Smallholders in highland regions of Southeast Asia: Agricultural land-use and transitions, farmer risk preferences, the effects of poor vision on economic farm performance, and a note on scientific publishing in the field of development studies

Dissertation

To attain the doctoral degree (Dr.sc.agr)

of the Faculty of Agricultural Sciences

University of Göttingen

Submitted by Frederik Sagemüller

Göttingen, January 2023

Name of supervisor: Prof. Dr. Oliver Mußhoff

Name of co-supervisor: Prof. Dr. Marcela Ibañez Diaz

Date of oral examination: April 27th, 2023

i. Summary

Smallholders play an important role in feeding a growing world population and contributing to the eradication of hunger and poverty. It is estimated that there are more than 475 million smallholders in the world and recent research shows that globally, smallholders are responsible for around 30 percent of total crop production. One hotspot region of smallholder farming systems is Southeast Asia (SEA) (Cohn et al., 2017). In contrast to the success story of the Green Revolution, which was mainly achieved in the highly productive areas in the SEA lowlands, the agricultural development in marginalized highland regions of SEA lags behind, with stagnating or even worsening poverty and environmental degradation (Rosegrant & Hazell, 2000). Recent developments in the region show widespread and rapid adoption of cash crops like cassava, coffee, maize or rubber. Depending on the context, adoption of these cash crops yield varied socioeconomic outcomes. These can range from extremes of rapid socioeconomic differentiation, polarization and dispossession, to inclusive patterns of development (Cramb et al., 2017; Euler et al., 2017; Fox & Castella, 2013b; Mahanty & Milne, 2016; Vicol et al, 2018).

Presently, we lack a sufficiently deep understanding of the underlying drivers of positive or negative socioeconomic outcomes. To add knowledge to the existing body of literature, the present dissertation contains three case studies on smallholder production systems from the highland regions of SEA. Furthermore, we present a study that is not directly related to smallholders, but rather takes a reflective gaze on scientific practice, in the belief that open access to scientific information is a cornerstone in delivering impact at scale through science.

In the first study, we discern the linkages between land use change and plot-level and household-level characteristics by contrasting these developments in a subsistence-oriented site and a market-oriented site. We find that land use dynamics vary strongly between the sites. In the subsistence-oriented site, 66 percent of the land use types were completely

replaced during the past 10 years. In the market-oriented site, only 15 percent of land use types were replaced. The associated key drivers of land use change also differed significantly: while market orientation of agricultural products was the main driver behind land use changes in the market-oriented site, mostly agronomic challenges like slope, soil tillage and agrochemical input use are associated to land use change in the subsistence-oriented site.

In the second study, we analyze risk preferences of smallholders in Cambodia and Lao PDR with an incentivized lottery design under the framework of Expected Utility Theory (EUT), Rank Dependent Utility Theory (RDU) and Cumulative Prospect Theory (CPT) and test the effect of household shocks on these risk preferences, specifically aversion to losses. First, we find that including loss aversion is essential in describing smallholder preferences. Second, we find that household shocks increase loss aversion.

The third study examines the prevalence of poor vision among rural smallholders in Cambodia and investigates if poor vision is associated with a loss in agricultural profitability of family-owned farms. First, we find that almost 30 percent of farmers in our sample suffer from poor vision. Second, we find that farmers with poor vision lose farm profits in comparison to farmers with good vision. We obtain robust results that estimate forgone profits in the range of 630 USD/year associated with having poor vision.

Lastly, the fourth study screens through 28 million download logs of the pirate website Sci-Hub. We look for downloads of papers in the field of development studies and we lay out trends for Sci-Hub use in this discipline, including geographic location of downloads and socioeconomic drivers. We find that Sci-Hub is used the most by researchers from the Global South, primarily from middle-income countries; whereas researchers from the poorest countries in the data set use Sci-Hub the least. Open access to research outputs is important for knowledge-building and effective policy development to reach the ambitious goals of the development agenda.

ii. Acknowledgements

First of all, I want to thank my supervisor Prof. Dr. Oliver Mußhoff for giving me the chance to work on this thesis and guiding me in putting my research ideas into practice. Also I thank Prof. Dr. Marcela Ibañez Diaz, Prof. Dr. Dr. Daniel Hermann and Prof. Dr. Stephan v. Cramon-Taubadel for reviewing this thesis.

I also want to thank my coauthors: Luise, Selina, Sabine, Dharani, Louis, Randall, Than, Seuth, Lyda and Long. It is such a privilege to have worked with you!

Special thanks go to the wonderful colleagues Sreymom, Punlork, Thara, Vin, Pisey, My, Khong, Keo, Amelie, Adrian and Poukham who made the field trips in SEA a fun adventure that I will never forget. Also, I am indebted to all the smallholders that participated with a lot of patience and provided so much hospitality to us. Kop chai and okun cheraown!

I also want to thank my family for their support, first of all my wife Neda and my daughters Fariba and Minoo for their unconditional love and endless support, I love you more than anything, Phillip for the brotherhood, and Angelika and Werner for putting me into this world and being always there for me, I love you guys!

Cheers go out to the Eiermaulversammlung for making me laugh in hard times and showing me that friendship is more important than work. Thanks William for the epic bike rides in the woods of Göttingen and beyond, that kept my head clear when I needed it. Thanks to Peter and Mick for reading and discussing my work.

Table of contents

List of appendices

iii. List of abbreviations

AIC	The Akaike Information Criterion
ANOVA	Analysis of Variance
ATT	Average Treatment effect on the Treated
CPT	Cumulative Prospect Theory
EUT	Expected Utility Theory
CAPI	Computer Assisted Personal Interviewing
CH	Central Highlands of Vietnam
CI	Complexity Index
CIAT	The International Center for Tropical Agriculture
CPK	Communist Party of Kampuchea
GAIC	Global Akaike Information Criterion score
GDP	Gross Domestic Product
GIS	Geographic Information System
HFIAS	Household Food Insecurity of Access Scale
KBM	Kernel Based Matching
MDM	Mahalanobis Distance Matching
NNM	Nearest Neighbor Matching
ODK	Open Data Kit
OLS	Ordinary Least Square regression
PPI	Progress out of Poverty Index
PSM	Propensity Score Matching
RDU	Rank Dependent Utility Theory
RHoMIS	Rural Household Multi-Indicator Survey
RUA	Royal University of Agriculture
SDG	Sustainable Development Goals
SEA	Southeast Asia
USD	United States Dollar
VIF	Variance Inflation Factor
XK	Xiangkhouang province

iv. List of tables

v. List of figures

1 General introduction

1.1 The importance of smallholders on a global scale

Agriculture is a livelihood strategy for hundreds of millions of people, of which most are farming on areas smaller than 2 hectares. These farms, even though they vary in structure, function and size both within and between countries, are generally referred to as smallholder farms. It is estimated that there are more than 475 million smallholder farms, accounting for approximately 80 percent of all farms, operating on roughly 12 percent of the global agricultural land (Lowder et al., 2016). Recent research shows that smallholders globally produce around 30 percent of total crop production (Ricciardi et al., 2018). In smallholder dense regions, 90 percent of food calories are produced by smallholders (Samberg et al., 2016). The importance of smallholders in feeding a growing world population and eradicating hunger and poverty is acknowledged by researchers and policy makers alike. In 2015, the UN member states adopted the 17 Sustainable Development Goals (SDGs) which set out a 15-year plan to achieve the core goals to eradicate hunger and poverty. To reach these goals, target 2.3 of the SDGs directly addresses smallholder production systems and sets out the goal that by 2030, "[to] double the agricultural productivity and incomes of small-scale food producers, in particular women, indigenous peoples, family farmers, pastoralists and fishers, including through secure and equal access to land, other productive resources and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment" (United Nations, 2015).

1.2 Smallholder production systems in Southeast Asia

One world region where smallholders are particularly prevalent is SEA (Cohn et al., 2017). Past agricultural development strategies in SEA have concentrated on irrigated agriculture and high-potential areas in the lowlands. Proactive public policies supporting a package of yield-boosting inputs led to an increase in food production which in turn spurred economic growth (McArthur & McCord, 2017). In contrast to the success story of this so-called green revolution stands the agricultural development in marginalized highland regions which depend on rainfed agriculture, often on relatively infertile soils. Low baseline productivity on these lands and slower growth rates have led to stagnating or even worsening poverty in these regions (Rosegrant & Hazell, 2000). In 2007-2008 came a surge in international land deals which some perceived as the return of large-scale agribusiness ventures in SEA. Since then, studies have focused on the actual and potential negative impacts on the welfare of rural populations (Bissonnette & Koninck, 2017). There is little doubt that widespread smallholder engagement with agricultural commodity chains in low-income regions is associated with more inclusive patterns of rural development, especially in comparison to large-scale land concessions that typically restrict and displace traditional rural livelihoods (Cramb et al., 2017). Much of the research on recent agrarian change in upland SEA focuses on the role of cash crops like cassava, coffee, maize, oil palm and rubber, as drivers of transformation of smallholder cropping systems. Depending on the context, the widespread and rapid adoption of these cash crops may yield varied socioeconomic and environmental outcomes. These can range from extremes of rapid socioeconomic differentiation, polarization and dispossession, to broadly inclusive patterns of development beneficial to most rural inhabitants (Cramb et al., 2017; Euler et al., 2017; Fox & Castella, 2013b; Mahanty & Milne, 2016; Vicol et al, 2018). For now, we lack a sufficient understanding of the underlying drivers of positive or negative socioeconomic and environmental outcomes of these land use transitions.

To add knowledge to the existing body of literature on smallholder systems in marginal mountainous areas of SEA, the present dissertation contains three case studies on smallholder production systems from the highland regions of Cambodia, Lao PDR and Vietnam. The following section will introduce specific challenges for smallholder farmers in these regions and describe knowledge gaps that are important to achieve a better understanding of smallholder systems in SEA.

1.3 Description of challenges and knowledge gaps

1.3.1 Understanding smallholder transitions

As described in section 1.2, agricultural systems in the study region have been undergoing a transition from subsistence to commercial agriculture in the past four decades (Ashraf et al., 2017; Goto & Douangngeune, 2017; Diez, 2016). The transition of agricultural production follows pathways differentiated by local and national contexts. Generally, traditional practices like shifting cultivation and subsistence farming are phased out with intensification, modernization and market orientation. However, these transitions are not absolute, and the classification into market-oriented and subsistence-oriented systems is scale dependent. Observations at the village level make it possible to identify a diverse set of farm types and their distribution across the village. For example, research shows that it is not uncommon for villages in the northwest of Cambodia to inhabit a mix of market-oriented farms, subsistence-oriented farms and farms that practice shifting cultivation (Milne, 2013).

Geographic Information System (GIS) plot level data allows us to zoom in beyond the household level. In some cases, both subsistence and commercial strategies exist within the same farm; some plots are used to cultivate a combination of cash crops and subsistence crops, while others are dedicated entirely for commercial production. Thus, most farms are moving back and forth along a gradient of transition pathways (from subsistence to market-oriented) rather than drifting towards one pole of the spectrum, as the regional narrative might suggest. Hence, land use change is not homogenous and irreversible, and it does not follow a unidirectional pathway.

Most studies that assess land use change are based on remote sensing GIS data products. Contextual knowledge, such as factors like input use or soil tillage, cannot be effectively gathered solely with the use of remote sensing data products, but they contribute land-use change. Mixed approaches, in which GIS analysis is combined with contextual knowledge obtained through surveys, have the potential to overcome limitations associated with GIS-only or survey- based approaches. A major challenge remains to discern linkages between land use change and plot-level and household-level characteristics, as well as performance indicators at the plot and household level. With the development of new statistical approaches such as sequence analysis, it is now possible to better leverage spatial and temporal dimensions of data which are collected using mixed approaches. One such example, which has not been applied to smallholder farming systems, is the Complexity Index (CI), which allows compressing historical patterns of land use into a single metric. This composite measure combines the number of transitions occurring on each plot across multiple land uses (i.e. states), across time (i.e. sequence), with the longitudinal entropy. A comprehensive understanding of complex land use patterns in the context of smallholder farming systems in SEA is typically lacking, however new statistical approaches and mixed methods can shed light on the underlying drivers of land use change.

1.3.2 Shocks, risk preferences and loss aversion

Given its reliance on rainfall, farming in the study areas is especially vulnerable to risks posed by climate change. Changes in rainfall patterns and temperature and increases in the frequency and intensity of extreme weather events due to climate change cause significant damage and losses to crop and livestock production (Amnuaylojaroen & Chanvichit, 2019; Asian Development Bank, 2021). For example, extreme heat stress during the reproductive stage of crop growth can diminish yields in certain crops, notably rice (Wang et al., 2019). The risks smallholder farmers are exposed to put their livelihoods and assets in jeopardy and deter investment. Apart from extreme weather events, common shocks to the smallholder are crop loss due to pest and disease, fluctuations of input and output prices, and demographic shocks like illness or death of a household member. As a consequence, smallholders engage in costly shock-coping (after a shock event) and risk-management (in anticipation of a shock) adjustments like dis-saving, emergency borrowing, sale of productive assets, emergency migration, use of child labor by taking children out of school, postponement of consumption expenditures, adopting low-risk but low-return technologies and engaging in income diversification at an efficiency cost (de Janvry & Sadoulet, 2020).

Based on this rationale, many studies state that avoiding risk is a key element that contributes to persistent poverty in vulnerable populations around the globe (Dercon and Christiaensen, 2011; Mosley and Verschoor, 2005). This rationale is used to explain low agricultural productivity (Rosenzweig and Binswanger, 1993) and technology adoption (Dercon and Christiaensen, 2011) in developing countries, and why smallholders might be locked into a poverty trap (Carter and Barrett, 2006). With this in mind, it is crucial to understand individual risk attitudes and how they are affected by household shocks to design effective policy instruments to support smallholder farmers.

Many studies are published around this topic, but the results are inconclusive. Although many studies reveal influences of adverse events on risk aversion, no consensus has been reached regarding the shape and the magnitude of effects. For example, some studies report a decrease in risk aversion after an adverse event, a result that is rather counterintuitive (Eckel et al, 2009; Li et al., 2011; Voors et al., 2012). These inconsistencies are explained most commonly by emotional behavior or behavioral heuristics, which can increase risk seeking after catastrophic events. Not clear however, are the impacts of aggregated, frequent small- and medium-scale household shocks. This form of long-term risk and vulnerability is a central feature in the lives of smallholders in the study region and has strong implications for their behavior under risk.

Another important research gap arises when it comes to the methods that are applied to measure risk preferences. A variety of methods have evolved to elicit individual risk attitudes, from econometric approaches (Just & Pope, 1978) to experimental approaches (Binswanger, 1980; Holt & Chavas, 2002) and hypothetical questionnaire-based approaches (Dohmen et al., 2011; Weber et al., 2002). While simple to understand, they are not incentivized, which raises doubt that the recorded preferences reflect an individual's true attitudes toward risk, particularly in the domain of financial decision-making (Charness et al., 2013). This is the main argument to directly elicit preferences via incentivized experiments. However, the interpretation of the risk preference data from experimental methods is challenging because there are competing models on how to specify the utility function. Most commonly used are EUT, RDU and CPT. Most research is focused exclusively on estimating risk preferences in the gain domain. Not known are the effects of household shocks on risk preferences in the loss domain. Loss aversion describes the tendency of individuals to interpret outcomes as gains and losses relative to a reference point in which individuals are more sensitive to losses than to equally-sized gains (Tversky & Kahneman, 1992). This behavioral insight is crucial in the understanding of smallholder farmers risk preferences, and ignoring it leads to a misinterpretation of smallholder risk preferences. Taking into account that rainfed farming systems in marginal areas are exposed to climatic shocks and those extreme weather events could ruin entire harvests, it becomes clear that farmers are frequently confronted with situations that include net losses. Policy that aims to support smallholders in risk reducing strategies, for example through crop insurance and risk-decreasing inputs, has to take into account the aversion to losses of smallholder farmers to be effective at scale.

