargrave and Kis-Katos (2012) and Boubacar (2012), among others, while focusing on subnational scale and aiming to contribute to the estimation of reference levels in a PES mechanism for REDD+.

To this end, we draw on previous spatial (Boubacar, 2012; Hargrave and Kis-Katos, 2012; Wheeler et al., 2013) and non-spatial (Combes Motel et al., 2009) econometric studies of deforestation (see Busch and Ferretti-Gallon (2017) for a review). In terms of scale, we follow Wheeler et al. (2013) and use Indonesian districts as units of analysis with actual deforestation, as opposed to forest clearing index in their case, as a dependent variable. In terms of data, we use a balanced panel data from 2003 to 2012 every 3 years (T=4); this enables us to capture heterogeneity between and dynamics within the districts. In terms of focus, we give special emphasis to the information that weighted average of neighbors' deforestation could carry in predicting deforestation while controlling for various relevant socioeconomic and climatic variables. We also test for the existence of a forest transition phenomena at the subnational scale (Rudel, 1998).

In the next section, we give an overview of the drivers of deforestation and some of the modeling issues. In section 3, we describe the methods and data used and explain the spatial econometric panel data model. In section 4, we present and discuss the main results from both non-spatial regression models and various random-effects specifications of the spatial panel regressions. The last section concludes on how deforestation in neighboring districts may turn out to be an important predictor of deforestation and its implication for REDD+.

2 Deforestation revisited

Nearly two decades since Angelsen and Kaimowitz (1999) rethought the causes of deforestation, what we know is still far from being adequate to enable policy makers design effective policies and incentive mechanisms such as PES. Despite the bulk of studies before and since their synthesis of the economic models of deforestation, there is neither a universal approach to its modeling nor a consensus on the set of relevant causes. There are both conceptual and practical sources of this disparity, which we elaborate in below.

2.1 Direct vs underlying causes

The first source of variation, if not of ambiguity, in the economic models of deforestation is the use of terms like 'causes', 'drivers', 'determinants', 'factors', 'predictors', etc. The main interest is in the distinction between direct or immediate causes and underlying causes (drivers) of deforestation (Geist and Lambin, 2002). In their conceptual framework, Angelsen and Kaimowitz (1999) argue that the actions of agents who make land use change decisions are the *sources* of deforestation (e.g., land clearing by shifting cultivators). The *immediate causes* of deforestation, in this framework, are the decision parameters (e.g., agricultural farm-gate prices), which affect agents' choice variables and are subject to the influence of overall macroeconomic dynamics and broader policies (e.g., trade policies)—the *underlying causes*. Recognizing these layers of variables coupled with the research question one poses serves as a basis for choosing an economic model and identifying a set of predictors of deforestation. Mixing variables at different levels in the same model will likely result in biased estimates (Angelsen and Kaimowitz, 1999).

The type of model of deforestation also depends on the scale of analysis (Angelsen and Kaimowitz, 1999). When households or local firms are the units of analyses, micro models may be more appropriate. In subnational studies, aggregate indicators are considered. This has its merits and drawbacks: we lose some information when we aggregate variables, but we also observe emergent properties. When countries, regions or districts are the units of analyses, macro-level models of deforestation become more preferable, and macro-economic indicators are used to explain deforestation. Examples of macro-level studies include Clark (2012), who shows that deforestation is correlated to population density and level of poverty in low-income countries. Empirical regularities such as the Environmental Kuznets Curve or the Forest Transition, which suggests a reversal in deforestation as income grows (Culas, 2012), as well as the existence of persistently high deforestation due to high dependence on natural resources in many low income countries (Boubacar, 2012) can only be uncovered at larger scales. These macro-level analyses focus on the underlying causes of deforestation (Angelsen and Kaimowitz, 1999; Geist and Lambin, 2002), which Combes Motel et al. (2009) refer to as structural variables.

Geist and Lambin (2002) systematically identified different clusters of both the underlying causes and proximate causes of deforestation after they thoroughly analyzed 152 subnational case studies of tropical deforestation. They argue that factors such as infrastructure extension, agricultural expansion, commercial and noncommercial wood extraction and other social and biophysical factors have direct effect on forest cover. This effect emanates from the active decisions of agents who undertake these development works intentionally. Demographic, economic, technological, cultural, political ana institutional factors may reinforce the aforementioned direct causes (Li et al., 2013). Some factors are mentioned more frequently in empirical studies and considered more important than others, but it should be noted that the interplay of this multitude of causes is always at work (Geist and Lambin, 2002; GFC, 2010).

The choice of the scale of analyses and the potential explanatory factors may be guided by the research question, or—as is usually the case—dictated by the availability of data. Macro level aggregate measures of demographic factors (e.g., population density and population growth rate) and socioeconomic factors (e.g., per capita GDP and agricultural commodity prices) are often used in these studies and have been shown to have a significant effect on deforestation. Köthke et al. (2013), Boubacar (2012) and Combes Motel et al. (2009), among others, applied panel data econometric methods to model the dynamics of deforestation using countries as the unit of analysis.

Micro-founded local studies tend to investigate the direct incentives that compel agents to convert forest land to other land uses (Babigumira et al., 2014). When the purpose of the study is to get a bigger picture of aggregate forest loss at national or subnational levels, factors that characterize these units—usually aggregate socioeconomic and biophysical indicators—are more appropriate. This explains the pattern in the choice between underlying and proximate causes: cross country studies use the former while local and micro analyses use the latter (Angelsen and Kaimowitz, 1999).

2.2 Spatial explicitness

The use of spatial analyses has increased, necessitated by the need for accurate prediction of the location of deforestation and facilitated by the better availability of data. Spatially explicit models may be more inclusive and informative compared with their non-spatial counterparts. Busch and Ferretti-Gallon (2017) distinguish between 'spatially explicit econometric studies', which identify the endogenous and exogenous factors of interest for the spatial units under study and apply econometrics to make inferences, and 'spatial econometrics studies', where the spatial interaction effects of neighboring units are also among the parameters to be estimated. We focus on the latter.

The most common justification behind this is derived from the first law of geography that "everything is related to everything else...", due to Tobler (1970, p.4). This is likely to hold also for forest systems (Boubacar, 2012; Hargrave and Kis-Katos, 2012; Wheeler et al., 2013). In empirical analyses of deforestation and its drivers, the advantage of recognizing and taking into account this 'relatedness' is twofold. First, it enables us to find estimates of the spillover effect of deforestation in neighboring districts, which is of policy significance as spatial units must consider the contagiousness of deforestation in their forest conservation strategies and seize the opportunity for synergistic performance (Boubacar, 2012). Second, including or controlling for the spatial interaction effects makes estimates of other explanatory variables in the model unbiased and consistent. This again translates to policy effectiveness by enabling the researcher to make accurate inferences and distinguishing between about factors which do and do not predict deforestation (Robalino and Pfaff, 2012).

According to Boubacar (2012), the causal effects of variables in an econometric model would be underestimated if spatial correlation is not accounted for. A spatial model is thus preferred to its non-spatial counterpart as the latter suffers from omitted variable problem. Location decision is one aspect where spatial dimension is relevant in the study of economic resource allocation. Nevertheless, spatial heterogeneity (i.e., variations in relationships depending upon location) and spatial dependence (also called spatial autocorrelation, due to Cliff and Ord (1970), or the existence of neighborhood effects) are the foundations of spatial econometric analyses (Anselin, 1988).

We suspect, also drawing on existing spatial econometric analyses of deforestation (Boubacar, 2012; Wheeler et al., 2013), that there is a systematic spatiotemporal pattern in the dynamics of deforestation among Indonesian districts. More precisely, deforestation is contagious: deforestation in one district is affected by its magnitude in neighboring districts, and it in turn affects deforestation in all its neighbors. This spatial dependence arises due to different reasons, for example, changes in local prices (of agricultural commodities and forest products) (Robalino and Pfaff, 2012) or infrastructure expansion (Aguiar et al., 2007), which makes encroachment into forests easier and transport costs lower, may drive more deforestation in neighboring districts.

2.3 The forest transition (FT) hypothesis

Forest transition, a term coined by Mather (1990), refers to a particular empirical pattern for how forest cover change during the course of the economic development of a country or region. The early stage is characterized by low deforestation rates and high forest cover. As the population and economy grow, demand for agricultural products (and new land) increases, and improved roads and infrastructure make new forest areas accessible for agricultural conversion. Deforestation increases and forest cover shrinks. Eventually, this development levels off, and we reach a turnaround where reforestation and afforestation overtake deforestation.

Angelsen and Rudel (2013) identify five factors that can contribute to this turnaround. First, scarcity of forest products, as a result of shrinking forest stocks and increasing demands. Second, scarcity of environmental services, i.e., (perceived and real) links between forest loss and environmental services. For example, high rates of forest loss have compelled governments to take reforestation measures that have stabilized forest cover. These first two explanations are referred to as the forest scarcity path (Rudel et al., 2005). Third, diminishing agricultural rents from continued deforestation due to either longer distances and lower soil fertility on marginal lands. Fourth, economic development and structural changes, linked to higher (labor) production costs, changes in demands of forest and agricultural products. Often referred to as the economic development path (Rudel et al., 2005), industrialization and urbanization bring better off-farm opportunities and thus forest recovers in abandoned land. Fifth, policy changes, such as direct lands use and tenure regulations and changes in agricultural taxes and subsidies.

The appropriate scale at which FT can be observed is not clear. The original analyses focused on the national level, but FT can occur at a subnational unit, country or regional level (Rudel et al., 2005). Some researchers argue that studies should be at larger scales as deforestation in one country may be displaced to other (Meyfroidt et al., 2010). Lower deforestation in Vietnam has been linked to more deforestation in neighboring countries, including Cambodia and Laos (Meyfroidt et al., 2010), and international trade can bring about a "globalization of the forest transition" (Angelsen and Rudel, 2013).

One might also argue for the merits of applying the forest transition to lower, sub-national scales. There are relevant local processes which are highly diverse and crucial to understand the forest cover dynamics (Perz and Walker, 2002). Historic land use dynamics, tenure and land distribution, often set the stage for how forest cover changes over time (Aguilar-Støen et al., 2011). In a large and culturally and socioeconomically diverse country as Indonesia, the argument for using sub-national units is particularly compelling.

3 Data and Methods

3.1 Data sources

This study focuses on sub-national deforestation patterns and district-level interactions in Indonesia. A balanced panel data of deforestation and forest cover was obtained from the Indonesian Ministry of Forestry (MoFor), while various potential explanatory variables were collected from different sources, mainly Statistics Indonesia and the World Bank (below). These data cover 190 districts in Indonesia and span between 2003 and 2012, where observations are taken every three years. The time dimension is four and the number of panels is 190, and hence the total number of observations is 760. Only 190 districts are included in this study partly because we did not find complete data for other districts.

We excluded districts in Bali and Java islands, as often done in deforestation analysis, as these islands are dramatically different from "the outer islands" in terms of, *inter alia*, population density, land cover and land use, with limited potential for deforestation. The total number of districts has been changing following decentralization and subdivisions in the late 1990s. There were 483 districts in 2008 (Burgess et al., 2012) , and our sample thus includes about 40% of them.

We extracted climate variables (precipitation and temperature) from CRU TS3.21[1]. Three years' average temperature and precipitation were computed at district level from the monthly observations between January 2000 and December 2012. Total GDP, agricultural GDP, poverty gap index, population, literacy rate and access to electricity were obtained from the World Bank's latest INDO-DAPOER (Indonesia Database for Policy and Economic Research)[2].

The dependent variable is annual deforestation per land area. We chose land area rather than forest area as the denominator for several reasons. First, in the forest transition theory it is change in forest cover (i.e., deforestation) that is of interest, and this is measured as share of total land area. Second, small absolute changes in forest area can result in high

[1] https://climatedataguide.ucar.edu/climate-data/cru-ts321-gridded-precipitation-and-other-meteorological-variables-1901.

[2] https://data.worldbank.org/data-catalog/indonesia-database-for-policy-and-economic-research

deforestation rates if defined relative to forest area, and the forest cover is low.

The explanatory variables include different socio-economic indicators and climate variables. Population density, measured as the natural logarithm of the number of people per square kilometers, is expected to affect deforestation as high population pressure increases the demand for more agricultural land and urban expansion. Total district level GDP and share of agricultural GDP capture the economic activities and the role of dependence on and nature of agriculture in a given district. The expected effects are not clear *a priori*. Poverty has been mentioned as one of the drivers of tropical deforestation, and to explore this effect we include poverty gap index, a district level measure of how far income is below the poverty line (Foster et al., 2010). Similarly, the literacy rate (among the population between ages 15 and 65) and access to electricity are included to capture the level of awareness and availability of alternatives (e.g., substitute electricity for firewood), respectively.

Regarding the spatial weighting matrix, a polygon shape file was obtained from the GADM database (http://www.gadm.org/download). Then first-order contiguity-based spatial weighting matrix was generated. The weighting matrix, W, is a square matrix with districts as rows and columns. Two districts are classified as neighbors, and hence their corresponding row-column entry is 1, if they are contiguous—i.e., they share a vertex along their border. We did not consider higher order contiguity, neighbors of neighbors, but we used distance based weighting matrix to check if results are robust.

3.2 Econometric model

This study applies a longitudinal (panel) data econometric model, which has the advantage of capturing the individual specific effects and identifying the intra-individual dynamics. Another practical advantage of using panel data is its simplicity in estimation when time invariant unobservable factors are suspected to affect the accuracy of parameter estimates. Cross sectional analysis gives biased estimates as it fails to account for individual heterogeneity (Arellano, 2007). It often suffers from omitted variable bias. Previous studies have used different measures of the dependent variable. Some use a continuous value of absolute forest area loss, or a proportion thereof relative to total forest or land area. The downside of absolute measure of forest loss is that it introduces bias owing to differences in total land area, as larger districts may have larger forest area and/or higher deforestation in absolute terms. Therefore, actual deforestation, adjusted for total area, was used as a dependent variable in our study.

A standard starting point in such analyses is to use the Ordinary Least Square (OLS) on the pooled data. However, OLS is likely to be biased and unable to control for heterogeneity and spatial interaction effects. A specification test indicates that more inclusive models may perform better, and we fitted a spatial panel data model. As such, we take advantage of available information at multiple points in time to control for the potential bias from ignoring district-specific fixed effects. Following Elhorst (2014), the most general setup of a spatial panel data model with spatiotemporal effects is given by the following vector notation:

$$Y_t = \rho W Y_t + \alpha l_N + X_t \beta + W X_t \theta + \mu + \xi_t l_N + u_t \tag{1}$$

$$u_t = \lambda W u_t + \varepsilon_t$$

Where $\mu = (\mu_1, ..., \mu_N)^T$
Y is an $N \times 1$ vector of observations and X is an $N \times K$ matrix of exogenous regressors while u is an $N \times 1$ vector of error terms. The parameter α is an intercept and μ is district-specific fixed effect which captures the effect of unobservable (or omitted) time-invariant variables. ξ_t is time specific fixed effects which is district-invariant factor such as macroeconomic shock which affects all districts alike. The parameters ρ, θ and λ represent the spatial autoregressive, spatial lagX and spatial autocorrelation coefficients respectively. Similarly, $\varepsilon_{it} = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})^T$ which is an $N \times 1$ vector of stochastic error terms where ε_{it} is i.i.d across $i = 1, 2, \ldots, N$ with $\varepsilon_{it} \sim N(0, \sigma^2)$. Finally, W is $N \times N$ spatial weights matrix which does not have to be the same in all cases, but it is assumed to be constant over time.

Table 1: Nomenclature of terms and parameters in the general spatial panel model.

Parameters	Description
ρ	Spatial autocorrelation coefficient on the spatially lagged dependent variable.
W	N x N matrix of spatial weighting coefficients. It can vary in different parts of Eq.(1).
α	Intercept
l_N	N x N identity matrix
β	The effects of own characteristics on deforestation.
θ	Spatial autocorrelation coefficient on the explanatory variables.
μ	District-level fixed effects
ξ	Time-period effects
u_t	Composite error term
λ	Spatial autocorrelation coefficient on the spatial error term.
ε_t	Idiosyncratic error

Source:Adapted from Burnett et al. (2013)

There are a number of issues to consider here. First, the nature of spatial dependence determines what kind of spatial econometric model should be fitted for the data at hand. Equation (1) is the most general formulation while empirical studies usually embark, based on the nature of the expected spatial dependence, on various simpler variants of it. If $\rho = 0$ and $\lambda = 0$, then one finds a spatial lagX model in the explanatory variables which Elhorst (2014) refers to as exogenous interaction effects model. If $\theta = 0$ and $\lambda = 0$, equation(1) reduces to a spatial autoregressive (SAR) model, which is also called the Spatial Lag Model (SLM), and there are said to be endogenous interaction effects. A spatial error model (SEM) has only $\lambda \neq 0$, while only $\lambda = 0$ implies a spatial Durbin model (SDM). The list of various specifications and corresponding restrictions on these spatial parameters are presented in Table 2.

We hypothesize that all variants of the GNSM that include spatial lags of deforestation on the right hand side are good candidates, as deforestation in neighboring districts is likely to affect deforestation in a given district. In particular, SAR appears to be relevant and interesting in the context of REDD+ where leakage is a significant concern. As for the

Table 2: Variants of the spatial panel model.

Acronym	Symbolic representation	Spatial parameters	Description
GNSM	$Y_t = \rho W Y_t + \alpha l_N + X_t\beta + W X_t\theta + \mu + \xi_t l_N + \lambda W u_t + \varepsilon_t$	$\rho \neq 0$, $\theta \neq 0$ and $\lambda \neq 0$	*General Nesting Spatial Model:* spatial interactions occurs in the dependent and the independent variables, and in the error term.
SAC	$Y_t = \rho W Y_t + \alpha l_N + X_t\beta + \mu + \xi_t l_N + \lambda W u_t + \varepsilon_t$	$\rho \neq 0$, $\theta = 0$ and $\lambda \neq 0$	*Spatial Autocorrelation Model:* spatial interactions occur in the dependent variable and in the error term.
SDM	$Y_t = \rho W Y_t + \alpha l_N + X_t\beta + W X_t\theta + \mu + \xi_t l_N + \varepsilon_t$	$\rho \neq 0$, $\theta \neq 0$ and $\lambda = 0$	*Spatial Durbin Model:* spatial interactions occur in the dependent and the independent variables.
SDEM	$Y_t = \alpha l_N + X_t\beta + W X_t\theta + \mu + \xi_t l_N + \lambda W u_t + \varepsilon_t$	$\rho = 0$, $\theta \neq 0$ and $\lambda \neq 0$	*Spatial Durbin Error Model:* spatial interactions occur in the independent variable and in the error term.
SLM	$Y_t = \rho W Y_t + \alpha l_N + X_t\beta + \mu + \xi_t l_N + \varepsilon_t$	$\rho \neq 0$, $\theta = 0$ and $\lambda = 0$	*Spatial Lag Model:* also called spatial autoregressive (SAR) model, and only the spatial lag of the dependent variable is significant.
SLX	$Y_t = \alpha l_N + X_t\beta + W X_t\theta + \mu + \xi_t l_N + \lambda W u_t + \varepsilon_t$	$\rho = 0$, $\theta \neq 0$ and $\lambda = 0$	*Spatial Lag X:* only the spatial lags of some or all of the independent variables are significant.
SEM	$Y_t = \alpha l_N + X_t\beta + \mu + \xi_t l_N + \varepsilon_t$	$\rho = 0$, $\theta = 0$ and $\lambda \neq 0$	*Spatial Error Model:* the spatial interaction effects occur through correlated errors.
OLS	$Y_t = \alpha l_N + X_t\beta + \mu + \xi_t l_N + \varepsilon_t$	$\rho = 0$, $\theta = 0$ and $\lambda = 0$	*Ordinary Least Squares:* no spatial interaction effects.

Source:Adapted from Elhorst (2014, p.9).

other spatial terms, the spatial lags of some explanatory variables may be relevant, but we can determine which only by testing them empirically. SDM, a model named after Anselin (2013) as it parallels the Durbin model in standard panel data models, could reflect effects of resource flows such as migration, higher demand in markets in neighboring districts and other strategic interactions across units. There is no rule of thumb to prefer one model to the other *a priori*, but Elhorst (2014) recommends that one start from the most general one and use statistical tests to determine if it can be reduced to other simpler variants.

Second, the choice of a spatial weights matrix should be carefully made to capture the true nature of the spatial dependence. As such, a queen[3] first order contiguity matrix was chosen based on whether districts share a border or a single vertex or not, because administrative boundaries are usually used to restrict resource flows in the form of budgets and implementation of projects by local governments notwithstanding the basis used when they are created.

Third, assumptions made about μ and ξ_t in Eq.(1) really matter as to whether the fixed-effects (district specific, time specific or both) or a random effects model is used. The fixed effects specification assumes that μ and ξ_t are time and district invariant respectively, while the random effects specification considers these effects as part of the stochastic error term and assumes each of these terms is idiosyncratic.

4 Results

4.1 Spatiotemporal dynamics

Deforestation occurs across space and over time, giving rise to spatiotemporal dynamics. Over time, there is intra-district variation where average deforestation per land area in the study districts increased from 0.77% in 2003 to 1.87% in 2009 before it dropped to 0.72% in 2012. Across districts, we observe variations that could be attributed to inherent contemporaneous differences in the district characteristics. Figure 1 illustrates these spatiotemporal differences and patterns in deforestation. In addition to (and potentially driving) the spatiotemporal changes in deforestation rates, many of the districts differ in terms of key structural socioeconomic characteristics (see Table 3 for key summary statistics of the relevant variables used in the subsequent regression analyses).

In Sumatra, we observe high deforestation rates throughout the period in the districts of Riau and Jambi provinces, which has long been a focus due to the rapid expansion of palm oil. But, we also note increasing rates in other provinces on the island. Moving east to Kalimantan (the Indonesian part of Borneo), the other main island of action, we observe high deforestation rates in Central and South Kalimantan provinces in 2003, but with comparatively low rates at the end of the period (2012). Central Kalimantan has been a REDD+ pilot province. In several other districts, i.e., those located in the other three Kalimantan provinces, the trend has been the opposite. In Sulawesi, we detect lower deforestation rates in the later years. Finally, in the two Papua provinces, sometimes referred to as the last frontiers in Indonesia, we observe a mixed pattern.

[3] Rook contiguity is an alternative approach where districts that share a common border are considered neighbors.

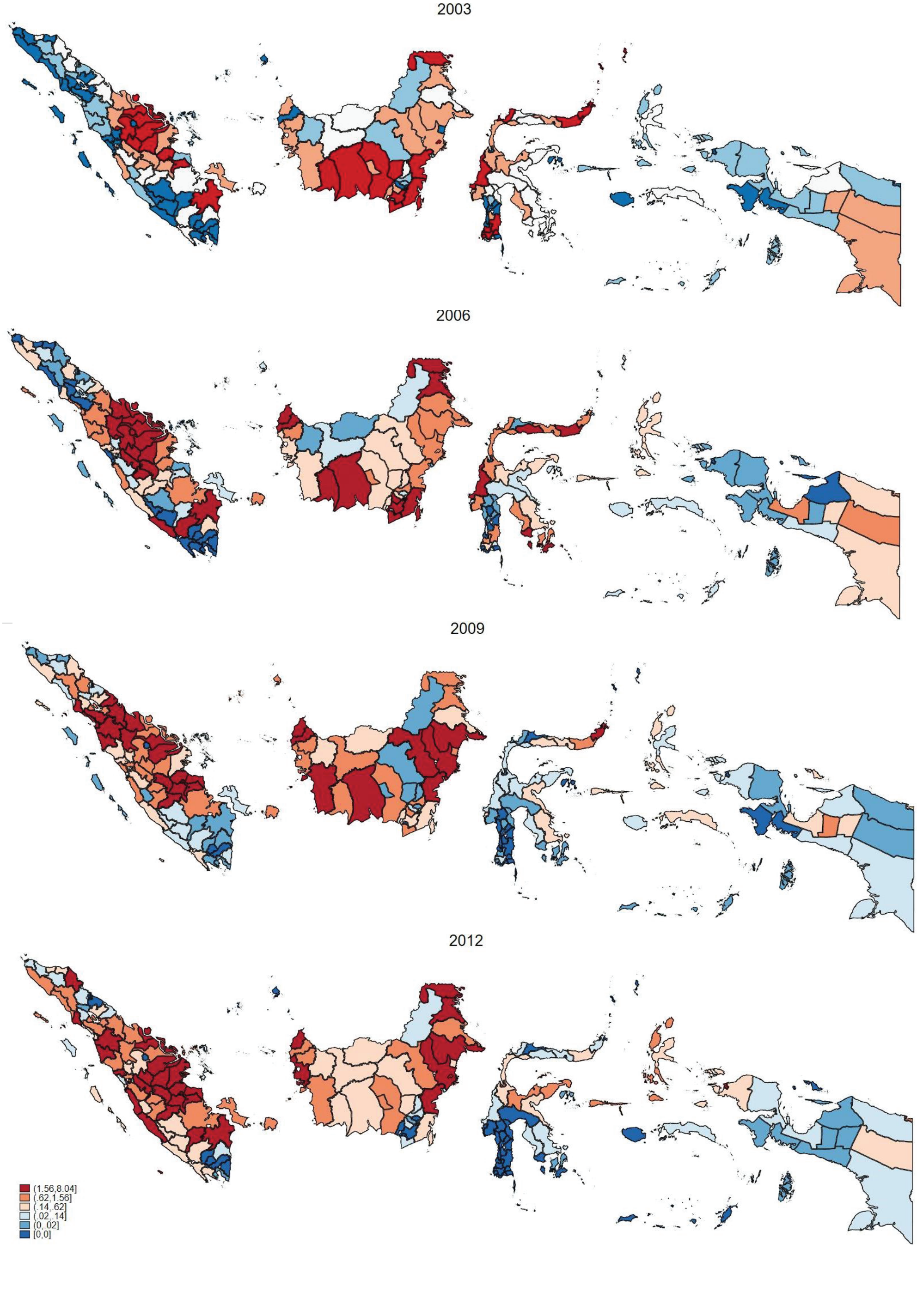

Figure 1 Spatio-temporal dynamics of deforestation (per land area) in 301 Indonesian districts (2003 to 2012)

The key characteristics we considered include population density, GDP, share of agricultural GDP, poverty gap index, literacy rate, percentage of households with access to electricity, and climate variables (precipitation and temperature). To test the forest transition hypothesis (section 2.3), we also included relative forest area (forest area per total land area, to adjust for size differences among districts) and forest area squared.

