Rural smallholders, income portfolios, and spillovers: Broadening the view for poverty alleviation strategies

Sytske C. Wijnsma and Paul A. Kattuman

Abstract

Extreme poverty continues to be overwhelmingly rural and smallholder farmers make up the majority of the world's poor. Much of the work on poverty alleviation through income interventions focuses on elevating specific income sources or targeting individual smallholders. Yet, smallholders typically combine multiple income sources in constructing their income portfolios, which creates a diverse set of income activities within villages. Interactions between sources and smallholders can cause spill-overs in the community, potentially creating unexpected links between poverty and smallholder portfolios. In this paper, this link is examined using data on over 4,000 smallholder households spanning 16 developing countries. We explore the role of smallholders' income portfolios in poverty in rural villages, contrasting the impact on the poverty of individual smallholders with poverty in the entire village. We distinguish between incomes derived from environmental, farm, and off-farm sources. We find that some portfolios are associated with reduced poverty while others exacerbate poverty. Notably, some income portfolios that might lift an *individual* smallholder out of poverty can be detrimental when featured prominently in a community. Our results highlight the need for studies to adopt a more holistic view of smallholder incomes as portfolios, rather than as isolated sources. This will improve the poverty alleviation potential of future livelihood interventions for smallholders and their broader community.

1 Introduction

Eradicating poverty is a fundamental objective of the Sustainable Development Goals (SDGs). As smallholder farmers comprise a large proportion of the world's poorest (World Bank 2018), finding ways to lift smallholder farmers out of poverty is gaining more attention from governments and the corporate sector alike.

Many poverty intervention strategies that have emerged remain focused on specific activities and their corresponding income sources – such as improving farming income through contract farming (pre-harvest agreement between farmers and buyers), or wage income through certifications that help products fetch a premium price (Federgruen et al. 2019, Thorlakson et al. 2018). An important characteristic of rural smallholders is, however, that they typically combine many economic activities and income sources contemporaneously: farm income, wages from employment, income from small businesses, and even income from environmental sources will all feature in the construction of a smallholder's income portfolio in some combination or the other (FAO 2015, World Bank Group 2016a). This raises the question about the role of the portfolio as a whole in determining poverty, knowledge of which is crucial for the design of interventions to reduce poverty.

Studies have shown evidence that certain livelihood sources and certain combinations can impact the incomes individual smallholders: Martin and Lorenzen (2016), Duchelle et al. (2014) on the extent of diversification of portfolios; and among source-specific studies, Angelsen et al. (2014), Angelsen and Dokken (2018) on reliance on natural resources, Bandiera et al. (2017) on livestock rearing, and Levi et al. (2020) on the selection of particular crops. These works largely find evidence that certain livelihood sources or combinations can impact the income of individual smallholders, but implications of smallholder portfolios for their communities remains empirically elusive (Meemken and Bellemare 2020). Smallholders typically live in tight-knit villages, with residents trading, interacting and even competing with each other. As smallholders are marked by significant diversity in their portfolios, their villages will have their own corresponding income source configurations that comprise of the diverse income portfolios of their smallholder residents. Both through economic linkages between livelihood activities and through interactions between smallholders that participate in various activities, larger community-level effects of income sources and portfolios can arise within these village economies (Taylor et al. 1996). This is an important, but often overlooked factor in poverty alleviation efforts in smallholder communities (Meemken and Bellemare 2020).

This work sets out to improve our understanding of the community-level impacts of income portfolios. We contribute by showing that the combination of sources matters for poverty through two avenues: directly at the level of the smallholder household, and indirectly at the level of the community. First, certain income combinations can be more successful at lifting a smallholder household out of poverty than others. For example, as a way to improve smallholder incomes in Comoros, Unilever and Oxfam found that rural farmers who started collecting and selling ylang ylang flowers (used by the fragrance industry) in fact benefited substantially. However in Haiti, the organizations introduced livestock to smallholder portfolios and saw the intervention backfire, causing declines in household income (Oxfam 2021). Other studies have found that poverty is reduced when the poor start participating in some off-farm activities, such as wage employment, as it can pick up any slack in agriculture (FAO 2015, Himanshu et al. 2013). These are examples of the direct effect: an individual smallholder can directly affect their own poverty status through their income portfolio.

Second, certain portfolios pursued by smallholders may help lift others out of poverty or, conversely, push them into poverty due to spill-overs (Meemken and Bellemare 2020). Such spill-overs are generated at the village-level by economic linkages between smallholders through their income activities and can take many forms. A famous example that shows how economic linkages can cause negative spill-overs is the tragedy of the commons, where the (over)exploitation of a specific resource negatively impacts others in the community. But economic linkages also exist across different income sources. For example, inputs that go into generating income from one source could be the output of another income source – farming can require hired labor (Neven et al. 2009) and a more diverse set of raw materials collected through e.g, forestry can engender small businesses built on the basis of those products (Tang 2018). Importantly, the resulting spillover effects may be positive or negative. For example, if too many smallholders embark on (unskilled) wage employment to complement their farming income, the result will be downward pressure on wages in the community, potentially exacerbating poverty (FAO 2015, Loayza and Raddatz 2010). The converse has also been noted in that households exiting employed labor for other non-farm activities resulted in reduced poverty as labor market tightening elevated the wages of those who remained (Bandiera et al. 2017, Himanshu et al. 2011). These indirect effects may vary between villages as within each village, the smallholder residents will have adopted different portfolios. We can thus analyze the effect of varying village configurations of portfolios, which depend on smallholder choices made, conditional on village characteristics.

In the presence of negative spill-overs, benefits gained by individuals through certain portfolios may not aggregate up to the community level. Admittedly, the deep structure of economic linkages between smallholders pursing diverse income-generating activities within a village is difficult to observe and measure. Yet, using granular data on livelihoods within communities, we can simultaneously analyze the direct and indirect effects of the income portfolios on smallholder and community-level poverty using the same dataset of over 4,000 smallholders across 80 villages. We first perform a smallholder-level analysis that examines the impact of a smallholder's portfolio choice on their own poverty status, thus capturing the direct effect. We contrast these results with a village-level analysis that examines the impact of all portfolio choices on community poverty, capturing both the direct and indirect effect. The consistency or dissonance between the poverty effects of these analyses will throw light on potential spill-overs.

2 Data and Method

Our analysis is based on survey data provided by the Poverty and Environment (PEN) global data set with income data of almost 8,000 smallholders for 2012-2013. The data distinguishes between income derived from the environmental, farm, and off-farm sources and is the result of the largest quantitative, global-comparative research project on forests and rural livelihoods to date, spanning study sites selected to be representative of smallholder-dominated (sub)tropical regions in Latin America, Asia, and Sub-Saharan Africa. Although the data have limitations in terms of the time dimension, its rich cross-sectional granularity helps open up the research direction we address.

We restrict the sample to households with reported incomes and to villages where 30 or more households were sampled. We removed the 2 villages that remained in Latin America. Although this means results cannot be generalized to Latin-America, the large majority of the rural poor live in Africa and Asia (World Bank Group 2016b). This process leaves us with approximately 4,100 smallholders in 80 villages spanning 16 developing countries.

Income is defined as the value-added of labor and capital minus the cost, and is reported in PPP adjusted USD per Adult Equivalent Unit (AEU). For agriculture and extractive activities, this means the gross value (quantity produced multiplied by price) minus the costs of purchased inputs (e.g., fertilizers, seeds, tools, hired labor, etc). Both the household's subsistence extraction and the production that generates cash income are accounted for. Four quarterly surveys collected data on income from seven major income sources: forest (Fo) and non-forest environmental income (No), cropping (Cr), livestock (Li), wage employment (Wa), business (Bu), and other income (Ot) which includes pensions and remittances. The differences between these sources can be found in *Materials and Methods*. Detailed information on the data collection process can be found in Angelsen et al. (2014).

Poverty: Poverty is our dependent variable where a smallholder's poverty status is defined as poor when they fall below the poverty line defined for their village. For the analysis at the smallholder level, we estimate a regression model with a binary dependent variable that represents the individual smallholder's poverty status. For the analysis at the village-level, we estimate a regression model with the well-established headcount ratio – the proportion of smallholders that fall under the poverty line for their village – as the dependent variable. The poverty line is set at 2/3 of median income of the village, following World Bank (2018). Using the village median income neutralises the considerable differences in incomes between countries, urban and rural communities, and even between (sub)tropical and other regions.

A conventional approach to understanding variation in poverty, between communities as well as over time, is to decompose the difference into an income level component and an income distribution component (Bluhm et al. 2018, Kalwij and Verschoor 2007). Such a decomposition is not exact and a residual term remains (Datt and Ravallion 1992). In the approach we follow essentially examines whether that residual can be substantially explained by the income portfolios (See *Materials and Methods* for the model specifications).

Portfolios: Our key independent variables are the income portfolios. To understand the impact of livelihood *composition* on poverty we examine the effect of the proportion that each source contributes to total income. We begin by identifying the distinctive income portfolios adopted by smallholders. We do this by applying K-Means clustering on the vectors containing the proportions of total income that each smallholder derives from the seven sources. The algorithm identifies seven clusters, which represent the set of distinct income portfolios that smallholders in our sample adopt. These portfolios can be interpreted as resulting from the decisions smallholders have made on the combinations of income sources most beneficial to them conditioned on their own household and village characteristics. The selection of seven centroids, the stability of those clusters, their generalizability, and their robustness to different data structures are discussed in Appendix B.