1.3.3 Farmer eye health as a burden to growth

Without a doubt, smallholder farmers depend on good health to master the numerous challenges they are confronted with on a day-to-day basis. Health is central to the global agenda of reducing poverty, as well as an important measure of human well-being in its own right (Dodd & Cassels, 2006). The productivity gains needed to achieve the SDGs which I describe in section 1.1, and the transitions towards more intensified and market-oriented farming systems which I describe in section 1.3.1, can only be achieved by healthy and productive smallholders. A common health problem in rural areas of low-income economies is eye health, with official data suggesting that roughly one-fifth of the global population is affected by poor visual acuity.

On an anecdotal note, when we conducted field work in in the target villages in SEA, we rarely encountered farmers that wore glasses. As someone who depends on glasses, I am acuity aware of how important good vision is for everyday life. So, how are smallholders dealing with this issue, since modern optometric services are scarce in the study regions? How can farmers with low and uncorrected visual acuity identify pests and disease on their crops? How can they apply the right amounts of fertilizers and pesticides? How can they follow a visual presentation on new technologies from an extension agent? How can they recognize faces and build relationships with the buyers of their produce? What are the overall effects of this on the income of smallholders?

The literature on this topic is very thin. Most studies look at the macroeconomic impacts of eye health on work productivity (Ajani & Ugwu, 2008; Audibert & Etard, 2003; Loureiro, 2009; Sabasi & Shumway, 2018). What remains unknown from these estimates are the losses that occur in the informal agricultural sector due to visual impairments. It is safe to say that this topic is widely overlooked. In order to be effective, the efforts to support smallholders through extension services should take into account the physical impairments of farmers.

1.3.4 A note on ethical issues in scientific practice and publishing

Lastly, I want to add a section on a topic that is indirectly connected to smallholder farmers in the region, and rather reflects the process of doing scientific research itself. The overarching aim of my dissertation is to contribute to the knowledge about marginalized smallholders in the belief that scientific discourse and discovery contributes to improving the livelihoods of resource poor and marginalized smallholders. The importance of science is acknowledged at the highest political level. The Scientific Advisory Board to the Secretary general of the United Nations states that “Science is a driver and enabler of inclusive and people-centered sustainable development […] and science will be one of the most critical means of implementation for the Agenda 2030” (Scientific Advisory Board, 2016). Achieving these goals hinges on the assumption that scientific discovery is available to policy makers, development practitioners and scientists themselves. However, recent trends in subscription fees of scientific journals raise doubts if scientists and practitioners from resource-poor institutions can actually access the literature they need to produce high-impact research themselves, or to give the most up-to-date policy recommendations.

The nature of scientific practice in the field of development opens up several gateways for discussion and reflection. To prepare this dissertation, I relied on collecting data in the rural areas of SEA. The results and conclusions of my research however, might end up behind a paywall, inaccessible for the people who need it the most. Thus, engaging in the studies of development from the position of privilege, I have to ask myself: How far am I complicit in upholding racial hierarchies that this field is inherently confronted with? A global north/south divide in research outputs is undeniable, and it is the responsibility of the research community to change this (Gibbs, 1995).

One way around the paywalls of scientific journals are piracy websites or shadow libraries that offer scientific content for free. The most prominent example of a shadow library is Sci-

Hub, which retrieves and distributes scholarly literature without regard to copyright. Sci-Hub has been growing rapidly since its creation in 2011; by March 2017, its database contained 85 percent of articles published in toll-access journals (Himmelstein et al., 2018). For my colleagues in SEA for example, accessing literature on a day-to-day basis is difficult and presents them with a moral dilemma: Is it ethical to use Sci-Hub, which violates copyrights of publishers, to access information for the greater good?

In 2017, data on the usage of Sci-hub was published, including the digital object identifiers and the locations of the downloader (Bohannon & Elbakyan, 2017). The data provides an opportunity to look at how the scientific community makes use of Sci-Hub, and to test the hypothesis that piracy is a way for researchers from low-income countries to bypass unjust hierarchies. Not known is the use of Sci-Hub in the discipline of development studies and the geographic distribution of downloads. Also unknown are the socioeconomic factors that influence the use of Sci-Hub.

1.4 Research objectives and structure of the dissertation

The previous sections highlight challenges for smallholder farming systems in SEA and address research gaps for each challenge. In this section, I want to state the research objectives and the outline of the dissertation.

Chapter 2 presents the paper titled "Drivers of land use complexity along an agricultural transition gradient in Southeast Asia", which is published in the journal *Ecological Indicators*. It address the challenges laid out in section 1.3.1 by discerning the linkages between land use change and plot-level and household-level characteristics and processes. The study reports data on plot-level land use history over a 10-year period, as well as farm management and farm performance indicators that were collected from 163 households in Xiangkhouang province in the northern Lao uplands (XK) and in the Central Highlands of Vietnam (CH).

We chose these two sites to contrast recent developments in land use change between a site that is subsistence oriented (XK) and a site that is market oriented (CH). The objectives of the study are (1) to describe plot-level sequence patterns of seasonal variation of land use over several years, (2) to apply a sequence dissimilarity metric, the CI, to measure land use transition in an agricultural system, and (3) to identify the key drivers of land use change and their linkages with farm performance indicators and plot-level characteristics through multi-dimensional analysis.

Chapter 3 presents the paper “Effects of Household Shocks on Risk Preferences and Loss Aversion: Evidence from Upland Smallholders of Southeast Asia”, which is published in the *Journal of Development Studies*. It addresses the challenges laid out in section 1.3.2. The risk preferences of 93 smallholders in Cambodia and 91 smallholders in Lao PDR are examined with an incentivized lottery design under the framework of EUT, RDU and CPT. This enables us to estimate parameters for loss aversion and to compare competing models. To identify the model that best explains the choices of smallholders, we conduct nested and non-nested hypothesis tests. In a second step, we measure the effect of shocks on risk preference parameters by including them in a variety of specifications. Furthermore, the study takes into account the manifold household shocks that farmers are confronted with and looks at how they influence the risk preference parameters.

Chapter 4 presents the paper “The effect of poor vision on economic farm performance: Evidence from rural Cambodia”, which was published in the journal *PLoS ONE*. The paper addresses the challenges laid out in section 1.3.3. The study presents the results of a standardized eye test with 288 farm managers in rural Cambodia. The objectives of the study were to quantify the prevalence of poor vision among rural smallholders in Cambodia and to investigate if poor vision is associated with a loss in agricultural profitability of family-owned farms.

Chapter 5 presents the paper "Where Can the Crow Make Friends? Sci-Hub's Activities in the Library of Development Studies and its Implications for the Field", which was published in the special issue "Decolonizing Open Access in Development Research" in the journal *Development and Change*. The paper addresses the challenges laid out in section 1.3.4. The aim of this study is to describe Sci-Hub's activities in the field of development studies. We achieve this by screening through 28 million download logs in Sci-Hub for articles in the field of development studies. We identify the geographic location of download requests and we use country metadata (Gross Domestic Product (GDP) and population) to reveal factors that influence the use of Sci-Hub. Furthermore, we take a critical look at current practices in scientific publishing and their implications for scientific conduct in this field.

Chapter 6 will synthesize conclusions and policy recommendations. Potential limitations of the study and important areas of future research are also discussed.

References

Ajani, O. I. Y., & Ugwu, P. C. (2008). Impact of Adverse Health on Agricultural Productivity of farmers in Kainji Basin North-Central Nigeria Using a Stochastic Production Frontier Approach. *Trends in Agricultural Economics*, *1*(1), 1–7.

Amnuaylojaroen, T., & Chanvichit, P. (2019). Projection of near-future climate change and agricultural drought in Mainland Southeast Asia under RCP8.5. *Climatic Change*, *155*(2), 175–193. https://doi.org/10.1007/s10584-019-02442-5

Ashraf, J., Pandey, R., & de Jong, W. (2017). Assessment of biophysical, social and economic drivers for forest transition in Asia-Pacific region. *Forest Policy and Economics*, *76*, 35–44. https://doi.org/10.1016/j.forpol.2016.07.008

Asian Development Bank. (2021). *Asian Development Outlook 2021 Update*. Asian Development Bank, Mandaluyong City, Philippines. Retrieved from : http://dx.doi.org/10.22617/FLS210352-3

Audibert, M., & Etard, J. F. (2003). Productive benefits after investment in health in Mali. *Economic Development and Cultural Change*, *51*(3), 769–782. https://doi.org/10.1086/367982

Binswanger, H. P. (1980). Attitudes toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics*, *62*(3), 395–407.

Bissonnette, J. F., & Koninck, R. De. (2017). The return of the plantation? Historical and contemporary trends in the relation between plantations and smallholdings in Southeast Asia. *Journal of Peasant Studies*, *44*(4), 863–883. https://doi.org/10.1080/03066150.2017.1311867

Bohannon, J., & Elbakyan, A. (2017). Data from : Who is downloading pirated papers? Everyone. *Dryad* dataset. https://doi.org/10.5061/dryad.q447c

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: Empirical and policy implications. *Journal of Development Studies*, *42*(2), 178–199. https://doi.org/10.1080/00220388.2013.785527

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization*, *87*, 43–51. Retrieved from http://dx.doi.org/10.1016/j.jebo.2012.12.023

Cohn, A. S., Newton, P., Gil, J. D. B., Kuhl, L., Samberg, L., Ricciardi, V., … Northrop, S. (2017). Annual Review of Environment and Resources: Smallholder Agriculture and Climate Change. *Annual Review of Environment and Resources*, *42*, 347–375. Retrieved from https://doi.org/10.1146/annurev-environ-

Cramb, R., Manivong, V., Newby, J. C., Sothorn, K., & Sibat, P. S. (2017). Alternatives to land grabbing: exploring conditions for smallholder inclusion in agricultural commodity chains in Southeast Asia. *Journal of Peasant Studies*, *44*(4), 813–841. https://doi.org/10.1080/03066150.2016.1242482

de Janvry, A., & Sadoulet, E. (2020). Using agriculture for development: Supply- and demand-side approaches. *World Development*, *133*, 105003. https://doi.org/10.1016/j.worlddev.2020.105003

Dercon, S., & Christiaensen, L. (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics*, *96*(2), 159–173. https://doi.org/10.1016/j.jdeveco.2010.08.003

Dodd, R., & Cassels, A. (2006). Health, development and the Millennium Development Goals. *Annals of Tropical Medicine and Parasitology*, *100*(5–6), 379–387. https://doi.org/10.1179/136485906X97471

Dohmen, T., Falk, A., Huffmann, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 371–589.

Eckel, C. C., El-Gamal, M. A., & Wilson, R. K. (2009). Risk loving after the storm: A Bayesian-Network study of Hurricane Katrina evacuees. *Journal of Economic Behavior and Organization*, *69*(2), 110–124. https://doi.org/10.1016/j.jebo.2007.08.012

Euler, M., Krishna, V., Schwarze, S., Siregar, H., & Qaim, M. (2017). Oil Palm Adoption, Household Welfare, and Nutrition Among Smallholder Farmers in Indonesia. *World Development*, *93*, 219–235. https://doi.org/10.1016/j.worlddev.2016.12.019

Fox, J., & Castella, J. C. (2013). Expansion of rubber (Hevea brasiliensis) in Mainland Southeast Asia: What are the prospects for smallholders? *Journal of Peasant Studies*, *40*(1), 155–170. https://doi.org/10.1080/03066150.2012.750605

Gibbs, W. W. (1995). Lost Science in the Third World. *Scientific American*, 273(2):92-99

Goto, K., & Douangngeune, B. (2017). Agricultural modernisation and rural livelihood strategies: the case of rice farming in Laos. *Canadian Journal of Development Studies*, *38*(4), 467–486. https://doi.org/10.1080/02255189.2017.1263553

Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Greshake Tzovaras, B., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *ELife*, *7*, 1–22. https://doi.org/10.7554/eLife.32822

Holt, M. T., & Chavas, J.-P. (2002). The Econometrics of Risk. In R. E. Just & R. D. Pope (Eds.), *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture* (pp. 213–241). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-3583-3_11

Just, R. E., & Pope, R. D. (1978). Stochastic specification of production functions and economic implications. *Journal of Econometrics*, *7*(1), 67–86. https://doi.org/10.1016/0304-4076(78)90006-4

Li, J. Z., Li, S., Wang, W. Z., Rao, L. L., & Liu, H. (2011). Are people always more risk averse after disasters? Surveys after a heavy snow-hit and a major earthquake in China in 2008. *Applied Cognitive Psychology*, *25*(1), 104–111. https://doi.org/10.1002/acp.1648

Loureiro, M. L. (2009). Farmers' health and agricultural productivity. *Agricultural Economics*, *40*(4), 381–388. https://doi.org/10.1111/j.1574-0862.2009.00385.x

Lowder, S. K., Skoet, J., & Raney, T. (2016). The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development*, *87*, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041

Mahanty, S., & Milne, S. (2016). Anatomy of a boom: Cassava as a 'gateway' crop in Cambodia's north eastern borderland. *Asia Pacific Viewpoint*, *57*(2), 180–193. https://doi.org/10.1111/apv.12122

McArthur, J. W., & McCord, G. C. (2017). Fertilizing growth: Agricultural inputs and their effects in economic development. *Journal of Development Economics*, *127*(February), 133–152. https://doi.org/10.1016/j.jdeveco.2017.02.007

Milne, S. (2013). Under the leopard's skin: Land commodification and the dilemmas of Indigenous communal title in upland Cambodia. *Asia Pacific Viewpoint*, *54*(3), 323–339. https://doi.org/10.1111/apv.12027

Mosley, P., & Verschoor, A. (2005). Risk Attitudes and the 'Vicious Circle of Poverty.' *The European Journal of Development Research*, *17*(1), 59–88. https://doi.org/10.1080/09578810500066548

Revilla Diez, J. (2016). Vietnam 30 years after Doi Moi: Achievements and challenges. *Zeitschrift Für Wirtschaftsgeographie*, *60*(3), 121–133. https://doi.org/10.1515/zfw-2016-0035

Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., & Chookolingo, B. (2018). How much of the world's food do smallholders produce? *Global Food Security*, *17*(May), 64–72. https://doi.org/10.1016/j.gfs.2018.05.002

Rosegrant, M. W., & Hazell, P. B. R. (2000). *Transforming the rural Asian economy: The unfinished revolution*. Hong Kong: Oxford University Press.

Rosenzweig, M. R., & Binswanger, H. P. (1993). Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *The Economic Journal*, *103*(416), 56–78. https://doi.org/10.2307/2234337

Sabasi, D., & Shumway, C. R. (2018). Climate change, health care access and regional influence on components of U.S. agricultural productivity. *Applied Economics*, *50*(57), 6149–6164. https://doi.org/10.1080/00036846.2018.1489504

Samberg, L. H., Gerber, J. S., Ramankutty, N., Herrero, M., & West, P. C. (2016). Subnational distribution of average farm size and smallholder contributions to global food production. *Environmental Research Letters*, *11*(12). https://doi.org/10.1088/1748-9326/11/12/124010

Scientific Advisory Board. (2016). Science for sustainable development: policy brief by the Scientific Advisory Board of the UN Secretary-General. Retrieved from http://unesdoc.unesco.org/images/0024/002461/246104E.pdf

Tversky, A., & Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/Doi 10.1007/Bf00122574

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development.* New York.