We suspected that excluding districts from Bali and Java might create biases in the estimation of spatial interaction effects, as sharing a border with districts where there is little or no deforestation can affect deforestation. As a robustness check, we executed the same regressions on 301 districts (i.e., included those from Bali (n=28) and Java (n=83)), but results do not change much (see Table 9).

Table 3: Descriptive statistics of deforestation and other relevant characteristics.

Variable	Variable definition	Obs.	Mean	Std. dev.	Min	Max
Defr	Deforestation/ total land area	760	1.09	2.534	0	35.89
logPopDen	Natural log of population density (persons/km2)	760	4.37	1.644	-0.002	9.11
logGDP	Natural log of district level GDP (million rupiahs)	760	7.63	1.008	4.80	10.63
AgriGDPshare	Agricultural GDP/total GDP	760	35.28	19.295	0.11	90.56
PGI	Poverty gap index (%)	760	5.21	8.468	0.34	41.27
LitRate	Literacy rate (%)	760	90.46	11.039	16.04	99.75
Electr	Access to electricity (% of households)	760	81.71	17.622	0.70	99.9
Precip	Precipitation (mm)	760	235.25	44.937	117.89	394.34
Temp	Temperature (°C)	760	26.28	1.238	21.019	28.32
FrstArea	Forest area/total area	760	49.73	26.192	0	98.57

4.2 Non-spatial models

To test whether there are significant spatial interaction effects, we first estimated a non-spatial pooled OLS model, and the fixed and random effects variants of the standard panel data model. These models explain deforestation in terms of the variation in socioeconomic and climate factors. They tacitly presume no significant spatial spillover effects and thus the spatial lag variables do not belong in the model. If this presumption is wrong, however, the estimates in these models will be biased and inconsistent. We then tested if there are potential spatial interaction effects in deforestation, by computing spatial autocorrelation statistics, i.e., the Moran's I test (Moran, 1950).

Table 4: Estimates from non-spatial models of deforestation and district characteristics.

Variables	OLS	FE	RE
Population density (log people/km2)	-0.39***	-0.16	-0.45***
	(0.09)	(0.36)	(0.11)
GDP(log)	0.13	-0.53	0.10
	(0.11)	(0.74)	(0.13)
Share of agricultural GDP	-0.02***	-0.02	-0.02***
	(0.01)	(0.01)	(0.01)
Poverty gap index	0.03	-0.05	0.03**
	(0.03)	(0.09)	(0.01)
Literacy rate (% of literate people age $\geq$ 15)	-0.02**	-0.05	-0.02**
	(0.01)	(0.04)	(0.01)
Access to electricity (% of households)	0.01**	0.01	0.01
	(0.01)	(0.01)	(0.01)
Precipitation (mm)	0.01***	0.02***	0.01**
	(0.002)	(0.006)	(0.003)
Temperature (°C)	0.13*	0.55	0.12
	(0.07)	(0.86)	(0.11)
Forest area	0.05***	-0.12	0.05***
	(0.01)	(0.09)	(0.02)
Forest area squared	-0.001***	-0.003***	-0.001***
	(0.0001)	(0.001)	(0.0002)
Year dummies:			
2006	0.23	0.29	0.24
	(0.23)	(0.30)	(0.24)
2009	0.86**	0.33	0.84***
	(0.36)	(0.42)	(0.25)
2012	-0.42	-1.27**	-0.45*
	(0.27)	(0.51)	(0.27)
Constant	-1.83	jun.31	-0.87
	(2.10)	(25.10)	(3.10)
Sigma_u		10.01***	0.91
Sigma_e		02.jul	02.jul
		0.96	0.16
Observations	760	760	760
R-squared	0.117	0.257	0.117
No. of districts		190	190

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The coefficient estimates of various drivers of deforestation from the non-spatial models (pooled OLS, fixed and random effects panel) are reported in Table 2. At this point, the focus is on whether these benchmark models can be improved by incorporating spatial lags. Our statistical test results in Table 5 show that there is significant (p-value< 0.00) spatial autocorrelation in the dependent variable at each point in time, corroborating that deforestation is contagious and models which include spatial lags of deforestation on the right-hand side will be preferred to those reported in Table 4.

Table 5: Global Moran's I test statistics for spatial autocorrelation of deforestation.

Year	2003	2006	2009	2012
Chi-square	13.12	10.89	27.39	23.27
p-value	0.0014	0.0043	0.000	0.000

Based on the evidence from the spatial autocorrelation test results in Table 5, we estimated the spatial models using the suite of commands for spatial panel analyses in Stata®15. Before the release of that version, we had done similar analyses using the user-written xsmle command (Belotti et al., 2016). Results of the spatial autocorrelation are strong in both cases, but the fixed effects specification does not return parameter estimates for most variables, except for the spatial lag terms, when we use the official Stata® commands. Now we tackle the question of which spatial model to use.

The Moran's I test statistic indicates whether a given variable has neighborhood effects, and is generated using a contiguity-based weighting matrix. In our case, we used contiguity weighting matrix, where two districts are considered as neighbors if they share at least a common vertex in their borders. A distance-based weighting matrix was then used to check the robustness of the results. Table 5 reports that the spatial autocorrelation of deforestation is strongly significant in all four time points. This justifies and necessitates the use of spatial models, as results in the non-spatial models (Table 4) may not be accurate.

4.3 The spatial models

One of the main questions which this paper seeks to address is whether there are spatial interaction effects and hence the spatial models fit the data better than do their non-spatial counterparts. The previous section showed the presence of positive and strong spatial autocorrelation, which justifies a spatial model. Since the particular spatial model is not known yet, we start with the most general form of the spatial panel data model (see Eq. (1)). In all alternative regressions, presented in Table 6, Table 8 and Table 9, there is significant spatial autocorrelation in deforestation; the spatial lag of the dependent variable was found to be highly significant and robust to changes in the choice of the spatial model as well as the sample. We included spatial lags of GDP and population density in the SDM model, but neither of them carries significant information about deforestation. This suggests that the SDM could be reduced to either SAR or SAC, i.e., the weighted averages of neighbors' GDP and population density do not have significant effect on deforestation in a typical district.

The fact that the parameter that captures the spatial interaction effects in the dependent variable, ρ, is strongly statistically significant suggests that we should keep the spatial lag of the dependent variable in the model. This rules out the SEM, which posits that the spatial interaction occurs only through the residuals. Yet we need to test if there is spatial correlation in the error term as well, i.e., whether λ in Eq. (1) is statistically different from zero. If this turns out to be the case, then SAC is the appropriate choice, since it captures the spatial autoregressive component, as do both SDM and SAR, while accounting for the spatial correlation in the error term. The SAR and SAC are non-nested models, and we choose either one using both the Akaike's and Bayesian Information Criteria (AIC and BIC). Results in Table 8 show that SAR has marginally lower AIC and BIC scores and is superior to SAC. Significant spatial interaction occurs only through the contagious and diffusive nature of deforestation.

The interpretations of the findings in this study are, therefore, based on the random effects SAR model, the second column of Table 6. The results in the third column of the same table are from the SAC model with random effects, and λ, the coefficient of the spatial autocorrelation in the error term, is not statistically significant. It is important to note that the coefficient on the spatial lag of the dependent variable is highly significant at 1% level in all specifications.

Besides the spatial lags of the dependent variable, other structural factors also predict deforestation in different ways. Population density and share of agricultural GDP are negatively related to deforestation. Higher poverty gap index indicates higher deforestation per land area, holding other factors fixed. Deforestation in districts with higher percentage of literate population is likely to be lower than it is in districts comparable in other characteristics but have lower literacy rate. The negative effect of population density is not surprising, as densely populated districts have low forest cover and thus low potential for deforestation. The year dummies show that 2009 had statistically significantly higher deforestation.

We included forest area and forest area squared to test the forest transition hypothesis, and an interesting result emerges. The coefficient estimates for forest area are positive and significant while the sign reverses in forest area squared. That is, deforestation and forest area have a quadratic relationship, with a turning point at a forest cover of 36.5% in the average district.

Table 6: Econometric results of SDM, SAR and SAC models.

Variables	SDM	SAR (RE)	SAC(RE)
Explanatory variables:			
Population density (log people/km2)	-0.224**	-0.193**	-0.201**
	(0.103)	(0.094)	(0.096)
GDP (log)	0.100	0.048	0.060
	(0.112)	(0.104)	(0.106)
Share of agricultural GDP	-0.009	-0.011*	-0.011*
	(0.006)	(0.006)	(0.006)
Poverty gap index	0.026**	0.028**	0.029**
	(0.011)	(0.011)	(0.011)
Literacy rate (% of literate people age$\geq$ 15)	-0.022**	-0.023**	-0.023**
	(0.010)	(0.010)	(0.010)
Access to electricity (% of households)	0.011*	0.010	0.010
	(0.007)	(0.010)	(0.006)
Precipitation (mm)	0.003	0.003	0.003
	(0.002)	(0.002)	(0.002)
Temperature (°C)	0.060	0.099	0.089
	(0.092)	(0.089)	(0.092)
Forest area	0.039***	0.0365***	0.038***
	(0.014)	(0.014)	(0.014)
Forest area squared	-0.0005***	-0.0005***	-0.0005***
	(0.0001)	(0.0001)	(0.0002)
Year dummies:			
2006	0.083	0.102	0.103
	(0.218)	(0.218)	(0.232)
2009	0.395*	0.468**	0.497**
	(0.239)	(0.234)	(0.250)
2012	-0.347	-0.278	-0.302
	(0.250)	(0.245)	(0.260)
Spatial terms:			
W*Deforestation	0.614***	0.598***	0.554***
	(0.052)	(0.051)	(0.081)
W*population density	-0.087		
	(0.153)		
W*GDP	-0.008		
	(0.094)		
W*Error			0.097
			(0.125)
sigma_u	0.634***	0.638***	0.664***
	(0.131)	(0.131)	(0.131)
sigma_e	2.061***	2.068***	2.068***
	(0.062)	(0.062)	(0.062)
Constant	-0.305	-1.065	-0.909
	(2.602)	(2.569)	(2.615)
Observations	760	760	760
Number of groups	190	190	190

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The auxiliary results in Table 6, sigma_u and sigma_e are also statistically significant at 1% level. This suggests that the random effects contribute significantly to the unobserved sources of variation. As the focus of the spatial panel analysis is to explore how neighboring interaction effects might enter the model that explains deforestation, we pay special attention to the coefficients of the spatial lag of the dependent variable (ρ). It is highly significant, and its magnitude is 0.598 (see Table 6), indicating that deforestation in a given district is 59.8% of deforestation in its neighboring districts, on average. This is higher than the 53.4% effect Boubacar (2012) finds in country level analysis. Arguably, spatial interactions among smaller units (districts) is expected to be stronger than those of countries. As pointed out in Section 2, a spatial panel specification has two advantages. First, they allow us to estimate the spatial interaction effects, and are thus helpful in predicting deforestation. Second, they increase the reliability of the estimates of the other coefficients by getting rid of part of the bias caused by the omitted variable (i.e., the spatial lag of deforestation). A quick inspection of the results in Table 4 and Table 6 illustrates that the magnitude of the coefficients of most of the explanatory variables is higher in the non-spatial case than it is after the spatial effect is accounted for, contrary to the observation by Boubacar (2012). We return to this in the discussion.

5 Discussion

Deforestation is a complex phenomenon, and its causes differ depending on, *inter alia*, the scale of analysis. This study explores drivers of deforestation in Indonesian districts, which makes the choice of potential predictors limited, in the context of the framework in Angelsen and Kaimowitz (1999), to structural and aggregate socioeconomic indicators as opposed to immediate causes and characteristics of agents. Three aspects have been assessed and are discussed below: choosing between spatial and non-spatial models, the nature of spatial interdependence, and the forest transition (FT) hypothesis.

5.1 Non-spatial or spatial

Before estimating the spatial panel models, we began with the standard non-spatial panel data models, and explored and found evidence, which justifies inclusive spatial models. The Moran's I test suggests significant spatial interaction effect and reducing the model to its non-spatial version will inevitably render unreliable results, as inferences will be based on erroneous standard errors. This is in line with similar findings in the literature (Boubacar, 2012; Hargrave and Kis-Katos, 2012; Wheeler et al., 2013). The implication is that excluding the spatial lags from the model would lead to erroneous inferences about the effect of the other variables in the regression. The magnitude of the coefficients of these variables has *decreased* in the spatial models of this paper, which is in contrast to what Boubacar (2012) observed, namely that non-spatial models underestimated the effects of the explanatory variables. Also, Mets et al. (2017) argue that spatial autocorrelation may lead to spurious effects of the exogenous variables. Overall, we cannot generalize that non-spatial models under- or over-estimate the effects of covariates in the presence of autocorrelation, but that the non-spatial models suffer from omitted variable bias seems evident.

Our findings support the hypothesis that patterns of forest cover change have a strong tendency of spatial clustering, i.e., there is positive autocorrelation of deforestation among neighboring districts. This pattern is also partly visualized in Figure 1. Spatially explicit models have recently been applied to not only better understand where and why deforestation occurs but also to improve the design, implementation and evaluation of incentivized mitigation policies such as REDD+ (Busch and Ferretti-Gallon, 2017). There is consistent support for the fact that spatial models provide better fit to data on deforestation and its potential drivers, be it at national (Boubacar, 2012) or subnational (Hargrave and Kis-Katos, 2012) levels of analysis. The lesson for future similar research is that it pays off to utilize available spatial data fully.

5.2 The spatial dependence

Once we find sufficient evidence that a spatial model must be fitted, the next question is about the form of the spatial econometric model to be used. The spatial interaction may occur through deforestation, other explanatory variables, the error terms, or combinations of those. When deforestation in a given district affects (and is affected by) deforestation in its neighboring districts, there is endogenous interaction effect. This was the strongest effect found, and it has the most salient implication for REDD+. The effect is positive, which may be a good or bad news depending on the trend of deforestation in neighboring districts. It is a bad news if the given district shares borders with deforestation hotspots. Deforestation is contagious, and it spreads across borders.

The positive correlation may, nevertheless, be good news as districts can reap synergistic effects in their efforts to reducing deforestation. There are positive externalities of conservation, much like the role of vaccination in lowering the risk for those not treated. This finding also suggests that we have *not* identified any problem of leakage at our scale of analysis, i.e., that conservation of forests in one district leads to more deforestation in neighboring districts.

Generally, the spatial interaction effects need not be limited to this strong autoregressive relationship, as there may be spatial correlation in the explanatory variables—a spatial interaction effect which is exogenous—and in the error term. However, we did not find sufficient empirical evidence that the spatial lags of either GDP or population density explains deforestation. Likewise, our findings show that there is no significant autocorrelation in the error terms of neighboring districts.

The answer to the question of which spatial model fits our data well is the SAR model. Given the panel structure of the data, either random effects or fixed effects versions of the SAR model can be estimated. However, the latter does not return coefficient estimates for most of the right hand side variables (except for the spatial interaction terms). This is probably because the variables are mostly aggregate values with limited variation to either over time or across districts but not both. In this study, results from the random effects versions of the spatial models are reported, among which the other post-estimation tests indicate that SAR model gives the best fit to the data at hand.

From the SAR model, we draw two important lessons. First, focusing on socioeconomic factors alone, as it has conventionally been the case in many non-spatial empirical studies, will not yield a reliable model that helps us to understand the dynamics in deforestation. Results from non-spatial econometric analyses should be scrutinized, as the conclusions will potentially lead us astray. Second, we have evidence that the spatial weighted average of deforestation is a strong predictor of deforestation. It is also robust to variations in the spatial model specification as well as the combination of other explanatory variables. The magnitude of the effect is also larger than the neighborhood effect that Boubacar (2012) found in his country level spatial panel analysis. As deforestation takes place more locally, its neighborhood effect is also likely to be felt among spatial units at a finer scale than among countries that have distinct policies and legal barriers to limit input flows.

The mechanism through which this strong spatial interaction effect occurs is not self-evident, nor can it be tested with the data we have. One explanation from the literature points to the social interaction among people belonging to similar ethnicity in neighboring countries may be behind the high forest depletion in Sub-Saharan Africa (Boubacar, 2012). However, this may not necessarily be the case in Indonesia. Within the confines of a country, subnational units are likely to interact more directly in their economic decisions. There is higher socioeconomic interdependence among subnational units than among countries, and thus stronger spatial interaction effect as well. Transmigration of people, as in such a policy set by the Indonesian government in the 1980's and 1990's to move people from Java to other islands, may also explain the spatial diffusion of deforestation.

The diffusive nature of deforestation may be explained in terms of legality and norms. Deforestation is often semi-legal, and it might be consider more acceptable and less risky if it is widespread. We also speculate that the effects of some variables not included because of lack of data might be reflected in the strong spatial autocorrelation. For example, indicators of infrastructure such as roads and market connectivity (not included in our formal analysis) may facilitate the spatial interaction.

5.3 The forest transition hypothesis

The forest transition hypothesis can be tested in different ways. One way would be to map forest cover to a long series of time points, where a U-shaped trajectory confirms a transition from high forest & low deforestation, to high deforestation and eventually low deforestation and forest recovery. This is the route that the original work by Mather (1990) followed, but requires forest cover time series for decades or – preferably – centuries. Another way, which follows the rationale of the Environmental Kuznets Curve, is to explore how forest cover is correlated to income, which permits the use of cross-sectional data (Culas, 2012). A third way, which this study makes use of, is to include quadratic terms—forest cover and forest cover squared—in a regression of forest cover change as the dependent variable.

The results of the SAR model show that the coefficients of forest area and forest area squared have statistically significant correlation with deforestation. Similar to the findings in Angelsen et al. (2013), we find a bell-shaped relationship between deforestation and forest area (both measured as share of total district land). These results are suggestive of the existence of forest transition, in line with arguments in the literature (Rudel, 1998; Rudel et al., 2005). The concept of forest transition is an 'empirical regularity' (Angelsen and Rudel, 2013) instead of a well-established theory. As such, it prevails over a long time, i.e., several decades. The time dimension of this study may not be long enough to provide the ultimate test of the existence of a long-term forest transition trajectory, but the findings are in line with the FT hypothesis.

6 Conclusion

Knowing where and how deforestation takes place is a part of the endeavor to mitigate it. We revisited the old problem of deforestation, aiming to identify potential drivers using district-level balanced spatial panel data from Indonesia. We draw three important conclusions. First, we found sufficient—and consistent with several findings in the literature—evidence that spatial models are superior to their non-spatial alternatives. This is a good news for predicting deforestation and thus setting reference levels. Second, the spatial interaction occurs through the dependent variable—deforestation is contagious. The converse is that there are synergistic gains from forest conservation efforts among neighboring districts. In the context of REDD+, this is a good news as there is room for positive leakage. Lastly, we found a forest transition pattern, which is important for policy differentiation (Angelsen and Rudel, 2013). It is also useful for predicting deforestation and setting reference levels in REDD+.

References

Aguiar, A. P. D., Câmara, G., and Escada, M. I. S. (2007). Spatial statistical analysis of land-use determinants in the brazilian amazonia: Exploring intra-regional heterogeneity. *Ecological Modelling*, 209(2):169–188.

Aguilar-Støen, M., Angelsen, A., and Moe, S. R. (2011). Back to the forest: Exploring forest transitions in candelaria loxicha, mexico. *Latin American Research Review*, 46(1):194–216.

Allen, J. C. and Barnes, D. F. (1985). The causes of deforestation in developing countries. *Annals of the association of American Geographers*, 75(2):163–184.

Angelsen, A., Ainembabazi, J., Bauch, S., Herold, M., Verchot, L., Hänsel, G., Schueler, V., Toop, G., Gilbert, A., and Eisbrenner, K. (2013). Testing methodologies for REDD+: Deforestation drivers, costs and reference levels. Report, Ecofys.

Angelsen, A. and Kaimowitz, D. (1999). Rethinking the causes of deforestation: lessons from economic models. *World Bank Res Obs*, 14(1):73–98.

Angelsen, A. and Rudel, T. K. (2013). Designing and implementing effective redd+ policies: A forest transition approach. *Review of Environmental Economics and Policy*, 7(1):91–113.

Anselin, L. (1988). Model validation in spatial econometrics: a review and evaluation of alternative approaches. *International Regional Science Review*, 11(3):279–316.

Anselin, L. (2013). *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media.

Arellano, M. (2007). Comments on: Panel data analysis—advantages and challenges. *Test*, 16(1):23–27.

Babigumira, R., Angelsen, A., Buis, M., Bauch, S., Sunderland, T., and Wunder, S. (2014). Forest clearing in rural livelihoods: Household-level global-comparative evidence. *World Development*, 64:S67–S79.

Belotti, F., Hughes, G., and Mortari, A. P. (2016). Xsmle: Stata module for spatial panel data models estimation. *Statistical Software Components*.

Boubacar, I. (2012). Neighboring effects of deforestation: a spatial econometric approach. *Environmental Economics*, 3(3).

Burgess, R., Hansen, M., Olken, B. A., Potapov, P., and Sieber, S. (2012). The political economy of deforestation in the tropics. *The Quarterly Journal of Economics*, 127(4):1707–1754.

Burnett, W. J., Bergstrom, J. C., and Dorfman, J. H. (2013). A spatial panel data approach to estimating u.s. state-level energy emissions. *Energy Economics*, 40:396–404.

Busch, J. and Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? a meta-analysis. *Review of Environmental Economics and Policy*, 11(1):3–23.

Clark, M. (2012). Deforestation in madagascar: Consequences of population growth and unsustainable agricultural processes. *Global Majority E-Journal*, 3(1):61–71.

Cliff, A. and Ord, K. (1970). Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, 46(sup1):269–292.

Combes Motel, P., Pirard, R., and Combes, J. L. (2009). A methodology to estimate impacts of domestic policies on deforestation: Compensated successful efforts for "avoided deforestation" (REDD). *Ecological Economics*, 68(3):680–691.

Culas, R. J. (2012). REDD and forest transition: Tunneling through the environmental kuznets curve. *Ecological Economics*, 79:44–51.

Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels.* Springer.

Foster, J., Greer, J., and Thorbecke, E. (2010). The foster–greer–thorbecke (fgt) poverty measures: 25 years later. *Journal of Economic Inequality*, 8(4):491–524.

Geist, H. J. and Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. *BioScience*, 52(2):143–150.

GFC, G. F. C. (2010). Getting to the roots: Underlying causes of deforestation and forest degradation, and drivers of forest restoration. *Global Forest Coalition, Amsterdam, Netherlands*.

Hargrave, J. and Kis-Katos, K. (2012). Economic causes of deforestation in the brazilian amazon: A panel data analysis for the 2000s. *Environmental and Resource Economics*, 54(4):471–494.

IPCC (2013). Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Report.

Köthke, M., Leischner, B., and Elsasser, P. (2013). Uniform global deforestation patterns — an empirical analysis. *Forest Policy and Economics*, 28:23–37.

Li, M., Wu, J., and Deng, X. (2013). Identifying drivers of land use change in china: A spatial multinomial logit model analysis. *Land Economics*, 89(4):632–654.

Mather, A. S. (1990). *Global forest resources*. Belhaven Press.

Mets, K. D., Armenteras, D., and Dávalos, L. M. (2017). Spatial autocorrelation reduces model precision and predictive power in deforestation analyses. *Ecosphere*, 8(5).

Meyfroidt, P., Rudel, T. K., and Lambin, E. F. (2010). Forest transitions, trade, and the global displacement of land use. *Proceedings of the National Academy of Sciences*, 107(49):20917–20922.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

Perz, S. G. and Walker, R. T. (2002). Household life cycles and secondary forest cover among small farm colonists in the amazon. *World Development*, 30(6):1009–1027.

Portela, R. and Rademacher, I. (2001). A dynamic model of patterns of deforestation and their effect on the ability of the brazilian amazonia to provide ecosystem services. *Ecological Modelling*, 143(1–2):115–146.

Robalino, J. A. and Pfaff, A. (2012). Contagious development: Neighbor interactions in deforestation. *Journal of Development Economics*, 97(2):427–436.

Rudel, T. K. (1998). Is there a forest transition? deforestation, reforestation, and development. *Rural sociology*, 63(4):533–552.

Rudel, T. K., Coomes, O. T., Moran, E., Achard, F., Angelsen, A., Xu, J., and Lambin, E. (2005). Forest transitions: towards a global understanding of land use change. *Global Environmental Change*, 15(1):23–31.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46(1):234–240.

Wheeler, D., Hammer, D., Kraft, R., Dasgupta, S., and Blankespoor, B. (2013). Economic dynamics and forest clearing: A spatial econometric analysis for indonesia. *Ecological Economics*, 85:85–96.