Figure 1 shows the pattern that individual smallholders follow within each of the seven portfolios identified by the clustering algorithm. Each portfolio is characterized by its dominant income source, with typically forest and/or cropping as the main complementary activity. These top three sources together explain on average 80% of total smallholder income. We will refer to each portfolio using the ordered sequence of the three top income sources. For example, the [Fo,Cr,Wa] cluster has Forest, Cropping, and Wage as the primary activities. There turn out to be two portfolios featuring environmental sources as most dominant, two where farming is dominant, and three where off-farm income is dominant.

For our small-holder level analysis, the portfolio variable is a dummy for each portfolio which equals one if the smallholder is seen to have adopted to that portfolio and zero otherwise. The estimates speak to the effect of smallholder's own income portfolio on their own poverty status (the direct effect), controlling for other factors.

We contrast these results from the smallholder level analysis with the results from the village-level analysis. Our objective is to identify whether and which portfolios differ in their effects between the level of the smallholder and the level of the community or village. Such differences, if any, would point to evidence of spillovers at the community level – the indirect poverty effect of income portfolios.

For the village-level analysis, we need a portfolio variable measured at the village-level. For each village, we calculate the proportions of smallholders that have adopted each portfolio. This represents the village-level configuration of income portfolios (Appendix A). Table 1 shows the proportion of smallholders in each portfolio averaged across villages. On average, villages have approximately between 10% and 20% of their smallholders residents in each portfolio. The standard deviation illustrates the substantial variation in configurations between villages.

Table 1: Village level portfolio configurations

	Environment		Farm		Off-farm		
	[Fo, Cr, Wa]	[No, Cr, Fo]	[Cr, Fo, Li]	[Li, Cr, Fo]	[Bu, Cr, Fo]	[Wa, Cr, Fo]	[Ot, Cr, Li]
Mean	18.10%	10.96%	23.52%	16.61%	16.56%	8.65%	5.60%
Standard Dev	(19.39%)	(12.00%)	(19.38%)	(15.48%)	(16.98%)	(9.57%)	(11.10%)

Controls:

Evidently, there are common factors that could influence both poverty and the smallholder's portfolio

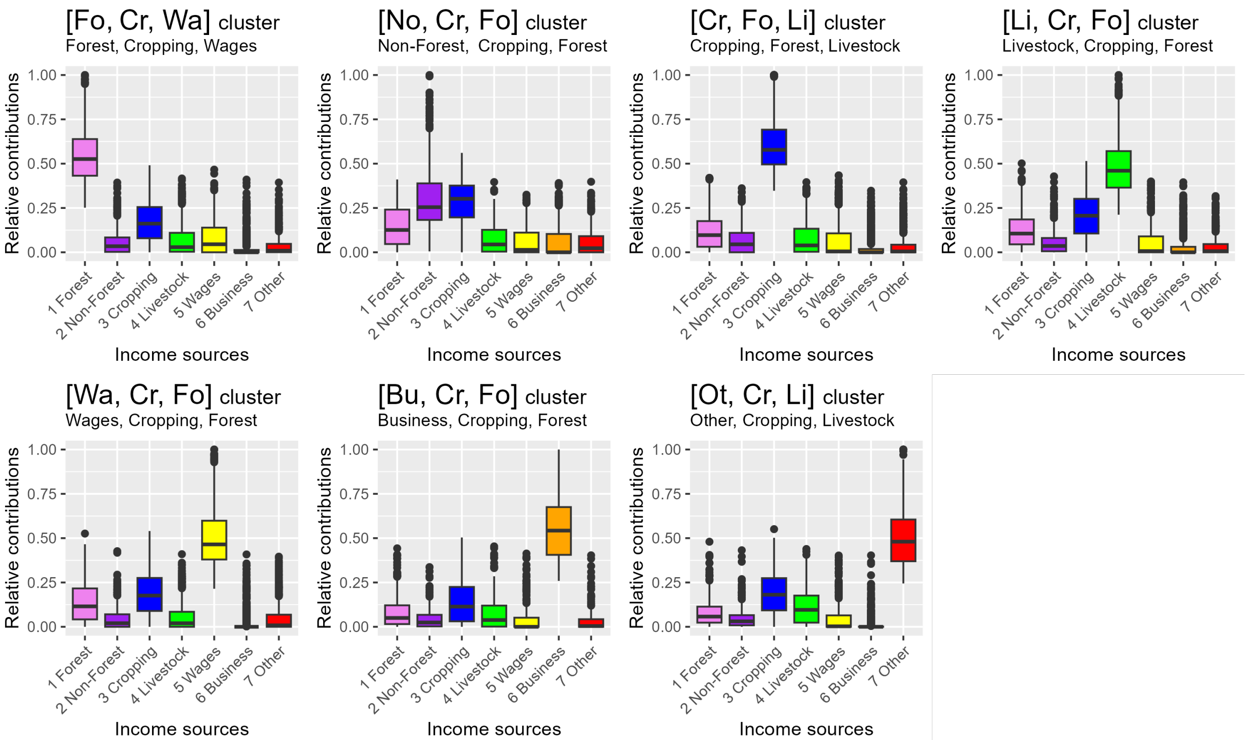


Figure 1: Smallholder level portfolio compositions

choice. For example, constraints in land ownership and education can restrict a smallholder's ability to introduce more remunerative sources into a portfolio Bandiera et al. (2017), Nielsen et al. (2013). We include a number of controls for the analysis at the smallholder level to account for confounding factors that simultaneously affect poverty and the ability of the *smallholder* to pursue a certain portfolio. These include the size of the household, gender, age, and education level of the household head, distances from home to the forest and to the village center, the extent of income diversification (operationalized through the commonly used inverse measure – the Herfindahl–Hirschman index), size of land and value of assets owned, and whether the smallholder is a member of a Forest User Group (groups committed to restricting the unsustainable use of natural resources). There will be other unobserved factors that we are unable to control for, but it is important to note that at the point of observation smallholders can be expected to have chosen the portfolio that is most beneficial to them, given the observed as well as unobserved characteristics of their households as well as the villages they reside in. We reiterate the point that our objective is not the identification of factors that drive smallholder choices per-se, but on whether the smallholder choices combine at the level of their villages to generate positive or negative effects for the community at large.

To this end we contrast the smallholder-level analysis with one at the village-level. The village-level analysis illustrates whether the prevalence of any specific portfolio is associated with higher or lower poverty levels in the village. This result does not depend on who in the village adopt those portfolios – rich or poor, high or low in education, etc. As such, to examine the association between poverty and portfolios in a village, we do not need to control for any particular individual smallholder's ability to pursue a specific portfolio. It is important to control for variables that could simultaneously affect poverty and the ability of the entire *community* to pursue certain portfolios. To this end, we control for the average value of assets held by smallholders within the community, adjusted for adult equivalent units (AEU). Assets are key productive assets that can correlate with poverty and the type of portfolios pursued in the community, some of which will require more assets. Second, we control for the average number of hectares of forest lands available to the community members, a key productive asset for environmental portfolios. [1] We further control for the proportion of community land in the village with respect to state- and privately-owned land. The tenure regime influences whether residents have the autonomy to manage their own resources and use the land based on what applications are needed from a community perspective (e.g., land for grazing, farming, or forestry), which could help reduce community poverty (FAO and UNEP 2020). Conversely, a greater share of community land can be disadvantageous as it tends to be over-exploited relative to state- or privately-owned land (Naughton-Treves et al. 2011), potentially exacerbating poverty. Whichever of those two opposing effects prevails, it may be moderated by the available forest land. For example, private forest lands are usually much closer to homesteads than for example state-owned forest lands and are therefore preferred for the collection of regularly used subsistence products (Jagger et al. 2014). These characteristics may influence both poverty and accessibility or attractiveness of certain portfolios. We therefore also include the interaction between available forest land and the share of community land.

There may be additional unobserved structural factors that exclude the entire village from accessing particular portfolios (absence of skills, too distant from cities, etc). This would lead to villages that do not have any smallholders participating in, for example, self-owned businesses. In 7 villages, we find that there is at least one income source that no smallholder is pursuing, which are mainly non-forest environmental income (5 villages) and business income (3 villages). To ensure robust inferences, we use a conditional modeling technique that separates out instances where there are no smallholder residents pursuing one of the portfolios (Section A). Our regression coefficients can therefore be interpreted as the portfolio effect for villages where there are smallholders pursuing that particular portfolio, i.e., it is accessible to the community.

3 Results

3.1 Smallholder poverty

We first investigate whether a smallholder's poverty status is significantly associated with their chosen portfolio. This will reveal the existence of any direct effects, i.e., whether a specific portfolio is beneficial or not for an individual smallholder's poverty status ceteris paribus, thereby leaving out potential community-level spillover effects.

Columns 1 and 2 in Table 2 present the results of our smallholder-level regression, with standard errors being clustered at the village-level. Column 1 reports the coefficients of the classic poverty decomposition: the average income in the smallholder's village, the poverty line and the inequality. In line with existing work, the income elasticity of poverty is negative and the inequality elasticity of poverty is positive. In other words, a lower mean income and a more unequal income distribution in villages corresponds to significantly higher poverty rates.