Vicol, M., Pritchard, B., & Htay, Y. Y. (2018). Rethinking the role of agriculture as a driver of social and economic transformation in Southeast Asia's upland regions: The view from Chin State, Myanmar. *Land Use Policy*, *72*(September 2017), 451–460. https://doi.org/10.1016/j.landusepol.2018.01.009

Voors, M. J., Nillesen, E. E. M., Verwimp, P., Bulte, E. H., Lensink, R., & Van Soest, D. P. (2012). Violent conflict and behavior: A field experiment in Burundi. *American Economic Review*, *102*(2), 941–964. https://doi.org/10.1257/aer.102.2.941

Wang, Y., Wang, L., Zhou, J., Hu, S., Chen, H., Xiang, J., … Zhang, Y. (2019). Research Progress on Heat Stress of Rice at Flowering Stage. *Rice Science*, *26*(1), 1–10. https://doi.org/10.1016/j.rsci.2018.06.009

Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. https://doi.org/10.1002/bdm.414

2 Drivers of land use complexity along an agricultural transition gradient in Southeast Asia

Abstract

Agricultural systems in SEA are rapidly transitioning from subsistence-oriented to market-oriented agriculture. Driven by the highly complex and variable decision processes of individual farm households, these transitions have produced a diverse landscape mosaic across the region. Elucidation and characterization of underlying decision-making processes, and the factors that influence land use choices, are thus essential for sustainable land use planning. To enable a study that seeks to understand these linkages, data on plot-level 10-year land use history, management and farm performance indicators were collected from 163 households in Lao PDR and in Vietnam, areas chosen to represent two extremes of the transition gradient. The objectives of the study were (1) to describe plot-level sequence patterns of seasonal variation of land use over several years, (2) to apply a sequence dissimilarity metric, the Complexity Index (CI), to measure land use transition in an agricultural system, and (3) to identify the key drivers of land use change and their linkages with farm performance indicators and plot level characteristics through multi-dimensional analysis. CI allowed compressing historical land use data and quantifying land use complexity in a simple and efficient manner. Land use dynamics varied strongly between the two sites, with 66 percent of the land use types in the Laos site being completely replaced by others during the recall periods, compared to only 15 percent in the Vietnam site. Associated key drivers of land use change also differed significantly: while end use of agricultural products was the main driver behind land use changes in the Vietnam site, a more complex relationship between topography and management vs. land use change was evident in the Laos site. Likewise, land use complexity does not exhibit the same relationship with farm performance in the two sites: in Vietnam, households with higher food availability are half as likely to

transition, while in the Lao uplands, land use complexity was significantly correlated with the Progress out of Poverty Index (PPI). Multidisciplinary studies remain necessary to assess the impact of innovative sustainable intensification options on system performance and environmental sustainability, before policies are enacted to support their dissemination in SEA smallholder agricultural systems. Context-specific CI thresholds associated with system quality indicators could support this by informing decision-makers on the pace of agricultural transformation and its environmental impacts.

This chapter is coauthored by Dharani Dhar Burra (DB), Louis Parker (LP), Nguyen Thi Than (NT), Phonepaseuth Phengsavanh (PP), Chau Thi Minh Long (CL), Randall S. Ritzema (RR), Frederik Sagemueller (FS) and Sabine Douxchamps (SD). The contributions to the article are: DB: conceptualization, indicator development, formal analysis, visualizations, writing. LP: survey methodology, data collection, visualization of spatial data. NT: data collection and curation. PP: site coordination in Laos PDR. CL: site coordination in Vietnam. RR: secondary data contribution, writing. FS: conceptualization. SD: conceptualization, supervision, project management, visualization, writing. The chapter is published in the journal *Ecological Indicators* with the following citation: Burra, Parker, Nguyen, Phengsavanh, Chau , Ritzema, Sagemüller, Douxchamps (2021). Drivers of land use complexity along an agricultural transition gradient in Southeast Asia. *Ecological Indicators*, https://doi.org/10.1016/j.ecolind.2021.107402.

2.1 Introduction

Agricultural activities account for 62 percent of the observed land use changes in Asia during the last decades (Song et al., 2018). In SEA, agricultural systems are undergoing a rapid transition from subsistence towards market-oriented agriculture (Diez, 2015; Ashraf et al., 2017; Goto and Douangngeune, 2017). Smallholder farmers replacing subsistence crops with cash crops typically follow market demands and economic opportunities (Alexander et al., 2017; Dawe, 2015; Green and Vokes, 1997; Rigg, 2012). However, transition processes are complex and nuanced: a transition to cash cropping systems is not a unidirectional pathway, but rather a dynamic process in which a return to subsistence farming may occur. These transitions are driven by complex and varying decision processes of individual farm households, producing a diverse mosaic across the landscape. It is not uncommon, for example, to find villages comprised of a mixture of market-oriented farms, subsistence-oriented farms and farms that still practice shifting cultivation (Milne, 2013), sometimes with seasonal cropping transitions that seem random. Although the transition towards intensive and market-oriented agriculture generally improves overall income for smallholders (Hettig et al., 2016), it often occurs at the expense of ecological and environmental sustainability, as well as livelihood security (Klasen et al., 2016; Dressler et al., 2017; Ditzler et al., 2019). Beyond its effects on key performance indicators such as smallholder income and food availability, land use transition is linked to plot management practices such as agrochemical use or soil tillage that influence soil health and fertility, as well as plot allocation to particular seasonal, annual, or permanent land uses. Just as the factors affecting adoption of sustainable management practices must be well understood to ensure positive impact at scale, the drivers of land use change at the farm household level must be considered by policy makers and development actors for sustainable land use planning. It is essential to understand and characterize the decision-making processes of smallholders, and the myriad of factors that influence their choices (Lambin et al., 2003; Southworth et al., 2012; Ashraf et al., 2017). In reality, this has often not been the case, and relevant government policy has often been misguided and even contradictory, with some policies e.g. encouraging upland farmers to replace swidden agriculture for monoculture (Dressler 2017), and others redirecting the trajectory of these transitions in a manner that attempts to balance financial stability and environmental sustainability (Fröhlich et al., 2013). Indeed, land use planning and other regulatory approaches to environmental services issues have had little success in SEA to date, and a more robust understanding of the linkage between policy and underlying biophysical and decision-making processes could help to expand the range of policy options for supporting sustainable land uses (Tomich et al., 2004). The aim of this study was to discern linkages between land use change and plot-level and household-level characteristics and processes: information that could ultimately inform land use policy. However, drivers of land use change are known to be highly context- and location-specific (Lambin et al., 2000). Therefore, the two sites for this study were chosen to represent two extremes of the transition gradient in SEA: (1) Xiangkhouang province in Lao PDR (XK), which is highly subsistence-oriented, has low levels of formal education, has poor accessibility, and garners little attention from policy makers (Hepp et al., 2019; Thanichanon et al., 2019); and (2) the Central Highlands of Vietnam (CH), an example of forward-looking, market-oriented agriculture that has a high degree of political involvement (Müller and Zeller, 2002). Most studies that assess land use change are based on remote sensing or GIS data products. Since open source remote sensing data products are of low spatial but relatively high temporal resolution, a comprehensive understanding of complex land use patterns is typically lacking, particularly in the context of smallholder farming systems (Kammerbauer and Ardon, 1999). Additionally, contextual knowledge, such as factors contributing to a particular land use, cannot be effectively gathered solely with the use of remote sensing data products. Mixed approaches, in which GIS analysis is combined with contextual knowledge obtained through surveys, have the potential to overcome challenges associated with GIS-only or survey-based approaches. With the development and increased use of new statistical approaches such as sequence analysis, it is now possible to better leverage spatial and temporal dimensions of data that is collected using mixed approaches. For instance, the temporal dimension of land use data can be used to construct time-series indicators, which can then be combined with contextual survey-based knowledge, to better understand drivers of land use (Ritschard and Studer, 2018). Multidimensional characterization of complex farming systems and associated land use is necessary to design intervention strategies that enhance sustainability across several community dimensions, such as financial, environmental, and health. The objectives of this study were therefore (1) to describe plot-level sequence patterns of seasonal variation of land use over several years, (2) to apply a sequence dissimilarity metric, the CI, to measure land use transition in an agricultural system, and (3) to identify the key drivers of land use change and their linkages with farm performance indicators and plot level characteristics through a multi-dimensional analysis.

2.2 Materials and methods

2.2.1 Site description

The study was conducted on the hillsides of the XK plateau in northern Laos, within a 40 km radius from Phonsavanh (19◦26′ 59.30″ N, 103◦13′16.43″ E), and in the CH, in DakLak and DakNong provinces. Sites XK and CH typify transitions in smallholder agricultural production systems of SEA. While farming systems in XK are currently transitioning from subsistence, low-intensive to market-driven, high intensive production, site CH underwent such a transition in the late 1980′ s and early 1990′ s. Site XK is approximately 1095 m above sea level, with two seasons a year: a cool and dry season between November and March, and a warm and rainy season from April to October. Site CH consists of several plateaus ranging from 500 to 1500 m above sea level, with an annual rainfall ranging from 1500 to 2400 mm.

The CH rainy season typically lasts from May to October, with April and May being the hottest months of the year.

2.3 Data acquisition surveys

To characterize farming systems in the study area, a baseline survey was conducted in December 2015 among 366 and 310 households selected randomly in site XK and site CH, respectively (Ritzema et al., 2019), using the Rural Household Multi-Indicator Survey (RHoMIS) tool (Hammond et al., 2017). RHoMIS questionnaire modules were administered using the Open Data Kit (ODK) and included questions on household level social and demographic characteristics, food security indicators, poverty, crops and livestock including yields, sale prices and inputs, and measures of off-farm incomes. Focus group discussions with local experts in each site, and exploration of the RHoMIS data using unsupervised clustering analysis, identified cumulative diversity (combinatorial counts of crop and livestock species diversity) and market orientation (a dimensionless ratio defined as relative importance of crop and livestock sales in generating potential total food energy) as the two main drivers of variation in the RHoMIS dataset (Epper et al., 2020).

Subsequently, a stratified sampling approach capturing contrasting levels of these two variables was used to sample a subset of the RHoMIS households. Seventy-two households and 91 households were sampled for a detailed temporal land use survey in sites XK and CH, respectively. The standardized Computer Assisted Personal Interviewing (CAPI)-based survey was conducted in March and May 2017 by teams of trained local enumerators. The survey instrument consisted of the following modules: a) a household datasheet capturing descriptive socio-economic indicators from the household head (household level data), (b) a field properties registration sheet, which contained geotags of all the plots belonging to the interviewed household, historical land use based on farmer recall for every plot and season for the time period spanning from 2007 to 2016 (plot level data, temporal, 10 years), and (c) a broad spectrum of socio-economic, biophysical and cultural factors that influence land use such as access to resources (e.g. markets, water etc.), slope, soil fertility management, irrigation, final use, transport, and source of planting material for the last season (plot level data). Land use was classified into 34 and 41 plot-level land uses for site CH and XK, respectively. Minor plot level land use types that were not present in the predetermined list of land use types (i.e. 34 and 41 for site CH and XK, respectively) were coded as “others”. The main fruit trees (cashew, mango, durian and avocado) were recorded separately, while the others were grouped under one land use type (fruit trees). The land use type “fallow” was used for plots without crops for both long and short durations between cropping periods, whereas “forages” included all material planted for grazing or livestock feeding using the cut-and-carry system.

2.3.1 Data analysis

2.3.1.1 Data processing and transition rate

Punctual data was classified into plot-level characteristics, household-level characteristics and system performance (Table 2.1) indicators, based on the assessment methodology of smallholder agricultural systems described in Hammond et al. (2017).

Table 2.1: Variables used to characterize households, plots and system performance in the regression analysis

Class	Variable Name	Unit	Type	Definition
Household level	Ethnic group	n.a.	categorical	Ethnic group of the household head. CH site: Kinh, Mnong, Tay or Thai; XK site: Loum or Hmong
	Origin	yes or no	categorical	Origin from current location or migrated from elsewhere (y/n)
	Off-farm employment	yes or no	categorical	Presence of income from non-farming activities
	Labor distribution	n.a.	categorical	Predominant source of labor for agricultural activities: family, contracted, community members, or combination
	Household size	AME[1]	continuous	Number of people in the household
	Education	n.a.	categorical	Education level of the household head: illiterate, primary, secondary or post-secondary
	Land cultivated	ha	continuous	Total area cultivated
	Fertilization	kg N/year $^{-1}$	continuous	Total chemical nitrogen inputs on the farm
	Crop diversity	n.a.	continuous	Number of different crop species cultivated
	Livestock diversity	n.a.	continuous	Number of different livestock species cultivated
	Number of plots	n.a.	continuous	Number of plots cultivated by the household
Plot level	Soil tillage	yes or no	categorical	Practice of tillage
	Agrochemical inputs	n.a.	categorical	Use of chemical fertilizers, pesticides or both
	Property	n.a.	categorical	Ownership type: collective, family owned or on rent
	Irrigation	n.a.	categorical	Irrigation source: canal, pump or rainfed
	Slope	n.a.	categorical	Slope estimation: flat, modest or strong
	Source of planting material	n.a.	categorical	Source of planting material: bought, subsidized, exchanged, own or a combination
	Final use	n.a.	categorical	Fate of crop products: for sale, home consumption or both
Performance	PPI	n.a.	continuous	Likelihood of household's total expenditure below the national poverty line[2]
	Total income	USD/year^{-1}	continuous	Sum of income from agricultural activities
	Food availability	kcal/AME/year^{-1}	continuous	Potential amount of food that can be generated from on and off-farm income[3]
	Food self-sufficiency	kcal/AME/day^{-1}	continuous	Capacity to fulfill the household energy requirements from food produced on-farm
	Food insecurity	n.a.	continuous	Household Food Insecurity of Access Scale (HFIAS) score[4]
	Total energy available	MJ	continuous	Sum of energy from crop and livestock products produced on-farm, as well as from food bought in the market

[1]Adult male equivalent (adult = 1; child = 0.5). [2]PPI score = Desiere, S., Vellema, W., D'Haese, M., 2015. A validity assessment of the Progress out of Poverty Index (PPI)(TM). Eval. Program Plan. 49, 10-18. [3]Ritzema et al., 2019. [4]HFIAS measures the frequency and severity of hunger (Coastes et al., 2007). Scores range from 0-27, where a score of 0 signifies that the respondent is 'food secure' and a score of 27 is the severest level of food insecurity. Obtained from Ritzema et al. (2019).

All analyses were performed in R statistical computing environment (v 3.4.1). For temporal data, and separately for the dry and wet season, the total number of plots and total area for each land use type were calculated from the field properties registration sheet, using R packages reshape2 (Wickham, 2012) and dplyr (Wickham et al., 2015). The resulting bar charts were produced using R package ggplot2 (Wickham, 2011). The land use type and the corresponding time dimension of recall data was used to develop a state-sequence for each plot, wherein the state corresponds to the land use type, and sequence corresponds to the sequence of land use types, across the recall period for each plot. This state sequence was further used to calculate a transition rate matrix, using the R package TraMiner (Gabadinho et al., 2011), which calculates the rate of probability of transition between all combinations of plot level land use types captured across time, thereby providing a proxy for stability of a specific land use type.

The transition rate between two states *si* and *sj* is calculated using the following formula:

$$p(sj|si) = \frac{\sum_{t=1}^{L-1} n_{t,t+1}\,(sj, si)}{\sum_{t=1}^{L-1} n_t\,(si)} \tag{2.1}$$

wherein $p(sj|si)$ is the probability of switching at a given position from state s_i to s_j, L is the maximum observed sequence length, $n_t(s_i)$ is the number of sequences that do not end in t with state si at position t, and $nt,t+1(si,sj)$ is the number of sequences with state s_i at position t, and state s_j at position $(t + 1)$. Each row in the resulting transition rate matrix provides the transition distribution from the originating state s_i in t, to the states in $(t + 1)$, such that each row equals to one, while the diagonal provides an assessment of the stability of each state. The transition rate matrix was visualized using a heat map produced using R package ggplot2 (Wickham, 2011).