Appendices

Table 7: Information criteria

Information criterion	SDM	SAR	SAC
AIC	3396.24	3394.92	3396.32
BIC	3484.27	3472.69	3479.72

Table 8: Results of SDM, SAR and SAC models with distance based weighting matrix

Variables	SDM	SAR (RE)	SAC(RE)
Explanatory variables:			
Population density (log people/km2)	-0.320***	-0.364***	-0.374***
	(0.097)	(0.097)	(0.100)
GDP (log)	0.143	0.121	0.122
	(0.110)	(0.109)	(0.112)
Share of agricultural GDP	-0.006	-0.009	-0.012*
	(0.006)	(0.006)	(0.006)
Poverty gap index	0.032***	0.037***	0.034***
	(0.012)	(0.012)	(0.012)
Literacy rate (% of literate people age$\geq$ 15)	-0.023**	-0.031***	-0.024**
	(0.010)	(0.010)	(0.011)
Access to electricity (% households)	0.013*	0.010	0.012*
	(0.007)	(0.007)	(0.007)
Precipitation (mm)	0.003	0.00136	0.004
	(0.00227)	(0.00226)	(0.002)
Temperature (°C)	0.064	0.089	0.10
	(0.097)	(0.095)	(0.097)
Forest area	0.042***	0.040***	0.045***
	(0.015)	(0.015)	(0.015)
Forest area squared	-0.001***	-0.001***	-0.001***
	(0.0002)	(0.0002)	(0.0002)
Year dummies:			
2006	-0.044	-0.045	0.158
	(0.231)	(0.231)	(0.390)
2009	-0.610**	-0.545**	0.306
	(0.278)	(0.257)	(0.411)
2012	-0.141	-0.140	-0.357
	(0.292)	(0.259)	(0.414)
Spatial terms:			
W*Deforestation	1.666***	1.596***	0.633***
	(0.099)	(0.091)	(0.129)
W*population density	0.085		
	(0.390)		
W*GDP	-0.262		
	(0.273)		
W*Error			0.504***
			(0.174)
sigma_u	0.656***	0.681***	0.700***
	(0.140)	(0.137)	(0.136)
sigma_e	2.174***	2.183***	2.186***
	(0.0648)	(0.0650)	(0.0652)
Constant	-0.147	-0.596	-1.166
	(2.764)	(2.724)	(2.845)
Observations	760	760	760
Number of groups	190	190	190

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9: Results of SDM, SAR and SAC models on a bigger sample (Bali and Java included)

Variables	SDM	SAR (RE)	SAC(RE)
Explanatory variables:			
Population density (log people/km2)	-0.221**	-0.260***	-0.265***
	(0.090)	(0.075)	(0.077)
GDP (log)	0.021	-0.027	-0.024
	(0.094)	(0.085)	(0.087)
Share agricultural GDP	-0.013**	-0.016***	-0.016***
	(0.005)	(0.005)	(0.005)
Poverty gap index	0.025**	0.029***	0.030***
	(0.011)	(0.011)	(0.011)
Literacy rate (% of literate people age$\geq$ 15)	-0.019**	-0.020**	-0.020**
	(0.009)	(0.008)	(0.008)
Access to electricity (% of households)	0.005	0.003	0.003
	(0.006)	(0.005)	(0.005)
Precipitation (mm)	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)
Temperature (°C)	0.068	0.091	0.087
	(0.080)	(0.079)	(0.081)
Forest area	0.027**	0.025**	0.025**
	(0.011)	(0.011)	(0.011)
Forest area squared	-0.0004***	-0.0004***	-0.0004***
	(0.0001)	(0.0001)	(0.0001)
Year dummies:			
2006	0.118	0.131	0.133
	(0.189)	(0.189)	(0.195)
2009	0.657***	0.707***	0.731***
	(0.201)	(0.199)	(0.211)
2012	-0.148	-0.085	-0.091
	(0.206)	(0.204)	(0.210)
Spatial terms:			
W*Deforestation	0.459***	0.453***	0.425***
	(0.050)	(0.049)	(0.074)
W*population density	-0.143		
	(0.117)		
W*GDP	0.048		
	(0.084)		
W*Error			0.052
			(0.099)
sigma_u	0.543***	0.558***	0.571***
	(0.135)	(0.132)	(0.132)
sigma_e	2.272***	2.273***	2.274***
	(0.054)	(0.054)	(0.054)
Constant	0.821	0.924	1.040
	(2.251)	(2.254)	(2.291)
Observations	1,204	1,204	1,204
Number of groups	301	301	301

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Chapter 3:
Performance payments and reference levels

Performance payment and reference levels: a framed field experiment on forest conservation*

Amare Teklay Hailu[1] and Arild Angelsen[1,2]

[1]School of Economics and Business, Norwegian University of Life Sciences (NMBU), P.O.Box 5003, 1432 Ås, Norway
[2]Center for Internatioanl Forestry Research, Bogor, Indonesia

Emails: amare.teklay.hailu@nmbu.no; arild.angelsen@nmbu.no

Abstract

Reference level (RL) is a key element in designing Payment for Environmental Services (PES), potentially an important tool for Reducing Emissions from Deforestation and forest Degradation (REDD+). We conducted a framed field experiment (FFE) in Ethiopia to assess the effect of different RLs on forest extraction, in a way that resembles the policy discussion on alternative REDD+ and PES design. Average pre-treatment harvest was half the Nash equilibrium, which suggests pro-environmental and/or pro-social preferences. The PES treatment reduced harvest by ca. 20 percentage points. While the magnitude of reduction is comparable across treatment groups, we found that the effect of high/low RL depends on the payment modality. With individual pay, the incentive effect of higher RL seems to dominate, giving a larger conservation effort. With group pay, a lower RL seems to invoke group dynamics and a strong anchoring effect, where the RL appears to serve as a norm and as an aspirational target.

Keywords: Framed field experiment, PES, REDD+, common pool resources, reference level
JEL Classifications: Q23, C93, Q57

*We would like to acknowledge funding from the Norwegian Agency for Development Cooperation [agreement no. GLO-3945 QZA 13/0545; grant on Learning from REDD+ an enhanced Global Comparative Analysis] through the Center for International Forestry Research (CIFOR).

1 Introduction

Designing Payment for Environmental Services (PES) schemes for Reducing Emissions from Deforestation and forest Degradation (REDD+) requires setting forest reference (emission) levels (FRL/FREL - henceforth simply RL). These serve as benchmark against which future performance is measured and/or financial transfers effected. The core challenge is to identify the counterfactual or business-as-usual (BAU) scenario, partly because predicting future deforestation is inherently difficult, and partly because of scant data. Despite numerous proposals on how RLs should be set, there is limited empirical evidence about the impact of different RLs on actual performance (Agrawal et al., 2011; Sheng et al., 2016).

We developed and implemented a framed field experiment (FFE) that portrays a realistic PES scheme. We observed forest extraction in a first, baseline session of five rounds, and then randomly set one of three alternative RLs in the second session of the experiment: historical (first session, pre-treatment) average, above historical average, and below historical average. In this way, we assess the effect of setting different RLs on forest conservation through PES.

We pose three questions. First, how does the RL affect performance? Tighter (lower) RLs limit the potential for compensation, as they are more difficult to meet and—if met—lower the payment for any given forest use. However, it is also possible that low RLs enhance effort through 'anchoring effects' and by serving as 'aspirational targets'. Second, (how) do these effects differ by the pay modality, i.e., whether the payment is based on individual or collective performance? Third, which type of RL is most cost efficient from the perspective of a principal (government agency or donor), i.e., gives the largest reduction (avoided harvest) per Ethiopian Birr paid to the participants. A high RL provides stronger incentives for participation and reduction, but is also costlier as the point from which payments are made is higher.

Setting RLs is a critical element of any performance based incentive system, which REDD+ intends to be. This paper contributes to the literature on setting RLs and brings a new perspective to this debate. As far as we know, this is the first paper to report on experimental tests of the impacts of varying RLs on forest conservation.

In the next section, we highlight the nature and purpose of RLs in REDD+, and present a formal model on the links between RLs and the common pool resource (CPR) problem, based on both neoclassical and behavioral perspectives. Section 3 describes the sampling and experimental design. Section 4 presents the main findings, which are then discussed in section 5. Section 6 concludes.

2 Reference levels and incentives

2.1 Reference levels in REDD+

The United Nations Framework Convention on Climate Change (UNFCCC), by decision 1/CP.16 paragraph 71(b), calls on developing countries to contribute to Reducing Emissions from Deforestation and forest Degradation, to conservation and enhancement of forest carbon stocks, and to sustainable management of forests (REDD+) as part of the global climate mitigation action. One of the four requirements under the Warsaw framework for REDD+ (UNFCCC, 2013) is the submission of a forest reference emission level or forest reference level (FREL/FRL). This should serve as a benchmark for both assessing its performance in implementing the activities pursued, and for performance-based or result-based payments (UNFCCC, 2010, 2012).

There are different meanings of the term 'reference level', depending on its purpose (Angelsen, 2008). First, it might refer to the BAU-scenario or counterfactual, used to measure the impact of a policy. Second, it might also refer to the benchmark for result-based payments, sometimes referred to as the 'financial incentive benchmark' or 'crediting baseline'. This might be set above or below the BAU-scenario, for example, based on the UNFCCC principle of "common but differentiated responsibilities". In the UNFCCC process, negotiators have opted not to make this distinction, as it will—in the words on one negotiator—"open up a can of worms" (pers.com during COP 19, 2013).

The main approach taken by UNFCCC is thus that FRELs/FRLs should represent the best estimate of future deforestation (and possibly also forest degradation) without a REDD+ policy, i.e., a forecast of a counterfactual, and that this also should form the basis for any future result-based payments.

There is also a third interpretation. According to the UN-REDD Programme (2014), RLs are also used to express countries' contribution to REDD+ actions under the UNFCCC. Linked to that, RLs have been referred to as "aspirational targets".[1] Clearly, these three meanings of RL (BAU, crediting baseline, aspirational target) are conceptually very different, and any analysis must make clear in which meaning the term is used.

Historical emissions are a fundamental component of a BAU forecasting exercise. It is also recognized, given the dearth of data in most developing countries, that a step-wise approach is useful (Herold et al., 2012; UNFCCC, 2012). As the future may not necessarily repeat the past, countries may also adjust their RL in accordance with their national circumstances. An exact definition of 'national circumstances' is not provided, but countries need to argue their case for adjustments of historical averages. By the end of 2017, 25 countries have submitted their FRELs/FRLs[2]. Almost all use historical averages as the point of departure, but with varying length of the historical period. Many have also adjusted upwards for various national circumstances. In a separate project, we analyze the possibility of 'gaming' behavior in the submissions, i.e., a tendency to submit unrealistically high RLs to obtain more funding, or to demonstrate success in their REDD+ efforts.

In local REDD+ projects with PES components, there are also guidelines and standards for how to set RLs. The most widely used is the Verified Carbon Standard (VCS, 2016), whose guidelines on baseline setting resembles the UNFCCC guidelines of (adjusted) historical averages. The debate on RLs, particularly in global fora, has been dominated by practical and political considerations, with hardly any debate on the incentive effects and cost efficiency of setting different RLs, neither at country nor project level. This motivates our paper.

2.2 The CPR model

We adapted the common pool resources (CPR) model of Rodriguez-Sickert et al. (2008) to guide our experimental design. Let N be the number of players who are endowed with a common-pool resource stock of trees, X_s. Each player $i\epsilon\{1, 2, \ldots, N\}$ decides to extract $x_{it}\epsilon\{0, 1, \ldots, x^{max}\}$ in round t. Let p and q represent the marginal private benefit of a

[1] Personal observation (AA) during UNFCCC meetings.

[2] http://redd.unfccc.int/fact-sheets/forest-reference-emission-levels.html

harvested tree and the collective benefit of a standing tree, respectively. Decisions are made privately (independently) and simultaneously. Collective benefits are shared equally among the players. Thus the monetary payoff (π_{it}) of each participant is:

$$\pi_{it} = px_{it} + \frac{q}{N}(X_s - x_{it} - \sum x_{-it}); \quad \text{where} \quad x_{it} \leq x^{max} \tag{1}$$

For simplicity, we do not include the cost of extraction in Eq. (1), but the maximum technical limit, x^{max} can be interpreted to partly reflect extraction costs. A rational and well-informed *homo economicus* tries to maximize profit by deciding his level of extraction, x_{it}. This formulation represents a social dilemma when the parameters are such that $\frac{q}{N} < p < qN$. It is rational for each a *homo economicus* user to harvest the maximum allowable amount, which leads to the well-known 'tragedy of the commons' (Hardin, 1968). The socially optimal level of harvest, i.e., the level which maximizes group (sum of individual) payoff, is when each player harvests zero. The overall collective benefit from a standing tree (qN) exceeds the marginal private benefit to an individual of harvesting a tree (p), but each individual receives only $\frac{1}{N}$ of that benefit ($\frac{q}{N}$) and therefore prefers to harvest (assuming others' harvest as given).

2.3 RL treatments

The social dilemma can be addressed by different treatments. PES seeks to solve the dilemma by compensating resource users for the forgone private benefit. Measuring individual performance relative to the benchmark (RL), we introduce a PES scheme as follows:

$$\pi_{it} = px_{it} + \frac{q}{N}(X_s - x_{it} - \sum x_{-it}) + \text{Max}\{r(x^{RL} - x_{it}), 0\}; \quad \text{where} \quad x_{it} \leq x^{\max} \tag{2}$$

The last term in Eq. (2) puts the condition that harvest x_{it} must be below RL (x^{RL}) in order for player i to qualify for PES reward (r) per unit. A player would find the scheme attractive on the margin if the PES reward plus his share of the collective benefit is above or equal to the harvest benefit: $\frac{r+q}{N} \geq p$.

Another variant of the PES scheme would be to make the reward at the group level. This has advantages in terms of lower transaction costs (not included in our formulation). This group PES modality gives the following payoff:

$$\pi_{it} = px_{it} + \frac{q}{N}(X_s - x_{it} - \sum x_{-it}) + \text{Max}\{r(x^{RL} - \frac{\sum_{i=1}^{N} x_{it}}{N}), 0\}; \quad \text{where} \quad x_{it} \leq x^{\max} \tag{3}$$

In the group-level version of PES, the requirement is that the average group harvest is below the benchmark in order to qualify for PES. A player will not only compare the marginal gains and losses of harvesting, but also decide based on expectations about the group condition being met. Thus, group level pay introduces an element of uncertainty about how much other members will harvest and whether the condition will be met (see Chapter 3).

Formally, let f_{it} be the subjective probability of player i at time t that average harvest is below the benchmark. The decision problem extends to a comparison at the margin of the certain outcome p and the uncertain outcome, $\frac{q}{N} + f_{it}r$. In a setting with rational, payoff maximizing players, to achieve the same conservation target, these lower expectations about the group performance should be matched with higher PES levels.[3]

This situation corresponds to experiments of public goods provision with a threshold, which has a long history in the experimental literature (e.g., Cadsby and Maynes (1999) and Dickinson (1998)). A high threshold lowers contributions, unless there is a 'money back guarantee', which is not relevant in our experimental setting. High rewards increase contributions and elicit convergence of contributions to the threshold (Cadsby and Maynes, 1999).

2.4 Extending the basic model

Unlike the standard economics prediction that resource users maximize own payoff, there is strong empirical evidence that *homo sapiens* do extract below the Nash equilibrium in CPR games. For example, Cardenas (2000) and Rodriguez-Sickert et al. (2008) found in the open access case—comparable to our baseline game—just 55 and 58% of the Nash equilibrium (max. allowable harvest). People have social and environmental preferences, which can modify or even reverse some of the predictions based on *homo economicus* assumptions. To include pro-social and pro-nature preferences, we borrow from the conceptual framework outlined in Levitt and List (2007). The central idea is that the utility function of individuals has both monetary and non-monetary components. It is assumed, for tractability, that these are additively separable.

The utility function without PES is given in Eq. (4). A utility-maximizing individual i chooses x_{it} based on not only the private and common pecuniary benefits, but also the non-pecuniary, pro-social and pro-nature preferences. The first part of Eq. (4) is the reduced form of Eq. (1) above, capturing the wealth effect (W). The second part of the equation is the moral payoff effect (M) (Handberg and Angelsen, 2016).

$$U_{it}(p, x_{it}, q, N, \sum x_{-it}) = W_{it}(x_{it}; \sum x_{-it}, p, q, N) + M_{it}\{\frac{x_{it}}{\frac{\sum x_{-it}}{N-1}}, x_{it} + \sum x_{-it}; p, q, N\} \quad (4)$$

We hypothesize that at least two effects are important in the moral payoff. First, a player is concerned about how own harvest compares to the average harvest of other group members. This relates to the social preference of individual i, where higher x_{it} relative to others' harvest can invoke a feeling of either guilt (Lopez et al., 2012; Rodriguez-Sickert et al., 2008) or retribution, while harvesting less can invoke feelings of either warm-glow (Andreoni, 1990) or betrayal, depending on their personal characteristics as well as the social setting.

Second, a player may also have preferences for the overall harvest, which determines both the group payment (social preferences) and the remaining forest stock (environmental preferences) (Handberg and Angelsen, 2016).

By introducing PES, we are throwing in additional factors to the utility function of the players. These factors include the benchmark against which we seek to measure performance, x^{RL}, the marginal reward per unit of reduction below the benchmark, r, and, in the case

[3] A more complete analysis should also include an analysis of behavior under risk, including that risk aversion makes a player demand even higher payment to compensate for the risk being introduced with group level payment.

of group pay, one's belief about the likelihood that the group average will be below the benchmark, f_{it}. These are given in Eq. (5) for individual pay and Eq. (6) for group pay.

$$U_{it}(.) = W_{it}(x_{it}; \sum x_{(-it)}, p, q, N) + M_{it}\{\frac{x_{it}}{\frac{\sum x_{-it}}{N-1}}, x_{it} + \sum x_{-it}; p, q, N, r, x^{RL}\} \tag{5}$$

$$U_{it}(.) = W_{it}(x_{it}; \sum x_{(-it)}, p, q, N) + E[M_{it}\{\frac{x_{it}}{\frac{\sum x_{-it}}{N-1}}, x_{it} + \sum x_{-it}; p, q, N, r, x^{RL}\}] \tag{6}$$

With PES, there is a potential wealth effect (if $x_{it} < x^{RL}$ in individual pay and $\frac{\sum x_{it}}{N} < x^{RL}$ in group pay) and the expected value of the moral payoff depends on how the individual perceives the size of x^{RL} and the way it has been set. It may also change the expectations about other players' harvest. In Eq. (6), the probability that the group will reach x^{RL} is relevant as each player is not informed about others' decision and hence cannot know beforehand whether the group will qualify, no matter how little one harvests unilaterally. As such, the parameters r, x^{RL}, and f_{it} will not only determine the expected wealth effect of the PES but also the moral payoff through the individual's perception.

The magnitude of the RL also has wealth effect as it determines whether it is costly to reach it and qualify for PES. The higher x^{RL} is relative to individual historical average, the higher its wealth effect will be, while the moral payoff is indeterminate, *a priori*.

Finally, participants may also succumb to anchoring (Tversky and Kahneman, 1974) their choice to x^{RL}. The anchoring effect is among the most robust cognitive heuristics ('mental shortcuts'), and documented in experiments in a variety of domains (Furnham and Boo, 2011). The effect can be caused by several mechanism, but one dominant explanation is confirmation bias, where participants engage in a "confirmatory search mechanism" (Ickowitz, 2006). Relatedly, Strack and Mussweiler (1997) argue that anchors can serve to adjust the boundaries of the range of plausible values.

Related to our experiment, the assigned RL might thus be interpreted by participants as an upper boundary for what is acceptable use. In our PES experiment, this *normative signal* can potentially be very strong as participants are told that they will only receive payment if their (individual or group) harvest is below the RL.

It is interesting to note the parallel between the usage of RL as an 'aspirational target' in the UNFCCC debate (section 2.1) and as an 'anchor' in behavioural economics.

2.5 Hypotheses

The combinations of RL formula (above, historical and below), its level (from one to five) and pay modality (individual vs. group) give rise to different treatments which may lead to variations in responses. Irrespective of the nature of the group in the first stage, the fact that historical performance is used to condition future rewards may entail different behavioral responses.

Based on the above discussion, we test the following hypotheses:

H1: Setting RLs below historical average leads to more extraction relative to RLs at or

above historical average.
H2: The effect in H1 is more pronounced in the case of group level pay, due to the uncertainty of the group reaching the RL.
H3: Setting the RL below historical average is less costly compared with setting it either at historical or above historical average.

3 Materials and methods

3.1 Sampling

A framed field experiment (FFE) was conducted in three zones of the state of Tigray, Ethiopia, in February-May 2016. We selected relatively forest rich zones: Western, North Western, and Southern zones. Within each zone, three *Woredas* (districts) were identified, and then within each *Woreda*, three *kebelles* (peasant associations, the administrative unit at the community level) were purposively selected to make sure that there is natural forest within a walking distance from the kebelle centers. A list of household heads was obtained from each *kebelle* administration, and a sample of 48 household heads was selected randomly within each *kebelle*.

Both men and women participate in forest use (decision) in a typical village in rural Ethiopia, but with some differences in the tasks: men typically are more engaged in clearing forests for agriculture while women usually collect firewood. As the list of resident household heads in a typical *kebelle* consists of both male and female heads of households, we sent invitation letters directly to the name recorded in the household list. About 72% of the participants were men. Participants were randomly assigned to groups of eight people, which were again randomly given: (i) individual or collective pay treatment and (ii) one of the three reference level formulas during the treatment rounds. At the end of the session, all participants were interviewed about their socioeconomic characteristics, actual forest use as well as their experience in and perception of the experiment (see the Appendix to Chapter 1).

3.2 Framing

Since the pioneering experimental work on CPR and social dilemmas by Ostrom et al. (1994) and Walker et al. (1990), several innovative approaches have been used to increase the internal and external validity of experimental results. Compared to lab and artefactual field experiments, FFEs stand out as having realism and validity to resemble the real world forest use of participants (Harrison and List, 2004). For instance, FFEs such as Cardenas et al. (2000) and Handberg and Angelsen (2015) took this one step further and improved the framing in a way that makes the experiment closer to real-life decisions. In the latter study, paper trees were used instead of tables of outcomes in the former.

In our study, we followed this tradition and conducted the experiment in the villages where participants make their real-life decisions of how many trees to cut. We replaced paper trees of Handberg and Angelsen (2015) with real tree branches. In each session, participants were endowed with a forest consisting of 60 real dry tree branches[4], each approx. 50 cm long.

[4] In pretesting, we also used forest stocks of 40 and 80. Setting the stock to 40 meant that the forest could be completely exhausted in one round ($5 * 8$), and this seemed to have an effect in terms of

Sitting in a circle, participants were asked to imagine that each tree branch represented an actual standing tree in a common forest near their village.

3.3 The experiment

The participants played a basic CPR game in groups of eight, each having the same information and all groups endowed with 60 branches of real trees. It is a one-shot game repeated 10 or 15 times, and participants were made to indicate how many trees they want to extract. Communication was not allowed. We announced group total harvest at the end of each round. Their decisions were simultaneously made, but the experimental history can be taken into account in participants' decisions in subsequent rounds. The information provided to the participants is given in the Appendix to Chapter 1.

The payoff for each combination of own and others' harvest to a given participant was determined based on Eq. (1). We calibrated the parameters as follows: p=1, q=2 and $X^{max} = 5$, and N=8. The conditions for a social dilemma are thus met (section 2.2). We described this and other conditions in a detailed instruction in the local language, and showed examples to ensure that all participants have similar level of understanding about the game.

Each harvested tree gives a constant private benefit of one Ethiopian Birr (ETB)[5], while the public benefit from each standing (not harvested) tree is 2 ETB, shared equally among the group members. This generates the social dilemma. A living tree generates a collective benefit twice the individual benefit of a harvested tree, but the individual benefit of a saved tree is only $\frac{1}{4}$ of a harvested tree. The dominant strategy is, therefore, to harvest the maximum allowable of five trees, which results in a symmetric Nash equilibrium total harvest of 40 trees where each participant earns 10 ETB each $(5 + 2(\frac{20}{8}))$. The socially optimal solution (maximizing aggregate payoff) is achieved when each player harvests nothing and earns 15 ETB $(2(\frac{60}{8}))$.

3.4 The treatments

We introduced two kinds of treatments before the sixth round. First, we varied the payment modality between individual and collective pay, the effect of which has been reported in Hailu and Angelsen (2017) (Chapter 4). Second, and the focus of this paper, we set three RLs in order to compare alternative proposals for setting RLs. The RLs were set either at the group level average of the first five rounds (termed 'historical average'), or at historical average plus/minus one. The treatment distribution of the 54 groups is presented in Table 1.

lowering the group's harvest.

[5] 1 ETB was aprox. 0.046 USD at the time of experiment (between February and May, 2016).

Table 1: Distribution of groups in different combinations of RL and pay modality treatments.

RL	PES system	No. of sessions	No. of participants
Above (H+1)	Individual	8	64
	Group	8	64
Historical (H)	Individual	11	88
	Group	11	88
Below (H-1)	Individual	8	64
	Group	8	64

The number of sessions where historical average was used as an RL is 22 and is larger than each of the other types of RL because we want to use this subsample for another study on crowding in/out effects of PES (see Chapter 5).

Once the participants knew the RL, the PES component of the payoff was added to that in the baseline stage. Besides the private benefit from harvested trees and collective benefit from standing trees, the payoff consisted of this time the reward computed as a product of $r = 0.75$ and the difference between the RL and own harvest for individual pay (or group average in the case of collective pay). Let the baseline RL be X_h^{RL} where the subscript h stands for historical. This is set at the group level historical average of the pre-treatment five rounds, as in Eq. (7).

$$X_h^{RL} = \frac{\left(\sum_{(t=1)}^{5} \sum_{(i=1)}^{8} x_{it}\right)}{40} \tag{7}$$

The remaining two variants of the RL treatment, denoted X_a^{RL} and X_b^{RL}, where a and b index above and below, are computed be adding and subtracting one tree to and from X_h^{RL} : $X_a^{RL} = \text{Min}(X_h^{RL} + 1, 5)$ and $X_b^{RL} = \text{Max}(X_h^{RL} - 1, 1)$. The RLs are rounded to the nearest integer.

4 Results

4.1 Descriptive statistics

Table 2 presents summary statistics of relevant characteristics of participants in each treatment group. Harvest rates declined sharply following the PES scheme introduced from the sixth round in all three main groups. The RL formula was randomly applied to the groups. Groups with above historical RL have a notably lower average harvest rate *on average*, in spite of the random allocation. As such, the pre-treatment average harvest rates matter, and we take note of this difference across the groups in the later analysis.

Once we set a benchmark, it is interesting to see how many participants reduce harvest to or kept it below the RL. The number of participants who harvest below their group historical average increased following the PES in all three RL regimes. By construction, it is easier for the average participant to qualify for PES when RL is set above historical average. We see that about the same percentage of participants in all three treatments have harvest levels below the first period harvest (73-77%). The most interesting case is when RL is set below the historical average. Many (73%) have lowered their harvest but 23% of them did not reach the RL and, therefore, did not quality for payments, suggesting the presence of an anchoring effect of the RL.

Table 2: Mean and standard deviation (in parentheses) of relevant characteristics of participants.

Variables	Forest reference level		
	H + 1	Historical (H)	H − 1
Group behavior:			
Average harvest rate in stage 1	0.44	0.53	0.52
	(0.36)	(0.34)	(0.34)
Average harvest rate in stage 2	0.23	0.36	0.32
	(0.27)	(0.31)	(0.32)
p1Hrvst - p2Hrvst	0.21	0.18	0.20
	(0.24)	(0.22)	(0.24)
Percentage reduction	53.74	38.53	40.80
	(42.57)	(38.35)	(42.08)
Share with harvest1<p1Hrvst	0.40	0.46	0.45
	(0.49)	(0.50)	(0.50)
Share with harvest2<p1Hrvst	0.77	0.77	0.73
	(0.43)	(0.42)	(0.45)
Share with harvest2<RL	0.93	0.77	0.56
	(0.26)	(0.42)	(0.50)
Reference levels:			
Group average in stage 1	2.13	2.64	2.63
	(1.22)	(1.07)	(0.70)
RL (number of trees)	3.13	2.64	1.63
	(1.22)	(1.07)	(0.70)
Group characteristics:			
Age (years)	41.90	42.18	42.62
	(11.38)	(12.93)	(12.2)
Sex (1= female, 0 = male)	0.24	0.31	0.27
	(0.43)	(0.46)	(0.45)
Education5	2.34	2.02	2.07
	(1.46)	(1.26)	(1.44)
Position (dummy)	0.36	0.18	0.36
	(0.48)	(0.39)	(0.48)
Fuelwood use (dummy)	0.91	0.96	0.99
	(0.28)	(0.21)	(0.09)
Charcoal use (dummy)	0.81	0.87	0.66
	(0.39)	(0.34)	(0.48)
Visit to forest (trips/week)	0.92	0.88	0.96
	(0.81)	(0.66)	(0.80)
Livestock (TLUs)6	3.90	2.86	2.96
	(7.62)	(3.30)	(2.72)
Observations	128	176	128

Notes: Education (1=illiterate, 2=read and write, 3= grade one to five 4=grade six to eight 5=high school or above)

LTU = tropical livestock unit.