Column 2 introduces our portfolio variables, highlighted in gray. Keeping the classic decomposition in the specification will render the portfolio coefficients income level- and inequality-neutral. The coefficients on the portfolio dummies are elasticities, to be interpreted as the change in a smallholder's probability of being classified as poor if they choose that portfolio instead of [Cr, Fo, Li], the reference category. We find that the

[1] We also controlled for agricultural land, which was not significant. A potential explanation is that forest land is more telling about the abundance of land, since it has not been converted for agricultural uses, and thus relates stronger to poverty, especially if it provides disproportionate subsistence support for the poor.

portfolios are all either poverty reducing or have no significant effect on poverty, with one exception. This is intuitive, as individual smallholders will have pursued the portfolios that make them best off given their household and village characteristics. As such, the smallholders' portfolio selection, given their observed and unobserved household and village characteristics, will have been aimed at reducing the probability of poverty.

One exception is the [No, Cr, Fo] portfolio, which corresponds to significantly higher smallholder poverty. Non-forest environmental income is used primarily for the collection of food ($\sim$ 50%), followed by fuel ($\sim$ 20%). It is not a particularly productive income source and is primarily used for subsistence: only 13.3% of the total value is derived from cash sales. In comparison, forest resources have a much wider range of uses and markets, including wood fuels (firewood, charcoal, etc, for $\sim$ 35%), food (fish, bushmeat, fruits, mushrooms for $\sim$ 30%), and structural and fibre products (poles, sawn wood, leaves, grass for $\sim$ 25%) which are either sold (26.6% of the total value) or used for own consumption. It is possible that the [No, Cr, Fo], given the high reliance on Non-forest environmental income, becomes some sort of poverty trap.

3.2 Village poverty

We contrast these results with an analysis that reveals the poverty effects of these same portfolios at the village-level. Specifically, we examine whether the portfolios that the smallholder residents in the village pursue, matter for village poverty, measured by the headcount ratio. This analysis captures both the direct effect and indirect (spill-over) effect.

Columns 3 and 4 in Table 2 present the results of our village-level regression, with standard errors clustered at the country level. We draw attention to Column 4 where we introduce the village-level portfolio elasticities, highlighted in gray. These coefficients are to be interpreted as percentage changes in village poverty in response to a 1 percent increase in the number of smallholders pursuing that portfolio, relative to the number pursuing the [Cr, Fo, Li] portfolio, our reference group (see *Materials and Methods*). Again, the portfolio effects are netted out after controlling for mean income and inequality, i.e., they are level- and inequality-neutral with respect to income.

Turning to the village level poverty effects of portfolios, our objective is to identify the portfolios that generate spillover effects at the village level that reinforce the poverty reducing effects of individual level portfolio choices; and conversely, those portfolios that generate spillover effects at the village level that work in the opposite direction, exacerbating poverty. For example, if a portfolio exhibits direct benefits for the poor and/or indirect benefits through spill-overs, we will see a significant poverty alleviating effect. Conversely, if a portfolio exhibits direct benefits for the poor, but shows poverty exacerbating effects at the village-level, it points to negative spill-overs.

On average, villages with a higher proportion of smallholders engaged in the [Fo, Cr, Wa] portfolio, relative to those engaged in the [Cr, Fo, Li] portfolio, have significantly lower poverty headcount ratio. Note that while smallholders in the [Fo, Cr, Wa] cluster also derive income from cropping, and those in the base cluster [Cr, Fo, Li] also have forest income, their primary livelihood activity and how it is combined with other sources are different. This highlights the importance of composition, i.e., the extent to which these sources are used to generate income matters for village poverty.

We also find that when more smallholders in a village are engaged in the other portfolio dominant in environmental income, [No, Cr, Fo], villages tend to be poorer. This opposite, significant, result suggests that not all types of environmental income are helpful additions to the portfolio for village poverty. We find the same trend in the effect of the farm and off-farm income categories on village poverty. Portfolios where the primary income component is one of the two main off-farm categories (wage and business) tend to provide the opposite results in terms of their effect on village poverty. Comparing portfolios dominated by farm income, we similarly find that villages with more smallholders in the [Li, Cr, Fo] portfolio, relative to those in [Cr, Fo, Li], have significantly higher poverty rates.

To understand the role of spill-overs at the village level, we contrast the direction of the coefficients of the two analyses. First, we find consistent results for the [Fo, Cr, Wa] and [Bu, Cr, Fo] portfolios. These portfolios, with forest and business as their dominant sources, are associated with significantly lower poverty in both the smallholder and village-level analyses. The previous analysis showed that the direct effects of the adoption of these portfolios lifted respective smallholders out of poverty. In addition, we see that spill-over effects at the village level may be working in the same beneficial direction: the prevalence of those portfolios within communities also corresponds to lower poverty. Members in the [Fo, Cr, Wa] and [Bu, Cr, Fo] clusters could thus also benefit others in the community, strengthening their poverty reducing potential at the village-level.

The most striking results are those of the [Li, Cr, Fo] and [Wa, Cr, Fo] portfolios. Recall that the probabilities of being poor are significantly *lower* for a smallholder in these portfolios relative to being in [Cr, Fo, Li]. In other words, at the smallholder level, adopting the [Li, Cr, Fo] or [Wa, Cr, Fo] portfolios significantly reduces the probability of being poor. At the village level, however, we see that configurations with more smallholders in these portfolios are associated with significantly *higher* village poverty. This is suggestive of the direct effect of the adoption of these portfolios, lifting respective smallholders out of poverty, being countered by a significant negative indirect effect, where increased prevalence of these portfolios corresponds a higher poverty rate within the village.

Table 2: Impact of Portfolios on Poverty

	Poverty at the smallholder-level Measured by dummy $D = 1$		Poverty at the village-level Measured by log(headcount ratio)	
	(1) Base	(2) Complete	(3) Base	(4) Complete
VARIABLES				
Average income (log)	-2.557***	-2.123***	-2.134***	-1.906***
	(0.294)	(0.516)	(0.341)	(0.172)
Poverty line (log)	2.580***	2.574***	2.116***	1.999***
	(0.296)	(0.518)	(0.343)	(0.199)
Gini (log)	2.969***	1.621***	2.031***	1.872***
	(0.201)	(0.402)	(0.256)	(0.136)
[Fo, Cr, Wa]		-0.0357**		-0.048*
		(0.159)		(0.024)
[No, Cr, Fo]		0.334**		0.098***
		(0.160)		(0.022)
[Li, Cr, Fo]		-1.002***		0.028**
		(0.201)		(0.013)
[Wa, Cr, Fo]		-0.282**		0.031*
		(0.149)		(0.018)
[Bu, Cr, Fo]		-1.023***		-0.069***
		(0.270)		(0.014)
[Ot, Cr, Li]		0.782		-0.022
		(0.317)		(0.019)
Village characteristics	-	Yes	-	Yes
Smallholder characteristics	-	Yes	-	-
Battese dummies	-	Yes	-	Yes
Country dummies	-	Yes	-	Yes
Constant	2.969***	-0.540	2.191***	1.523***
	(0.462)	(0.946)	(0.561)	(0.212)
Observations	4,114	4,114	80	80
(Pseudo) R-squared	0.013	0.135	0.608	0.861
	Errors clustered at the village-level		Errors clustered at the country-level	

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

3.3 Potential Mechanisms

The fact that the individual level poverty effects of many portfolio choices are not directionally consistent with the village level poverty effects points to potential negative spill-overs that operate at the community level. We conjecture and test possible explanations (analyses can be found in Appendix C).

One explanation is related to the tragedy of the commons, which dictates that shared resources can create negative externalities due to overexploitation. For example, community grazing land that is freely accessible (i.e., has the character of a commons) can be expected to be overused as more villagers keep livestock. We can extend this notion to other income sources, such as wages. As more villagers pursue wage labor, the lack of co-ordination among wage seekers, absence of labor laws and unions, can depress wags and have the same underlying character as the tragedy of the commons. Our findings showed that a single smallholder who pursues livestock or wage as a primary income source is less likely to be poor, but on a village-level, this effect changes. One explanation could thus be that more animal husbandry leads to for instance overgrazing and more smallholders in wage employment drive down wages, making everyone with those income sources worse-off. Consequently, this could lead to higher village-level poverty as more smallholders adopt the livestock and wage income portfolios, even though at the individual level this is not immediately apparent. As such, it may not be about the portfolio per se, but rather that a high percentage of income derived from wage and livestock is detrimental for a community. To examine the possibility that individual choices are leading to collective detriment due to lack of coordination, we need to complement the above analysis by examining the poverty effects of income sources individually.

To test this conjecture, we re-run our village-level regression, but use the average percentage of income that smallholders in a village derive from each source, in the place of portfolios. If the underlying mechanism is collective failures such as overgrazing and downward wage pressure, we should find that a greater average reliance on wage and livestock within a village (regardless of their combination with other sources) is associated with higher poverty, similar to the findings from the previous model. Insignificant poverty effects at the community level, contrary to poverty reducing effects at the individual level suggest negative spill-overs. However, since income sources at the village level are not associated with greater poverty, in contrast to the village-level portfolios analysis, we need to look beyond the tragedy of the commons type explanations. We can conclude it is not just the community level over-reliance on wage and livestock income that causes a positive association with poverty, but also the way in which those sources are combined with the others.