2.3.1.2 Complexity Index

To compress historical patterns of land use (i.e. state-sequence), comprised of each crop/crop combination (as a state), and the recall period (as a sequence) into a single metric, a CI was calculated for each plot (Gabadinho et al., 2015). This composite measure combines the number of transitions occurring on each plot across multiple land uses (i.e. states), across time (i.e. sequence) with the longitudinal entropy. CI is calculated using the following equation:

$$C(x) = \sqrt{\frac{l_d(x)}{l(x)} \frac{h(x)}{h_{max}}} \tag{2.2}$$

wherein *C(x)* is the CI of a given plot *x*, h_{max} is the theoretical maximum value of the entropy given the state i.e. h_{max} = log (a). The entropy is a measure of the diversity of states at site level at a given position in the sequence. A CI score of 0 is reached by a sequence with a single distinct state; i. e. with no transition and an entropy of 0, meaning a constant land use type between 2007 and 2016 for the present study. CI reaches a maximum of 1 only if the sequence *x* is such that (1) *x* contains each of the states of the total states, (2) the same time $\ell(x)/a$ is spent in each state, wherein *a* is the size of each state and (2) the number of transitions is $\ell(x) - 1$. This is the case when there has been a different plot level land use type every season during the recall period.

2.3.1.3 Linkages between CI, performance, plot and household characteristics and assessment of within-site variation

The linkages between CI, performance, and plot- and household-level characteristics were explored through logistic regression for site CH and a generalized additive model for site XK. The analyses were performed separately for each of the sites, as they were distinct in terms of land use, biophysical conditions and topography, and different drivers of land use change were expected at each site. For the regression with household characteristics and performance variables, CI was averaged across all plots for each household, and this averaged value (agg-CI) was used as a response variable instead of CI. For site CH, the distribution of CI and agg-CI showed extreme left-skew, and was converted into a binomial response variable, with all CI and agg-CI values equal to 0 in one group (n = 115 plots, n = 24 households) and all scores above 0 (n = 82 plots, n = 43 households) placed in another group. Summary statistics, particularly counts of unique occurrences of variable combinations across both groups (i.e. transition and no-transition outcomes), were performed using R package dplyr. Logistic regression analysis was performed to identify variables that significantly influence CI and agg-CI. A stepwise (forward and backward) model building strategy was employed, using a 'full model' containing all explanatory variables and the response variable, and a 'null model' containing only the intercept and the response variable. The Akaike Information Criterion (AIC) was used for model selection. The selected model was further assessed for fit by performing a comparative Analysis of Variance test (ANOVA) on the residuals of the selected, the full and the null model. Pseudo-R squared values are most commonly used to assess fit of logistic regression models (Hu et al., 2006; Marmolejo-Ramos et al., 2020). In this case multiple pseudo-R squared indices (Cox and Snell and Nagelkerke index) were calculated for both the full and the selected model, using the LogregR2 function in package descr in R statistical environment (Aquino, 2018). Only those models that displayed significantly higher pseudo-R squared values across both the indices, in comparison to the full model were selected and described. To confirm the model selection results, a single term deletion analysis was performed to quantify the impact of the presence or absence of each variable used in the model selection on the AIC score of the model, using Chi-square tests. Additionally, diagnostic checks of the residuals, such as quantile-quantile plots of the residuals and the fit of residuals against the fitted values, of the selected model were performed. The coefficients of the selected model were exponentiated to obtain the odds ratio.

For site XK, summary statistics (i.e. mean and standard deviation) of CI and agg-CI disaggregated by plot-level, household-level, or performance-based characteristics were calculated using R package dplyr. For characteristics that were continuous in nature, the Pearson correlation coefficient was calculated between agg-CI and each characteristic. All variables, including agg-CI, were log-transformed to compute the Pearson correlation coefficient. In order to model the relationship between plot-level, household-level and performance-based characteristics with CI and agg-CI for site XK, generalized additive regression modelling was performed using the GAMLSS package in R. The GAMLSS framework addresses the skewed distribution of the response variable (i.e. CI and agg-CI), enables modelling of response variables that do not belong to an existing set of exponential families, and allows for modelling of multiple parameters (i.e. location, shape and scale) of the response variable distribution (Stasinopoulos and Rigby, 2008). Selection of an appropriate distribution function of the response variable was obtained by fitting multiple continuous distributions defined on the real line, using the fitDist function in GAMLSS, and the distribution that obtained the lowest Global Akaike Information Criterion score (GAIC), was used for subsequent analysis. This analysis showed that for both CI and agg-CI, sin-arcsinh distribution had the lowest GAIC score, and hence provided the best fit, in comparison to all continuous distribution families tested. The histDist function was used to develop histograms that overlay sin-arcsinh distribution over the distribution of either CI or agg-CI (Figure A 1). Model selection followed the same steps as for the CH site. Models with interactions between explanatory variables did not converge and resulted in a poor fit compared to models with no interactions, hence interactions were not included. Additionally, Cragg-Uhler and Cox-Snell pseudo-R squared values were derived for the selected model and were compared with the same values derived from the full model. Models that either had higher pseudo-R-squared values or had similar values but with fewer explanatory variables in comparison to the full model were selected and described. The pseudo R-squared values for the GAMLSS models were calculated using the Rsq function in GAMLSS package in R (Rigby and Stasinopoulos, 2005). Multicollinearity among explanatory variables in the final model was checked, and variables with a variance inflation factor (VIF) lower than 2 were selected in the final model. Selected models were subjected to additional diagnostics based on the residual distribution, and only those models whose mean and coefficient of skewness were closest to 0, a variance closest to 1, and a coefficient of kurtosis closest to 3 were selected and described in detail. Model diagnostics based on residual distribution for the selected models (for plot-level, household-level and performance-based variables) are presented in the supplementary materials (Table A 1). Model outputs were visualized using the package ggplot2 in R (Wickham, 2011), and presented using the package stargazer (Hlavac, 2014). Site CH included six villages, and five villages constitute site XK. To assess if CI and agg-CI differed between villages in each site, a Pearson's chi-square test was performed in the case of site CH, and the Kruskal-Wallis rank sum test (the non-parametric equivalent to ANOVA) was performed in the case of site XK, with CI and agg-CI as the response variable and the village as the explanatory variable.

2.3.1.4 Spatial data analysis

To complement Pearson's chi-square test and Kruskal-Wallis rank sum test, an assessment of spatial relationships between plots with differing CI values was performed. A spatial point pattern analysis identified differences in distances between plots that belonged to specific CI categories. For site CH, similarly to the regression analysis, plots with CI equal to 0 were grouped into one category, while those above 0 were grouped into another. For site XK, plots were categorized into groups by subjecting CI values to the Jenks natural breaks classification method (Rabosky et al., 2014). The optimal number of breaks was identified based on a goodness-of-fit measure using the GmAMisc package in R (Alberti, 2020). The highest goodness-of-fit score was obtained by using three breaks. Based on these results, plots in site

XK with CI between 0 and 0.3 were categorized into Group 1, while Group 2 consisted of plots with CI between 0.3 and 0.6, and Group 3 consisted of plots with CI above 0.6. Centroids were extracted from each plot polygon using ArcMap 10.7 and the ESRI default satellite base layer. The CI values were added to the centroids based on the common plot identification numbers, using the join function. The point shapefile was exported and the attribute table was saved in spreadsheets format for further analysis. Subsequently, pairwise distance (in km) between the centroids was calculated separately for each category of plots (i.e. two categories for site CH, and three for site XK) using the gDistance function in R package rgeos (Bivand and Rundel, 2020), and was then used to calculate the mean distance between each plot and all other plots belonging to the same category. The mean distance measure was used to produce density-based histograms for each category of plots, for sites CH and XK using the R package ggplot2.

2.4 Results

2.4.1 System performance characteristics

In both sites, most households reported no off-farm employment and relatively similar cultivated areas and crop diversity (Table 2.2), but the two sites differed in several other key characteristics. Compared to site XK, households in site CH had more plots per household (averaging 74 percent higher), more migrants, and higher education levels of household heads. They reported higher nitrogen fertilizer use, higher levels of soil tillage, and higher prevalence of pump-based irrigation systems despite 52 percent of the plots being located on flatlands. A greater proportion of CH farmers bought seeds, and production was market-oriented. CH households indicated a higher average PPI and total mean food availability score than XK households, which had higher modal values for livestock diversity, higher household sizes and higher food self-sufficiency scores.

Table 2.2: Farming system characteristics in CH and XK

Class	Variable name	Unit	CH Mean ± SD or %[1]	XK Mean ± SD or %[1]
Household level[2]	Ethnic group- Minority	n.a.	10	74
	Origin- Migrated	yes or no	91	16
	Off-farm employment- Yes	yes or no	29	42
	Labor- No contracted labor	n.a.	38	68
	Household size	AME[3]	3.30 ± 1.23	5.04 ± 2.5
	Education- Post secondary	n.a.	21	2
	Land cultivated	ha	2.40 ± 1.78	2.25 ± 1.52
	Fertilization	kg N/ year	93.0 ± 74.5	0.09 ± 0.80
	Crop diversity	n.a.	4.4 ± 1.5	4.3 ± 1.8
	Livestock diversity	n.a.	2.1 ± 1.1	3.1 ± 1.4
	Number of plots	n.a.	2.5 ± 1.5	1.78 ± 0.5
Plot level[4]	Soil tillage- Yes	yes or no	87	50
	Agrochemical inputs- No chemicals	n.a.	0	86
	Property- Owned	n.a.	99	89
	irrigation- Pump	n.a.	74	0
	Slope- Flat	n.a.	52	22
	Source of planting material- Bought	n.a.	89	18
	Final use- Sale	n.a.	83	12
Performance[2]	Progress out of poverty	n.a.	70.4 ± 18.8	55.16 ± 12.13
	Total income	USD/year	15,843 ± 84,083	8,980 ± 67,094
	Food availability	kcal/AME/year	159,975 ± 739,967	44,421 ± 271,645
	Food self-sufficiency	kcal/AME/day	4,179 ± 5,286	5,736 ± 6,075
	Food insecurity (HFIAS)	n.a.	5.14 ± 4.46	2.25 ± 3.51
	Total energy available	MJ	892,847 ± 4,641,712	366,276 ± 2,449,482

[1]Mean ± Standard Deviation (SD) for quantitative variables; % of occurrence for the most striking of the categories for categorical variables. [2]n = 68 for site CH, n = 76 for site XK. [3]Adult Male Equivalent (adult = 1; child = 0.5). [4]n = 211 for site CH, n = 153 for site XK.

2.4.2 Land use changes and transition rate

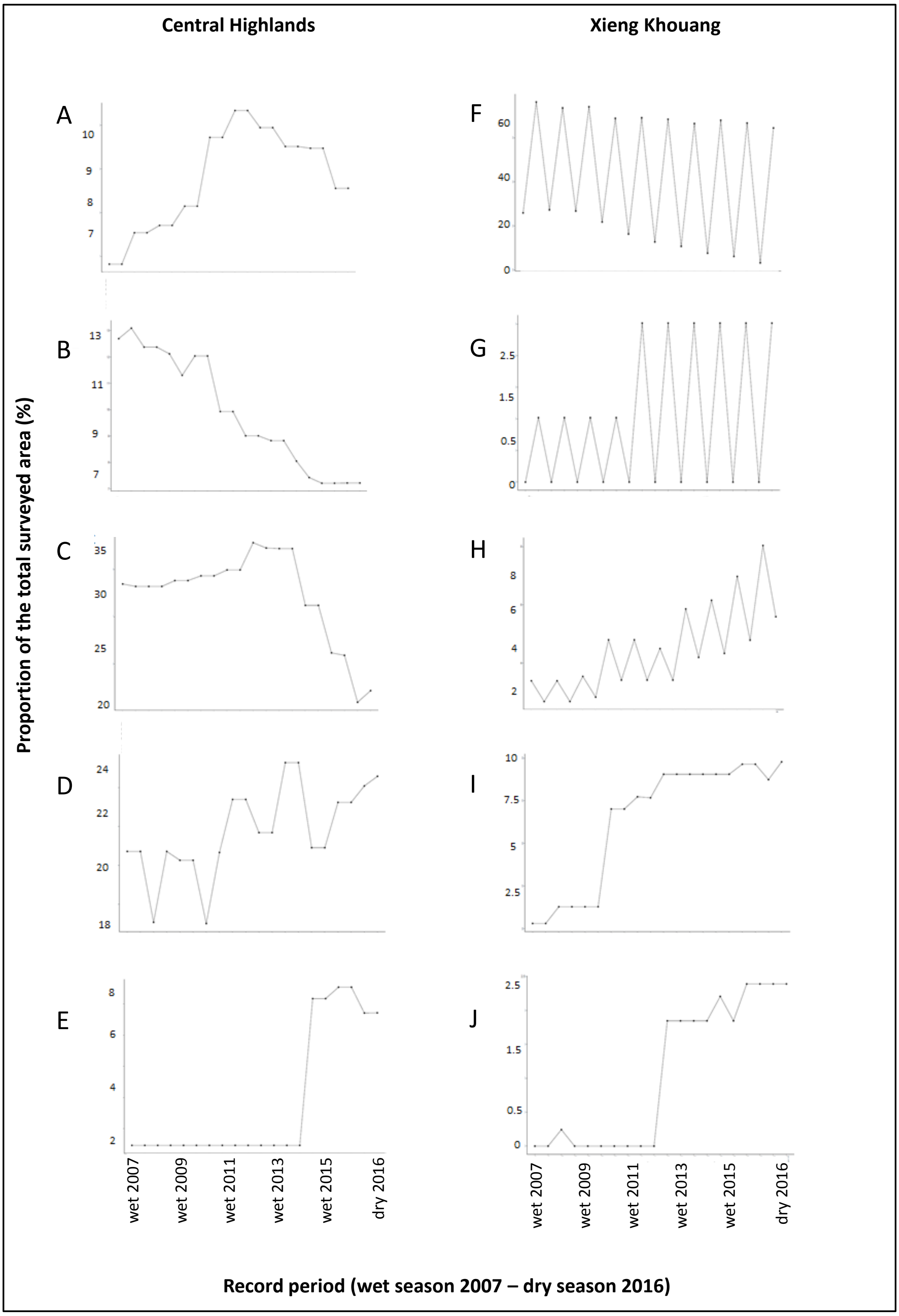
2.4.2.1 Site CH

Thirty-four different plot level land use types were identified in site CH over the recall period. Figure 2.1 (A-E) shows the temporal changes in land size for the top five land use types, in terms of total area occupied in the dry season of 2016. Since 2007, the land area dedicated to mixed cropping systems increased, with cashew-coffee land use constituting 2 percent of the total surveyed area in 2007, and 9 percent in 2016 (Figure. 2.1, A). However, when planted as monocultures, cashew and coffee declined by 7 percent and 10 percent, respectively, after 2007 (Figure 2.1, B and 2.1, C). Although not a major crop in terms of total area for the dry season of 2016, pepper, as a monoculture system, increased from 1 percent to 6 percent of the total surveyed area between 2007 and 2016. The increase of pepper in mixed cropping systems has been comparatively higher, from 3 percent to 16 percent of total surveyed area between 2007 and 2016. Monoculture plots of annual crops showed a slight increase: from 21 percent to 26 percent of the total surveyed area in the case of sugarcane (Figure 2.1, E), and from 2 percent to 7 percent in the case of cassava (Figure 2.1, D). Maize production area decreased in both monoculture (8 percent) and mixed cropping systems (12 percent). Similarly, fallow-based plot level land use decreased by 3 percent, as a proportion of total area surveyed, between 2007 and 2016. Average plot size varied significantly between plot level land use types but was relatively consistent across the recall period. Rice as an annual monocrop was associated with smaller plots (approximately 0.2 ha), while sugarcane, as an annual monocrop, was associated with larger plots (approximately 1.8 ha). Tree-based perennial plot level land use types such as cashew, coffee and their respective combinations in mixed cropping systems were associated with plots with an average size of 0.6 ha. All plots surveyed without agricultural use in 2007 were replaced by cashew-based land use (Figure. 2.2, A). Likewise, all plots with forages progressively shifted towards fruit trees. Transition rates of 100 percent were also observed from cassava-maize to rice-cassava. The inclusion of pepper in diverse tree-based land uses (plots including cashew, coffee and avocado) was observed in 50 percent of the cases, while coffee-based systems (intercropping with cashew, pepper or both) were stable. Cashew was replaced by rubber in a few instances, while 5 percent of plots returned to fallow land from cassava-maize.