4.2 Simple comparisons of harvest reductions

Using a mean comparison test, we found that average harvest rates are statistically significantly higher when RL is set at H-1 than at H+1. Setting RLs above historical average seems to promote more conservation relative to the other two formulas, while we do not find enough evidence to claim that setting RL below historical would increase harvest compared with RL set at historical average. We also tested these within similar pay modality groups; the difference of lower RLs are significant only in collective PES. When the PES design is individual pay, harvest rates do not have statically significant difference across RL regimes.

The RL level carries information about the degree of (unspoken) cooperation among group members during the first stage (higher group average before treatment gives rise to relatively higher RL amount for a given RL formula). Norms of cooperation or defection seem to persist after introducing the treatment. Figure 1 shows higher harvest rates after treatment are associated with higher RLs. This could be attributed to either norms of (non-)cooperation established in the first stage, heterogeneity across groups (which could be related to the first factor), or the anchoring effect of the size of the RLs, or a combination of these factors.

Though it is not easy to disentangle these effects, there are suggestive results from groups with historical average of four or five and above historical RL (Figure 1). The RL is five in both cases and all possible harvest levels are equally attractive, and yet many participants stick to high harvest rates, which indicates the presence of an anchoring effect. Norms of defection in the case of groups with high harvest would be irrelevant once the incentives to harvesting anywhere between zero and five are the same.

To test whether there is anchoring effect or it is merely a reflection of the norms established before treatment, we compared average harvest in stage two with the size of RL (so that significant difference shows no anchoring) and with the group historical averages (to see if there is only some norm). These values are the same for groups with historical average RL, thus the test is trivial. For the remaining two treatment groups, however, we found that average harvest was statistically closer to the size of RL when the RL is below historical average.

The response of the six different treatment groups are depicted in Figure 1, and ordered by the size of the RL. Note that with 8 (or 11) groups in each of the six groups, the number of groups for each level of the RL becomes small. In particular, in the case of below historical RL and individual pay, no groups had a $RL > 2$.

Yet some broad patterns emerge. For H+1, the reductions appear to be largest for individual pay, while the opposite is true for the H-1 case. We test this pattern further in the next section.

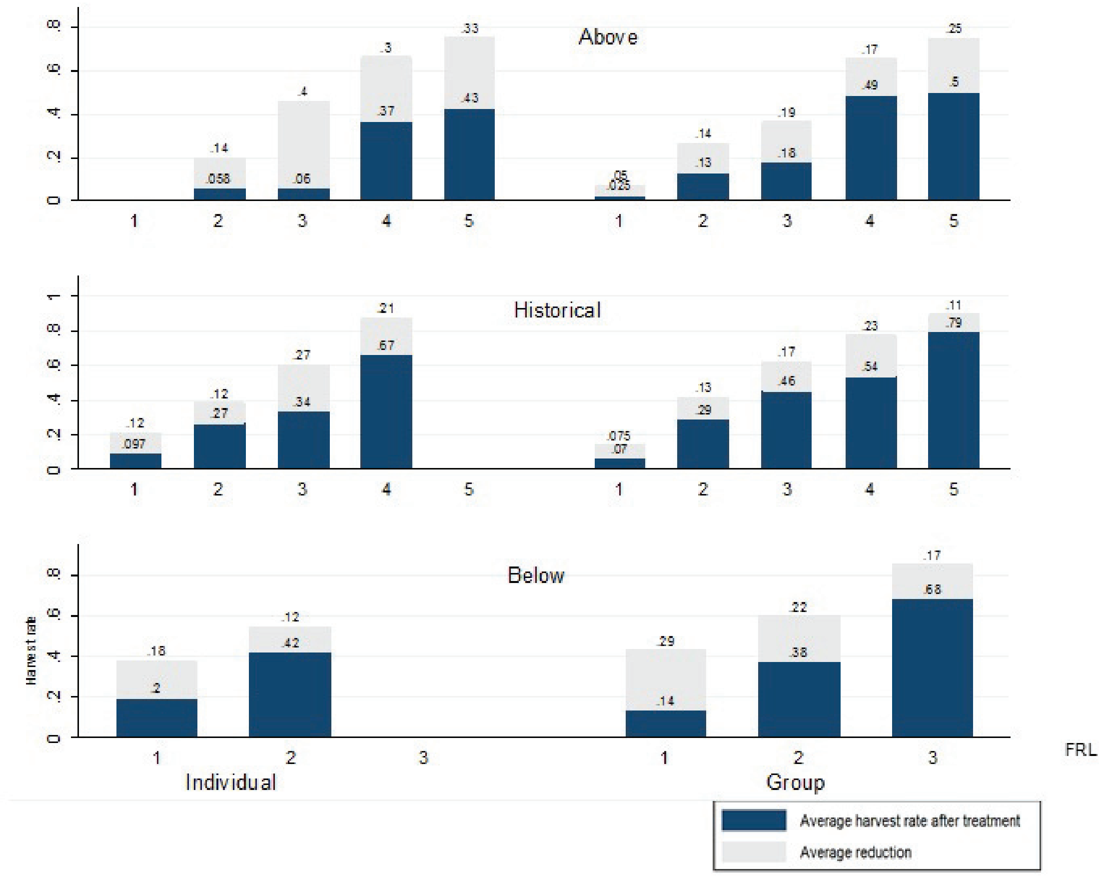

Figure 1: Harvest rate before and after treatment across different amounts and types of RLs.

4.3 Regression model

We now use regression analysis to identify the effects of the treatments while controlling for the both the individual and group level harvest in the first, pre-treatment period. These pre-treatment results should control for individual characteristics relevant for the harvest level (as expressed in the first period harvest), as well as group characteristics and possibly norms developed. The group average during the first period, also determines—together with the RL regime—the actual RL in the second period.

Table 3: Random effects Tobit estimates of the effects of RL and other control variables on period 2 harvest.

Variables	All	Individual	Group
Above (1=yes)	-0.141**	-0.142*	-0.0317
	(0.0679)	(0.0730)	(0.0657)
Below (1=yes)	0.0622	0.0582	-0.165***
	(0.0665)	(0.0717)	(0.0610)
Reward (1=group)	0.106*		
	(0.0610)		
Above * group	0.0991		
	(0.0972)		
Below * group	-0.228**		
	(0.0937)		
Group avg. harvest per. 1	0.536***	0.557***	0.516***
	(0.118)	(0.190)	(0.148)
Ind. avg. harvest per. 1	1.241***	1.188***	1.293***
	(0.0747)	(0.112)	(0.0995)
Constant	-0.800***	-0.788***	-0.711***
	(0.0700)	(0.0988)	(0.0881)
Sigma_u	0.331***	0.364***	0.296***
	(0.0186)	(0.0287)	(0.0242)
Sigma_e	0.381***	0.381***	0.382***
	(0.0106)	(0.0155)	(0.0146)
Observations	2,16	1,08	1,08
Number of participants	432	216	216

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results of the Tobit (panel, random effects) models are reported in Table 3. Figure 1 suggested interaction effects between the RL regime and the payment modality (individual vs. group payment), thus we introduced interaction terms, as well as running separate models for the sub-two groups.

Overall, the pattern visualized in Figure 1 is confirmed by the regression analysis. Setting the RL above the historical (first period) average leads to less harvesting, suggesting a strong incentive effect. This effect is, however, only significant with individual pay.

When the RL is set below the historical average, we get no significant effect in the case of individual pay, while the effect is *negative* on the harvest rates in the case of group pay. This is perhaps the most surprising result of this field experiment, and we return to possible interpretations in section 5.

If higher RL is associated with more harvest, are participants using the RL as an anchor? Table 4 shows that the median values of harvest is positively correlated with the size of RL. Though some groups, which had relatively higher pretreatment averages and thus higher RL, may continue to harvest high amount, it is also possible that many participants target the RL and hence anchor their choice to it, in which case a lower anchor induces more conservation.

Table 4: Median harvest (number of trees) across different size of RL.

RL	Round 6	Round 7	Round 8	Round 9	Round 10
1	0	0	0	0	0
2	0	1	0	0	0
3	1	1.05	2	1	1
4	2	3	2	3	2.05
5	2	2.05	3	3	3

To see if the low harvest in group PES with below historical RL is associated with some sense of determination to reach the common RL and win as a group, we run a random effects Tobit model with individual harvest in the treatment rounds as the dependent variable. The results are reported in Table 5. We identified each participant by their PES achievement in each round after the sixth round. Then we generated lagged indicators showing whether the participant received PES in the preceding round. In column (1) of Table 5, the negative and statistically significant effect of getting an RL below historical average and not getting PES in the previous round indicates that the average participant made efforts to further reduce harvest and reach the RL.

As illustrated in Figure 1, setting RL below historical average has an interesting effect depending on the reward modality. The results in columns (2) and (3) support this. When the average participant finds himself in a group-PES with an RL set below their historical average, and they learn that the group has not made it in the previous round, they make additional conservation effort to reach and help their group reach the RL and thus get the PES reward.

Table 5: A Tobit regression model capturing the reactions of different subgroups to PES reward status in preceding rounds (dependent variable: individual harvest in treatment rounds).

Variables	All	Individual	Group
Above#got$^{\psi}$	-0.221*	-0.251	-0.198
	(0.116)	(0.169)	(0.159)
Historical#didnot	-0.0497	-0.219	0.0595
	(0.117)	(0.176)	(0.160)
Historical#got	-0.0867	-0.262*	0.0406
	(0.103)	(0.147)	(0.147)
Below#didnot	-0.179***	-0.123	-0.243***
	(0.0685)	(0.109)	(0.0923)
Age	-0.00117	0.000997	-0.00180
	(0.00335)	(0.00447)	(0.00471)
Sex	0.0298	-0.0384	0.0717
	(0.0878)	(0.109)	(0.129)
TLU	0.00236	0.000630	0.0232
	(0.00790)	(0.00761)	(0.0203)
Pre-PES Avg. Harvest	0.0284	-0.136**	0.0927*
	(0.0397)	(0.0608)	(0.0540)
Forest visit	0.213***	0.254***	0.155**
	(0.0523)	(0.0711)	(0.0745)
Position	-0.217**	-0.290***	-0.125
	(0.0877)	(0.110)	(0.130)
Constant	0.0666	0.420	-0.148
	(0.202)	(0.280)	(0.285)
Observations	1,256	584	672
Number of participants	314	146	168

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
$^{\psi}$This indicates an interaction factor showing the type of RL (Above, Historical or Below) and whether the participant got PES reward in the previous round.

4.4 Cost efficiency

Our results in the previous section show that setting RLs above historical average leads to lower harvest, on average. To the funding agent sponsoring PES programs, the cost efficiency of alternative RL regimes also matters. Importantly, in this paper we define cost efficiency from the funding agency's perspective, and only include the payments in the PES scheme.[6] Efficiency is thus defined as the overall harvest reduction divided by the total PES transfers. Table 6 presents the total avoided harvest in each treatment group over the last five rounds along with the corresponding aggregate PES reward that participants earned.

[6] A social efficiency analysis would be quite different, for example, we would have to include the opportunity costs of conservation, while the PES payment is just a transfer (and not a cost, except that there are costs in terms of efficiency losses in raising public revenue).

Table 6: Cost efficiency for different treatments.

RL	Reward	(1) Avg.RL	(2) $\sum(RL - x_{it})$	(3) $\sum(R_h^{RL} - x_{it)}$	(4) PES (ETB)	(5=4/3) (ETB/tree)
Above (H+1)	Individual	3.4	672	376	543.00	1.44
	Group	2.9	538	232	402.00	1.73
Historical (H)	Individual	2.4	397	397	418.50	1.05
	Group	2.9	326	326	252.00	0.77
Below (H-1)	Individual	1.4	-12	308	158.25	0.51
	Group	1.9	19	326	60.00	0.18

Notes:
Column 2 is aggregate avoided harvest relative to RL.
Column 3 is aggregate avoided harvest relative to historical average, i.e., the real reduction.
Column 4 is the total PES made to those who qualified for payments.
Column 5 is our measure of cost efficiency.

The aggregate number of trees that was saved is highest when RLs are set at H+1, and more so when participants got individual pay. Further, we note—in line with previous results—that individual pay leads to more conservation, except for RL below historical average case where the reduction is slightly higher with group pay.

In terms of cost efficiency, a clear pattern emerges. Setting a high RL is costly. In spite of higher reductions in harvesting, the higher PES program costs outweigh this effect. The cost efficiency of individual vs. group pay is not consistent, and varies depending on the RL formula applied.

5 Discussion

5.1 The impact of the reference level: incentives vs. anchoring

The theoretical discussion identified two different effects of varying the RL. With a high RL (above historical average), more players will be eligible for payments, starting off at their pre-treatment harvest levels. In other words, fewer players will have to do a costly and uncompensated reduction of their harvest, before they qualify for PES. We refer to this as the incentive effect of high RLs.

Behavioural economics also suggests a possible anchoring effect, i.e., that the RL indicates a norm for what is an acceptable harvest level, or can be an aspirational target. Choices will therefore gravitate towards the RL. This effect thus suggests that a higher (lower) RL also leads to higher (lower) harvest levels, and thus might pull in the opposite direction of the incentive effect.

Which effect dominates? Both the descriptive and the regression analyses suggest that the answer depends on the modality of the PES (individual vs. group payment), and whether the RL is set below or above the pre-treatment group average.

When the RL is set above historical average and with individual PES, the incentive dominates. This corresponds to a classical PES scheme in field experiments, i.e., payments are made to individuals and the reference level is set at the maximum harvest such that any reductions below that are incentivized. As such, our finding confirms earlier results in the literature of a significant conservation effect of introducing PES (e.g., Handberg and Angelsen (2016)).

The case of not significant harvest reduction of a higher RL and with group pay is more challenging to explain. One possible explanation is the following. When the group is informed that the RL is set above their pre-treatment level, they feel sure that the target will be reached. Thus there is limited motivation for each player to ‘help the group’ (by reducing own harvest further) achieve that level. Together with the anchoring effect, there is no significant reduction compared with the case when RL is set at the historical harvest level.

This picture changes when RL is set below the pre-treatment level. With individual pay, we found no significant reductions. The incentive effect is assumed to be weaker compared to the default case (RL = historical), but the anchoring affect is stronger. The non-significant result suggests that these opposing effects roughly cancel out each other.

As already noted, the most surprising effect is the large and significant reduction in harvest levels when the RL is set below historical average and with group payment, based on aggregate performance. Following the reasoning above, this result suggests that the anchoring effect dominates, but this is not the full story. Based on both field observations and analysis of the group behaviour (Table 5), we propose that this treatment did invoke a group dynamics that was not present in the other treatments.

In short, the tight RL and group pay set an ambitious target for the whole group, which they collectively aimed to achieve. During the experiment, verbal communication was not permitted. Yet we observed a different dynamics with individual vs. group pay. During the group-pay modality, when the RL was not reached some participants would signal or even give comments outload such as “we still couldn’t make it”. Similarly, they would appear relieved or make small comments like “we achieved it” when the aggregate harvest got below the RL, and the group therefore qualified for pay.

This experience may be taken as a flaw in the experimental design (some communication was possible)—jeopardizing internal validity. Alternatively, it may be seen as a good sign in the way that field experiments better reflect the real-life mechanisms and dynamics (external validity). Yet, to what extent such group dynamics and aspirational target would be relevant in a real-world PES scheme, remains a question. Based on empirical work of Ostrom et al. (1994) and others, this would depend on factors like group size, group heterogeneity, and information about others’ harvest and shared norms. Yet, an important finding is that group level pay in some contexts can invoke a strong group dynamics of achieving the anchor or target that a RL appears to represent.

The result also relates to a broader debate on how the design of institutions and policy instruments (mechanisms) can affect preferences and thus choices. The standard assumption of economics is that social preferences are independent of the mechanism design, referred to as the separability assumption (Bowles and Polania-Reyes, 2012). But, pecuniary incentives and social preferences might be substitutes (crowding in) or complements (crowding out) (see also Chapter 5). One possible interpretation of our results is thus that group level pay (relative to individual pay) activates social preferences. Similar effects have been found in other studies. For example, Agrawal et al. (2015) report on how participants that received private rather than communal benefits in a sustainable development intervention in Northern India gradually changed from an environmental (social) to an economic (individual) motivation.

5.2 The high cost of a high reference level

Cost efficiency is defined as the number of trees saved per Ethiopian Birr spent on the program, while effectiveness refers to the number of trees saved only. For a sponsor of a PES program seeking to maximize cost efficiency, the theoretical trade-off is simple. A high RL should give higher reductions, although our results are not as straightforward as just seen (the 'nominator' effect). On the other hand, it also means that the point from which payments start is higher, and some participants are paid beyond what is needed to achieve a given reduction (the 'denominator' effect).

Our experimental results are clear: the cost increasing (denominator) effect of high RLs dominate the effectiveness (nominator) effect. The results is in part due to the strong anchoring and group dynamics effects that we identified.

In discussions about setting RL in REDD+ and PES programs, cost efficiency is just one of several considerations. There is a large discussion on benefit sharing in REDD+ (e.g., Luttrell et al. (2014)), where the assumption is that (poor) forest users should have a net gain from their participation in REDD+ and their efforts to help mitigate climate change. Yet, there are often real trade-offs between them. Making PES schemes efficient by targeting those that are likely to deforest is essentially an exercise in setting a correct RL for each PES program participant. Failing to do so yields lower effectiveness and cost efficiency. This remains a major challenge in PES schemes. Mohebalian and Aguilar (2016), for example, in a study of Ecuador's Socio Bosque (PES) program found that less than 1% of the payments provided additional conservation. Our experimental results point in the same direction; high RLs represent a waste of conservation spending.

6 Conclusion

In this paper, we explored the conservation effect of setting reference levels (RLs) above, at or below historical average, and how this effect varies by pay modality (individual or group). We framed this in terms of a positive incentive effect and a negative anchoring effect of a higher RL. The results were nuanced. With individual pay, a higher RL increases conservation. With group pay we get the opposite result, possibly due to the limited pecuniary incentives provided with group pay combined with a strong anchoring effect. When the RL was set below the pre-treatment (historical) average, we found no effect under individual pay, but a large and significant conservation effect under collective pay. We propose that this is due to an anchoring effect by which the RL is being interpreted as an 'aspirational target' for the group, which then collectively aimed to achieve it by reducing their harvest. The combination of group dynamics and the anchoring and normative effects outweighed the weak incentive effect of a low RL. Further, our results also lend to support to claims of the PES design affecting the extent to which social/environmental preferences are being expressed.

In terms of cost efficiency, our experiment demonstrated the high costs to the PES scheme funder of a high RL. This is simply due to the fact that the starting point for payment is higher (with some non-additional payments) as well as a generally weak net incentive effect of higher RL. This result points to a general problem in real-world PES design and implementation.

PES programs are debated, but still considered favorably by many for their attractive theoretical features (Wunder, 2007), while the "devil is in the details" in terms of dozens of design and implementation issues (Engel, 2016). The effectiveness and cost efficiency implications of setting different reference levels have hardly been explored in the empirical and experimental literature, and our field experiment represents a first attempt. Further replication and extension of such experiments would provide a more solid basis for practical recommendations on PES and REDD+ design and implementation.

References

Agrawal, A., Chhatre, A., and Gerber, E. R. (2015). Motivational crowding in sustainable development interventions. *American Political Science Review*, 109(03):470–487.

Agrawal, A., Nepstad, D., and Chhatre, A. (2011). Reducing emissions from deforestation and forest degradation. *Annual Review of Environment and Resources*, 36:373–396.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477.

Angelsen, A. (2008). *How do we set the reference levels for REDD payments?* Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.

Cadsby, C. B. and Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics*, 71(1):53–73.

Cardenas, J. C., Stranlund, J., and Willis, C. (2000). Local environmental control and institutional crowding-out. *World Development*, 28(10):1719–1733.

Cardenas, Juan C. Tranlund, J. W. C. (2000). Local environmental control and institutional crowding-out. *World Development*, 28(10).

Dickinson, D. L. (1998). The voluntary contributions mechanism with uncertain group payoffs. *Journal of economic behavior & organization*, 35(4):517–533.

Engel, S. (2016). The devil in the detail: A practical guide on designing payments for environmental services. *International Review of Environmental and Resource Economics*, 9(1-2):131–177.

Furnham, A. and Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1):35–42.

Handberg, Ø. N. and Angelsen, A. (2015). Experimental tests of tropical forest conservation measures. *Journal of Economic Behavior & Organization*, 118:346–359.

Handberg, Ø. N. and Angelsen, A. (2016). Pay little, get little; pay more, get a little more: A framed forest experiment in tanzania. *Ecological Economics*.

Hardin, G. (1968). The tragedy of the commons. *Journal of Natural Resources Policy Research*, 1(3):243–253.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4).

Herold, M., Angelsen, A., Verchot, L. V., Wijaya, A., and Ainembabazi, J. H. (2012). A stepwise framework for developing REDD+ reference levels. *Analysing REDD+: Challenges and choices*, pages 279–300.

Ickowitz, A. (2006). Shifting cultivation and deforestation in tropical africa: critical reflections. *Development and Change*, 37(3):599–626.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives*, 21(2):153–174.

Lopez, M. C., Murphy, J. J., Spraggon, J. M., and Stranlund, J. K. (2012). Comparing the effectiveness of regulation and pro-social emotions to enhance cooperation: Experimental evidence from fishing communities in colombia. *Economic Inquiry*, 50(1):131–142.

Luttrell, C., Resosudarmo, I. A. P., Muharrom, E., Brockhaus, M., and Seymour, F. (2014). The political context of REDD+ in indonesia: constituencies for change. *Environmental Science & Policy*, 35:67–75.

Mohebalian, P. M. and Aguilar, F. X. (2016). Additionality and design of forest conservation programs: Insights from ecuador's socio bosque program. *Forest Policy and Economics*, 71(Supplement C):103–114.

Ostrom, E., Gardner, R., and Walker, J. (1994). Rules, games, and common-pool resources.

Rodriguez-Sickert, C., Guzmán, R. A., and Cárdenas, J. C. (2008). Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior & Organization*, 67(1):215–227.

Sheng, J., Miao, Z., and Ozturk, U. A. (2016). A methodology to estimate national REDD+ reference levels using the zero-sum-gains dea approach. *Ecological Indicators*, 67:504–516.

Strack, F. and Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of personality and social psychology*, 73(3):437.

Tversky, A. and Kahneman, D. (1974). *Judgment under uncertainty: Heuristics and biases*, pages 141–162. Springer.

UNFCCC (2010). Framework convention on climate change: report of the conference of the parties on its 16th session, held in cancún from 29 november to 10 december. Report, Technical report, FCCC/CP/2010/7/Add. 1.

UNFCCC (2012). Report of the conference of the parties on its seventeenth session, held in durban from 28 november to 11 december 2011. United Nations Framework Convention on Climate Change: Bonn.

UNFCCC (2013). Warsaw framework for REDD-plus.

VCS (2016). Standards for a sustainable world. http://www.v-c-s.org/. accessed January 28, 2017.

Walker, J., Gardner, R., and Ostrom, E. (1990). Rent dissipation in a limited-access common-pool resource: Experimental evidence'. *Environmental Economics and Management*, 19:203–211.

Wunder, S. (2007). The efficiency of payments for environmental services in tropical conservation. *Conservation biology*, 21(1):48–58.

Chapter 4:
Pay individuals or groups

Pay individuals or groups to conserve forests? Experimental evidence from Ethiopia*

Amare Teklay Hailu[1] and Arild Angelsen[1,2]

[1]School of Economics and Business, Norwegian University of Life Sciences (NMBU), P.O.Box 5003, 1432 Ås, Norway
[2]Center for Internatioanl Forestry Research, Bogor, Indonesia

Emails: amare.teklay.hailu@nmbu.no; arild.angelsen@nmbu.no

Abstract

Payment for environmental services (PES) aims to correct a market failure and internalize externalities using a 'pay the provider principle', in lieu of the 'polluter pays principle' and command-and-control policies. We conducted a framed field experiment (FFE), using a random sample of 432 forest users in Ethiopia, to explore performance under individual- vs. group-based reward alternatives for conserving forests. We found that the mean harvest rate in group-based pay is 29% higher than in individual-based PES. Besides free-riding behavior with group pay, we tested—by varying the reference levels for payment—for an uncertainty effect: in the absence of credible communication, each participant may not know whether the group will reach that benchmark and be eligible for payment. We conclude that programs such as REDD+ may more effectively achieve conservation targets with individual pay unless the transaction cost saving from a collective pay is large.

Keywords: Payment for environmental services, REDD+, framed field experiment, deforestation, conservation, common pool resources

JEL codes: Q23, C93, Q57

*We would like to acknowledge funding from the Norwegian Agency for Development Cooperation [agreement no. GLO-3945 QZA 13/0545; grant on Learning from REDD+ an enhanced Global Comparative Analysis] through the Center for International Forestry Research (CIFOR).

1 Introduction

Payment for environmental services (PES) is gaining popularity as a policy tool to promote sustainable use of common pool resources (CPR). PES aims to correct a market failure and internalize externalities using a modified PPP principle: 'pay provider principle' in lieu of the 'polluter pays principle' (Engel et al., 2008). Performance-based payment such as PES is considered as an integral part of current efforts under the umbrella of Reducing Emissions from Deforestation and forest Degradation (REDD+).

A key design issue in PES is who to pay: individuals or groups? This involves a number of considerations. First, individual pay may have stronger extrinsic incentive effects as performance-based payment is tied directly to one's own decisions and the reward is known for all possible strategies. Second, the role of community organization, trust, and cooperation may be advanced by group pay, because individuals have intrinsic motivation and social preferences (Rodriguez-Sickert et al., 2008; Ryan and Deci, 2000). Third, there are practical and cost considerations, such as property rights, indivisibility and type of environmental service (e.g. wildlife conservation as in Zabel et al. (2013)) and monitoring costs, which often favour group level pay (Engel, 2016).