A second explanation could have to do with the level of specialization. Specifically, the degree of specialization (or diversification) has been tied to the level of smallholder income Martin and Lorenzen (2016), Duchelle et al. (2014). However, this diversification effect on income may not be the same for an individual as for a village. For example, a high level of specialization can be advantageous for the individual smallholder due to economies of scale. However, if smallholders are active in a few income sources that grant these economies, it can be problematic for local communities that rely on having residents in many different activities for the purposes of e.g., trading. The directional change in coefficients may therefore be driven by the diversification (or lack thereof) within portfolios and villages, rather than the specific sources. It is possible that perhaps the wage and livestock portfolios coincidentally have higher or lower specialization rates than the other portfolios. We re-run our village-level regression, now controlling for different measures of the level of income diversification within villages, including the Herfindahl-Hirschman index and the Jacquemin-Berry (J-B) entropy measure. We find however that these coefficients are not significant and can therefore not be the driver of the results.

These analyses suggest that income portfolios and how they combine in a village generate spill-overs through mechanisms beyond what has been suggested by existing literature. We hope this opens up the opportunity for future field studies on poverty to explore these deep structures of economic linkages between smallholders in a community through the income portfolios adopted by them.

4 Practical implications

Our analysis shows that there is substantial diversity of income strategies adopted by smallholders, and that the resulting income portfolios, as well as a village's constituent income portfolios, are associated with poverty outcomes. The multifaceted nature of smallholder incomes has resulted in different sources being targeted for poverty alleviation strategies, which has led to a wide range of income interventions. Farming, being a key component in the livelihoods of rural smallholders, has seen private and public sector involvement to improve productivity (IFC 2019), market access (Levi et al. 2020), or reduce price risks (Federgruen et al. 2019). Social enterprises are rapidly emerging with the intention to improve living standards by linking small businesses selling arts and crafts to global markets (Sodhi and Tang 2014). At the same time, small and medium forest enterprises (SMFEs) are being established to help the rural poor generate income from forest-related activities (e.g., by procuring products from the Açai palm and Baobab tree for cosmetics manufacturers) (FAO and UNEP 2020).

Although these efforts have the intention to reduce poverty in the entire community, they do so by targeting a specific income source. The results in this study point to the importance of monitoring compositional changes in incomes when poverty targeted livelihood interventions are applied. First, income portfolios relate to poverty in different ways and even those with income sources within the same broad category (environment, farm, off-farm) often act in opposite directions. Adding off-farm or environmental income to smallholder portfolios may therefore not unambiguously reduce poverty. Second, income portfolios that might lift an *individual* smallholder out of poverty can be detrimental to community level poverty due to negative spill-overs between them.

Governments, multinationals, and non-profits working to reduce and alleviate poverty should therefore inform their strategies with detailed considerations of income portfolios. There are mechanisms at play that can create negative spill-overs of specific livelihood combinations at the community level. Although exploring the exact mechanisms is beyond the scope of this study, agencies can already take these insights to heart when implementing a specific livelihood intervention. For example, instead of exclusively focusing on the impact of on the participating smallholder's wealth, agencies can monitor broader effects: changes to the participating smallholders' livelihood compositions, as well as impacts on community poverty. This becomes especially relevant as more firms and agencies source goods from specific sources directly from smallholders, but independently from each other. Coordinating demands between these organizations for smallholders' livelihood activities is needed to reduce the risk of inadvertently creating poverty exacerbating portfolios.

5 Discussion and Future Research

This work has provided initial evidence that the way in which income is constructed matters for poverty, and those portfolio effects can impact individual smallholders differently than the broader community due to spill-overs. Although it is beyond the scope of this study to understand the precise mechanisms through which income portfolios impact community poverty, we believe this to be a fruitful research direction for future work. One avenue is the design of targeted studies that test possible mechanisms. Another is the adoption of a more holistic view of smallholder incomes when assessing the effectiveness of livelihood interventions. For example, when targeting poverty alleviation through a particular income source, studies could monitor compositional changes in the portfolios of the participating smallholders, as well as those in the broader community, and relate those impacts back to poverty outcomes.

Although this work focuses on the poverty impacts of income portfolios, there can be an extended study incorporating whether the sustainability of the harvested resources is impacted. The importance of this question is evident from both an environmental and social point of view: if forest resources disappear due to unsustainable extraction, smallholders will have fewer income portfolios to choose from and likely become worse off. A recent report by the Food and Agriculture Organization of the United Nations suggests that the development of forest-based livelihoods through a diversified portfolio of forest products and services can help preserve these resources by increasing their value for local communities (FAO and UNEP 2020). Further research is necessary on the environmental front, but this would imply that certain portfolios could facilitate the sustainable use of forest resources and thereby advance environmental as well as poverty goals.

Many studies mentioned in this work provide evidence that portfolios impact the income of individual smallholders, but there is also work that explores the individual smallholder's decision-making process behind building a portfolio, which could include the level of wealth (e.g., Nielsen et al. (2013)). Evidently, a question that may linger is whether poverty can cause certain portfolio choices, i.e., portfolios may also be an outcome of poverty, rather than just a lever to influence it. Although we cannot examine this possibility, it does not undermine our results. The aim of our analysis is to understand whether there is a significant link between income portfolios and poverty at the community level, and whether that relationship ever differs directionally when contrasted with the poverty effects of portfolios held by individual smallholders. For instance, argued from a reverse causality standpoint, some poor smallholders may not have the means to purchase cattle, excluding them from livestock-dominant portfolios (Bandiera et al. 2017). This would naturally aggregate up: communities with high poverty rates would then also be less likely to have smallholders in livestock-dominant portfolios, ceteris paribus. Importantly however, we find that the direction of the effect changes between the individual and village levels of analyses, pointing to spill-over effects. The presence or absence of reverse causality at the individual level is not a factor in our conclusions about spillovers.

On the data front, it is important to note that although household level survey data has several advantages as explained in the paper, there are two drawbacks. The first is that obtaining panel data at this micro-scale is costly. The set-up up and collection of the cross-sectional PEN data by the Center for International Forestry research (CIFOR) in collaboration with universities and regional and international institutions from various countries took 10 years to complete. Needless to say, conducting several replications of a study of this scale is not feasible. However, the insights derived from cross-sectional analysis of this large data set can point the way to new lines of inquiry and guide smaller, more focused studies, to understand the most important mechanisms that work with specific income portfolios and subsequently move from a static observation to a dynamic approach.

The second drawback is one of data integrity. All surveys suffer from the fact that households may under-report incomes. Studies have shown that income derived from business is generally under reported due to tax concerns (Hurst et al. 2014). Since business income is associated with lower poverty rates, under reporting of business income would likely strengthen our main results. Forest income has also been found to be under reported by rural households extracting at least 25% of their income from this source; this may lead to upwardly biased national poverty estimates (Parvathi and Nguyen 2018). In our analysis, we focus exclusively on rural poverty and find that a higher share of forest income is typically associated with less poverty among rural smallholders. Thus, upward income corrections in forest income are also likely to strengthen these results. However, there is a possibility that the *extent* of under reporting is greater for the rural poor than the rural non-poor, given a certain income share that is extracted from forestry. If the poor systematically under report their forest income to a greater extent than the non-poor, inferences could become weaker. Further research into reporting differences between socio-economic groups within the rural community could be useful in accounting for possible biases in studies using rural survey data.

Materials and Methods

Data. We use CIFOR's Poverty and Environment (PEN) global data set. PEN uses a standardized set questionnaires to collect household-level socio-economic and village-level contextual data from a diverse sample of smallholder households, i.e., households that operate under a small-scale agricultural model. In the PEN research, a village is defined as the lowest administrative unit in an area and is typically under the jurisdiction of a village leader/council. A household is defined as a group of people (normally family members) living under the same roof, and pooling resources (labor and income). The study sites were selected to be representative of smallholder-dominated (sub)tropical regions (Latin America, Asia, and Sub-Saharan Africa) with moderate-to-good access to forest resources. While PEN sites were selected according to explicit stratification criteria, the within-site selection of households followed random sampling. The survey has been used extensively in environmental livelihood studies (e.g. Angelsen and Dokken (2018), Duchelle et al. (2014), Watmough et al. (2016)). Detailed information on the data collection process can be found in Angelsen et al. (2014).

Total household income is derived from a combination of 7 major income sources that in turn originate from 3 broad categories: environment, farm, and off-farm. Environmental income is derived from forest and non-forest sources. Income is considered from forests if that income depends on the existence of forest cover. Wood products, fish from rivers inside forests, and payments for ecosystem services are classified as forest income. Mineral extraction and natural products caught or harvested outside of forests are classified as non-forest environmental income. Environmental resources are typically used for a number of different applications. In the PEN dataset, wood fuels (firewood, charcoal, etc) are the dominant category of forest income ($\sim 35\%$). The second largest category is food ($\sim 30\%$), which can be fish, bushmeat, fruits, mushrooms, etc. and the third is structural and fibre products ($\sim 25\%$) such as poles, sawn wood, leaves, grass, etc. The composition of non-forest environmental income is slightly different, with food as the dominant category ($\sim 50\%$), followed by fuel ($\sim 20\%$) (Angelsen et al. 2014).

We differentiate farm income between cropping and livestock income. Crop income comes from cropping on agricultural or agroforestry land. Livestock income consists of the consumption or sale of animal products and services (e.g. renting out horsepower). Lastly, off-farm income consists of wage income from employed labor, income from self-owned businesses, and other income. The latter is derived from remittances, pensions, gifts, and any other sources not captured by the previous categories. Self-owned businesses can encompass offering specific services, as well as selling hand-made products like arts and crafts. These may be obtained from forest products, in which case the raw material (e.g. timber) is considered income from forest resources and the added value (e.g. sculpting the timber) is considered income from self-owned businesses.