2.4.2.2 Site XK

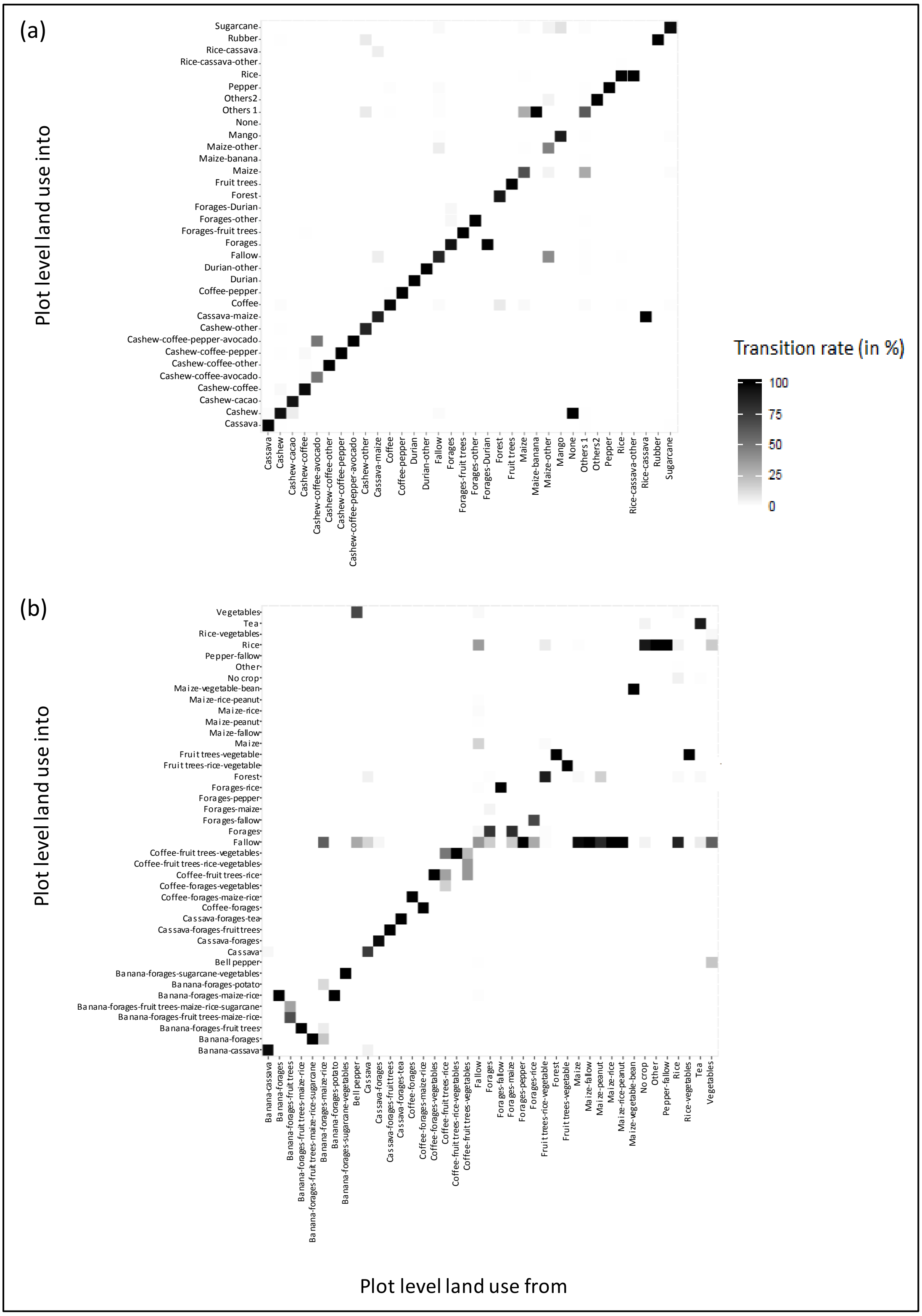
In site XK, 41 different land use types were captured over the recall period. Fallow land remained relatively constant, contributing approximately 60 percent of the total surveyed cropping area in the dry seasons (Figure 2.1, F). During the wet seasons, the area under fallow was minimal, as many plots were used for rice or maize. Rice area constituted an average of 34 percent of the total surveyed area between 2007 and 2016, and maize areas increased from 15 percent in the wet season of 2007 to 21 percent in the wet season of 2016. Interestingly, the area under fallow in the wet seasons decreased by 15 percent between 2007 and 2016. A small proportion of these fallow areas were improved with planted forages, particularly from 2011 with an increase of 3 percent (Figure 2.1, G). At the same time, forage plots indicated a considerable increase in area during the wet seasons, from 3 percent proportional contribution to total surveyed cropping area in the 2007 wet season, to 9 percent in the 2016 wet season (Figure 2.1, H). Tea emerged in 2008, and showed a strong increase in total proportional contribution to the surveyed cropping area, contributing to 10 percent of the total surveyed cropping area in 2016 (Figure 2.1, I). The contribution of cassava as a monocrop increased similarly by 3 percent between 2007 and 2016 (Figure 2.1, J), similar to site CH. Compared to site CH, less variation in average plot size was observed between the major plot-level land use types in the dry season of 2016, and across the recall period, indicating consistently diverse land use types in site XK. Seasonality, however, was evident, specifically with land use types such as forage-fallow. Land use transition was much more dynamic in site XK compared to site CH (Figure 2.2, B). High transition rates to fallow were observed, with 90 percent from maize plots, 80 percent from maize-peanut plots, and 70 percent from rice plots, highlighting seasonal rotations. Most of the remaining maize-peanut plots returned to forage. The remainder of the rice plots were converted to vegetable production or back to forage, while forage is then replaced in 10 percent of the cases by rice. Maize and rice, also intercropped with banana and forages, disappeared from the rotation in 5 percent to 20 percent of the cases and were replaced by potato or fruit trees. Upland rice also disappeared progressively from home gardens, as 80 percent of the fruit-rice-vegetables land was converted into fruit-vegetable plots. Likewise, 80 percent of the pepper plots were replaced by vegetable production. Few cassava plots transitioned into forest, fallow, or cassava-banana intercropping.

Figure 2.1: Temporal changes for the top five land use types



Land use types recorded by the survey (dry season of 2016), i.e. (A) Cashew-Coffee, (B) Cashew, (C) Coffee, (D) Sugarcane, (E) Cassava in CH, and (F) Fallow, (G) Forages-Fall

Figure 2.2: Transition rate between plot level land use types from 2007 to 2016

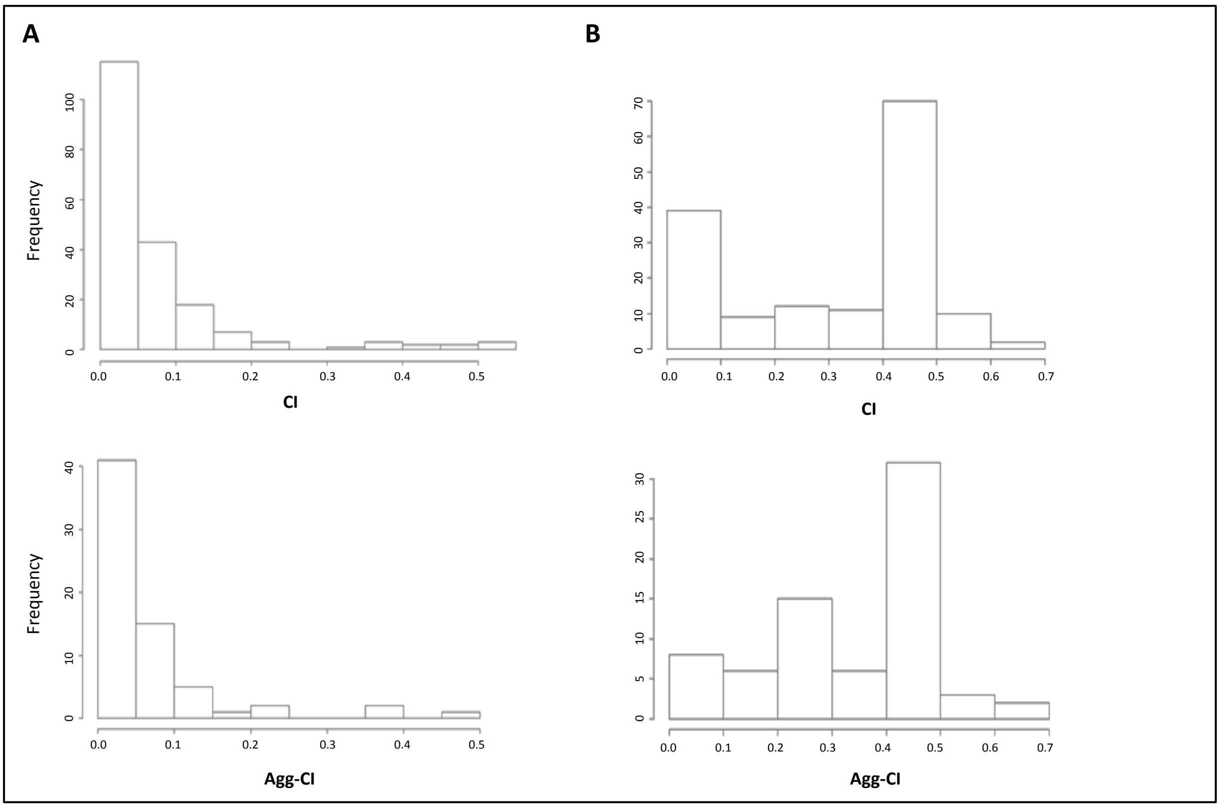


Panel (a) represents the CH, and (b) XK. Minor plot level land use types that were not present in the Pre-determined list of land use types (i.e. either the 34 and 41 for site CH and XK respectively) were coded as Other_1, while plot level land use types that were inter-cropped with Other_1, and were not in the predetermined list of land use types were coded as Other_2.

2.4.3 Complexity Index

In site CH, both CI and agg-CI display a left-skewed distribution pattern, with several plots (n = 115) and households (n = 24) showing a CI of 0 (Figure 2.3, A). In site XK, 46 percent of the plots and 44 percent of the households score between 0.4 and 0.5 (Figure 2.3, B). The highest CI are 0.51 (Figure 2.3, A) and 0.65 (Figure 2.3, B) for site CH and site XK, respectively. The highest agg- CI scores are 0.47 and 0.65 in site CH (Figure 2.3, A) and site XK (Figure 2.3, B), respectively.

Figure 2.3: Distribution of the plot level and household level CI



Panel (A) represents CH, and panel (B) XK. CI stands for CI per plot and Agg-CI for aggregated CI per household.

Pearson chi-square test results for site CH, and Kruskal-Wallis rank sum test for site XK show that although each site consists of several villages, inter-village differences in CI and agg-CI are insignificant. At site CH, most of the plots were tilled, whether showing transition (88 percent) or not (90 percent; supplementary materials, Table A 3). Topography was also similar between transitioning and non-transition plots, with 45 percent and 55 percent of land area, respectively, consisting of flat lands, and 15 percent and 12 percent of land area, respectively, consisting of steep slopes. However, the final use of agricultural production differed, with 20 percent of the non-transitioning plots and only 5 percent of the transitioning plots designated for home consumption. Ninety one percent of the transitioning households were from the Kinh ethnic majority group, which does not originate from the CH, whereas 17 percent of non-transitioning households identified as ethnic minorities (Table A 3). Most household heads had a secondary and post-secondary level of education, both for transitioning (77 percent) and non-transitioning (83 percent) households. Household size and total number of plots per household were higher on average for transitioning households (Table A 3). However, household welfare appeared higher for non-transitioning households, as indicated by better food availability, income, and Progress out of Poverty scores (Table A 4). At site XK, transitioning plots were tilled more, were flatter, and were managed using higher levels of fertilizers and irrigation than non-transitioning plots (Table A 5). These plots were also more often rented than collectively owned or owned by the household. Households from the Lao ethnic majority, as well as households with post-secondary education (Table A 6), tended to transition more than households from the Hmong minority. A relatively strong positive relationship was observed between land use complexity and crop and livestock diversity, while a negative relationship was observed with the total number of fields, household size, and total area cultivated (Table A 6). In terms of performance, a positive relationship was observed between land use complexity and the PPI, food availability and total energy available (Table A 7).

2.4.4 Relationship between CI and plot level characteristics

The best-fit model for site CH, based on both stepwise variable selection and single term deletion analysis, revealed that only production for the market was associated with CI (Table 2.3; Table A 8). More plots dedicated for home consumption did not transition. This was different for plots dedicated to market production, with 92 plots showing no transition and 78 exhibiting transition. The odds ratio calculation revealed that a plot dedicated for market production was four times more likely to display transition than a plot dedicated to home consumption. In site XK, the best-fit model revealed that CI was significantly associated with degree of tillage, irrigation use, agro-chemical use, planting material and slope (Table 2.4). Plots on flat land or gentle slopes that were managed with tillage and irrigation systems had higher CI scores compared to rainfed plots on steep slopes that are not tilled. Fertilized plots had significantly higher CI scores compared to untreated plots or plots receiving pesticides only. Finally, relatively lower CI scores were observed for plots growing planting material that was obtained through subsidies, in comparison to plots with owned or gifted planting material.

Table 2.3: Logistic regression between CI, plot- ,household- and performance level variables in CH

Response Variable[1]	Explanatory variables	Estimate	Standard Error	t-value	Odds ratio	2.5% CI	97.5% CI	p-value	Cox and Snell R^2	Nagelkerke/ Cragg-Uhler R^2
CI (n=197)	Plot level variables								0.071	0.095***
	Final use for sale	1.58	0.56	2.81	4.87	1.78	17.15	<0.05		
Agg-CI (n=67)	Household level variables								0.206	0.285***
	Number of fields	0.69	1.18	-2.47	0.05	0.03	0.45	<0.05		
	Household size	0.64	0.33	1.91	1.89	1.04	3.96	<0.1		
	Performance variables								0.064	0.088***
	Food availability	-6.5	3.34	-1.94	0.001	1.16E-06	0.72	0.05		

[1]Within-site variation was not significant (Pearson's chi-squared test p-value < 0.0001 for both CI and agg-CI

2.4.5 Relationship between agg-CI and household-level characteristics

At site CH, both the stepwise variable selection approach and single term deletion analysis showed that agg-CI was significantly associated with the total number of plots owned by the household (Table 2.3, Table A 9). Odds ratio analysis revealed that households with more plots and with a larger household size were twice as likely to show transition. While the stepwise variable selection approach showed a trend in the relationship between household size and agg-CI, single term deletion analysis revealed significant association between the two, wherein households with higher male adult equivalence have higher mean agg-CI values (Table 2.3, Table A 9). The converse was true at site XK where the relationship between agg-CI and the number of plots owned was negative (Table 2.4).