We conducted a framed field experiment (FFE) in Ethiopia on common pool resource (CPR) management to assess performance differential in individual and collective pay in terms of limiting forest use. A random sample of 432 smallholder farmers, who make forest use decisions in their daily lives, participated in the experiment. The framing adds relevance to the task subjects accomplish, and the farmers bring hands on experience into the game (Harrison and List, 2004).

The paper aims to compare the harvesting effect of choosing a particular reward modality. Drawing on several experimental studies, we presume that PES effectiveness on conservation outcomes is not only dependent on the level of payment, but also on who is paid. Some studies show that individual pay is more effective than collective pay, e.g., in the context of agrobiodiversity (Midler et al., 2015; Narloch et al., 2012). The exploration of an uncertainty effect with group pay is a novel aspect of our paper. Payments are made in the experiment only if the harvest (individual or aggregate) is below a predefined benchmark, which resembles experiments of public goods provision with a threshold (Cadsby and Maynes, 1999; Dickinson, 1998), and mirrors a typical PES-REDD+ scheme. A forest reference (emission) level (FRL/FREL, or RL for short) is set and used as a benchmark for program evaluation and financial transfers (Busch et al., 2009). We introduce three different reference levels, creating different levels of uncertainty, as participants do not know whether the group will reach the reference level and be eligible for payment.

In the following section, we put the research question in context by briefly discussing the definition of and core design issues in PES, as well as arguments in favor of individual or group pay. Section 3 gives a detailed description of the experimental design, sampling and implementation of the FFEs. Sections 4 and 5 present the empirical results and a discussion of the mechanisms through which individual and collective pay play a role in determining forest conservation. Section 6 concludes.

2 PES and forms of payment

2.1 PES definitions and design issues

Payment for Environmental Services (PES) involves a voluntary transaction between an environmental service provider and a buyer, where the payment is conditional on the provision of a well-defined environmental service (Wunder, 2005). Given that the exact service might be hard to define and measure, Engel (2016, p.3) broadens the definition and considers "PES as a positive economic incentive where environmental service (ES) providers can voluntarily apply for a payment that is conditional either on ES provision or on an activity clearly linked to ES provision."

A critical element in any PES design is who to pay: individuals or groups? This choice may impact both participation and performance in the PES scheme (Barr and Mourato, 2014; Newton et al., 2012; Wunder, 2005). Within these two broad payment modalities, there are several design options, for example, whether participants should receive the same payment, whether they should be paid per unit of service provided, whether payments should be differentiated based on costs, or whether payments should be made in cash or in kind.

The choice of reward modality is also partly guided by the type of service in question, e.g., to what extent the individual contributions can easily be identified and measured. Consider the case of deforestation. Identifying individual forest conversion is much more costly than assessing the overall deforestation for a community. Furthermore, PES aims for additionality, which implies identifying a realistic counterfactual scenario (reference level), which again is much more challenging at the individual than at the community level. Engel (2016), therefore, points out that it normally is easier, cheaper and more manageable to give contracts to groups.

While transaction costs and other practical considerations clearly will play a role in real-world design of PES schemes (Mahanty et al., 2013), such cost differences are not part of the experiment reported in this paper. Instead, we focus on the pecuniary incentives and non-pecuniary elements from the behavioral and experimental economics literature, relevant for the performance of PES with individual vs. group pay. Notable studies of common pool resource management comparing these forms of payment are Travers et al. (2011a) in Cambodia focusing on fishery, and Midler et al. (2015) and Narloch et al. (2012) on biodiversity. The empirical findings are in favor of individual pay with explanations ranging from crowding in or out of individual motivation (Narloch et al., 2012) to interactions of pay and social ties (Midler et al., 2015). Our paper will contribute to the testing of how robust these findings are to a framing of forest resources in developing countries where cutting trees is a necessity (Sumiya, 2016).

2.2 Pecuniary incentives and uncertainty

A starting point for a comparison of individual vs. collective PES is a simplified neoclassical model of *homo economicus*, the rational actor maximizing only own pecuniary (material) benefits (Kerr et al., 2012; Vatn, 2009). The individual pay format is then superior to collective pay. With collective pay equally distributed among the group members (N), each participant will receive only $\frac{1}{N}$ of his contribution to the service provision. If N is large, the incentive approaches zero, and the Hardin (1968) 'tragedy of the commons' problem arises (Castillo and Saysel, 2005).

Collective pay may create an additional problem. PES can be conditional on reaching a certain benchmark, i.e., a reference level (RL) in the case of REDD+. This creates an element of uncertainty: the reference level may be set such that participants are not sure whether the group will reach that level. Any contribution made risks being squandered. This situation corresponds to experiments of public goods provision with a threshold, which has a long history in the experimental literature (e.g.,Cadsby and Maynes (1999); Dickinson (1998)). With individual pay, the participants have full certainty of what the PES will be for different levels of their contribution.

2.3 Pro-social and pro-nature preferences

Decisions on how to use common pool resources (CPR) are not guided solely by own material benefits, but also by the impacts on other resource users and the environment. Behavioural and experimental economics suggest that most humans have pro-social preferences, for instance in the form of impure altruism or norms of reciprocity (Rustagi et al., 2010). Experiments on social dilemmas in CPR management have demonstrated that individuals harvest less than predicted by the *homo economicus* model (Cardenas et al., 2000; Midler et al., 2015; Ostrom, 2000), which suggests the presence of pro-social and pro-nature preferences. Ignoring these dimensions of human motivation results in incomplete models of human behavior (Bowles, 2008).

The existence of pro-social and pro-nature preferences also suggests that the difference between individual and collective pay may not be as large as the *homo economicus* model predicts. The more important question is yet: how does the mode of pay affect such preferences? Could group pay strengthen such motivations and narrow the gap even further or individual pay crowd in/out these preferences? Distinguishing between intrinsic and extrinsic sources of motivation (Ryan and Deci, 2000), individual PES may have strong and positive effect on the latter but mixed and indeterminate effect on the former. Its effect on intrinsic motivation depends on whether the amount of compensation is large enough (Gneezy and Rustichini, 2000a), or it gives guilt-relief (Narloch et al., 2012), or it is perceived as supportive of pro-environmental behavior (Vollan, 2008). The net effect determines whether participants are likely to conserve (or extract) more.

Group level pay often involves more interaction and communication among group members. Communication typically increases cooperation by reducing the level of distrust (Midler et al., 2015), although this effect might be smaller with large groups (Zabel et al., 2013) or when social norms are weak initially. Yet, positive reciprocity, altruism and trust may increase in collective PES; people's inherent social preferences could get enhanced in group settings where they can contribute to a common good (Midler et al., 2015; Narloch et al., 2012).

2.4 Trust and payment modality

In repeated games, individuals will have the chance to signal pro-social behavior by being nice first, which may be followed by strong reciprocity and could evolve into coordination game (Gintis et al., 2001). If they trust that other participants will reciprocate positively, collective pay may overcome—at least partly—the free rider problem. Lack of trust will, of course, have the opposite effects.

Lack of trust might also have another effect that may reduce the performance of collective PES. Distrust in the administrative system will leave participants unsure whether they will receive their collectively earned money. This 'distrust effect' can be due to 'elite capture' (Zabel et al., 2013). Nonetheless, it may reduce the performance of all forms of PES and not only of collective pay.

3 Materials and methods

3.1 Description of study area

The Federal Democratic Republic of Ethiopia (FDRE) is divided into nine National Regional States and two city administrations. Each region is then subdivided into administrative zones named *Woredas* (districts). The lowest administrative unit is a *kebelle* (peasant association). The constitution of the FDRE states that land belongs to the state while it guarantees individuals the right to freely use, bequeath and transfer land. The Regional States prepare their own specific regulations to ensure fair use rights of their inhabitants from land. As such, there is customary land tenure in many places. In the National Regional State of Tigray, where we conducted our framed field experiment (FFE), one finds private agricultural land, communal grazing areas and state protected areas.

The Northern part of Ethiopia being relatively degraded, area enclosures and planting trees in private land have been part of the effort to rehabilitate the area (Gebremedhin et al., 2003; Gebremeskel et al., ress). The question of individual or collective pay may be framed as a question of rewarding efforts to plant and protect trees in private plots or preserving the forest in common grazing areas. Alternatively, it is possible to regulate the use of communal land through individual PES payments; the reference level can be seen as a form of an (asymmetric) harvest quota, with reductions beyond that level being compensated.

3.2 Sampling

A framed field experiment (FFE) was conducted in three zones of the state of Tigray, Ethiopia. We selected relatively forest rich zones: Western, North Western, and Southern zones. Within each zone, three *Woredas* were identified, and then within each *Woreda*, three *kebelles* were purposively selected to make sure that there is natural forest within a walking distance from the *kebelle* centers. A list of household heads was obtained from each *kebelle* administration, and a sample of 48 household heads was selected randomly within each *kebelle* (See Table 5).

Both men and women participate in forest use (decisions) in rural Ethiopia, but with some differences in the tasks: men typically are more engaged in clearing forests for agriculture while women usually collect firewood. As the list of resident household heads in a typical *kebelle* consists of both male and female heads of households, we sent invitation letters directly to the name recorded in the household list. About 72% of the participants were men. Participants were randomly assigned to groups of eight people, which were again randomly given either individual of collective pay treatment. At the end of the session, all participants were interviewed about their socioeconomic characteristics, actual forest use as well as their experience in and perception of the experiment.

3.3 The experiment

3.3.1 Framed field experiment (FFE)

Since the pioneering experimental work on CPR and social dilemmas by Ostrom et al. (1994) and Walker et al. (1990), several innovative approaches have been used to increase the internal and external validity of experimental results. Compared to lab and artefactual field experiments, FFEs stand out as having realism and validity to resemble the real world forest use of participants (Harrison and List, 2004). For instance, FFEs such as Cardenas (2004) and Handberg and Angelsen (2015) took their precedents one-step further and improved the framing in a way that makes the experiment closer to real-life decisions. In the latter study, paper trees were used instead of tables of outcomes in the former.

In our study, we followed this tradition and conducted the experiment in the villages where participants make their real-life decisions of how many trees to cut. We replaced paper trees in Handberg and Angelsen (2015) with real tree branches. In each session, participants were endowed with a forest consisting of 60 real dry tree branches[1], each approx. 50 cm long. Sitting in a circle, participants were asked to imagine that each tree branch represented an actual standing tree in a common forest near their village.

3.3.2 Basic game structure

The basic CPR game and its corresponding social dilemma in this experiment draw from Cardenas et al. (2000), Walker et al. (1990), Cardenas (2004) and Handberg and Angelsen (2015). The social dilemma can arise in different ways. There might be resource growth such that individual harvesting reduces the forest stock for all participants in future rounds, as in Handberg and Angelsen (2015). Alternatively, as chosen in our experiment, the forest stock can produce services, such as improved water quality, that are shared equally among the participants (Cardenas, 2004). The PES treatment fits easily within the latter, and makes the experiment easier to understand by participants.

A particular session begins with a group of randomly assigned eight participants, collectively endowed with 60 tokens trees. A research assistant would read aloud a detailed instruction in Tigrigna, the local language in all study *kebelles*. An example was also shown to make sure that they understood the procedure. Participants were informed that the experiment would have 10 to 15 rounds so that all groups would have similar initial information until the end of the tenth round.[2]

[1] In pretesting, we also used forest stocks of 40 and 80. Setting the stock to 40 meant that the forest could be completely exhausted in one round ($5 * 8$), and this seemed to have an effect in terms of lowering the group's harvest.

[2] This paper was part of a bigger experiment with different treatments, and there were groups who would play five more rounds after finishing the first ten, with no treatment and for a different purpose (testing for any crowding out effects by removing the PES treatment).

The participants made their decisions simultaneously. They were not allowed to communicate throughout the session. Allowing communication ('cheap talk') in collective pay may mask the inherent trust and intrinsic motivation, and may be seen as a treatment in itself, making group pay with communication incomparable to individual pay without communication.

Participants indicated their choice by marking a symbol showing, both in Arabic numbers and in pictures, the number of trees harvested (see Table 4 in Chapter 1). This approach made it easier for them to understand the process as it partly mimics the way they vote during elections. As most participants are illiterate, using visual symbols helped them make choices independently of the research assistant.

At the end of each round, the research assistant would add up the group harvest, announce the total to the group, and remove the harvested branches from the forest stock. This enabled participants to reflect about the total impact on the forest stock and their collective payoff in that round. Before the next round began, the stock was reset to 60 branches. The stock size was kept constant throughout for simplicity and to make the mental calculation of payoffs easier for the participants.

The payoff, π_{it}, to participant i of cutting x_{it} trees in round t is given by Eq. (1). A description of this formula was also given in the instructions, and participants were given the chance to ask questions to ensure that they understood the payoff structure.

$$\pi_{it} = x_{it} + 2\left(\frac{60 - \sum x_{it}}{8}\right), \quad \text{where} \quad x_{it} \leq 5 \tag{1}$$

The first term (x_{it}) is the direct benefit individual i receives from own harvest in round t, while the second term is the collective benefit from the forest stock at the end of the round, and is shared equally among the group members. We put an upper harvest limit at five trees per person per round. This is justified by real world limitations such as increasing opportunity cost of labor, the scarcity of forests and the nature of the demand for forest products.

The payment was calibrated such that a typical pay-off would be 1-2 daily rural wages. Each harvested tree gives a constant private benefit of one Ethiopian Birr (ETB[3]), while the public benefit from each standing tree is 2 ETB. This creates a social dilemma. A living tree generates a collective benefit twice the individual benefit of a harvested tree, but the benefit to the individual of saving one more tree is only $\frac{1}{4}$ of the benefit from harvesting it. The dominant strategy is, therefore, to harvest the maximum allowable of five trees. The resulting symmetric Nash Equilibrium is a total harvest of 40 trees and each participant earns 10 ETB (5+2∗(20/8)). The socially optimal (aggregate payoff maximizing) solution is when each player harvests nothing and earns 15 ETB (2∗(60/8)). The neoclassical prediction is that a tragedy of the commons arises because each participant has a strong self-interest to harvest the maximum allowable amount. We conducted five rounds of baseline CPR games in all sessions.

[3] 1 ETB was aprox. 0.046 USD at the time of experiment.

3.3.3 The PES treatments

After the initial five rounds with payoffs as described in Eq. (1), all 54 groups received an offer to reward them for saving trees. We introduced randomly two different treatments, both individual and group pay, and three different reference level formulas (benchmark for PES). In this paper, we focus on the impact of individual vs. group pay. As such, 27 groups were assigned to group pay (PES calculated at the group level and then shared equally, provided the group's average harvest being below the reference level), while the remaining 27 groups were given individual PES pay if the individual harvest was below the reference level.

At the end of the fifth round, we informed the participants about the group-level average harvest in all five rounds, and how the reference level was computed. We used three different reference levels: the average group harvest over the first five rounds (historical RL), the average minus one tree (below historical RL), and the average plus one tree (above historical RL).[4] We did not inform the participants in the start of the experiment that their harvest would be used to set the RL to avoid any strategic behavior.

Once the participants knew the benchmark, we told the group with individual pay that they would receive 0.75 ETB, in compensation for each individual reduction below the RL, which is the same for all members of a group. In the collective reward case, each member of the group receives the same PES amount, based on the *group average* harvest below the RL. To improve their understanding, we gave examples and demonstrated that we would pay them 0.75, 1.5 or 2.25 ETB if the reduction is 1, 2 or 3 trees, respectively.

Formally, the payoff function of each participant under individual PES becomes:

$$\pi_{it} = x_{it} + 2\left(\frac{60 - \sum x_{it}}{8}\right) + 0.75 * \text{Max}(0, RL_k - x_{it}), \quad \text{where} \quad x_{it} \leq 5 \tag{2}$$

Note that the reward was calibrated such that if a participant harvests below RL_k, he would receive the same benefit from harvesting and saving a tree. With group-based PES, the payoff function is:

$$\pi_{it} = x_{it} + 2\left(\frac{60 - \sum x_{it}}{8}\right) + 0.75 * \text{Max}(0, R_k - \frac{\sum x_{i=1t}}{8}), \quad \text{where} \quad x_{it} \leq 5 \tag{3}$$

Each individual's payoff is the sum of the direct private benefit, benefit from standing trees as well as his share of the group's PES reward. If the average group harvest is below RL_k, he receives a benefit of 0.34 (2.75/8) from saving a tree, compared to the benefit of one from harvesting.

The group-pay treatment maintains the social dilemma. The individual conservation incentives are lower in the group pay, although the total PES levels are the same for a given harvest. Thus, in the group reward system, free riding behavior is (still) being rewarded, and the Nash Equilibrium is as without any PES treatment (i.e., maximum harvest). Further, the average group harvest in the current round is unknown when the harvest decision is made, introducing an additional uncertainty in the way that the participant might just receive 0.25 (2/8) by not harvesting a tree.

[4] We reserve further discussion for an accompanying paper (see Chapter 3), but control for this in our statistical analysis.

3.4 Data analyses

In most instances, attention is given solely to the outcome after the intervention, either because the pre-treatment outcome is presumed to be similar across different randomly assigned groups, or because control groups are used as the baseline (Midler et al., 2015; Narloch et al., 2012). Our random assignment of treatments means that we can start with a simple mean comparison test. Regression based analyses may, however, be more flexible to estimate robust standard errors and to test for the impact of covariates. Moreover, there is slight but persistent pre-treatment difference between individual and collective pay groups. It is not uncommon, however, to find a difference in a small sample. One way of using this before and after treatment information would be to apply a difference-in-difference estimation, but its common trend assumption does not seem to hold in our case. Therefore, we used a two-limit random effects Tobit model to assess the difference in performance, by controlling for pre-treatment heterogeneity. We chose Tobit because the dependent variable, defined as harvest rate ($\frac{x_{it}}{5}$), is bounded both from below and from above, at zero and one respectively. We chose random effects because we may benefit from the panel nature of the data.

The control variables include socioeconomic characteristics of participants such as age, gender, livestock and position in the *kebelle* administration, as well as *woreda* dummies. We also included indicators related to the experiments such as average harvest before treatment, the type of RL, and inverse of round number (to control for any learning effect) (Midler et al., 2015).

4 Results

4.1 Descriptive statistics

To check whether the two groups are comparable, we look at the distributions of selected socioeconomic characteristics, most of which are similar in both groups (Table 2). The mean harvest rate before treatment was 47% and 53% in the experiments with individual and group pay, respectively. In spite of the random treatment allocation, this difference is significant at 10% level. This may have implications for the subsequent analyses. The reduction in harvest rate after the treatment is about 20 percentage points (hereafter: pp) (or 43%) in individual pay and 18 pp (or 34%) in collective pay, and the difference is significant at the 5% level.

Table 1: Mean and standard deviation (in parentheses) of key variables by treatment.

Variable definition	Treatment	
	Individual	Group
Average harvest rate in stage 1	0.47 (0.337)	0.53 (0.347)*
Average harvest rate in stage 2	0.27 (0.291)	0.35 (0.316)**
P1Hrvst - p2Hrvst	0.203 (0.241)	0.181 (0.219)
Age (years)	42.29 (12.127)	42.16 (12.462)
Sex (1= female, 0 = male)	0.30 (0.460)	0.26 (0.437)
Educationa[a]	2.16 (1.363)	2.09 (1.411)
Position (1 = has some position, 0 = no)	0.282 (0.451)	0.292 (0.456)
Fuelwood use (dummy)	0.95 (0.211)	0.95 (0.211)
Charcoal use (dummy)	0.73 (0.444)	0.85 (0.361)***
Visit to forest in winter (trips/week)	0.95 (0.782)	1.1 (0.968)
Visit to forest in summer (trips/week)	0.82 (0.820)	0.83 (0.803)
Stated reward preference (1 = group)	0.37 (0.483)	0.29 (0.456)
Livestock (in TLUs)[b]	3.57 (0.334)	2.83 (0.333)
Farm size (in Timad)[c]	7.02 (9.445)	6.60 (9.348)
Size of the RL (number of trees)	2.37 (1.225)	2.59 (1.133)

Notes:
*$p < 0.1$ **$p < 0.05$ ***$p < 0.01$.
[a] Education is a categorical variable with 5 categories where higher value indicates higher level of education.
[b] TLU stands for tropical livestock units (Jahnke and Jahnke, 1982).
[c] One *Timad* is about a quarter of a hectare. Figures are excluding five observations who have land size of more than 25 hectares.

In the post-experiment interview, we asked the participants several questions about their actual forest use, including the number of visits to forests in the wet and dry seasons. Farmers prepare for the rainy season by cutting trees and making agricultural tools during the dry season. Similarly, women collect more fuelwood when they have more free time; their opportunity cost of labor increases in the harvesting season (summer). Average number of visits to forests is higher in the dry season than in the wet season for both treatment groups. We also note that the dummy for charcoal use is different between the two groups, highlighting the potential importance of controlling for covariates and the need for going beyond simple mean comparisons.

We also asked participants if they preferred that payments be made individually or in groups. Only 37% and 29% of the participants in the individual and group pay treatments, respectively, stated that they would prefer group pay, while the rest preferred individual pay. Interestingly, the preference for group pay is lower among participants that experienced that treatment.

4.2 Treatment effects

Figure 1 shows the average harvest rate for both groups over the ten rounds of the experiment. During the first five rounds, the overall average harvest per round is around 50% of the maximum allowable limit and with weak trend of declining harvest rates. Following the treatment (from round six), we observe a significant decline in the harvest rates, and then they stabilize. The effect of PES is similar in both treatment groups: average harvest rate sharply drops to 31% and 26%, respectively, and remain above (below) 30% for the collective (individual) pay group.

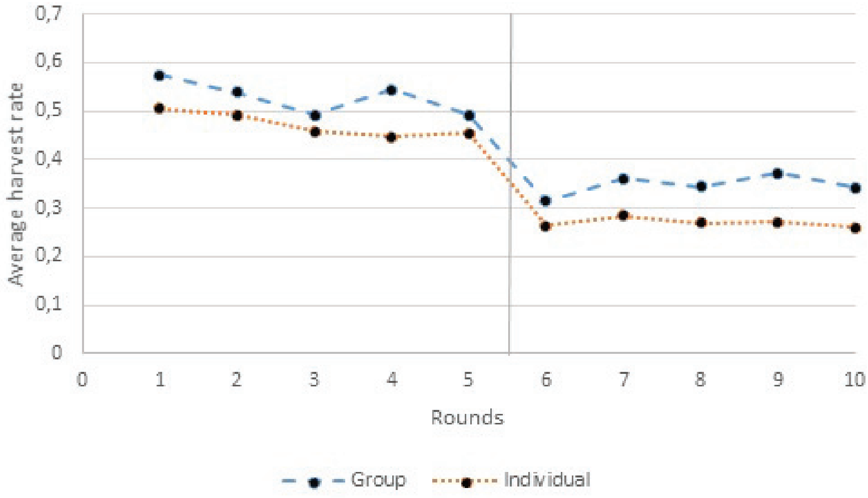

Figure 1: Trend of harvest rates before and after PES by pay modality

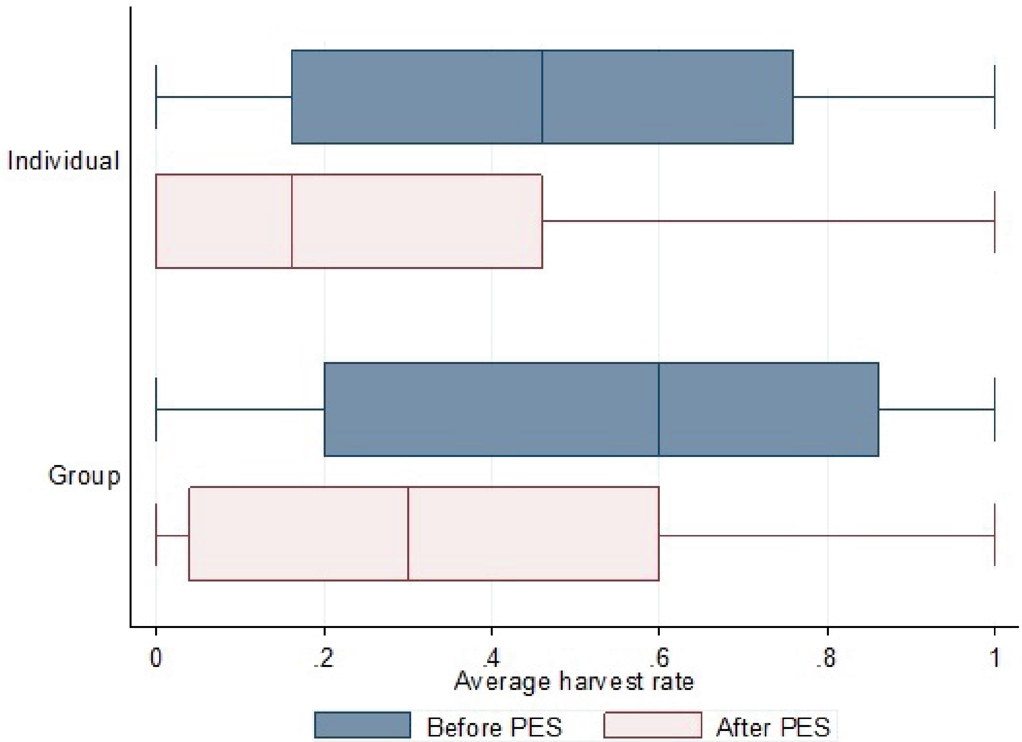

Figure 2: Box plot of individual and group PES with and without PES

To get more insight about how the reduction has been realized, we look at the distribution of the harvest rates before treatment more closely. As depicted in Figure 2, at least the first quartile of the participants with individual pay has zero average harvest rate (i.e., no whisker to the left of the box in individual pay after PES) making the median (and also the mean) lower than it is under collective reward. For individual reward, a quarter of the participants reduced (or kept) their harvest to zero. This is an interesting difference. In the collective pay, participants chose on average lower harvest than in the pre-treatment case, but not quite as low as under individual pay.

4.2.1 Simple group testing

A mean comparison test (t-test) is used to check the significance of the difference in harvest rates between the treatment groups. The distribution of harvest in our data set is not nicely bell-shaped, but the relatively large sample size ($n > 30$) enabled us to do a t-test. We also compared performance using the Mann-Whitney (nonparametric) test and found similar results. Table 2 presents evidence that avoided forest extraction under individual pay is significantly higher than that of under collective reward. Yet, caution is warranted here; the performance differential may be overstated, in light of the pre-treatment difference.