Approximately 85% of smallholders have at least 2 major income sources. Of smallholders with farming as a major livelihood source (major being defined as comprising at least 20% of total income), almost 60% engage in some type of off-farm work and over 75% complement that farm income with environmental sources.

Measurement of Portfolio Variables. With smallholder income derived from 7 sources, it can be defined as $y = w_1 + .. + w_7$. The composition of the smallholder's income is then $\omega = \left[\frac{w_1}{y}, .., \frac{w_7}{y}\right]$. Based on these vectors, the $K - means$ algorithm clusters the smallholders into 7 distinct groups (Appendix B), which are used as the independent portfolio variable in our smallholder-level analysis.

We subsequently analyze the impact of portfolios at the village-level. Different village-level configurations of income portfolios emerge as in each village, portfolios will be adopted to different extents by the resident smallholders.

Let n_{i1} be the number of smallholders n in village i that adopted Portfolio category 1 [Fo, Cr, Wa]. We then capture the configuration of income portfolios in village i by the vector $\left[\frac{n_{i1}}{n_i}, .., \frac{n_{i7}}{n_i}\right]$.

For the village-level regressions, we apply a number of methods to ensure robust estimation. First, the aforementioned vector is a compositional data vector: each element represents a proportion with a non-zero constraint and each row has a constant sum constraint (proportions must sum up to 1). This complicates traditional multivariate analysis and the interpretation of coefficients. We employ the alr transformation (Aitchison 1982) (see Appendix A.2). To resolve the unit sum constraint, the elements are transformed into the proportions of smallholders in that portfolio relative to those in the [Cr, Fo, Li] portfolio, which functions as our base category. To resolve the non-zero constraint, we take the logarithm of those proportions. Results are thus to be interpreted as a change in the proportion of smallholders in portfolio p relative to those in [Cr, Fo, Li] (our base).

Second, we combine methods that combat the problem of missing values, which is a common challenge in survey data from developing countries. Within a village-level portfolio configuration, a zero value occurs when the clustering algorithm did not assign any smallholder residents to a particular portfolio. The choice of method is based on the nature of those zero values: essential or rounded (please see Appendix A). Since our independent variables are converted using the log of ratios, it is important that the employed methods preserve the original ratio. Therefore, rounded zeros, those that can be attributed to the sampling process, are replaced using the ratio-preserving Bayes-Laplace posterior estimate. Essential zeros are treated in the regression using the Battese dummy technique (Battese 1997), which is a conditional modeling technique that separates out the zeros.

Empirical framework and Statistical Analysis. For the poverty analysis at the smallholder level we estimate logistic regressions, with individual smallholder j's poverty status as the dependent variable. Smallholder j is classified as poor if their income falls below the poverty line, set at 2/3 of the median income in the village they reside in. The key independent variables are the categorical variables $D_{.}$ denoting 6 income portfolios, with the [Cr, Fo, Li] portfolio serving as the omitted reference.

$$\log\left(\frac{\Pr(\text{Poor}=1)}{1-\Pr(\text{Poor}=1)}\right) = \hat{\alpha}_0 + \hat{\gamma}_1 \ln \overline{y}_i + \hat{\gamma}_2 \ln I_i + +\hat{\gamma}_3 \ln l_i + \sum_{k=1}^{6} \hat{\beta}_k D_{kj} + \text{controls} + v_j \tag{1}$$

Controls include village characteristics and individual smallholder characteristics.

For the poverty analysis at the village level we specify the following model and obtain estimates of poverty elasticity with respect to the configuration of income portfolios within the village. We use logratios $\mathbf{Z}_{ir}$ for village i and portfolio r (relative to the [Cr, Fo, Li] portfolio) as the key covariates:

$$\ln P_i = \hat{\alpha} + \hat{\gamma}_1 \ln \overline{y}_i + \hat{\gamma}_2 \ln I_i + \hat{\gamma}_2 \ln l_i + \sum_{r=1}^{6} (\hat{\eta}_r D_{ir} + \hat{\beta}_r \ln \mathbf{Z}_{ir}) + \text{controls} + v_i \tag{2}$$

The poverty measure P_i is the headcount ratio for village i. On the explanatory side we include the mean income in the village $\overline{y}_i$, inequality measure I_i operationalized through the Gini coefficient, and the village level poverty line l_i. The $\hat{\beta}_r$ coefficient measures the poverty elasticity of portfolio r pursued in a village. A positive and significant coefficient implies that villages where more smallholders pursue portfolio r relative to the [Cr, Fo, Li] portfolio are characterized by lower poverty. The Battese dummies D capture villages that do not have any smallholders in portfolio r. While the poverty measure is theoretically bounded between 0 and 1, in our data the headcount ratios never approach these bounds.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160.

Angelsen, A. and Dokken, T. (2018). Climate exposure, vulnerability and environmental reliance: A cross-section analysis of structural and stochastic poverty. *Environment and Development Economics*, 23(3):257.

Angelsen, A., Jagger, P., Babigumira, R., Belcher, B., Hogarth, N. J., Bauch, S., Börner, J., Smith-Hall, C., and Wunder, S. (2014). Environmental income and rural livelihoods: a global-comparative analysis. *World Development*, 64:S12–S28.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, 132(2):811–870.

Battese, G. E. (1997). A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics*, 48(1-3):250–252.

Battese, G. E., Malik, S. J., and Gill, M. A. (1996). An investigation of technical inefficiencies of production of wheat farmers in four districts of pakistan. *Journal of Agricultural Economics*, 47(1-4):37–49.

Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470.

Bluhm, R., de Crombrugghe, D., and Szirmai, A. (2018). Poverty accounting. *European Economic Review*, 104:237–255.

Brusco, M. J., Singh, R., Cradit, J. D., and Steinley, D. (2017). Cluster analysis in empirical OM research: survey and recommendations. *International Journal of Operations & Production Management*.

Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (2006). *Compositional data analysis in the geosciences: from theory to practice*. Geological Society of London.

Datt, G. and Ravallion, M. (1992). Growth and redistribution components of changes in poverty measures: A decomposition with applications to brazil and india in the 1980s. *Journal of Development Economics*, 38(2):275–295.

Duchelle, A. E., Zambrano, A. M. A., Wunder, S., Börner, J., and Kainer, K. A. (2014). Smallholder specialization strategies along the forest transition curve in Southwestern Amazonia. *World Development*, 64:S149–S158.

FAO (2015). *The economic lives of smallholder farmers: An analysis based on household data from nine countries*. Food and Agriculture Organization of the United Nations Rome.

FAO and UNEP (2020). *The State of the World's Forests 2020. Forests, biodiversity and people*. FAO and UNEP.

Federgruen, A., Lall, U., and Şimşek, A. S. (2019). Supply chain analysis of contract farming. *Manufacturing & Service Operations Management*, 21(2):361–378.

Godichon-Baggioni, A., Maugis-Rabusseau, C., and Rau, A. (2019). Clustering transformed compositional data using K-means, with applications in gene expression and bicycle sharing system data. *Journal of Applied Statistics*, 46(1):47–65.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52(1):258–271.

Himanshu, Lanjouw, P., Murgai, R., and Stern, N. (2013). Nonfarm diversification, poverty, economic mobility, and income inequality: a case study in village india. *Agricultural Economics*, 44(4-5):461–473.

Himanshu, H., Lanjouw, P., Mukhopadhyay, A., and Murgai, R. (2011). Non-farm diversification and rural poverty decline: A perspective from indian sample survey and village study data.

Hurst, E., Li, G., and Pugsley, B. (2014). Are household surveys like tax forms? evidence from income underreporting of the self-employed. *Review of Economics and Statistics*, 96(1):19–33.

IFC (2019). *Working with Smallholders: A Handbook for Firms Building Sustainable Supply Chains*. Washington, DC: World Bank Group.

Jagger, P., Luckert, M. M. K., Duchelle, A. E., Lund, J. F., and Sunderlin, W. D. (2014). Tenure and forest income: Observations from a global study on forests and poverty. *World Development*, 64:S43–S55.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666.

Kalwij, A. and Verschoor, A. (2007). Not by growth alone: The role of the distribution of income in regional diversity in poverty reduction. *European Economic Review*, 51(4):805–829.

Levi, R., Rajan, M., Singhvi, S., and Zheng, Y. (2020). The impact of unifying agricultural wholesale markets on prices and farmers' profitability. *Proceedings of the National Academy of Sciences*, 117(5):2366–2371.

Loayza, N. V. and Raddatz, C. (2010). The composition of growth matters for poverty alleviation. *Journal of development economics*, 93(1):137–151.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14):281–297.

Martin, S. M. and Lorenzen, K. (2016). Livelihood diversification in rural Laos. *World Development*, 83:231–243.

Martín-Fernández, J.-A., Hron, K., Templ, M., Filzmoser, P., and Palarea-Albaladejo, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets. *Statistical Modelling*, 15(2):134–158.

Meemken, E.-M. and Bellemare, M. F. (2020). Smallholder farmers and contract farming in developing countries. *Proceedings of the National Academy of Sciences*, 117(1):259–264.

Naughton-Treves, L., Alix-Garcia, J., and Chapman, C. A. (2011). Lessons about parks and poverty from a decade of forest loss and economic growth around kibale national park, uganda. *Proceedings of the National Academy of Sciences*, 108(34):13919–13924.