2.4.6 Relationship between agg-CI and system performance

Models fitted with both stepwise variable selection and single term deletion analysis revealed that only food availability had a significant impact on agg-CI in site CH (Table 2.3, Table A 10). Households with higher food availability are half as likely to show transition, compared to receiving an agg-CI score equal to 0 (Table 2.3). Households with higher food availability are also less market-oriented. The selected model for site XK showed contrasting characteristics: agg-CI was not associated with food availability and a significant positive relationship between agg-CI and the PPI was apparent (Table 2.4).

Table 2.4: Generalized additive regression between CI, plot- ,household- and performance level variables in XK

Response variable[1]	Explanatory variables	Estimate	Standard Error	t-value	p-value	Cox and Snell R2	Nagelkerke/ Gragg-Uhler R^2
CI (n=153)	**Plot level variables**					-42.58	0.001
	Soil tillage_Yes	3.40E-01	6.10E-06	55209.62	<0.05		
	Input agrochemicals_No	3.20E-05	5.20E-06	6.13	<0.05		
	Input agrochemicals_Pesticides	-9.20E-03	1.20E-05	-755.41	<0.05		
	Irrigation_rainfed	8.40E-05	4.00E-06	20.68	<0.05		
	Slope_Modest	-1.00E-04	-3.30E-06	-31.84	<0.05		
	Slope_Strong	-8.70E-03	8.10E-06	-1085.33	<0.05		
	Planting material_combination	1.50E-04	7.40E-06	20.54	<0.05		
	Planting material_exchange	-8.40E-02	1.30E-05	-6328.11	<0.05		
	Planting_material_Gift	4.00E-01	1.30E-05	29774.98	<0.05		
	Planting material_own	1.10E-04	6.90E-06	15.32	<0.05		
	Planting material_subsidized	-4.30E-01	1.20E-05	-35408.85	<0.05		
	Transport_Motorbike	5.94E-02	1.79E-05	3325.473	<0.05		
	Transport_Tractor	8.45E-02	1.26E-05	6720.49	<0.05		
	Transport_Walking	8.44E-02	1.21E-05	6989.305	<0.05		
Agg-CI (n=72)	**Household level variables**					38.36	0.31
	Number of fields	-0.08	0.04	-2.08	<0.05		
	Crop diversity	0.01	0.01	1.35	0.17		
	Performance variables					-37.47	0.3
	Progress out of poverty index	0.46	0.03	12.34	<0.05		
	Farm income	0.02	0.01	2.62	0.2		
	Food availability	-0.01	0.02	1.27	0.55		
	Food self-sufficiency	-0.01	0.01	-0.59	0.87		
	Total energy available	0.01	0.03	-0.16	0.6		

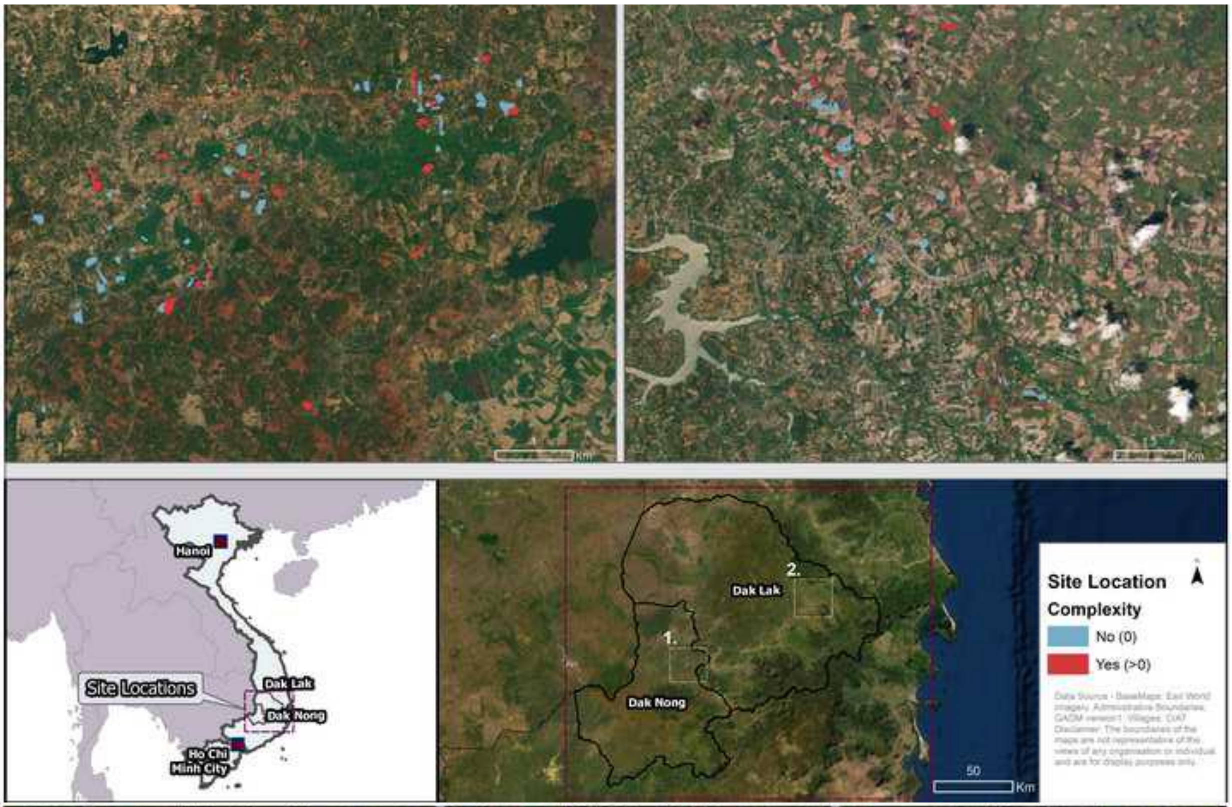
[1]Within-site variation was not significant (Kruskal-Wallis rank sum test p-value < 0.001 for both CI and agg-CI).

2.4.7 Spatial distribution of CI

Satellite views of the two sites show contrasting features, with a rather flat topography and crowded land use in site CH, and a hilly landscape with still large forest patches in site XK (Figure 2.4). In both districts of site CH, around 40 percent of the plots showed transition (43 percent in Dak Lak and 37 percent in Dak Nong). Their spatial distribution does not seem to follow any pattern. In site XK, the majority of plots have a low (40 percent) or medium (59 percent plots) complexity with only 1 percent of the plots falling in the high complexity category. Low complexity plots tend to be more on forested hill slopes and slightly more remote areas, while medium complexity plots prevail in the valley lowland area, especially for the third group. Spatial point analysis revealed opposing patterns between the two sites (Figure 2.5). In site CH, non-transitioning plots were homogenously distributed with a mean pairwise distance of 0.4 km, while transitioning plots had a relatively larger spatial spread around two groups of similar size, peaking at 0.36 km and 0.45 km. The pattern was more complex in site XK, with heterogeneous plot distribution for the three categories. Low- and medium-complexity plots (i.e. plots with $0 \leq CI \leq 0.3$ or $0.3 \leq CI \leq 0.6$) had a more heterogeneous spatial distribution than high-complexity plots (i.e. $0.6 \leq CI$). They were distributed in two groups, one important and compact group with mean average pairwise distances of <0.2 km, and one smaller group, more sparsely distributed, with mean pairwise distance of more than 0.8 km. The high complexity plots had mean pairwise distances of 0.4 and 0.7 km. Therefore, while low-CI plots are homogenously distributed in site CH, they show heterogeneous distribution in site XK.

Figure 2.4: Spatial distribution of the CI in (A) CH and (B) XK

A.



B.

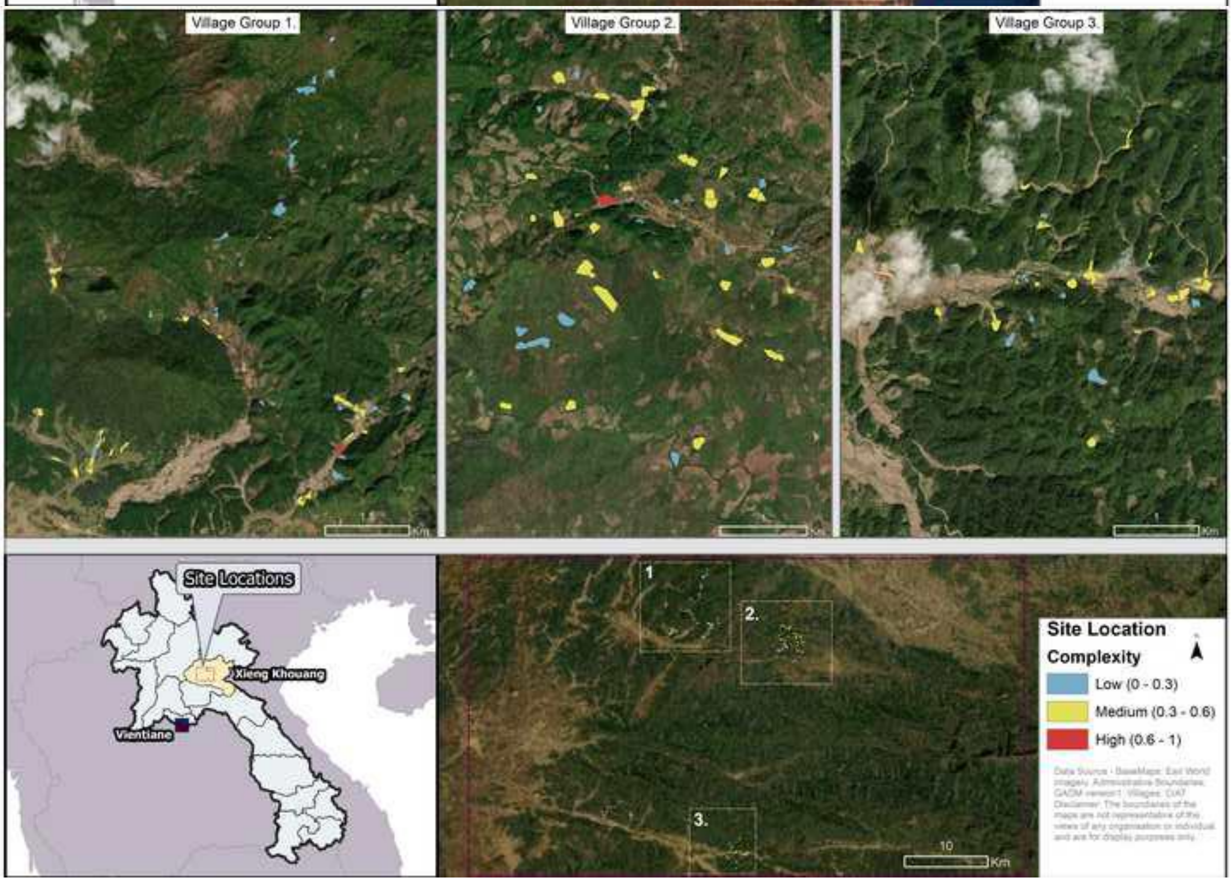
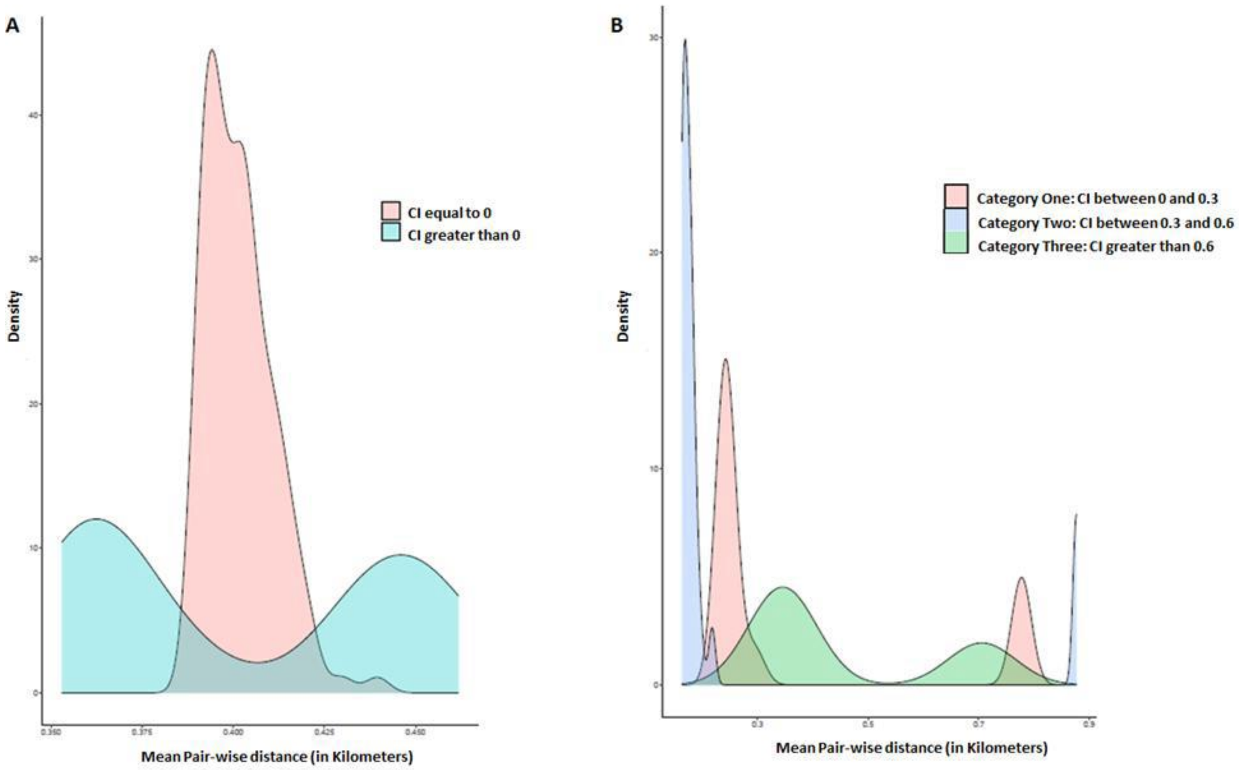


Figure 2.5: Density histograms of the mean pairwise distance between plots for different CI categories



Panel (A) represents CH, and panel (B) XK. The categories were produced for each site by subjecting CI to Jenks natural break classification method.