Since we used random selection of subjects and assignment of groups to treatment, the observed pretreatment difference is likely due to chance. Despite that, using the pre-treatment data points could improve the accuracy of our estimate of the impact. Given the between-group design with before and after data, we checked for the significance of the difference-in-difference estimate, but it is not statistically significant (Table 2).

Table 2: Mean comparison of average harvest rate before and after treatment by pay modality.

	Before	After	Difference
Group	0.53	0.35	0.18***
	(0.02)	(0.02)	(0.02)
Individual	0.47	0.27	0.20***
	(0.02)	(0.02)	(0.02)
Difference	0.06*	0.08***	0.02
	(0.02)	(0.03)	(0.02)

*$p < 0.1$ and ***$p < 0.01$

We thus have mixed results so far: the t-test shows significant difference while the difference-in-difference analysis says otherwise. The latter could be because there is no common trend between the two groups, in which case a difference-in-difference method will not be valid. Comparing the median values to their corresponding means (Figure 2), it is apparent that there is disproportionate response to treatment, i.e., half of those in the collective reward treatment harvest above 30%, on average, as opposed to the 16% second quartile in the individual reward. In fact, 25% of them (the third quartile) harvest 60% or more, indicating a substantial disparity in the responses.

4.2.2 Regression model

To further control for, *inter alia*, pre-treatment differences across the two groups, we ran a individual level random effects Tobit model with different control variables (Table 4). The dependent variable is harvest rate. The main independent variable of interest is the treatment, but we also included socioeconomic factors, revealed behavior in the first five rounds (unconditional cooperation, reciprocity and free riding), group average for the first five rounds, and RL dummies as controls. The random effects specification intends to take care of unobservable individual heterogeneity.

The three regression models in Table 4 are a Tobit model on the pooled data, an unrestricted (with controls to capture pretreatment performance) random effects Tobit model, and a restricted random effects Tobit model. The last column shows that failing to control for pre-treatment variation may overstate the effect of reward treatment. With the appropriate controls, there is about eight pp difference between harvest rates of those who received individual pay and collective pay. Harvest is higher in the latter, despite the sharp reduction in harvest for both groups following the introduction of PES. In other words, PES induces conservation, and more so with individual than collective pay.

Table 3: Regression of harvest rate on reward treatment and other covariates.

Variables	Tobit (Pooled)	Random effects Tobit	
		With pre. control	Without control
Reward (1=group, 0=individual)	0.0711***	0.0805*	0.134***
	(0.0273)	(0.0436)	(0.0488)
Age	-0.0163**	-0.0132	-0.0200*
	(0.00652)	(0.0106)	(0.0120)
Age2	0.000215***	0.000186	0.000259**
	(7.07e-05)	(0.000115)	(0.000130)
Sex (1=female, 0=male)	0.0653**	0.0617	0.0366
	(0.0310)	(0.0498)	(0.0561)
Position (1=some position, 0= no)	-0.0174	-0.0247	-0.0739
	(0.0314)	(0.0501)	(0.0555)
Round number (1/round)	0.183	0.0976	0.0767
	(0.562)	(0.425)	(0.426)
Livestock (in TLUs)	0.00892***	0.00884*	0.0102*
	(0.00312)	(0.00484)	(0.00532)
Uncond. Coop. (UCC[a])(1=yes, 0=no)	-0.494***	-0.463***	-0.651***
	(0.0462)	(0.0670)	(0.0729)
Free riding (1=yes, 0=no)	0.423***	0.442***	0.459***
	(0.0330)	(0.0538)	(0.0608)
Reciprocity (1=yes, 0=no)	0.0556*	0.0733	0.0594
	(0.0299)	(0.0480)	(0.0540)
Visit to forest (avg. per week)	0.0597***	0.0536*	0.0758**
	(0.0188)	(0.0301)	(0.0337)
Pre-treatment group average	0.275***	0.265***	-
	(0.0188)	(0.0293)	-
Historical RL	0.0959***	0.0927*	-
	(0.0338)	(0.0540)	-
Below historical RL	0.0799**	0.0579	-
	(0.0370)	(0.0594)	-
Woreda (1= Kafta Humera)	0.0461	0.0218	-0.269***
	(0.0421)	(0.0672)	(0.0643)
Woreda (1= Raya Azebo)	-0.138***	-0.131**	-0.0235
	(0.0338)	(0.0547)	(0.0592)
Constant	-0.524***	-0.558**	0.368
	(0.173)	(0.259)	(0.268)
sigma_u	0.534***	0.371***	0.433***
	(0.0141)	(0.0201)	(0.0223)
sigma_e		0.383***	0.383***
		(0.0107)	(0.0107)
Rho		0.485	0.562
Left censored		980	980
Uncensored		932	932
Right censored		248	248
Observations	2,16	2,16	2,16
Number of participants		432	432

Standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.

[a] UCC stands for unconditional cooperation and is defined as harvesting zero in the first round (i.e., cooperation without observing how others behave).

Several other factors can explain the variation of the harvest rate. Participants with more livestock cut more trees. Actual forest use (using average weekly visits to forest as a proxy) is also associated with higher harvest in the experiment. Unconditional cooperation, as revealed by harvesting zero level in the first round, has persistent negative harvest effect throughout. Participants who can be classified as 'free riders', those who harvested above group average in each of the first five rounds, harvest consistently more during the treatment rounds. The coefficient of the pre-treatment group average harvest shows that participants who belong to groups that start with high historical average continue to harvest more, which may point to norm-creation within each group during the experiment. The type of RL has the expected effect on harvest rates, but we are mainly interested in how the effect varies by the reward modality.

As discussed in section 2.3, collective reward creates uncertainty about whether the group will reach the RL and attain compensation. In order to test if this plays a role in causing the difference in reward performance, we run three regressions on subsamples (based on how the RL was set). We hypothesized that lower RL (relative to group average in pre-treatment) gives higher uncertainty in group-based reward and thus the coefficient of the reward treatment will be higher.

Table 5 shows that the difference in performance between individual and collective reward is about 15 pp when the RL is above historical average, and increases to about 18 pp and 27 pp when RL is equal to and below historical average, respectively. These results are in line with our expectation.

Table 4: Regression of average harvest rate on reward treatment and covariates for each group of RL treatment.

Variables	RL treatment groups		
	Above	Historical	Below
Reward (1=group, 0=individual)	0.147*	0.179**	0.272**
	(0.0811)	(0.0714)	(0.108)
Age	-0.0353*	-0.00546	-0.0335
	(0.0209)	(0.0165)	(0.0247)
Round number (1/round)	0.797	-1.433**	1.740*
	(0.702)	(0.632)	(0.915)
Livestock (TLUs)	0.00830	0.0282**	0.0294
	(0.00525)	(0.0111)	(0.0207)
UCC (1 = yes, 0 = no)	-0.468***	-0.546***	-0.678***
	(0.106)	(0.110)	(0.154)
Free riding (1=yes, 0=no)	0.285***	0.577***	0.552***
	(0.0931)	(0.0900)	(0.123)
Visit to forest (avg. per week)	0.106**	0.0474	0.00709
	(0.0447)	(0.0549)	(0.0706)
Woreda (1= Kafta Humera)	-0.448***	-0.0192	-0.372**
	(0.106)	(0.0939)	(0.189)
Woreda (1= Raya Azebo)	-0.0392	0.241***	-0.416***
	(0.105)	(0.0882)	(0.127)
Constant	0.507	0.155	0.382
	(0.467)	(0.372)	(0.539)
Sigma_u	0.331***	0.391***	0.465***
	(0.0334)	(0.0310)	(0.0464)
Sigma_e	0.334***	0.374***	0.439***
	(0.0175)	(0.0156)	(0.0237)
Rho	0.496	0.523	0.529
Left censored	339	348	293
Uncensored	262	416	254
Right censored	39	116	93
Observations	640	880	640
Number of participants	128	176	128

Standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$.
Age squared, position and sex were included in the regression but not reported, as they were not significant.

We included *woreda* dummies to control for overarching factors that determine the overall public perception about such programs and capture local governance practices, which are unique to certain communities. People from Raya Azebo *woreda* harvest less than those from Asgede Tsibla. This is in line with what the Natural Resource Conservation experts at the regional office told us before we started the experiments; there is a strong, indigenous conservation practice in the southern part of Tigray. The social sanction in some areas is so strong and consequences so severe that a young man who cuts trees illegally may be denied of getting married to a girl within the community. The circumstances in Asgede Tsimbla and Kafta Humera are similar; there is high deforestation in both areas, and more so in the latter. Many new settlers moved to Kafta Humera as part of a resettlement program during the last two decades. Informal conversations with elders revealed that there is a latent conflict between incumbents and newcomers. According to them, the latter destroy the forests and do not feel sense of ownership, to which the former react by cutting more trees as the newcomers would cut them anyway.

After controlling for all relevant covariates, specific to the experiments, individuals, or communities, we found that the difference in harvest between individual and group pay has weak statistical and modest substantive significance. The regression models give about 8 pp difference in harvest rate between individual and collective pay. The effect presented in the last column of Table 4 looks much stronger, but this overstates the impact, as pre-treatment differences are not controlled for.

5 Discussion

5.1 Overall harvest levels

The overall pattern of harvest before treatment reflects neither the Nash Equilibrium prediction of maximizing self-interest nor the social optimal level. We found that average harvest rate in the first five rounds was lower than it is in previous findings of CPR experiments (Cardenas, 2004; Ostrom et al., 1994; Rodriguez-Sickert et al., 2008), but slightly higher than what Handberg and Angelsen (2015) found. This reflects differences in the payoff structure, framing, and local context. Our results contribute, nevertheless, to the growing and consistent evidence of strong pro-social or pro-nature preferences among CPR users. The fact that many continue to harvest under individual PES is a sign of intrinsic forest use values, i.e., participants feel they need forest products such as fuelwood and cannot stop collecting easily (Handberg and Angelsen, 2016).

5.2 Individual vs. group pay

In terms of conservation, we found evidence that individual pay outperforms collective pay. This is partly explained by free-riding behavior, which undermines collective action, for instance, as conditional cooperators reciprocate negatively. This result seems to be robust across different types of resource use and environmental service: Travers et al. (2011b) studied fish harvesting, Midler et al. (2015) and Narloch et al. (2012) agrobiodiversity, and our paper forest use and conservation.

Another interesting source of the performance difference is the 'uncertainty effect' of collective pay: whether and how much a participant is compensated under different own harvest rates depends on whether and how much others reduce their harvest. This uncertainty effect is our main explanation of the rising gap between individual and collective pay harvest with increasing stringency of the benchmark (lower RL). The expectation of others' (reduction in) harvest will, in part, reflect the level of community trust. The low stated preference for group reward supports the '(dis)trust effect' as people may be afraid of 'elite capture' (Zabel et al., 2013).

Choices in social dilemma settings reflect both intrinsic and extrinsic motives. The type of reward system they face affects one or both of these sources of motivation, usually and arguably in opposite ways. The possibility of crowding out of intrinsic motivation makes Gneezy and Rustichini (2000b) advise that either "pay enough or do not pay at all". How effectively a program that pays individuals actually induces more sustainable resource use depends on both the level of payment and the strength of any crowding-out effect. We did not test this further, e.g., by having treatments with different levels of payment (Handberg and Angelsen, 2016). What we can conclude is that if there are any crowding out effects present, they are not dominant (see also Chapter 5).

Further, our experiment is (as most other FFEs) based on private decisions, i.e., the individual harvest is not public knowledge. This does not fully reflect real-life situations. Promoting transparent practices might reduce the uncertainty of group pay and increase social pressure. Further experimental field research on alternative arrangements would be useful to test the robustness of our findings, including social sanctions of free-riding or making community members choose the form of payment and other rules before the game starts. Our study was based on externally imposed institutions; some evidence suggests that endogeneity of institutions also changes behavior (Rodriguez-Sickert et al., 2008), although Handberg and Angelsen (2016) argues that weak participation has no significant effect on resource use.

5.3 External validity

External validity is one of the challenges of economic experiments. FFEs are by design meant to have higher external validity than either laboratory experiments or artefactual experiments. We used tree branches to make the product real, and selected *kebelles* where there is forest within a walking distance from the *kebelle* centre to help the participants relate the task to their daily lives.

People are generally aware that cutting trees is illegal, but they also admit that they cannot live without forest products. Conventional surveys may therefore underreport actual forest use, and we did indeed find inadequate responses to the post-experiment interview questions regarding sales of fuelwood and charcoal. This might point to the usefulness of experiments to elicit behavior, as compared to data based on household questionnaires.

To test for correspondence between real world and experimental harvest, we used average weekly visits to collect some forest product as a proxy for actual use. We found a strong positive correlation, and stronger than what Handberg and Angelsen (2015) found in their FFE in Tanzania. Participants clearly related the experiment to their forest use, to the extent that several participants thought they were given training on forest conservation, and told us that the experiment showed them how they were destroying their forests. (Such perception might have lowered the overall harvest rate in the experiment.)

Another illustrating anecdote is that in some villages people asked if cutting some branches of a standing tree would be counted as harvesting the tree (the working definition was that cutting a tree means cutting it from its roots). When we investigated such questions further, we realized that people would usually have a specific purpose in mind, for example, making a yoke, before they decide to cut a tree, and they knew from experience that some could be fulfilled with selective cutting from the branches of a big tree. Tailoring the working definition to the daily use patterns in the community could further increase external validity, but at a cost of higher complexity and possibly lower internal validity.

6 Conclusion

Framed field experiments (FFEs) are useful to test out different design elements of PES, including 'who to pay'. We found that individual pay leads to more forest conservation. There are several explanations of this. Although forest users in Ethiopia are not pure *homo economicus*, the direct pecuniary incentives individual pay provides remains a strong source of motivation. Group pay opens up the possibility of free riding, and introduces uncertainty as to whether the aggregate group harvest will be sufficiently low to qualify for payment. Lowering own harvest for the common good might be squandered. Finally, group level pay demands that you trust the authorities to distribute payments in a fair way to all participants.

Nevertheless, the difference is perhaps not as large as expected. Group PES does not change the participants' dominant strategy, which is still maximum harvest after the introduction of the treatment. Yet there is a substantial reduction, indicating significant pro-social norms.

Our findings lend support to the overall idea of REDD+, namely that the benefits should go as directly as possible to those who actually reduce deforestation and forest degradation, and individual pay performs better than collective pay. However, collective reward could reduce transaction costs due to lack of (or costly) information about individuals. Ill-defined property rights and indivisibility of the resource might add to the challenge of introducing individual pay. In contexts where the existing social capital is high, the incentive and uncertainty arguments against collective pay would be smaller. Deforestation hotspots tend, however, to have the opposite characteristic, presenting an argument for individual pay where the deforestation problem is most urgent.

References

Barr, R. F. and Mourato, S. (2014). Investigating fishers' preferences for the design of marine payments for environmental services schemes. *Ecological Economics*, 108:91–103.

Bowles, S. (2008). Policies designed for self-interested citizens may undermine" the moral sentiments": Evidence from economic experiments. *science*, 320(5883):1605–1609.

Busch, J., Strassburg, B., Cattaneo, A., Lubowski, R., Bruner, A., Rice, R., Creed, A., Ashton, R., and Boltz, F. (2009). Comparing climate and cost impacts of reference levels for reducing emissions from deforestation. *Environmental Research Letters*, 4(4).

Cadsby, C. B. and Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics*, 71(1):53–73.

Cardenas, J.-C. (2004). Norms from outside and from inside: an experimental analysis on the governance of local ecosystems. *Forest Policy and Economics*, 6(3-4):229–241.

Cardenas, J. C., Stranlund, J., and Willis, C. (2000). Local environmental control and institutional crowding-out. *World Development*, 28(10):1719–1733.

Castillo, D. and Saysel, A. K. (2005). Simulation of common pool resource field experiments: a behavioral model of collective action. *Ecological Economics*, 55(3):420–436.

Dickinson, D. L. (1998). The voluntary contributions mechanism with uncertain group payoffs. *Journal of economic behavior & organization*, 35(4):517–533.

Engel, S. (2016). The devil in the detail: A practical guide on designing payments for environmental services. *International Review of Environmental and Resource Economics*, 9(1-2):131–177.

Engel, S., Pagiola, S., and Wunder, S. (2008). Designing payments for environmental services in theory and practice: An overview of the issues. *Ecological Economics*, 65(4):663–674.

Gebremedhin, B., Pender, J., and Tesfay, G. (2003). Community natural resource management: the case of woodlots in northern ethiopia. *Environment and Development Economics*, 8(1):129–148.

Gebremeskel, G., Gebremicael, T., and Girmay, A. (2017 (in press)). Economic and environmental rehabilitation through soil and water conservation, the case of tigray in northern ethiopia. *Journal of Arid Environments*.

Gintis, H., Smith, E. A., and Bowles, S. (2001). Costly signaling and cooperation. *Journal of theoretical biology*, 213(1):103–119.

Gneezy, U. and Rustichini, A. (2000a). A fine is a price. *The Journal of Legal Studies*, 29(1):1–17.

Gneezy, U. and Rustichini, A. (2000b). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810.

Handberg, Ø. N. and Angelsen, A. (2015). Experimental tests of tropical forest conservation measures. *Journal of Economic Behavior & Organization*, 118:346–359.

Handberg, Ø. N. and Angelsen, A. (2016). Pay little, get little; pay more, get a little more: A framed forest experiment in tanzania. *Ecological Economics*.

Hardin, G. (1968). The tragedy of the commons. *Journal of Natural Resources Policy Research*, 1(3):243–253.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4).

Jahnke, H. E. and Jahnke, H. E. (1982). *Livestock production systems and livestock development in tropical Africa*, volume 35. Kieler Wissenschaftsverlag Vauk Kiel.

Kerr, J., Vardhan, M., and Jindal, R. (2012). Prosocial behavior and incentives: Evidence from field experiments in rural mexico and tanzania. *Ecological Economics*, 73:220–227.

Mahanty, S., Suich, H., and Tacconi, L. (2013). Access and benefits in payments for environmental services and implications for REDD+: Lessons from seven PES schemes. *Land Use Policy*, 31:38–47.

Midler, E., Pascual, U., Drucker, A. G., Narloch, U., and Soto, J. L. (2015). Unraveling the effects of payments for ecosystem services on motivations for collective action. *Ecological Economics*, 120:394–405.

Narloch, U., Pascual, U., and Drucker, A. G. (2012). Collective action dynamics under external rewards: Experimental insights from andean farming communities. *World Development*, 40(10):2096–2107.

Newton, P., Nichols, E. S., Endo, W., and Peres, C. A. (2012). Consequences of actor level livelihood heterogeneity for additionality in a tropical forest payment for environmental services programme with an undifferentiated reward structure. *Global Environmental Change*, 22(1):127–136.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3):137–158.

Ostrom, E., Gardner, R., and Walker, J. (1994). *Rules, Games, and Common-pool Resources*. University of Michigan Press.

Rodriguez-Sickert, C., Guzmán, R. A., and Cárdenas, J. C. (2008). Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior & Organization*, 67(1):215–227.

Rustagi, D., Engel, S., and Kosfeld, M. (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *science*, 330(6006):961–965.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67.

Sumiya, B. (2016). Energy poverty in context of climate change: What are the possible impacts of improved modern energy access on adaptation capacity of communities? *International Journal of Environmental Science and Development*, 7(1):73–79.

Travers, H., Clements, T., Keane, A., and Milner-Gulland, E. (2011a). Incentives for cooperation: The effects of institutional controls on common pool resource extraction in cambodia. *Ecological Economics*, 71:151–161.

Travers, H., Clements, T., Keane, A., and Milner-Gulland, E. J. (2011b). Incentives for cooperation: The effects of institutional controls on common pool resource extraction in cambodia. *Ecological Economics*, 71:151–161.

Vatn, A. (2009). Cooperative behavior and institutions. *The Journal of Socio-Economics*, 38(1):188–196.

Vollan, B. (2008). Socio-ecological explanations for crowding-out effects from economic field experiments in southern africa. *Ecological Economics*, 67(4):560–573.

Walker, J., Gardner, R., and Ostrom, E. (1990). Rent dissipation in a limited-access common-pool resource: Experimental evidence'. *Environmental Economics and Management*, 19:203–211.

Wunder, S. (2005). Payments for environmental services: some nuts and bolts. Bogor, Indonesia: CIFOR. Report.

Zabel, A., Bostedt, G., and Engel, S. (2013). Performance payments for groups: The case of carnivore conservation in northern sweden. *Environmental and Resource Economics*, 59(4):613–631.

Appendices

Table 5: Distribution of treatments in the study *Woredas* and *Kebelles*.

Woreda (District)	Kebelles (Peasant associations)	Reward treatment		Total
		Individual	Group	
	Mizan	8	40	48
Asgede Tsimbela	May Feres	32	16	48
	Mayshak	32	16	48
Subtotal		72	72	144
	Maykadra	24	24	48
Kafta Humera	Rawyan	32	16	48
	Adebay	16	32	48
Subtotal		72	72	144
	Tsigea	16	32	48
Raya Azebo	Hawelti	24	24	48
	Eibo	32	16	48
Subtotal		72	72	144
Total		216	216	432

Chapter 5:
Motivation crowding and forest conservation

Forest conservation and motivation crowding: a framed field experiment*

Amare Teklay Hailu

School of Economics and Business, Norwegian University of Life Sciences (NMBU), P.O.Box 5003, 1432 Ås, Norway

email: amare.teklay.hailu@nmbu.no

Abstract

This study explores the nature of motivation crowding that results from the introduction and later removal of Payment for Environmental Services (PES), and whether this differs between groups with individual and collective pay. Data are from a framed field experiment (FFE) with a random sample of 176 farmers in Northern Ethiopia and framing the commodity to real tree branches and real incentives. Results show that the PES program induced more conservation during the policy treatment stage, and this persisted even after the program phased out. Properly designed incentives may thus reinforce (crowd in, rather than crowd out) intrinsic motivation for collective action in nature conservation. But, the response is heterogeneous, and close to a quarter of the participants increased their post-treatment harvest compared with their pre-treatment levels. There are no significant differences between groups that received individual PES and groups with collective PES.

Keywords: Framed field experiment, motivation crowding, common pool resources, intrinsic motivation, PES

JEL Classifications: Q23, C93, Q57

---

*I would like to acknowledge funding from the Norwegian Agency for Development Cooperation [agreement no. GLO-3945 QZA 13/0545; grant on Learning from REDD+ an enhanced Global Comparative Analysis] through the Center for International Forestry Research (CIFOR). I am grateful to Arild Angelsen and Samantha Marie Copeland for their invaluable inputs and comments.

1 Introduction

Economic incentives such as Payment for Environmental Services (PES) have gained popularity in attempting to incentivize and compensate providers of environmental services (ES). ES such biodiversity, climate regulation and soil conservation would, otherwise, not be supplied du ket failure (Börner et al., 2017; Engel et al., 2008). Nature-dependent resource users in developing countries extract more resources to get short-term, personal material gains at the expense of more long-term and collective benefits. The underlying presumption of PES schemes is that resource users should gain a material payoff from participation in and compliance with these programs, as compared with business-as-usual behavior. In this way, PES schemes intend to make resource users internalize thus-caused externalities into their day-to-day decisions.

Yet, there is ample evidence from behavioral and experimental economics studies that the users may have pro-nature and pro-social preferences as well (Benabou and Tirole, 2003; Bowles, 2008; Bowles and Polania-Reyes, 2012; Deci, 1971). Originating in social psychology literature and being increasingly embraced by economists is the duality of the motives: intrinsic motivation—doing something inherently satisfying—and extrinsic motivation—doing something that brings results with external value (Deci, 1971; Deci and Ryan, 1975). The external material incentives interact with intrinsic sources of motivation, i.e., they may have a multiplicative instead of an additive role.

The key issue of this paper is whether the introduction and design of economic incentives have an impact on social preferences or not. Bowles and Polania-Reyes (2012) refer to this as the "seperability assumption", where economic incentives and social preferences are independent. In addition to weighing the material costs and benefits of each course of action (i.e., participating or not participating), the particular (design of the) policy intervention may have an impact on the agents' social preferences. For example, whether the amount of proposed compensation is large enough (Gneezy and Rustichini, 2000), or whether it signals high or low value of the environment, and whether the monetization of efforts to conserve the resource shifts their perspectives (Rode et al., 2015), will determine how agents respond. The effect of the intervention on intrinsic motivation cannot be determined *a priori*, i.e., it can reinforce or discourage pre-existing efforts.

The theoretical predictions and empirical evidence on motivation crowding theory are equivocal. There are both reports of motivation crowding out effects of economic instruments. Festré and Garrouste (2015), Cardenas et al. (2000), Kerr et al. (2012) are among those who found crowding out effect. On the other hand, Narloch et al. (2012) and Rodriguez-Sickert et al. (2008) are examples of studies where external incentives have crowding in effect. Several studies found no significant effect, i.e., neither crowding out nor crowding in effect (see Rode et al. (2015) for a comprehensive review and synthesis), and there also in this area likely to be a publication bias in the way the non-significant findings are less likely to be published. I give a brief overview of relevant empirical results in section 2.

This paper contributes to this ongoing debate concerning whether external monetary reward, in the form of PES, leads to motivation crowding in, crowding out, or neither. It is based on data collected from a framed field experiment conducted in Northern Ethiopia. In a stylized PES scheme that resembles key elements of the global effort for Reducing Deforestation and forest Degradation (REDD+), participants first harvest trees in a common pool resource game. In the second stage PES is introduced, and then removed in the third stage. The crowding in/out effect is measured as the difference between pre- and post-treatment harvest. During the PES stage, participants received their conditional reward either individually or collectively.

This paper addresses three research questions. First, is there motivation crowding in or crowding out because of introducing and revoking PES? I test whether the PES has motivational crowding out effect, or whether it has 'prescriptive effect' (Rode et al., 2015) on forest conservation. In either case, the setting of the experiment allows for exploring if the impact of the PES changes or persists beyond the program stage as well. Second, is there a systematic difference in terms of motivation crowding in the post-treatment stage between those who receives individual and collective PES? Third, is there heterogeneity in motivation crowding in the post-treatment stage between subgroups, i.e., based on individual characteristics, other than the pay modality, of the participants?