Neven, D., Odera, M. M., Reardon, T., and Wang, H. (2009). Kenyan supermarkets, emerging middle-class horticultural farmers, and employment impacts on the rural poor. *World development*, 37(11):1802–1811.

Nguyen, T. T., Do, T. L., Bühler, D., Hartje, R., and Grote, U. (2015). Rural livelihoods and environmental resource dependence in Cambodia. *Ecological Economics*, 120:282–295.

Nielsen, Ø. J., Rayamajhi, S., Uberhuaga, P., Meilby, H., and Smith-Hall, C. (2013). Quantifying rural livelihood strategies in developing countries using an activity choice approach. *Agricultural Economics*, 44(1):57–71.

Oxfam (2021). *Enhancing Livelihoods Fund: Final Report*. Oxfam.

Parvathi, P. and Nguyen, T. T. (2018). Is environmental income reporting evasive in household surveys? evidence from rural poor in Laos. *Ecological Economics*, 143:218–226.

Pawlowsky-Glahn, V. and Buccianti, A. (2011). *Compositional data analysis: Theory and applications*. John Wiley & Sons.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sodhi, M. S. and Tang, C. S. (2014). Supply-chain research opportunities with the poor as suppliers or distributors in developing countries. *Production and operations management*, 23(9):1483–1494.

Tang, C. S. (2018). Socially responsible supply chains in emerging markets: Some research opportunities. *Journal of Operations Management*, 57:1–10.

Taylor, J. E., Taylor, J. E., Adelman, I., and Golan, E. H. (1996). *Village economies: The design, estimation, and use of villagewide economic models*. Cambridge University Press.

Thorlakson, T., de Zegher, J. F., and Lambin, E. F. (2018). Companies' contribution to sustainability through global supply chains. *Proceedings of the National Academy of Sciences*, 115(9):2072–2077.

Vrecko, D. and Langer, T. (2013). What are investors willing to pay to customize their investment product? *Management Science*, 59(8):1855–1870.

Watmough, G. R., Atkinson, P. M., Saikia, A., and Hutton, C. W. (2016). Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. *World Development*, 78:188–203.

World Bank (2018). *Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle*. Washington, DC: World Bank Group.

World Bank Group (2016a). *Forest Action Plan FY16–20: The WBG Contribution to the Forest Agenda*. World Bank Group.

World Bank Group (2016b). *Poverty and Shared Prosperity 2016: Taking on Inequality*. World Bank Group.

Appendix

A Portfolio variables

In this section, we outline the approach for obtaining smallholders' income portfolios, the village-level configuration of those portfolios (i.e., community portfolios), and the appropriate statistical techniques for analyzing compositional data.

A.1 Smallholder income portfolios

Consider a household pursuing a combination of sources, which are mutually exclusive and add up to a household's total income: $y = w_1 + .. + w_Q$. The composition of the household's income is $\omega = \left[\frac{w_1}{y}, .., \frac{w_Q}{y}\right]$, with ω differing between smallholders. Since smallholders are marked by significant diversity in terms of income compositions, we need to classify those compositions into a finite number of commonly adopted income portfolios. Clustering and component extraction techniques are frequently used in livelihood analyses (Martin and Lorenzen 2016, Nguyen et al. 2015) as well as in portfolio selection problems (Vrecko and Langer 2013) to uncover common configurations within a large set of possibilities. We use K-means clustering on the income composition vectors (ω) to find the common income portfolios of the smallholders in our data set. The approach for deriving these clusters, the validity and stability analyses, as well as a more detailed characterization, are discussed in Appendix B.

The clustering process produces 7 clusters representing the most common combinations, to which we refer as the income portfolios. Table 3 summarizes the cluster means, or cluster centers, of each portfolio. One of the 7 income sources is the main contributor to household income, combined with a secondary and tertiary source that are often cropping income and forest income. On average, the top 3 income sources make up around 80% of total income. For convenience, we will refer to the cluster by their primary source, which are listed at the bottom of Table 3.

Table 3: Cluster Means (%)

Income category		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Environment	Forest	54.77	14.67	11.53	12.51	14.01	8.61	7.87
		(0.56)	(0.51)	(0.32)	(0.41)	(0.42)	(0.53)	(0.40)
	Non-forest	5.78	30.74	6.67	5.61	4.52	4.76	4.84
		(0.27)	(0.81)	(0.23)	(0.28)	(0.22)	(0.35)	(0.31)
Farm	Cropping	17.40	28.08	60.98	20.69	18.70	13.79	18.50
		(0.44)	(0.59)	(0.50)	(0.51)	(0.47)	(0.67)	(0.60)
	Livestock	7.25	7.82	7.93	48.17	5.68	7.82	11.13
		(0.35)	(0.39)	(0.30)	(0.63)	(0.29)	(0.55)	(0.50)
Off-farm	Wage	8.44	6.43	6.39	5.93	50.26	4.85	4.90
		(0.37)	(0.39)	(0.31)	(0.37)	(0.61)	(0.51)	(0.42)
	Business	2.65	6.33	2.92	3.51	2.12	56.48	2.18
		(0.24)	(0.42)	(0.20)	(0.29)	(0.22)	(1.05)	(0.29)
	Other	3.71	5.93	3.59	3.58	4.71	3.70	50.58
		(0.23)	(0.36)	(0.20)	(0.24)	(0.28)	(0.38)	(0.86)
Observations		716	490	932	593	748	318	383
Portfolio Characterization								
		Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5	Portfolio 6	Portfolio 7
Primary		Forest	Non-forest	Crop	Livestock	Wage	Business	Other
Secondaries		Crop/Wage	Crop/Forest	Forest/Livestock	Crop/Forest	Crop/Forest	Crop/Forest	Crop/Livestock

Note: Cluster means are reported as percentages and the standard error of the mean is reported in brackets.

For our main question, we wish to determine the responsiveness of poverty in villages to the different sets of income portfolios that are pursued within those villages. To do so, we need to understand how these portfolios combine in a village and obtain their village-level configuration. This is discussed in the next section.

A.2 Village-level Configuration of income portfolios

Different village-level configurations of income portfolios emerge as in each village, portfolios will be adopted to different extents by the resident smallholders. Specifically, let n_{ip} be the number of smallholders n in village i that adopted portfolio p, numbered from 1 to 7. Then we can capture the configuration of income portfolios in village i by the vector $\boldsymbol{\pi}_i = \left[\frac{n_{i,1}}{n_i}, .., \frac{n_{i,7}}{n_i}\right]$ with $p = 7$ elements.[2]

In the village-level analysis, the variables of interest are so-called 'compositional data' and therefore subject to non-negativity and constant sum constraints: each element of the vector $\boldsymbol{\pi}_i$ must be equal or greater than zero and their sum must equal 1. These constraints complicate traditional multivariate analysis and the interpretation of coefficients: the data are non-normally distributed since they must always lie between zero and one and must sum to one and regression specifications with predictor variables that sum to unity suffer from multicollinearity. The latter is typically overcome by eliminating one variable to function as the base group. However, it is known that this formulation is subject to an inconsistency (Pawlowsky-Glahn and Buccianti 2011). The essential consequence of compositional data is that changes in one component are not invariant to changes in other components. Standard statistical techniques that are devised for unconstrained random variables and that use the covariance or correlation matrix of vectors of observations cannot be used for the compositional data (Aitchison 1982, Buccianti et al. 2006).[3]

In our approach, we map the configurations to real numbers by using the additive logratio (alr) transformation, originally developed by Aitchison (1982). The advantage of the additive logratio is that it retains parameter interpretations as well as the logical consistency argument that shares are restricted to unit sum. The technique removes the restrictions and transforms compositional data to a scale where they follow a multivariate normal distribution.

The approach takes the log of a vector of compositional ratios ($\boldsymbol{\pi}_i$) with respect to a base cluster, for which we select the [Cr, Fo, Li] cluster. Over 95% of the smallholders in our data set derive some of their income from cropping and in all of the villages there are smallholders with cropping income, so this source can be seen as the most accessible. The cropping portfolio is also the most widely adopted with members in all but 8 villages. Our vector of explanatory variables thus becomes $\log(\frac{n_{i,1}}{n_i}/\frac{n_{i,B}}{n_i}), ..., \log\frac{n_{i,p-1}}{n_i}/\frac{n_{i,B}}{n_i})$, or: $\log(\frac{n_{[Fo,Cr,Wa]}}{n}/\frac{n_{[Cr,Fo,Li]}}{n}), ..., \log\frac{n_{[Ot,Cr,Li]}}{n}/\frac{n_{[Cr,Fo,Li]}}{n}$. Results are thus to be interpreted as the proportion of smallholders in portfolio p relative to those in base cropping (B). Note that the denominator (total number of villagers) cancels out and this vector can be simply interpreted as $\ln\left(\frac{n_{i,1}}{n_{i,B}}, ..., \frac{n_{i,p-1}}{n_{i,B}}\right)$.