2.5 Discussion

2.5.1 Evaluation of CI as a metric for land use complexity

The application of sequence data analysis to agriculture proved to be an efficient and useful method to compress and quantify land use complexity. This compression results into a single metric, based on the number of transitions occurring on a single plot over time and the longitudinal entropy. Although, in consequence, the transformation impedes the study of spatial–temporal land use patterns, it allows making linkages with indicators that are typically not captured over time. CI calculation is simple, with a low level of parameter uncertainty, and can be applied at a variety of levels. Indeed, the level of definition of transition depends on the user: for example, if the inclusion of hedgerows in a field qualifies for a new land use type, the resulting CI will be very different than if these types of land use are considered equal. This makes CI very flexible in its use and able to capture a wide variety of transition types, but also might restrict its potential for meta-analysis if the level of definition varies significantly between studies. When CI is mapped, it highlights trends at the landscape level. Similarly to the definition of land use, the size of the unit of observation will influence the type and scale of the drivers of change observed. Compared to land use changes observed based on satellite data (Huang et al., 2019; Yang et al., 2020), the level of definition is much more precise, and more appropriate to capture crop sequences and seasonal changes. Developing context-specific CI thresholds associated with system quality indicators would be useful as ecological engineering control to inform decision makers on the pace of agricultural transformation and its environmental impacts. The disadvantage is that CI integrates seasonal information without consideration for the novelty of the crops: rice-fallow rotations are treated equivalently to switching from one crop to a completely new one (from coffee to sugarcane, for example). This has no consequence on the complexity of the system, but care must be taken in not translating this complexity as a measure of diversity, which is best considered by looking at Figure 2.2. CI reflects crop transition at the plot level and agg-CI is useful to evaluate transition at the household level. In this study, performance indicators and household characteristics are defined at one point in time (2017) whereas agg-CI is the aggregate result of 10 years. This assumes that there has been no major alteration in the households and that the present situation comprehensively summarizes all happenings on the farm during the last ten years. This was deemed to be a reasonable assumption for the purposes of this study. Future studies could bring more time variant perspective into household and plot level characteristics if some form of agricultural census data is available, but it would be difficult to do so on a recall base, as farmers might have difficulties remembering some of this information precisely for remote years. Still, some form of aggregation would need to take place for a regression with CI. The use of plot level and agg-CI in this context although is unique, and the models obtained are parsimonious and provide a significantly better fit than the full model. Regression analysis suggests that, in order to obtain robust results, there is a need to calculate CI scores from a larger set of households and their plots, to obtain significantly higher resolution between households, plots and their locations based on the derived CI scores.

2.5.2 Land use complexity and drivers of change

Land use complexity in sites XK and CH contrast starkly. While site XK shows a much more dynamic, heterogeneous and complex land use pattern, site CH has an established and developed landscape. Indeed, as displayed in Figure 2.2, only 15 percent of land uses show significant changes in site CH, while high transition rates from one land use type into another are more frequent in site XK, with 66 percent of the land use types being completely replaced by others during the recall period. Site CH presents a core area under stable tree and shrub cultivation with only small changes in composition, such as the inclusion of cashew plants in coffee plantations. Intensive production systems are not yet a reality for XK farmers, who rely exclusively on rainwater for irrigation. Thus, the typical cropping system in XK follows clear biannual patterns of paddy and maize-based land use in the rainy season to produce for household consumption (Table 2.2), with some additional dry season crops such as vegetables and forages, but at a much smaller scale. Since no farm households in our XK sample have water pumps, dry season production area is limited to 40 percent of the total area. Thus, most of the observed intensification is constrained to the wet season, accompanied by a decrease in rainy season fallow area of 15 percent over the past 10 years. Results reveal that the most prevalent dry season crop is improved forage production (Fig. 2.1, H), highlighting the importance of integrated crop-livestock systems for site XK. Other crops which showed increased uptake by farmers are cassava, maize and peanuts, which are all known to be relatively suitable in dry conditions. The agronomic challenges of these changing systems are readily apparent in this analysis, as all significant plot-level explanatory variables concern agronomy (Table 2.4). Factors associated with reduced land use complexity in XK are the degree of sloping topography, pesticide use, and planting material from sources beyond farmers' control (subsidies or exchanges). In contrast, at site CH, the only plot-level explanatory variable for land use change is market orientation (Table 2.3, "final use for sale"). A plot dedicated for home consumption is four times less likely to show transition than a plot dedicated to market crops, suggesting that CH farm households adapt quickly to local market conditions. For example, pepper production has increased mainly within existing land use types via progressive inclusion in coffee and cashew plantations, reflecting Vietnam's place as one of the world's leading exporters of pepper. However, the land use changes were not reflected in dietary changes: the crops designated to household consumption stayed the same (Ritzema et al., 2019). In CH, the binding constraints to changing systems are the number of plots in the household and availability of family labor (Table 2.3). Thus, if labor and plots are abundant, households can allocate these resources to on-farm experimentation of new mixed crop-tree systems. Households with higher food availability are also those that are less market-oriented: cash crops result in less food energy than staple crops such as tea, pepper and vegetables. These households tend to avoid experimentation and do not follow market trends, as seen in Table 3, where households with higher food availability typically do not implement complex land use sequences. Those who have low food availability, i.e. households that do not manage to obtain much food energy from both on-farm activities and purchased items, also have more incentive to adapt and innovate, especially in a market-driven environment. Households characterized by high complexity were in most cases from the Kinh ethnic majority, which has better access to markets and education. In XK, poverty is much higher, with close to half of the households situated below the national poverty line, compared to only 30 percent in CH (Table 2.2). Poor farmers must allocate resources carefully, illustrated by substantial differences in plot management and experimentation. Indeed, households showing better PPI values had more complex land use patterns (Table 2.4). The transition rate was also strongly affected by seasonal changes via the staple crop: rice. Farmers seek first to secure their consumption with more intensive management strategies such as irrigation and labor-intensive paddy production, while other secondary crops bolster self-sufficiency. Rice production is also a primary focus for government subsidies and support in most SEA countries, e.g. substantial investments in Laos in the development of direct rice seed planting (Xangsayasane, 2018; Laiprakobsup, 2019). In parallel, we see that land area devoted to low-input, low-risk cash crops such as wild tea or forages have been increasing in the area, consistent with other studies that note the risk-averse perspective of Lao farm households (Sagemüller and Musshoff, 2020). Topography is a key driver of complexity in site XK. Tillage, irrigation, and flat lands are characteristics of the lowlands and were associated with high transition rates (Table 2.4), a finding further supported by CI spatial distribution (Fig. 2.4). Transitioning households are mainly from the lowlands-based Lao ethnic majority while the Hmong minority occupies the highlands. In the relatively flat plateau of site CH, topography is less relevant whereas national policies have greatly influenced land use over the past decades, as lowland people were encouraged to colonize the area and invest in coffee production (Doutriaux et al., 2008). Market prices also incentivize land use change, when farmers are well-informed of price changes through strong market connectivity. However, it is ultimately the farmer's own capacity for change that will determine if a change occurs, and this therefore drives the arising pattern of complexity in land use. Innovation capacity and likelihood to take risks are further factors that should be explored in future studies.

2.5.3 Agricultural system transformation in Southeast Asia

The historical and political contexts in Laos and Vietnam are very different (Ritzema et al., 2019), and so are the drivers of land use change. In the CH site particularly, falling coffee prices in the last decade have induced smallholders to partially reconvert coffee or pepper farms into more diversified livestock-crop-tree-fish systems (D'Haeze et al., 2005). Intensification in Laos in the last decade is due to other factors, including land use regulations introduced by the government to reduce forest clearings in the uplands to protect natural resources from overly-intensive short-cycle shifting cultivation practices. These regulations, in tandem with recent demographic growth (Bouahom et al., 2004), placed increased pressure on available land for cultivation (Lestrelin and Giordano, 2007; Lestrelin, 2010). Site XK is also relatively isolated in comparison to other areas of Laos, with less access to markets and education than the lowlands. Indeed, while the Lao lowlands are further along the 'transition gradient' towards market-oriented and intensive farming, the uplands are still in an early stage of transition and face numerous structural constraints, such as low access to high-quality seeds, insecure land tenure and limited access to advisory services, irrigation, finance and markets (Thongmanivong and Fujita, 2006; Heinimann et al., 2013; Southavilay et al., 2013; Hirota et al., 2014; Castella et al., 2018). Land use intensity and structural complexity of landscapes are separate landscape level factors, at small spatial scales (Persson et al., 2010).

This study identified plots that were used to cultivate a combination of cash crops and subsistence crops, and others that were dedicated entirely to commercial production. Thus, both subsistence, home-oriented and commercial, market-oriented farming systems can coexist within the same farm. Most farms oscillate along a transition pathway (from subsistence to market-oriented) rather than following a linear path towards one end of the spectrum, as the regional narrative suggests. Hence, land use change is not homogenous and irreversible, and it does not follow a consistent progression, in accordance with the findings of Ritzema et al. (2019). These agrarian transitions are not ubiquitous in both northern Laos and in the CH. There were indeed several farmers and plots in the study sample that have not yet made the transition from subsistence to cash crop production. In addition, some farmers previously cultivated a combination of both food and cash crops, while others dedicated all plots entirely to commercial production. Therefore, the current landscape of agrarian systems across both these sites, and in general across SEA, is diverse, and both traditional and intensive farming systems co-exist simultaneously. At the plot level, transition patterns contrast to those at the household level: this study shows that no transition is expected for home consumption plots in Site CH (Table 2.3), and only minimal transition for plots on sloping land in site XK (Table 2.4). Sloping plots generally have poorer soil fertility, and investment in rice terraces is only economically attractive if water is available for irrigation (Castella, 2012). Alternatively, less water-demanding crops like maize and cassava can be used in the uplands, assuming that some soil conservation measures, such as contour farming, are simultaneously implemented (Castella, 2012). But such measures require substantial investment in agricultural extension. Therefore, whenever innovative land use options for sustainable intensification are discussed with farmers, chances for adoption are much higher when the plots under consideration are not allocated to staple crops (unless it complements rather than replaces them) and are not located on steep slopes. Innovative land use options would be successful on slopes only if complemented by appropriate soil and water conservation measures that are acceptable from a capital and labor investment point of view. Although the transition from extensive to more intensive forms of agriculture seems to have overwhelmingly negative impact for farmers' livelihoods and the environment (Dressler et al., 2017), there may be little choice in view of increasing climate, population and market pressures. Declining farm sizes and increasing wage rates are further constraints for production efficiency and will need to be addressed with strong policies to ensure that agricultural production in SEA keeps a comparative economic advantage to other regions (Fan and Chan-Kang, 2005; Otsuka et al., 2016). In the long term, it is unclear whether intensification will plateau. In the future, the Lao uplands and the CH may have similar complexities and land use dynamics, or alternatively might evolve separately.

2.6 Highlights and conclusions

This study has described plot-level sequence patterns of seasonal variation of land use in two sites of Laos and Vietnam, characterized by contrasting stages of intensification and market orientation, using CI, a sequence dissimilarity metric. CI allowed compressing historical land use data and quantifying land use complexity in a simple and efficient manner. In the CH, relatively well-educated migrants have increasingly applied intensive agricultural practices in market-oriented intercropped tree-based systems during the 2006–2017 period, accompanied by a decrease in monoculture production and fallow land. These 'highly transitioned' systems result in relatively high income and food availability. In XK, self-sufficient farmers are in the initial stages of reducing forest and fallow land to increase their rain-fed low-input staple crop production, with progressive inclusion of vegetable crops and forages as well as several cash crops. Land use dynamics vary strongly between the two sites, with 66 percent of the land use types in site XK being completely replaced by others during the recall periods, compared to only 15 percent in site CH. Associated key drivers of change also differed significantly: while end use of the agricultural products is the main driver behind land use changes at site CH, the relationship is more complex at site XK, with changes associated with topography and management. Households with higher food availability in site CH are less likely to show transition, while in site XK, complexity was significantly correlated with the Progress out of Poverty index. For smallholder farming systems already showing high levels of intensification, innovative land use options have a higher likelihood of adoption when these include market-oriented crops that are not labor-intensive and that do not replace staple crops. For low-input subsistence farming systems, these options should target low risk plots first, ensuring seed availability and avoiding sloping lands and areas where intensive management can be difficult. Multidisciplinary studies remain necessary to assess the impact of innovative sustainable intensification options on system performance and environmental sustainability, before policies are enacted to support their dissemination in SEA smallholder agricultural systems. Context-specific CI thresholds associated with system quality indicators could support this by informing decision-makers on the pace of agricultural transformation and its environmental impacts.

References

Alberti, G., 2020. GmAMisc: 'Gianmarco Alberti' Miscellaneous. R package version 1 (1), 1. https://CRAN.R-project.org/package=GmAMisc.

Aquino, J., 2018. descr: Descriptive Statistics. R package version 1 (1), 4. https://CRAN. R-project.org/package=descr.

Ashraf, J., Pandey, R., de Jong, W., 2017. Assessment of bio-physical, social and economic drivers for forest transition in Asia-Pacific region. *Forest Poicy Economics*, 76, 35–44.

Bivand, R., Rundel, C., 2020. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-3. https://CRAN.R-project.org/package=rgeos.

Bouahom, B., Douangsavanh, L., Rigg, J., 2004. Building sustainable livelihoods in Laos: untangling farm from non-farm, progress from distress. *Geoforum*, 35, 607–619.

Castella, J.-C., 2012. Agrarian transition and farming system dynamics in the uplands of South-East Asia. The 3rd International Conference on Conservation Agriculture in Southeast Asia. CIRAD, NOMAFSI, University of Queensland, Hanoi, Vietnam.

Castella, J.-C., Sysanhouth, K., Saphangthong, T., Victor, M., Ingalls, M., Lienhard, P., Bartlett, A., Sonethavixay, S., Namvong, S., Vagneron, I., Ferrand, P., 2018. Adding Values to Agriculture: A Vision and Roadmap for Sustainable Development in the Lao Uplands. Lao Uplands Initiative, Vientiane, Laos.

D'Haeze, D., Deckers, J., Raes, D., Phong, T.A., Loi, H.V., 2005. Environmental and socio- economic impacts of institutional reforms on the agricultural sector of Vietnam: Land suitability assessment for Robusta coffee in the Dak Gan region. *Agriculture, Ecosystems & Environment*, 105, 59–76.

Diez, J.R., 2015. Vietnam 30 years after Doi Moi : achievements and challenges. *Zeitschrift Für Wirtschaftsgeographie*, 60, 121-133.

Ditzler, L., Komarek, A.M., Chiang, T.-W., Alvarez, S., Chatterjee, S.A., Timler, C., Raneri, J.E., Carmona, N.E., Kennedy, G., Groot, J.C.J., 2019. A model to examine farm household trade-offs and synergies with an application to smallholders in Vietnam. *Agricultural Systems*, 173, 49–63.

Doutriaux, S., Geisler, C., Shively, G., 2008. Competing for Coffee Space: Development- Induced Displacement in the Central Highlands of Vietnam. *Rural Sociology*, 73, 528–554.

Dressler, W.H., Wilson, D., Clendenning, J., Cramb, R., Keenan, R., Mahanty, S., Bruun, T.B., Mertz, O., Lasco, R.D., 2017. The impact of swidden decline on livelihoods and ecosystem services in Southeast Asia: A review of the evidence from 1990 to 2015. *Ambio*, 46, 291–310.

Epper, C.A., Paul, B., Burra, D., Phengsavanh, P., Ritzema, R., Syfongxay, C., Groot, J.C. J., Six, J., Frossard, E., Oberson, A., Douxchamps, S., 2020. Nutrient flows and intensification options for smallholder farmers of the Lao uplands. *Agricultural Systems*, 177, 102694.

Fan, S., Chan-Kang, C., 2005. Is small beautiful? Farm size, productivity, and poverty in Asian agriculture. *Agricultural Economics*, 32, 135–146.

Fröhlich, H.L., Schreinemachers, P., Stahr, K., Clemens, G. (Eds.), 2013. Sustainable Land Use and Rural Development in Southeast Asia: Innovations and Policies for Mountainous Areas. Springer-Verlag, Berlin Heidelberg.

Gabadinho, A., Ritschard, G., Mueller, N.S., Studer, M., 2011. Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40, 1–37.

Goto, K., Douangngeune, B., 2017. Agricultural modernisation and rural livelihood strategies: the case of rice farming in Laos. *Canadian Journal of Development Studies / Revue canadienne d'etudes du d´eveloppement*, 38, 467–486.

Hammond, J., Fraval, S., van Etten, J., Suchini, J.G., Mercado, L., Pagella, T., Frelat, R., Lannerstad, M., Douxchamps, S., Teufel, N., Valbuena, D., van Wijk, M.T., 2017. The Rural Household Multi-Indicator Survey (RHoMIS) for rapid characterisation of households to inform climate smart agriculture interventions: Description and applications in East Africa and Central America. *Agricultural Systems*, 151, 225–233.