In the following section, I discuss motivation crowding theory vis-à-vis the psychological concepts of intrinsic and extrinsic motivation, and give examples of situations where external economic incentives could either reinforce or undermine intrinsic motivation. Section 3 describes the framing of the experiment and data. Section 4 presents the main results, and section 5 discusses their implications their implications. Section 6 gives concluding remarks.

2 An overview of motivation crowding theory

2.1 Social psychology theories

Motivation crowding theory is contentious, as there is a debate in the literature as to what constitutes intrinsic and extrinsic motivation (Reiss, 2005). Motivation crowding originates in theories within social psychology. One of these is *self-determination theory* (SDT), which suggests that external incentives might crowd out intrinsic motivation, as the agent in question feels that their locus of control changes from the self to an external force (e.g., a principal), and their autonomy is degraded (Frey and Jegen, 2001). Another is *cognitive evaluation theory* (CET), where motivation crowding out is ascribed to diminished feelings of competence and confidence following an external reward (Deci and Ryan, 1975; Frey and Jegen, 2001).

These theories are based on the underlying distinction between intrinsic and extrinsic sources of motivation (Deci, 1971; Deci and Ryan, 1975; Festré and Garrouste, 2015). Intrinsic motivation refers to a situation where the activity involved is thought of as an end and choosing that course of action by itself renders satisfaction, whereas extrinsic motivation describes a situation where the activity leads to a separate external outcome that brings satisfaction.

There is also a debate on which outcomes matter. The rational, self-interested agents of neoclassical economics theory are expected to respond only to changes in monetary pay-offs, whereas humans in real life have social preferences and also consider outcomes in terms of, for examples, fairness and equity when making decisions. There is mounting evidence that economic incentive mechanisms that intend to change behavior should not only attempt to create external material reward but also anticipate the likely effects of these incentives on intrinsic motivation. In some cases, the adverse effect of the external incentive motivation outweighs (and results in crowding out), and in other cases it reinforces pre-existing intrinsic motivation and bolsters it (motivation crowding in) (Kaczan et al., 2016; Rode et al., 2015).

2.2 Cases for motivation crowding in

The standard (neoclassical) economics choice theory predicts that rational self-interested agents will try to balance the costs and benefits of a set of alternatives and choose the one that brings them the highest material payoff. In choice settings involving risk, the neoclassical view is that decision-making agents maximize expected payoffs (expected utility theory). However, mounting evidence suggests a long list of 'anomalies' to this behavior in real life, e.g., social preference, pro-nature preference, inequity aversion and altruism. These are embedded in the intrinsic motivation discussed earlier (Bowles, 2008).

When external economic incentives take into account these behavioural aspects, they will reinforce intrinsic motivation and result in motivation crowding in effect. Rode et al. (2015) review 18 empirical papers on motivation crowding and summarize the main psychological mechanisms (see also Bowles and Polania-Reyes (2012)) in which motivation crowding in could occur. Motivation crowding in typically occurs when the external incentives are sufficiently large, and are perceived as supportive of pre-existing behavior. They can also have a reinforcing effect when they signal as to what is the desired course of action as well as how others are likely to respond to the incentive (Narloch et al., 2012).

By enhancing environmental and social awareness about the type of behavior that is more desirable, properly designed economic incentives could signal the value of cooperation, and individuals come to learn that cooperation pays off. When players learn about others' cooperative behavior, it results in convergence of individual decisions towards more socially optimal points. In cases with little cooperation initially, the economic intervention may be perceived as a good way to deter defectors and will in turn motivate conditional cooperators (Narloch et al., 2012).

In a best-case scenario, rewarding intrinsically motivated pro-social and pro-nature behavior reinforces them further and increases cooperation. Individuals enjoy good self-esteem and better social recognition, which give them instant 'warm glow' (Andreoni, 1990). It may also build more trust and promote sustainable behavior. These are, however, highly dependent on the local context and individual motivation. For example, local communities may value non-economic measures that recognize their traditional conservation mechanisms as much (or more) as (than) external financial incentives (Van Hecken and Bastiaensen, 2010).

2.3 Cases for motivation crowding out

Titmuss (1971) was the first to contend, intuitively, that paying individuals for socially desirable activities that they perform out of moral concern, such as blood donation, may be counterproductive. At first glance, this argument contrasts the economic prediction that any additional positive payment given to individuals for any action they do for free will make them do no less, if not more, of it. However, the motivation crowding out effect has been identified and recognized as an important anomaly (Festré and Garrouste, 2015). External incentives, through their interactions with internal commitments and moral sentiments, may lower motivation to engage in pro-social behavior.

Once again, perception serves as an interface between the parameters of the external incentive and the internal motivating factors of the individuals. When a principal uses external reward to entice an agent into doing a certain action, the latter reacts to the size and meaning of the incentive, learns about the intents and purposes of the principal and forms a belief about the belief the principal might have formed about the agent. The intrinsic motivation of the agent diminishes when such measures are perceived as controlling (against his autonomy) (Frey and Stutzer, 2006; Vollan, 2008).

Studies show that motivation crowding out occurs for a number of reasons: first, control averse individuals may find it controlling and against their need for self-determination when an external mechanism is imposed to change their pre-existing behavior (d'Adda, 2011). Second, and related to control aversion, individuals may feel frustrated when they perceive that a policy is put in place to correct their behavior because they could not be trusted (Gneezy et al., 2011). This is again related to a reduced self-image as one's social commitments may now be attributed (or could at least not be separated from that which is due) to the external incentives (d'Adda, 2011; Kerr et al., 2012). Third, paying individuals for environmental services gives them 'guilt relief' when they do not act in the sustainable way, because it only means they prefer not to comply (Cardenas et al., 2000; Rodriguez-Sickert et al., 2008). Fourth, introducing economic incentives may shift the frame of reasoning and change mindsets of individuals such that they focus on economic reasoning instead of their moral concerns. The monetization of social and environmental values creates a new long term perspective where people start to make decisions on monetary basis rather than moral sentiments, and the intrinsic motivation to choose more sustainable strategies ends up being crowded out by the newly hatched economic motives (Fisher, 2012; Greiner and Gregg, 2011).

The literature seems to present more cases of crowding out than crowding in effects (Rode et al., 2015). Notable studies include Frey and Oberholzer-Gee (1997) who found that price incentives may crowd out people's 'civic duty'. In biodiversity conservation, Narloch et al. (2012) examined the effect of rewards for using traditional crop varieties in Bolivia and Peru and found crowding out in collective reward and crowding in individual reward. Fisher (2012) studied a payment scheme for tree planting in Uganda, and the results suggest that financial incentives are less effective. In the context of forest management in particular, García-Amado et al. (2013) argue that a shift in focus from conservation to monetization may explain crowding out effect of external incentives.

Two other relevant studies based on FFE in East Africa (Tanzania) are worth mentioning here. Kaczan et al. (2013) investigated motivational crowding in forest conservation and concluded that their PES treatment did not crowd in or crowd out motivation significantly. Only in subsamples in their study did they find suggestive results that both crowding in and out effects coexisted. Similarly, Handberg and Angelsen (2016) found little evidence to support motivation crowding out. They argue that paying little yields only little conservation, which suggests more conservation could only be realized by paying more, not because paying little crowds out motivation but because it simply is not enough.

2.4 Research questions

Given the issues raised in the above discussion, I designed an experimental study intending to answer the following questions.

Q1: Does introducing PES and later removing it crowd out pre-existing motivation to conserve forests? The design of the PES scheme in the experiment involves three stages: stage one (baseline, pre-treatment), stage two (introducing PES, treatment) and stage three (removing PES, post-treatment). The answer to this research question will be positive if the average harvest rate in stage three turns out to be statistically significantly higher than it is in stage one. Whether there will be crowding out or crowding in effect is an empirical question, because it is not possible to predict it *a priori*.

Q2: Does the type of motivation crowding depend on whether the PES have been made individually or collectively? During stage two, the participants were offered individual or group based rewards for reducing tree harvest beyond a stipulated benchmark. It is not self-evident whether and how the pay modality during the PES stage would make a difference in how participants react to the removal of the PES.

Q3: Participants may have different preferences towards other participants, the environment and the sponsors of the program. Are there heterogeneous responses to the removal of the PES program, and what factors explain them?

3 Data and methods

3.1 The study area

Ethiopia is the second most populous country in Africa, with high competition among various land use options. The government is striving to realize a Climate Resilient Green Economy (CRGE) (FDRE, 2011) through both domestically promoted and externally supported environmental rehabilitation programs, including both afforestation and reforestation. Following the call for voluntary participation in REDD+ by the UNFCCC, Ethiopia has developed its national strategy, and there are several REDD+ projects operating mainly in the Southern part of the country.

The Tigray National Regional State, located in the North, has been exposed to severe environmental degradation over the past several decades. But, extensive environmental rehabilitation efforts have been implemented widely with significant results, also earning international recognition.[1] Despite the increasing environmental awareness and massive participation in environmental programs, the poor rural farm households in Tigray, as in the rest of the country, remain the primary agents of deforestation and forest degradation.

3.2 Data collection and sampling

We conducted a framed field experiment between February and June 2016. The region has six administrative zones,[2] and *kebelles* (peasant associations, the lowest administrative units) were selected for the study from three of those zones: the southern, north western and western zones. Three *kebelles* were selected from each zone to give a total of nine. The

[1] https://www.reuters.com/article/us-climatechange-accord-temperature/this-year-to-be-among-three-hottest-on-record-extraordinary-weather-u-n-idUSKBN1D619E accessed on 07.11.2017

[2] According to the Ethiopian Federal system, Regional States are divided into administrative zones, which consist of districts, which in turn consist of *kebelles*.

existence of natural forest close to the *kebelle* center was the main criterion for selection, while the zones were selected such different ecological and socio-cultural are represented in the sample.

Once the *kebelle* was identified, a team of enumerators visited the *kebelle* administration and requested for a full registry of the *kebelle* residents. As most local administrators would not readily permit such studies, we obtained letters of cooperation from *woreda* (district) forest and natural resource management experts each of whom had, in turn, been contacted by their regional coordinator. From each *kebelle* registry, 48 households were randomly selected, and invitation letters were sent to the person registered as household head. The total sample size is thus 432 household heads who participated in a bigger experiment, of which this study is a part. For this paper, data from 176 subjects, who were randomly assigned to the *historical average* reference level (RL)[3] treatment, were used.

In most families, men are registered as household heads, unless it is a female-headed household. Therefore, most of the participants (about 69%) turned out to be men. As the invitation letter clearly states participation is voluntary, turning up at the stipulated time and place was taken to be an expression of consent. Each participant was randomly assigned to a session consisting of eight players.

3.3 Experimental design

Participants played a common pool resource (CPR) game, i.e., they face a social dilemma between private benefit from harvesting more, or a larger common benefit shared among all players if they do not harvest. The experiment is done in three stages. The first stage is without PES, and participants face the CPR dilemma. Second, we introduced PES conditional on harvest being lowered from a benchmark (reference level), which was set at the historical average (i.e., group average in the baseline stage). We randomly assigned groups to two types of PES: individual or group based pay. In the former, individuals compare their harvest with the historical average (reference level) and decide whether to participate or not. If their harvest in a given round is below the reference level, they receive PES for the amount of reduction. In the case of group level pay, the PES requires the group's average to be below the historical average, and the payment is shared among the participants. This introduces an element of uncertainty that is further discussed in Chapter 4 of this thesis.

In the third stage, we removed the PES and made participants play with the same pay-off structure as in the first stage, the only difference being that they now have experienced a PES treatment stage.

We read out detailed instructions in the local language and showed demonstration examples. There was a practice session to make sure all participants understood the process. At the end of each round, we announced the group total harvest and picked the same number of tree branches from the stock on display. This helped the participants to consider their contribution relative to group averages, to learn about how group behavior evolves, and to see the effect of their collective behavior on the forest resource. We believe that the real incentives we gave the participants coupled with the real tree branches we used to frame the

[3] Analyses of the effect of the PES vis-à-vis varying forest reference level (RL) restrictions and alternative pay modalities (individual vs. group pay) have been done in accompanying papers.

forest make them reveal their actual behavior better than, for example, just writing their choice (number of trees to harvest) on a piece of paper.

The payoff structure for the first and third stages is as follows:

$$\pi_{it} = x_{it} + 2\left(\frac{60 - \sum x_{it}}{8}\right), \quad \text{where} \quad x_{it} \leq 5 \tag{1}$$

makes him better off than not harvesting ($1 > 0.25$). Rational individuals are expected to harvest the maximum allowable amount, $x_{it} = 5$, while the social (aggregate group) payoff is maximized when everyone harvests nothing, $x_{it} = 0$.

During the treatment stage, participants either received an individual or collective PES if the harvest of the individual participant (or the group average) was below the historical (i.e., stage 1) average harvest of the group, denoted $\bar{\mathrm{H}}$. The payoff function during stage 2 is thus:

$$\pi_{it} = \begin{cases} x_{it} + 2\left(\frac{60-\sum x_{it}}{8}\right) + 0.75 * \mathrm{Max}(0, \bar{\mathrm{H}} - x_{it}), & \text{individual pay.} \\ x_{it} + 2\left(\frac{60-\sum x_{it}}{8}\right) + 0.75 * \mathrm{Max}(0, \bar{\mathrm{H}} - \frac{\sum x_{it}}{8}), & \text{group pay.} \end{cases} \tag{2}$$

where $x_{it} \leq 5$.

Harvesting one more tree yields one ETB as a direct benefit in both individual and collective PES. The payoff from conserving one tree is still one ($\frac{2}{8} + 0.75$) in individual PES, but $0.34(\frac{2.75}{8})$ in collective PES. In the latter, the participants face uncertainty about whether the group will reach the benchmark and qualify for PES or not, which implies that one may forgo harvest and end up earning only $0.25(\frac{2}{8})$. If the group qualifies while certain members harvest beyond the benchmark, they are paid despite their free riding behaviour. The impact of this incentive discrepancy has been analyzed in Chapter 4 of this thesis.

The payoff function in stage three is the same as Eq. (1). As the decision problems in each round of stage 3 remains the same as those of stage 1, there is no real economic incentive to change behavior apart from (positive or negative) behavioral reactions to and rational expectations based on their experience. This experience does not affect the pecuniary gain available for each choice set.

3.4 Data analysis

We started with a simple mean comparison test within the group in the pre- and post-treatment stages. For a robustness check, we conducted nonparametric tests as the simple student's t-test may be sensitive to the normality assumption. We also compare between group means by PES category to see if paying individuals or paying groups results in different motivation crowding-related conclusions.

The overall average treatment effect can give us an overview of the dominant responses but we have to find out specific mechanisms that determine how different participants respond to the change in the rules of the program. We have potential sources of variation such as reward type, gender, age, actual forest use, etc. Additional regression results have been reported to control for these factors and check the consistency of the results in the mean comparison tests. In a difference-in-difference analysis, we take advantage of the panel nature of the experimental data and test for the significance of the indicator variables for different stages of the sessions. Both the standard panel and two-limit-Tobit models were used, and the latter is more appropriate as the dependent variable (harvest rate) is censored both from below (at zero) and from above (at one). Then Ordinary Least Squares (OLS) was used to assess heterogeneity in the degree of conservation crowding, measured as the difference between average harvest rates in stages 3 and 1.

4 Results

4.1 Descriptive statistics

Table 1 presents summary statistics of relevant characteristics of the participants. As noted earlier, the majority of the participants were men, while close to a third were women. Being a household head, i.e., that one's name is on the *kebelle* registry, was the only criterion used to select subjects. There was considerable range in the age of the participants. The average age is 42 years, which is representative of the typical middle-aged farmer who is physically active enough to engage in deforestation to get agricultural land and in forest products extraction. In terms of education, 52% of the sample household heads cannot read and write, while the remaining 48% either had formal education or can read and write because they had religious education.

Participants were asked how many times a week, on average, they visit the forest to collect some forest product. This varies by season: in the rainy season, the opportunity cost of time spent in the forest is high since they have more productive work to do on the farm. Most women state that they collect firewood for the rainy season during the dry season, and most men prepare farm tools, build houses and animal shades and clear land for agriculture during the late months of the dry season. Overall, the average participant stated that they visit the forest about once a week. An important advantage of having a measure of forest harvesting in real life is to check the validity of the observed behavior in the experiment in representing the real behavior in the day-to-day lives of the subjects. We found that the correlation between the two was positive and significant both in the whole experiment and in this subsample.

We are interested in the harvest rate, calculated as the ratio of actual harvest to the maximum allowable amount ((harvest/5)*100). Generally, average harvest rate is lower in stage 3 than it is in stage 1. We take the difference between the two as a measure of conservation crowding, which ranges from negative (crowding in) to positive (crowding out). The former dominates in the whole sample, with 9 percentage points (pp) difference. Depending on whether this difference is negative, zero or positive, we crudely classified participants to three indicators of crowding status.[4] A participant's crowding status is said to be crowding

[4] This classification might be sensitive to the definition of crowding in/out. Any positive or negative difference between stage 3 and stage 1 is strictly labeled as crowding out or crowding in respectively. However, there may be cases of no crowding a certain fraction of standard deviations from zero on both sides. In fact, 60% of the participants had their differences within one standard deviation on both sides

out if average harvest in stage three exceeds his average harvest in the baseline stage [5]. This hints at the existence of some heterogeneity of responses to the introduction and later removal of the PES scheme, as nearly a quarter of the participants have crowding out status.

Table 1: Summary statistics of selected characteristics of the sample.

Variable	Mean	Std. dev.	Min	Max
Participant characteristics:				
Age	42.18	12.96	20	80
Gender (1=male)	0.31	0.46	0	1
Education (1=read and write or above)	0.48	0.5	0	1
Forest visit (weekly average)	0.88	0.66	0	3.5
Experimental results:				
Avg. harvest rate in stage 1	0.53	0.34	0	1
Avg. harvest rate in stage 2	0.36	0.31	0	1
Avg. harvest rate in stage 3	0.44	0.36	0	1
Avg. PES reduction (stage 1 – stage 2)	0.18	0.22	-0.4	0.92
Avg. reversal (stage 3 – stage 2)	0.09	-0.52	0.2	0.96
Avg. crowding (stage 3 – stage 1)	-0.09	0.19	-0.8	0.44
Crowding status:				
Crowding in (1=yes)	0.53	0.5	0	1
Crowding out (1=yes)	0.24	0.43	0	1
No change (1=yes)	0.23	0.42	0	1

Comparing all three stages, tree harvest was highest in stage one, though only 53% of the Nash Equilibrium level, then it dropped sharply by 18 pp to 36% following the introduction of PES. When PES was revoked, average harvest reversed but only by about half compared with the stage one harvest rate. Thus, on average we see a crowding in effect of about 9 pp.

of zero.

[5] This is only suggestive of the heterogeneous effects of the withdrawal of the PES, as average of five rounds might be affected, for instance, by accidentally high choice in one round.

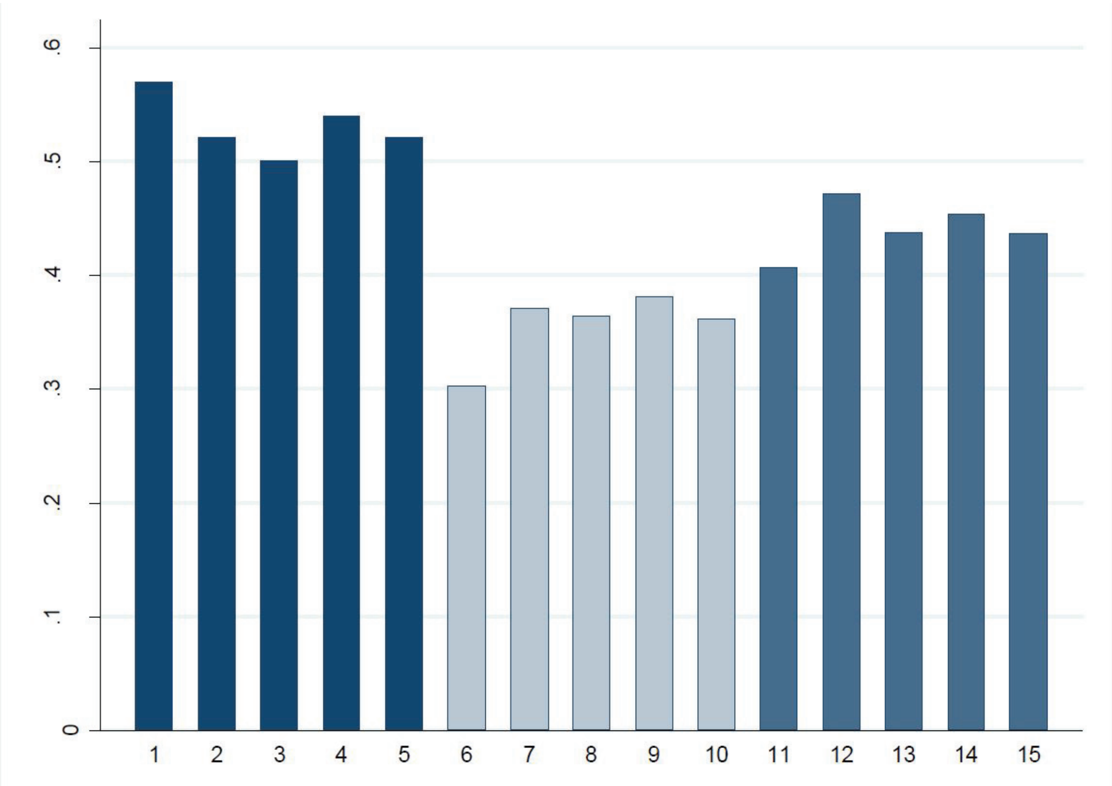

Figure 1: Trend of average harvest rate over 15 rounds for the whole sample

Figure 1 visualizes this by showing the trend over all 15 rounds (with 5 rounds for each of the three stages). We noted a large reduction in the harvest rate just after PES is introduced, while the increase is comparatively small when PES is removed. We tested for the statistical significance of the differences further, and results are reported in Table 2.

The second hypothesis related to differences between the sub-groups receiving PES based on individual or collective performance. Figure 2 depicts the mean and median values for groups in each pay type at all three stages of the game. The overall pattern looks similar: harvest rate is higher before the PES in both individual and collective PES. Then it drops by about 19 and 16 pp in individual and collective PES, respectively. Eventually, it goes up again in response to the removal of the performance based reward by 10 pp in individual PES and 7 pp in collective PES. The statistical test for significance of this difference is included in Table 2.

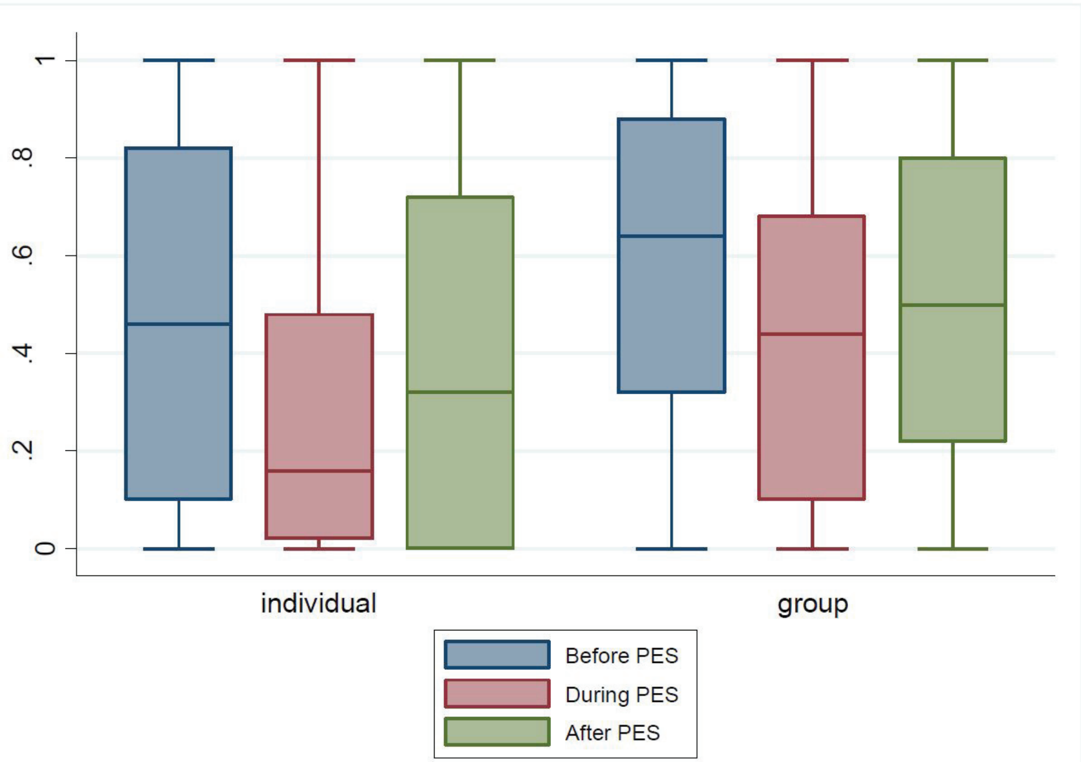

Figure 2: Comparison of harvest rates before, during and after treatment by pay modality

4.2 Statistical tests and regression results

The data consists of subgroups of the sample based on pay modality, and within (sub)group records of harvest rate over different (stage 1 and stage 3) stages. These allow for both within and between group mean comparison tests. Table 2 presents several combinations of pairwise mean comparison tests. First, test results of the comparison between overall average harvest rates in stage one and stage three, show that there is a statistically significant difference between the two. As suggested by a quick inspection of the figures in the previous section, the former is significantly higher.

Table 2: Pair-wise mean comparison tests (within and between group t-test).

Comparison	Variables	Mean	Std. Err	p-value
Within (full sample):				
	Avg1 (Average harvest in stage 1)	0.53	0.007	
	Avg3 (Average harvest in stage 3)	0.44	0.007	
	Difference (Avg1 – Avg3)	0.09	0.004	0.00
Within (Individual PES):				
	Avg1 (Average harvest in stage 1)	0.48	0.038	
	Avg3 (Average harvest in stage 3)	0.39	0.039	
	Difference (Avg1 – Avg3)	0.095	0.022	0.00
Within (Group PES):				
	Avg1 (Average harvest in stage 1)	0.58	0.035	
	Avg3 (Average harvest in stage 3)	0.50	0.036	
	Difference (Avg1 – Avg3)	0.08	0.019	0.00
Between (individual vs. group PES):				
	Individual reduction (Ind.Avg1 - Ind.Avg2)	0.194	0.026	
	Group reduction (Grp.Avg1 – Grp.Avg2)	0.155	0.019	
	Difference	0.039	0.033	0.23
	Individual reversal (Ind.Avg3 – Ind.Avg2)	0.099	0.024	
	Group reversal (Grp.Avg3 – Grp.Avg2)	0.072	0.019	
	Difference	0.027	0.031	0.36
	Individual net (Ind.Avg1 – Ind.Avg3)	0.095	0.022	
	Group net (Grp.Avg1 – Grp.Avg3)	0.083	0.019	
	Difference	0.012	0.029	0.68

Note: average harvest rate in stage 1 is higher in collective pay despite the random selection and assignment to treatments of participants. This is the shortcoming of having relatively small sample size.