Zero values: Special care must be taken with this type of data transformation as data with many disaggregated categories, especially in developing country contexts, often take 0 values (i.e., when a village has no smallholders in a particular portfolio category). A logarithmic transformation would then exclude villages with at least one portfolio membership of 0, which is the major drawback of compositional data analysis. In

[2] An alternative as the explanatory variable can be constructed from the average proportions of income that smallholders in the village derive from the different income sources (the aggregate of vector ω). But aggregating income compositions of households this way ignores the considered portfolio constructions made by individual smallholder households, i.e., the importance of how income activities are typically combined. In Appendix C we report results from using village-level income proportions in place of village-level portfolio configurations (i.e., the share of smallholders in each portfolio). There are no inconsistencies, but the village-level income proportions do not capture all of the significant effects. The variable representing portfolio configurations thus provide more reliable results.

[3] Note that this includes principal component analysis, hierarchical clustering, and factor analysis, but excludes clustering techniques based on geometrical separation like k-means.

the instance where there are no smallholders in the base group, the transformation would fail. This event is fortunately rare. There are only 8 villages for which the share of smallholders in the cropping portfolio is zero. These can be handled using a zero replacement technique, which is a common approach in count data when the zero observation can be rationalized as an artefact of the sampling process (Pawlowsky-Glahn and Buccianti 2011), i.e., if it is likely that that component is represented in the population, but by coincidence or due to limited sample size, not in the sample. Over 95% of smallholders in our sample have cropping income and even in the villages where no households belong to the cropping cluster, households derive income from cropping. In fact, there are households that derive over 30% of their income from cropping in these 8 villages. It is highly likely that a slightly larger or different sample would have allocated at least one smallholder to the cropping portfolio in the villages where none were allocated.

Before applying the additive logratio, we use the Bayesian-multiplicative technique to treat the zeros of the base group B, which replaces zero values from compositional count data by its posterior Bayesian estimate. Importantly, this modification does not distort the covariance structure since it modifies these nonzero parts in a multiplicative way to preserve the original ratios between the elements as well as the total sum representation of the vector (Martín-Fernández et al. 2015). We assume a uniform prior, which is the most common estimate, but differentiate the prior by village: outcomes of identical and independent trials can fall in any of the mutually exclusive portfolio categories *that already have members* in the village. For instance, a village with smallholders in all portfolios but the base will have a prior of 1/6. We further apply the well-known Bayes-Laplace posterior estimate by applying the following expression

$$b\boldsymbol{\pi}_{ip} = \begin{cases} \boldsymbol{\pi}_{ip} & \text{if } \boldsymbol{\pi}_{iB} > 0 \\ \frac{1}{n_i+v_i} & \text{if } \boldsymbol{\pi}_{iB} = 0 \text{ and } p = B \\ \boldsymbol{\pi}_{ip}(1 - \frac{1}{n_i+v_i}) & \text{if } \boldsymbol{\pi}_{iB} = 0 \text{ and } k \neq B \end{cases} \tag{3}$$

where there are seven portfolios p, one of which functions as the base B, of which the village-level configuration is captured by the vector $\boldsymbol{\pi}_i$. n_i is the total count of the vector, i.e., number of smallholders in village i, and v_i is the number of portfolio categories in the village that have members (i.e., the nonzero values) plus the base group category B. This gives us the number of elements that will be modified by the substitution. As explained next, we do not modify any zero values that occur in portfolio categories other than the base, but treat those in the regression. We find that one village is an influential outlier, exacerbated by the subsequent logarithmic transformation, and remove it from the data set.[4]

To economize on notation, going forward we denote the logratios of the portfolio configurations of village i after replacement of the zero values by the matrix **Z**:

$$\boldsymbol{Z} = \begin{bmatrix} \frac{b\boldsymbol{\pi}_{i1}}{b\boldsymbol{\pi}_{iB}} & \cdots & \frac{b\boldsymbol{\pi}_{ir}}{b\boldsymbol{\pi}_{iB}} \\ \vdots & \ddots & \\ \frac{b\boldsymbol{\pi}_{N1}}{b\boldsymbol{\pi}_{NB}} & \cdots & \frac{b\boldsymbol{\pi}_{Nr}}{b\boldsymbol{\pi}_{NB}} \end{bmatrix} \tag{4}$$

Where element Z_{ir} gives for village $i = 1, ..., N$ the logratio $r = 1, ..., 6$ of the number of smallholders in a portfolio with respect to those in the base category.

Finally, villages may also face zeros in the portfolios categories other than the base group. Replacement

[4]We take a simple example to illustrate the transformation. A village with a positive proportion of smallholders in all but the cropping cluster, has $v_i = 7$. The cropping cluster will be increased by the amount in the second term of the expression. To preserve the ratios and unit sum, the other positive components will be reduced to the amount given by the third term.

techniques may not be appropriate in these cases since it is more difficult to rationalize that these zeros are not part of the data generating process. As mentioned, cropping is pursued by almost all smallholders, but this cannot be said for all the other income sources. It is therefore not obvious that if we, for example, took a larger or slightly different sample of a village that currently has zero values, we would find that all portfolios are populated. These zero observations must then be classified as essential zeros and should not be replaced. To address essential zeros, we use the method proposed by Battese (1997) for our estimation, which was applied in an agricultural setting in Battese et al. (1996). The author introduces a form of conditional modeling that separates out the zeros. Using dummy variables that represent the occurrence of zero values, the effect on poverty can be decomposed into two components according to the value of Z_{ir}:

$$(Z_{ir} > 0)\ln(Z_{ir})\beta_{Z_{ir}>0} + (Z_{ir} = 0)\beta_{Z_{ir}=0} \tag{5}$$

The combination of the additive logratio, the Bayesian-multiplication, and the Battese zero value techniques allows us to estimate efficient and unbiased portfolio estimators with OLS using the full data set of compositional data.

B Cluster Analysis

In this section, we will discuss our clustering approach and stability assessments.

B.1 Centroid Selection

K-means clustering is an unsupervised learning method that starts the grouping process with K number of centroids or starting seeds, which are randomly selected observations. Each of the remaining observations is assigned to the centroid that minimizes the chosen distance measure between the observation and cluster mean. After all are assigned, the new mean value of each cluster is calculated and every observation is checked to see if within-cluster variation can be reduced by reassigning it to a different grouping. This variance minimization technique is an iterative process until convergence is achieved and the observations are organized into a set of K clusters.[5]

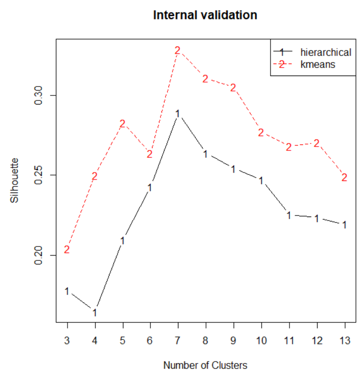
One of the key decision variables for clustering is the determination of the number of clusters or centroids, K. We use the following formal methods and informal approaches to select K and subsequently assess the stability of the clusters. First, we run the K-means MacQueen algorithm with 3 to 13 starting seeds. This technique assigns (and reassigns) observations based on the Euclidean distance between the coordinate and the centroid, which is the mean of the 'belonging' cases (MacQueen et al. 1967).[6] A clear natural hierarchy emerges in the clusters: a cluster is added by splitting another while the rest remains relatively stable. The number of centroids is selected via several approaches. First, we use the selection method employed by Baudry et al. (2012), which is based on the slope heuristics that minimizes a nonasymptotic penalty criterion and is recommended by Godichon-Baggioni et al. (2019) for the clustering of compositional data. The method suggests 7 clusters for our data. To further verify the internal validity of using 7 centroids, we also look at the average Silhouette Width, which compares the tightness (average within-cluster dissimilarity) and separation (average between cluster dissimilarity) of clusters (Rousseeuw 1987). A high silhouette score

[5]Note that because the variables are measured in the same units and on the same scale (percentages) there is no need to standardise them as is usually required in clustering.

[6]To ensure the validity of this algorithm, we also use the K-means algorithm by Hartigan and Wong (1979), which reassigns points based on the within cluster sum of squares, and find that on average fewer than 5 households are grouped differently.

means that the within dissimilarity is small compared to the smallest between dissimilarity, implying that the observation is 'well-clustered'. Though this method is not specifically adapted to compositional data, it also suggests that 7 centroids are most appropriate (see Figure 2).[7]

Figure 2: Silhouette Width of K-means and Ward



Note: The Silhouette Width is the average of each observation's Silhouette value. The Silhouette value measures the degree of confidence in a particular clustering assignment and lies in the interval [-1,1], with well-clustered observations having values near 1 and poorly clustered observations having values near -1. K-means lies strictly above Ward's hierarchical algorithm, with $K = 7$ leading to the best clustering outcome, attesting to the internal validity of our centroid selection as well as the relative validity of using K-means.

Seven centroids are also theoretically meaningful, ensuring greater validity of the groups and of their interpretation. We have seven income sources on which we cluster and, as we will see, with seven centroids each income source represents the highest average proportion in a cluster. We refer to this as the primary livelihood strategy of the cluster (though it may not be the primary strategy for each smallholder in that group, it is on average across smallholders).

B.2 Cluster Stability

Cluster stability is measured as the amount of variation in the clustering solution and we apply a number of methods to assess this. First, the stability of the clusters can be affected by the K-means method as the results are realized from just one set of seed points and thus may achieve only a local optimum. We ensure the global optimality of the solution by performing 5,000 runs with different sets of randomly drawn initial seed values, per the suggestion of Brusco et al. (2017). Running multiple restarts only results in the re-allocation of a handful of households. Thus, the clustering process is consistent and stable in the sense that only a few different partitions are found across a broad spectrum of starting solutions.