Second, the same mean comparison test is applied to each group in the two pay modalities, and we find, likewise, a strong statistical difference between pre- and post-PES average harvest within each group. Third, we dig deeper and examine the difference in the dynamics suggested by Figure 2. The reduction in average harvest between stage one and stage two is higher in individual PES (see Chapter 4). Nevertheless, average harvest in individual PES also reverses by a relatively higher amount than its counterpart in collective PES does. The net effect is, thus similar between the two: the higher response to the PES in individual PES seems still to reverse by high response to the removal of the PES too, making the final average amount comparable to collective PES, which is relatively more stable. Though the changes in individual PES appear to show more variability, there is no statistically significantly difference between the two groups in this respect.

Table 3: Difference-in-difference regression of both standard panel and two-limit Tobit models.

Variables	Standard Panel (RE)	Tobit (RE)
Participant characteristics:		
Age	-0.003	-0.005
	-0.009	-0.02
Sex (1=female)	0.148	0.144
	-0.261	-0.55
Position (1= yes)	-0.277	-0.791
	-0.309	-0.655
Visit forest (average no. of visits/week)	0.383**	0.943**
	-0.18	-0.379
Treatments:		
Pay type (1=group)	0.422*	0.739
	-0.246	-0.515
PES stage (1=yes)	-0.970***	-2.140***
	-0.078	-0.167
Post-PES stage (1=yes)	-0.475***	-1.114***
	-0.078	-0.164
payType*PESstage	0.195*	0.677***
	-0.11	-0.226
payType*postPESstage	0.059	0.328
	-0.11	-0.225
Auxiliary results:		
Constant	2.227***	1.771*
	-0.49	-1.035
sigma_u		3.193***
		-0.201
sigma_e		1.987***
		-0.046
Observations	2,640	2,640
Number of participants	176	176

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We further affirm this evidence with regression results from a difference-in-difference formulation and interaction terms between pay modality and stage of the experiment. Table 3 reports the random effects regressions of the standard panel and two-limit Tobit models.

Harvest rates are statistically significantly lower during and after the PES relative to the pre-PES stage, but the difference is smaller when PES has been removed. Despite the slight reversal in harvest following the termination of PES, the program has a crowding in effect. There is not enough evidence to claim any difference between pay modalities, as the coefficient for the interaction term between pay type and post-PES stage is not statistically significant (see also Figure 4). The coefficient for the average number of visits to forests supports the claim that outcomes in the experiment reflect real behavior, and thus indicates the external validity of the experiment.

4.3 Heterogeneous responses

Average treatment effect (ATE) is a reliable measure of the effect of an intervention when there is a nearly homogeneous effect of a treatment among participants. Otherwise, extreme values might create substantial bias, and the ATE no longer serves as an accurate estimate of the most likely treatment effect in the population. Two observations in the experimental data pointed to the need to further scrutinize and search for potential heterogeneity.

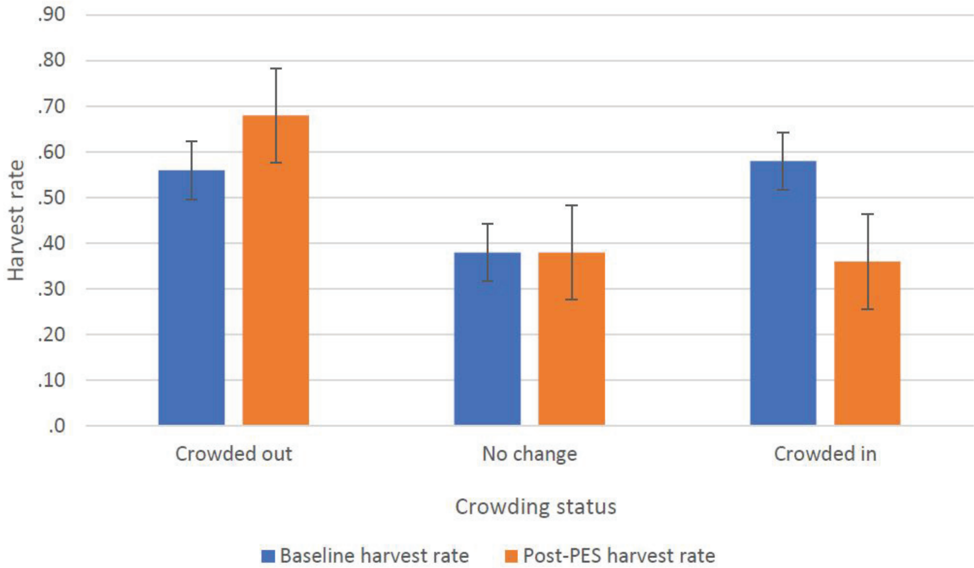

Figure 3: Comparison of harvest rates before and after treatment by crowding status

First, most of the data points on harvest rate appears to pile up either around zero or close to the maximum value. Second, 24% of the participants increased their average harvest from stage one to stage three (see Figure 3). These raise questions about whether this tendency to stick to extreme values is random or driven by some systematic underlying mechanism, and—more generally—if this confirms heterogeneity in the crowding among participants .

For the group of participants where there was a tendency toward crowding out, harvest rate increased from 56% in stage 1 to 68% in stage 3. As a higher proportion of the participants decided to save more trees after the PES was withdrawn, the amount of conservation crowded out has been more than offset. The crowding in effect dominates as the majority (53%) of the participants reduced harvest by about 22 pp while only about a quarter of the participants increased their harvest by 12 pp.

The next question concerns whether these differences in the response of participants to the removal of the PES are systematic or random. I took the difference between average harvest rates in stage 3 and stage 1 as a measure of conservation crowding. The distribution of this variable is close to bell-shape, and its values range from negative (crowding out) to positive (crowding in). Thus, OLS regression was used to explore what factors explain this variation. Pay modality, sex, forest reliance (as proxied by average weekly trips to collect forest products), position in the local administration, and average harvest rate in the baseline stage were regressed on the measure of crowding. Table 4 reports the OLS regression results. Though it is difficult to make causal inference from the relationship between average harvest in stage 1 and the measure of crowding, including this in the regression can help to control for unobserved individual characteristics, which affect harvesting behavior from the very beginning.

Table 4: Linear regression of conservation crowding.

Variables	Coefficients	Std. Errors (Robust)
Participant characteristics:		
Pay modality (1=group pay)	-0.028	0.029
Sex (1=female)	-0.053*	0.032
Position (1= have some position)	0.003	0.033
Forest reliance (Avg. weekly trips to forest)	-0.035*	0.021
Experimental results:		
Avg. harvest in stage one	0.501***	0.139
Sq. Avg. harvest in stage one	-0.365**	0.141
Constant	0.03	0.024
Observations	176	
R-squared	0.108	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Given our measure of conservation crowding, a positive (negative) coefficient means that the variable is associated with a crowding in (out) effect. Table 4 shows that removing PES is likely to trigger more crowding out behavior among female participants than among males, all other factors being equal. Similarly, participants who rely more on forest products in daily life are more likely to revert to their pre-treatment harvest levels, sometimes exceeding their harvest in the baseline stage.

The higher the harvest rate is in stage 1, the stronger is the positive correlation with the measure of conservation crowding. This is not surprising, as the computation of our measure of crowding includes harvest rate in stage 1. The interpretation of this relationship is not straightforward. Although this may suggest that the potential reaction of an average participant to the removal of a PES can be predicted using their behavior in the baseline stage, the real cause may be another factor affecting behavior. A high harvest rate at stage 1 predicts a higher propensity to crowd out, with a turning point when harvest in stage 1 is 68.6%. This relationship is depicted in Figure 5.

The statistically insignificant effect of the pay modality comes as no surprise given the analysis in the previous section. Having been paid individually or collectively does not seem to matter for the crowding behavior. An indicator variable for some position in the local administration was included to capture if participants who are close to development agents possess additional motivation and feel more responsibility, but it has no significant effect.

5 Discussion

5.1 Crowding effect

It has been well established, in theory as well as with a plethora of experimental evidence, that resource users are not selfish money maximizing *homo economicus* beings (e.g., Ostrom (1999)). Designing PES schemes merely on the basis of neoclassical predictions is likely to yield inefficient outcomes, and possibly backfire. That is, there have been conflicting empirical results regarding the effect of an external economic incentive. In some cases, external incentives have detrimental effects on conservation (Cardenas et al., 2000; García-Amado et al., 2013; Vollan, 2008), while in others they increase collective action and hence achieve their intent (Narloch et al., 2012; Rodriguez-Sickert et al., 2008).

In contrast to the view that PES may crowd out pre-existing moral sentiment and intrinsically motivated commitment to adhere to sustainable and other-regarding behavior, the results in the present study suggest the opposite. The aggregate effect of introducing performance-based payments turns out to be in favor of them. During the PES stage, as reported in two other accompanying papers (Chapters 2 and 3), there was substantial reduction in tree harvest. This is unsurprising, because the PES has real incentive effect while triggering some group dynamics in collective pay modality. Even when the PES scheme phases out, average tree harvest remained below those of pre-PES stage despite the fact that the payoffs were the same.

This is good news to programs like REDD+, where the effect has to be permanent while it is hard to pay forest users forever. The reason why a rational resource user, facing exactly the same payoff function, would cut fewer trees after PES than he used to before PES is not obvious. We can only speculate and allude to what Rode et al. (2015), Rodriguez-Sickert et al. (2008) and Bowles and Polania-Reyes (2012) would call the *prescriptive effect* of the program. The regional and local governments have been trying to create environmental awareness, implement legal restraints on cutting trees and promote good practices by rewarding environmental stewardship. When the experiment takes place with this backdrop, it may not be difficult for the participants to consider it as a continuation of another twist on such large-scale ongoing efforts. Indeed, some participants mentioned in exit interviews that the experiment "was a very good training which showed them their actual behavior". Even though we explicitly told them that it was an experiment to observe their relationship with their natural forest resources, they thought they were being given training, and this seems to be reflected further in their behavior after the PES phased out. The focal point theory (Schelling, 1980) may explain some of the tendency to become more pro-environmental.

The framing of the experiment may also matter, as outcomes are always context dependent. We introduced a PES scheme that mimics REDD+ in our field experiment, and we found significant reduction in harvest, not least because the reward was perceived as supportive of the participants' efforts to conserve as much as they can. The motivation to conserve that persisted beyond the project stage was stronger evidence favoring motivation crowding in effect than similar studies had found. For example, Kaczan et al. (2016) reported that they did not find persistent motivation crowding out, but that there were both crowding in and crowding out effects simultaneously.

5.2 Motivation crowding and pay type

As relevant as the question of 'performance of performance-based incentives' under individual and collective PES is how they fare after the termination of the scheme. Both Figure 2 and the accompanying statistical tests indicate that the reductions in harvest rates are similar between those offered individual and collective pay modalities. This is different from previous findings such as Narloch et al. (2012). The pay modality does not affect the payoff functions of participants in both the first and third stages, and these are the outcomes of interest here. This suggests that this information is relevant only during the PES stage where it has real incentive effect both directly, as others' harvest determines whether the condition is met, and indirectly by triggering social anomalies like free riding, or even fear thereof among those who do not free ride but expect others to do so. Narloch et al. (2012) also found that the nature of the motivation crowding effect was different for individual and collective pay treatments: crowding in in the former and crowding out in the latter. These results are context-dependent, and we could not replicate that result in our study.

In a short questionnaire survey administered after each session, most the participants stated that they trust their neighbors, and they knew many participants because they live in the same village. Since participants were randomly assigned to sessions, we did not collect data on and formally test for social ties, but these could be indicators of why being exposed to group pay settings leads to equivalent motivation crowding in to that of participating in individual pay. The local context might have had significant effect on this: poor farmers in our study area are aware of the adverse effect of destroying natural forests, and they work closely with development agents and agricultural extension workers to grow trees in their plots. However, they also admit that they cannot live without forest products, such as farm tools, fuelwood, fences and hay storage. We witness high interest to conserve even when there is no apparent PES, but we also observe how difficult it is to reduce harvest to zero even with PES.

5.3 Heterogeneity of responses

A third question this study tackles is whether the analysis can yield additional insights if we go beyond the Average Treatment Effect (ATE) and look at potential heterogeneity of responses. The descriptive and graphical explorations gave suggestive results for the existence of heterogonous responses to the withdrawal of the PES. In fact, the positive effect on forest conservation dominates the crowding out effect observed among nearly a quarter of the participants.

We generated a measure of conservation crowding—without specifying the nature of the crowding effect—and attempted to support the descriptive results. Though significant at only 10% significance level, we found that female participants − more than male participants - tend to increase their harvest and exceed their baseline averages if a PES scheme ends. The reasons are not obvious, but one can speculate that women have responsibilities such as cooking, which are relatively immediate and more pressing than the need to make agricultural tools for the next harvesting season. That is, there is greater reliance on forest products, when it comes to firewood—and this drives women to harvest more. These results might reflect that forest visit, the proxy for actual forest use, is not sufficiently capturing forest reliance.

The positive correlation between more forest harvesting in real life and in the experiment is an encouraging result. Perhaps this indicates that the framed field experiment was designed to represent the decision problem in real life, and the participants took it seriously. Even within the experiment, there is conservation crowding out effect among participants who harvested at the higher end of the distribution in the baseline scenario. Together with the effect of visits to the forest in real life, this result suggests that introducing and then removing a PES scheme is likely to have crowding out effect among villagers who rely more on short-term direct benefits from the forests in real life.

Generally, results are indicative of the possibility that both crowding in and crowding out effects co-exist, as Kaczan et al. (2016) argued. Nevertheless, the relatively small sample and lack of choice scenarios for establishing inherent classes of forest users are limitations of this study. In respect to these limitations, it is worth noting that the seemingly small proportion of participants who increased their harvest after we removed the PES may turn out to be in the order of thousands in a population of tens of thousands of households. Consider that even a small amount of reversal in harvest of 40% of a population of thousands of forest users is big enough to destroy thousands of hectares of forest. Further, the long-term feedback effect, though not explored in this paper, may be speculated to affect the behavior of conditional cooperators such that when some segment of the community extracts more, they will likely follow suit. More research is needed to identify and characterize the participants and understand what motivates them to increase their harvest in the last stage despite exactly the same payoff function as in the baseline scenario.

6 Conclusion

Given the unsettled debate about the long-term effect of external incentive programs on collective action, it is not possible to know the effect of a PES program on motivation to conserve forests beforehand. We designed this study to answer three main questions. First, we explored whether introducing and then revoking an incentive program crowds in or crowds out motivation relative to a baseline. Second, we varied the pay modality and compared the nature of motivation crowding between individual and collective pay. Third, we assessed if there is heterogeneity in responses to the withdrawal of the PES based on observable characteristics of the participants.

We found strong evidence that the PES increased motivation to conserve. Though harvest rate went up after PES relative to their levels during the program, they remained below the baseline case on average. These results are similar across groups of pay modality as well. Regarding other sources of heterogeneity, we found that women are relatively more likely to increase their harvest when they learn that a previously existing PES has ended.

The lesson to PES program proponents is that external incentives could work. PES can indeed create more environmental awareness and hence induce a lasting pro-nature behavior. These conclusions are always context dependent, and this study provides yet more evidence that PES schemes can be effective, and that the effect can last after the program has terminated. Particularly, the issue of permanence of the reductions without having to pay resource users forever is salient. The ES from reduced deforestation is carbon—and has to be sequestered permanently—while the funding options are limited. Given the finite resources available to implement PES, integrating PES with training in environmental awareness could create a lasting sense of responsibility in communities where there is strong intrinsic motivation.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477.

Benabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520.

Bowles, S. (2008). Policies designed for self-interested citizens may undermine" the moral sentiments": Evidence from economic experiments. *Science*, 320(5883):1605–1609.

Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.

Börner, J., Baylis, K., Corbera, E., Ezzine-de Blas, D., Honey-Rosés, J., Persson, U. M., and Wunder, S. (2017). The effectiveness of payments for environmental services. *World Development*, 96:359–374.

Cardenas, J. C., Stranlund, J., and Willis, C. (2000). Local environmental control and institutional crowding-out. *World Development*, 28(10):1719–1733.

d'Adda, G. (2011). Motivation crowding in environmental protection: Evidence from an artefactual field experiment. *Ecological Economics*, 70(11):2083–2097.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, 18(1):105.

Deci, E. L. and Ryan, R. M. (1975). *Intrinsic motivation*. Wiley Online Library.

Engel, S., Pagiola, S., and Wunder, S. (2008). Designing payments for environmental services in theory and practice: An overview of the issues. *Ecological Economics*, 65(4):663–674.

FDRE (2011). Ethiopia's climate-resilient green economy, green economy strategy. Report, Federal Democratic Republic of Ethiopia.

Festré, A. and Garrouste, P. (2015). Theory and evidence in psychology and economics about motivation crowding out: A possible convergence? *Journal of Economic Surveys*, 29(2):339–356.

Fisher, J. (2012). No pay, no care? a case study exploring motivations for participation in payments for ecosystem services in uganda. *Oryx*, 46(01):45–54.

Frey, B. S. and Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Survey*, 15(5).

Frey, B. S. and Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, 87(4):746–755.

Frey, B. S. and Stutzer, A. (2006). Environmental morale and motivation.

García-Amado, L. R., Pérez, M. R., and García, S. B. (2013). Motivation for conservation: assessing integrated conservation and development projects and payments for environmental services in la sepultura biosphere reserve, chiapas, mexico. *Ecological Economics*, 89:92–100.

Gneezy, U., Meier, S. ., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–209.

Gneezy, U. and Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, 29(1):1–17.

Greiner, R. and Gregg, D. (2011). Farmers' intrinsic motivations, barriers to the adoption of conservation practices and effectiveness of policy instruments: Empirical evidence from northern australia. *Land Use Policy*, 28(1):257–265.

Handberg, Ø. N. and Angelsen, A. (2016). Pay little, get little; pay more, get a little more: A framed forest experiment in tanzania. *Ecological Economics*.

Kaczan, D., Swallow, B. M., and Adamowicz, W. L. (2013). Designing a payments for ecosystem services (PES) program to reduce deforestation in tanzania: An assessment of payment approaches. *Ecological Economics*, 95:20–30.

Kaczan, D. J., Swallow, B. M., and Adamowicz, W. L. (2016). Forest conservation policy and motivational crowding: Experimental evidence from tanzania. *Ecological Economics*.

Kerr, J., Vardhan, M., and Jindal, R. (2012). Prosocial behavior and incentives: Evidence from field experiments in rural mexico and tanzania. *Ecological Economics*, 73:220–227.

Narloch, U., Pascual, U., and Drucker, A. G. (2012). Collective action dynamics under external rewards: Experimental insights from andean farming communities. *World Development*, 40(10):2096–2107.

Ostrom, E. (1999). Collective action and the evolution of social norms. *Journal of Natural Resources Policy Research*, 6(4):235–252.

Reiss, S. (2005). Extrinsic and intrinsic motivation at 30: Unresolved scientific issues. *The Behavior Analyst*, 28(1):1–14.

Rode, J., Gómez-Baggethun, E., and Krause, T. (2015). Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*, 117:270–282.

Rodriguez-Sickert, C., Guzmán, R. A., and Cárdenas, J. C. (2008). Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior Organization*, 67(1):215–227.

Schelling, T. C. (1980). *The strategy of conflict*. Harvard university press.

Titmuss, R. M. (1971). *The gift relationship: from human blood to social policy*. Pantheon Books, New York.

Van Hecken, G. and Bastiaensen, J. (2010). Payments for ecosystem services in nicaragua: Do market-based approaches work? *Development and Change*, 41(3):421–444.

Vollan, B. (2008). Socio-ecological explanations for crowding-out effects from economic field experiments in southern africa. *Ecological Economics*, 67(4):560–573.

Appendices

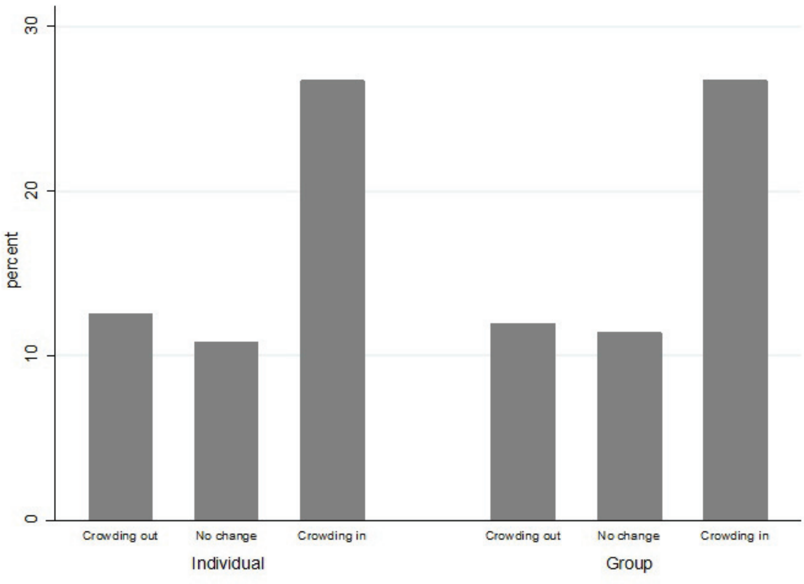

Figure 4: Percentage of participants by crowding status and pay modality.

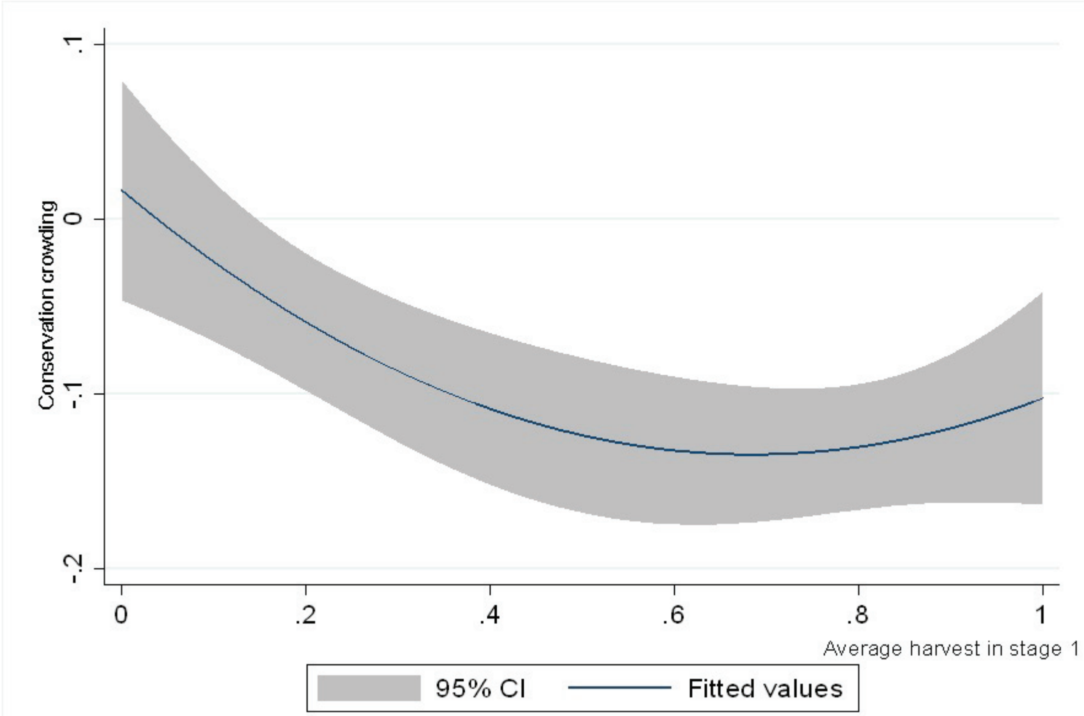

Figure 5: Relationship between harvest rate in the baseline stage and the direct and magnitude of conservation crowding. There is only suggestive trend that as average harvest in the pre-treatment stage increases, so will the difference between average harvest rates in the two scenarios.

Amare Teklay Hailu

School of Economics and Business, Norwegian University of Life Sciences (NMBU)
P.O.Box 5003
N-1432 Ås, Norway
Telephone: +47 67231100
email: hh@nmbu.no
http:/www.nmbu.no/hh

Amare Teklay Hailu was born on January 5, 1985 in a village called Nefasit, in Tigray National Regional State, Northern Ethiopia. He earned his BA degree in Economics in Arba Minch University (Ethiopia) in 2006, and his MSc in Economics (Natural Resources) from Mekelle University in 2011. He also holds a BSc Degree in Software Engineering from MicroLink Information Technology College in Mekelle (2012). Amare had worked as a graduate assistant and as a lecturer in Samara University and Mekelle University (Ethiopia) before he was enrolled as a PhD student at HH in August 2013. His thesis focuses on incentivized forest conservation for Reducing Emissions from Deforestation and forest Degradation (REDD+). The thesis consists of four self-contained papers: paper I uses spatial econometrics panel data analysis to explore spatial interaction effects and predict deforestation across districts in Indonesia. Results show that deforestation is contagious, and conservation policies may have synergistic effect (i.e., positive leakage), much like the effect of vaccination in reducing risk for those not treated.

Papers II-IV are based on framed field experiments conducted in Ethiopia, and seek to assess *ex ante* impact of various conservation payment (PES) designs. High reference levels (RLs) induce more conservation, but low RLs trigger sense of group achievement in collective PES and are also more cost efficient. Paying individuals leads to more conservation, while collective pay has uncertainty effect pushing resource users away from socially optimal level. When PES is introduced and then revoked, conservation remains above the baseline level, indicating no crowding out effect. Generally, the spatial analysis suggests *leakage* may not be a big problem, while the experimental evidence indicates *additionality* can be ensured by lowering RLs and the effect of PES lasts beyond the project period (there is *permanence*).

Supervisors: Arild Angelsen and Stein Holden

Contacts:
amare.teklay.hailu@nmbu.no
amareteklay@gmail.com
www.amareteklay.com

ISBN: 978-82-575-1501-0
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

98003 / ANDVORDGRAFISK.NO