Second, the clusters are considered stable if the solution does not vary much over different samples drawn from the input data (Jain 2010). We use the Jaccard coefficient as a measure of cluster stability, which is a similarity measure between matched clusters resulting from different samples drawn from the data. One of the advantages of this approach is that we can assess cluster-wise stability. Instead of having an overall stability measure, having a coefficient for each cluster shows to what extent the clustering process finds meaningful and stable patterns in the data, as well as whether this applies to some but not all of the clusters. The stability is assessed by the (non-parametric) bootstrap distribution of the Jaccard coefficient

[7] Since the elements on which we cluster are linear dependent due to the compositional nature of the data, we must note that not all validity assessment techniques are appropriate (i.e. those that assess stability by removing a column or that are based on the correlation matrix).

and every single cluster is compared to the most similar cluster in the bootstrapped data sets. The bootstrap introduces some bias and variation, because in reality no true underlying distribution and no true clustering is known. We run 100 re-sampling runs and find that the mean of the Jaccard coefficent is higher than 0.75 for each of the 7 clusters. Note that over 0.6 indicates a pattern, whereas valid and stable clusters yield values of 0.75 or more (Hennig 2007). We arrive to the same conclusion when we use subsetting as an alternative resampling method. We set the size of the subsets at 2,090 observations, which constitutes 1/2 of our data. Using 100 reruns, we calculate the mean of the Jaccard coefficient and find that 6 clusters score above 0.75, and the remaining cluster only slightly below. The results are reported in Table 4.

Table 4: Jaccard coefficients of cluster resampling

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Bootstrap	0.950	0.761	0.883	0.910	0.910	0.923	0.891
Subsetting	0.944	0.715	0.857	0.902	0.918	0.924	0.902

Note: The mean of the Jaccard similarity index is reported over 100 reruns of either a non-parametric bootstrap or resampling by subsetting. If patterns exist, the coefficient yields at least 0.6; highly stable clusters typically yield 0.75.

B.3 Generalizability

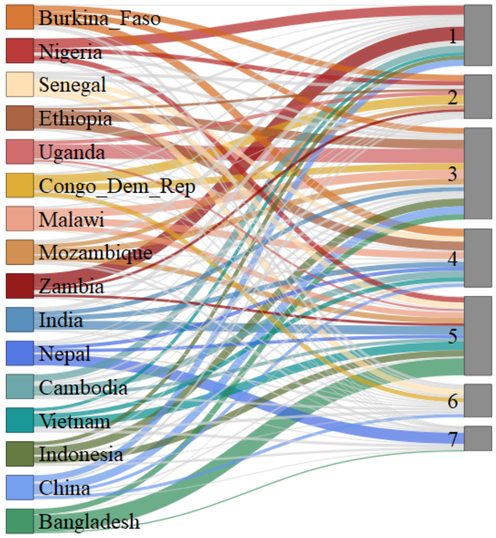
We want to ensure that our results can be robustly interpreted using this clustering process. This means that the clusters should be representative of the group we analyze (smallholders in the (sub)tropics with moderate-to-good access to forest resources). We already found with the re-runs and bootstrapping that within our sample, the clustering results are highly stable. However, as you may recall, the clustering solution was derived from a subset of the complete sample as we selected households from villages where more than 30 surveys were conducted. We therefore perform the same K-means algorithm on the complete sample of 7,696 households. We match the clusters from the full sample with the clusters from the subsample based on the values of their centroids. The matching falls out in a straightforward way as the patterns and centroids are very close. We find that on average only 2% of the households from the subsample are allocated to a different group when clustering the complete sample, which attests to the stability of the clusters out of sample and builds trust in the ability to generalize our findings.

Since we analyze livelihood patterns in a number of different countries, we also assess how well represented the countries are in the clusters, or whether, for example, some livelihood patterns are specific to a country. Figure 3 shows the diversity with which the countries map to the seven clusters, with the three most sizable flows of each country colored in. The Herfindahl-Hirschmann concentration index and Shannon's entropy measure show more formally that clusters are populated by a broad range of countries. In terms of the HHI, only cluster 7 is relatively concentrated having a significant presence of Nepal although this representation does not lead the cluster to score low on Shannon's entropy.

B.4 Data Structure

Finally, we must recognize the structure of the data we are clustering. Our compositional data is made up of Q proportions and the sample space of compositional data is the simplex, which we refer to as $\mathbb{S}^Q$. For clustering compositional data, one can consider adopting the Aitchinson distance measure, which highlights the relative difference between compositions on the simplex, instead of the Euclidean distance, which is based on absolute differences between compositional vectors. However, the Aitchinson transformation breaks down

Figure 3: Cluster Concentration



	HHI^{-1}	Shannon
Cluster 1	0.865	0.505
Cluster 2	0.860	0.483
Cluster 3	0.896	0.436
Cluster 4	0.856	0.492
Cluster 5	0.897	0.465
Cluster 6	0.899	0.441
Cluster 7	0.653	0.554

Note: The Sankey diagram shows the relative number of households from each country allocated to the 7 clusters. The three largest flows from each country are colored in. The table reports the inverse Herfindahl-Hirschmann concentration index and Shannon's entropy measure of each of the clusters. The indices are bound between 0 and 1 and take on a value of 0 when the cluster is only present in a single country.

when compositions contain a zero value since relative distances cannot be calculated. Zero replacement techniques have been found to cause incorrect clustering because the distance of a composition with values tending to zero from others, will tend toward infinity. The use of these techniques can therefore cause algorithms to group together profiles with several (near) zeros in common rather than those with a strong non-zero coordinate in common (Godichon-Baggioni et al. 2019). To assess stability of the clusters for this relative distance measure, we subset our data to be populated exclusively by households that have positive contributions from each income component, limiting the sample space to the reduced simplex. Despite being just below 20% of our original sample (833 versus 4,180 households), the clusters in the reduced set follow the same patterns and characterization as our main sample, regardless of distance measure. Even household level clustering outcomes within the reduced set only vary slightly between the Euclidean and Aitchinson distance measures. For instance, matching the clusters based on their centroids shows that 87.4% of the households are allocated to the same group when using the different distance measures. We therefore conclude that we can safely use the Euclidean distance for our sample set.

C Mechanisms

In this section, we analyze whether explanations by existing literature could reveal potential underlying mechanisms. First, we include different measures of diversification as a control variable to examine if the portfolio effects are merely due to differing levels of diversification within the portfolios, rather than the underlying sources. We calculate the Herfindahl-Hirschman Index (HHI) and the Jacquemin-Berry (J-B) entropy measure of each smallholder's portfolio, and take the average of these measures within the villages. We find that neither measures are significant, with p-values $p = (0.287, 0.415)$ for the HHI and J-B measures respectively, and the portfolio elasticities continue to be significant even by adding these measures to the specification.

Second, we consider village-level income proportions instead of portfolios to understand whether the effects are an artifact of the reliance on a particular source, rather than their combination. In column 1 of Table 5, we report the regression results using the log ratios of income proportions as independent variable, with standard errors clustered at the country level. For convenience, column 2 lists the full results of the analysis in our main paper using cluster membership as the independent variable. There are no inconsistencies compared to our main findings (i.e., there is no reversal in significance). In a few cases, estimates that are significant in our main results are insignificant when analysed through the lens of income proportions (e.g., livestock and wage). This illustrates that the poverty effect not just depends on how much income is derived from any particular source, but also on the way those sources are combined with others, i.e., the portfolio.

Table 5: Comparison of portfolio measures

	Income Proportions	Cluster Proportions
	(1)	(2)
Average income (log)	-1.90***	-1.91***
	(0.27)	(0.17)
Poverty line (log)	1.04***	1.20***
	(0.29)	(0.20)
Gini (log)	1.88***	1.87***
	(0.23)	(0.14)
[Fo, Cr, Wa]	-0.06**	-0.05*
	(0.03)	(0.02)
[No, Cr, Fo] (log)	0.02	0.10***
	(0.02)	(0.02)
[Li, Cr, Fo]	0.02	0.03**
	(0.02)	(0.01)
[Wa, Cr, Fo]	0.02	0.03*
	(0.02)	(0.02)
[Bu, Cr, Fo]	-0.06**	-0.07***
	(0.03)	(0.01)
[Ot, Cr, Li]	0.01	-0.02
	(0.01)	(0.02)
Dummy $Z_{\text{Forest}} = 0$	0.08	0.04
	(0.06)	(0.05)
Dummy $Z_{\text{Non-forest}} = 0$	-0.32*	-0.10**
	(0.07)	(0.04)
Dummy $Z_{\text{Livestock}} = 0$	-	-0.01
	-	(0.06)
Dummy $Z_{\text{Wage}} = 0$	-	-0.07
	-	(0.05)
Dummy $Z_{\text{Business}} = 0$	0.07	0.25***
	(0.09)	(0.06)
Dummy $Z_{\text{Other}} = 0$	0.13	-0.02
	(0.08)	(0.06)
Village characteristics		
% Community Land [CL]	-0.41***	-0.36**
	(0.14)	(0.15)
Mean Forest Land (log) [FL]	0.10***	0.11***
	(0.03)	(0.03)
CL × FL	-0.11***	-0.11**
	(0.03)	(0.04)
Mean Assets	0.00***	0.00***
	(0.00)	(0.00)
Constant	1.80***	1.52***
	(0.43)	(0.21)
Country dummies	Yes	Yes
Observations	80	80
R-squared	0.835	0.861

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