Heinimann, A., Hett, C., Hurni, K., Messerli, P., Epprecht, M., Jørgensen, L., Breu, T., 2013. Socio-Economic Perspectives on Shifting Cultivation Landscapes in Northern Laos. *Human Ecology*, 41, 51–62.

Hepp, C.M., Bech Bruun, T., de Neergaard, A., 2019. Transitioning towards commercial upland agriculture: A comparative study in Northern Lao PDR. NJAS – Wageningen. *Journal of Life Sciences*, 88, 57–65.

Hettig, E., Lay, J., Sipangule, K., 2016. Drivers of Households' Land-Use Decisions: A Critical Review of Micro-Level Studies in Tropical Regions. *Land*, (5), 32.

Hirota, I., Koyama, T., Ingxay, P., 2014. Mountainous Livelihood in Northern Laos: Historical Transition and Current Situation of a Swidden Village. In: Yokoyama, S., Okamoto, K., Takenaka, C., Hirota, I. (Eds.), Integrated Studies of Social and Natural Environmental Transition in Laos. Springer Japan, Tokyo, pp. 39–59.

Hlavac, M., 2014. Stargazer: LaTeX/HTML code and ASCII text for well-formatted regression and summary statistics tables. Harvard University, Cambridge.

Hu, B., Shao, J., Palta, M., 2006. Pseudo-R 2 in logistic regression model. *Statistica Sinica*, 16, 847–860.

Huang, A., Xu, Y., Sun, P., Zhou, G., Liu, C., Lu, L., Xiang, Y., Wang, H., 2019. Land use/ land cover changes and its impact on ecosystem services in ecologically fragile zone: A case study of Zhangjiakou City, Hebei Province, China. *Ecological Indicators*, 104, 604–614.

Kammerbauer, J., Ardon, C., 1999. Land use dynamics and landscape change pattern in a typical watershed in the hillside region of central Honduras. *Agriculture, Ecosystems & Environment*, 75, 93–100.

Klasen, S., Meyer, K.M., Dislich, C., Euler, M., Faust, H., Gatto, M., Hettig, E., Melati, D. N., Jaya, I.N.S., Otten, F., P´erez-Cruzado, C., Steinebach, S., Tarigan, S., Wiegand, K., 2016. Economic and ecological trade-offs of agricultural specialization at different spatial scales. *Ecological Economics*, 122, 111–120.

Laiprakobsup, T., 2019. The policy effect of government assistance on the rice production in Southeast Asia: Comparative case studies of Thailand, Vietnam, and the Philippines. *Development Studies Research*, 6, 1–12.

Lambin, E.F., Geist, H.J., Lepers, E., 2003. Dynamics of Land-Use and Land-Cover Change in Tropical Regions. *Annual Review of Environment and Resources*, 28, 205–241.

Lambin, E.F., Rounsevell, M.D.A., Geist, H.J., 2000. Are agricultural land-use models able to predict changes in land-use intensity? *Agriculture, Ecosystems & Environment*, 82, 321–331.

Lestrelin, G., 2010. Land degradation in the Lao PDR: Discourses and policy. *Land Use Policy*, 27, 424–439.

Lestrelin, G., Giordano, M., 2007. Upland development policy, livelihood change and land degradation: interactions from a Laotian village. *Land Degradation & Development*, 18, 55–76.

Marmolejo-Ramos, F., Tejo, M., Brabec, M., Kuzilek, J., Joksimovic, S., Kovanovic, V., Gonz´alez, J., Ospina, R., 2020. Distributional regression analysis of learning analytics and educational data. In press, DOI: 10.31235/osf.io/r8azq.

Milne, S., 2013. Under the leopard's skin: Land commodification and the dilemmas of Indigenous communal title in upland Cambodia. *Asia Pacific Viewpoint*, 54, 323–339.

Müller, D., Zeller, M., 2002. Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. *Agricultural Economics*, 27, 333–354.

Otsuka, K., Liu, Y., Yamauchi, F., 2016. The future of small farms in Asia. *Development Policy Review*, 34, 441–461.

Persson, A.S., Olsson, O., Rundlöf, M., Smith, H.G., 2010. Land use intensity and landscape complexity—Analysis of landscape characteristics in an agricultural region in Southern Sweden. *Agriculture, Ecosystems & Environment*, 136, 169–176.

Rabosky, D.L., Grundler, M.C., Anderson, C.J., Title, P.O., Shi, J.J., Brown, J.W., Huang, H., Larson, J.G., 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, 5, 701–707.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. *Journal of Applied Statistics*, 54, 507–554.

Ritschard, G., Studer, M., 2018. Sequence Analysis: Where Are We, Where Are We Going? In: Ritschard, G., Studer, M. (Eds.), Sequence Analysis and Related Approaches: Innovative Methods and Applications. Springer International Publishing, Cham, pp. 1–11.

Ritzema, R.S., Douxchamps, S., Fraval, S., Bolliger, A., Hok, L., Phengsavanh, P., Long, C. T.M., Hammond, J., van Wijk, M.T., 2019. Household-level drivers of dietary diversity in transitioning agricultural systems: Evidence from the Greater Mekong Subregion. *Agricultural Systems*, 176, 102657.

Sagemuller, F., Musshof, O., 2020. Effects of Household Shocks on Risk Preferences and Loss Aversion: Evidence from Upland Smallholders of Southeast Asia. *Journal of Development Studies*, https://doi.org/10.1080/00220388.2020.1736280.

Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. *Nature*, 560, 639–643.

Southavilay, B., Nanseki, T., Takeuchi, S., 2013. Analysis on policies and agricultural transition: Challenges in promoting sustainable agriculture in Northern Laos. *Journal of the Faculty of Agriculture, Kyushu University*, 58, 219–223.

Southworth, J., Nagendra, H., Cassidy, L., 2012. Forest transition pathways in Asia – studies from Nepal, India, Thailand, and Cambodia. *Journal of Land Use Science*, 7, 51–65.

Stasinopoulos, D.M., Rigby, R.A., 2008. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, 23.

Thanichanon, P., Schmidt-Vogt, D., Epprecht, M., Heinimann, A., Wiesmann, U., 2019. Balancing cash and food: The impacts of agrarian change on rural land use and wellbeing in Northern Laos. *PLoS ONE*, 13, e0209166.

Thongmanivong, S., Fujita, Y., 2006. Recent Land Use and Livelihood Transitions in Northern Laos. *Mountain Research and Development*, 26 (237–244), 238.

Tomich, T.P., Thomas, D.E., van Noordwijk, M., 2004. Environmental services and land use change in Southeast Asia: from recognition to regulation or reward? *Agriculture, Ecosystems & Environment*, 104, 229–244.

Wickham, H., 2011. Ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3, 180–185.

Wickham, H., 2012. Reshape2: Flexibly reshape data: a reboot of the reshape package. Harvard, Cambridge.

Wickham, H., Francois, R., Henry, L., Mueller, K., 2015. dplyr: A Grammar of Data Manipulation., Vienna.

Xangsayasane, P., 2018. Rice Breeding and Mechanization for Value Addition in Laos PDR. In: Food and Fertilizer Technology Center for the Asian and Pacific Region – Policy Platform (Ed.), Promoting Rice Farmers' Market through value-adding Activities, Kasetsart University, Thailand.

Yang, C., Zhang, C., Li, Q., Liu, H., Gao, W., Shi, T., Liu, X., Wu, G., 2020. Rapid urbanization and policy variation greatly drive ecological quality evolution in Guangdong-Hong Kong-Macau Greater Bay Area of China: A remote sensing perspective. *Ecological Indicators*, 115, 106373.

3 Effects of household shocks on risk preferences and loss aversion: Evidence from upland smallholders of Southeast Asia

Abstract

Avoiding risk in financial decisions is credited to be a key contributor to persistent poverty and poverty traps. In spite of this, the methods used to measure behavior under risk rarely reflect an adequate representation of the lives of smallholders in low income economies. We estimate risk preferences and their determinants by including two key aspects: aversion to losses and exposure to long term risk and vulnerability. We examine risk preferences of 93 smallholders in Cambodia and 91 smallholders in Lao PDR with an incentivized lottery design under the framework of EUT, RDU and CPT. We find that CPT best explains our data, but parameter values vary to those most commonly found in the literature. We report that the experience of household shocks have a significant effect on choice behavior in the loss domain, even when we control for a large set of socio-economic and demographic variables.

This chapter is coauthored by Frederik Sagemüller (FS) and Oliver Mußhoff (OM). The contributions to the article are: FS conceptualized the idea, collected the data, analyzed the data, and wrote the article. OM reviewed the article. This chapter is published in the Journal of Development studies with the following citation: Frederik Sagemüller & Oliver Mußhoff (2020) Effects of Household Shocks on Risk Preferences and Loss Aversion: Evidence from Upland Smallholders of Southeast Asia, The Journal of Development Studies, 56:11, 2061-2078, DOI: 10.1080/00220388.2020.1736280

3.1 Introduction

Avoiding risk is credited to be a key element contributing to persistent poverty in vulnerable populations around the globe (Dercon and Christiaensen, 2011; Mosley and Verschoor, 2005). According to theoretical models that describe mechanisms of persistent poverty, living under adverse and uncertain conditions puts a mental tax on the poor by increasing their sensitivity to risk. Consequently, this decreases the probability that small businesses reach their productive potential (Sandmo, 1971). This rationale is used to explain low agricultural productivity (Rosenzweig and Binswanger, 1993) and technology adoption (Dercon and Christiaensen, 2011) in developing countries, and why poor agricultural households might be locked into a poverty trap (Carter and Barrett, 2006).

Within this framework, the notion of changing risk preferences is acknowledged by state dependent preferences, not by an inherent preference shift (Stigler and Becker, 1977). The power of this assumption lies in the ability to assign causation between changing opportunity sets and choices in comparative statics exercises. Without this assumption, the comparative statics tests would become joint tests of the effect of changing opportunities and changing preferences (Andersen et al., 2008). Thus, while maintaining the assumption of the neo-classical economic model which states that preferences are stationary, we can define the arguments of the utility function as state contingent. This framework allows choices to vary with various states of nature, as long as these states are exogenous to the choices of the agent.

Especially for smallholders in developing countries, the states of nature are unstable and subject to drastic change. In most low income economies, rural populations depend on rain-fed cropping and livestock systems which are prone to climatic shocks and which are further aggravated by climate change (FAO, 2011). Often, markets for insurance fail, social welfare systems are not present and coping strategies are expensive. Thus, household shocks continuously endanger consumption and welfare of rural populations (Dercon, 2004).

Accordingly, the effects of household shocks on smallholders is two-fold: A direct effect by destroying material assets and an indirect effect by changing behavior under risk, thus further aggravating path dependence and poverty traps. Therefore, it is crucial to understand the determinants of risk preferences, as well as the vulnerability of farmers to shocks to effectively design support programs in low income economies.

Although many studies reveal influences of adverse events on risk aversion, no consensus has been reached regarding the shape and the magnitude of effects. Some studies find that adverse events decrease risk aversion (Eckel et al. 2009; Li et al. 2011; Voors et al. 2012), others find that adverse events increase risk aversion (Berg et al. 2009; Cameron and Shah 2015). These inconsistencies are explained most commonly by emotional behavior or behavioral heuristics, which can increase risk seeking after catastrophic events.

Not known are the impacts of cumulative household shocks on loss aversion, even though this aspect of risk aversion is credited to be fundamental to smallholders in developing countries (Abdellaoui et al. 2008; Nguyen and Leung 2009; Yesuf and Bluffstone 2009; Fafchamps 2010). Loss aversion describes the tendency of individuals to interpret outcomes as gains and losses relative to a reference point, where individuals are more sensitive to losses than to equally sized gains. This behavioral insight is of great relevance to smallholder farmers, since many farm management decisions include the possibility of net losses. Several empirical studies have found evidence of loss aversion (Tversky and Kahneman 1992; Tanaka, Camerer, and Nguyen 2010; Nguyen 2011; Liu 2013; Liebenehm and Waibel 2014) and loss aversion can also explain a variety of field data (Liu and Huang 2013; Holden and Quiggin 2016; Shimamoto, Yamada, and Wakano 2017). We argue that household shocks disrupt congruence of actual outcomes and expected outcomes, which in turn increases loss aversion. Thus, aversion to losses is tightly linked to the reinforcing nature of poverty.

We chose smallholders in the rural highlands of Cambodia and Lao PDR as the target group for our research because of their high exposure to uninsured shocks. Smallholders in our research area experience an array of agricultural and climatic shocks such as droughts and floods, as well as demographic shocks such as diseases and accidents. In both countries, institutional support is lacking, which drastically decreases opportunities for smallholders to buffer such shocks. Therefore, the risk exposure of the population is very high and impacts of household shocks are pertinent to the target group.

In the light of these issues, it is the aim of this study to determine risk preference parameters according to the most prominent models which are EUT, RDU and CPT. This enables us to estimate parameters for loss aversion and to compare competing models. To identify the model that best explains the choices of smallholders, we conduct nested and non-nested hypothesis tests. In a second step, we measure the effect of shocks on risk preference parameters by including them in a variety of specifications. In our analysis we control for important socio-economic determinants of risk preferences. We are the first to (1) correlate a comprehensive list of past household shocks with loss aversion in a developing country and (2) experimentally measure risk preference parameters in Cambodia and Lao PDR and to compare models of CPT, RDU and EUT.

The remainder of the paper is structured as follows. In section 3.2 we review the literature on how shocks and adverse events impact risk preference. In section 3.3 we give an overview on the area of research, the data collection process, as well as the conceptual framework and the estimation strategy of risk preference parameters. Section 3.4 presents the results and gives a discussion on key findings. Section 3.5 gives recommendations for future studies and Section 3.6 concludes the results with brief policy implications.

3.2 Literature review

3.2.1 Household shocks and risk preferences

A small but growing body of literature examines the direct effect of independent household shocks on risk preferences. It is commonly hypothesized that household shocks trigger a change in behavior, which can be measured as a change of an individual's utility function.

Empirical support for an increase in risk aversion after a shock comes from Said et al. (2015). The researchers investigate the impact of floods on risk preferences of villagers in rural Pakistan and find that people, who live in affected villages exhibit significantly higher risk aversion. Similarly, van Berg et al. (2009) use experiments to elicit impacts of natural disasters on risk preferences of farmers in Nicaragua and Peru. They conclude that the experience of disasters makes people more risk averse. Cameron and Shah (2015) investigate risk preferences of a rural population in Indonesia. They use 50-50 gambles to calculate risk preferences and find that people who experience a flood or an earthquake exhibit more risk aversion. Furthermore, affected individuals believe that the probability for a future shock is higher and the expected impact of such shocks to be more severe. Thus, individuals get more risk averse after a negative event. This view is backed by empiric studies which document that macro-economically unfavorable conditions increase risk aversion (Guiso and Paiella, 2008; Malmendier and Nagel, 2011).

Contrary to this, empirical support for an increase in risk seeking behavior after an adverse event is presented by Voors et al. (2012). The researchers use a series of field experiments to measure the impact of exposure to violent conflict on risk preferences and find that the affected population exhibit more risk-seeking behavior. Eckel et al. (2009) examine risk preferences of people affected by a storm in the United States. They find that evacuees display risk loving behavior compared to the control groups which display risk-averse behavior. They attribute this effect to short term stress that is caused by the disaster and they show that the effects dissipate one year after the event. The assertion of a short term increase in risk seeking after an adverse event is supported by studies in psychological sciences. The effects of emotional arousal on risk preferences (Mellers et al. 1997; Kaufman 1999), as well as in the concepts of the availability and representativeness heuristic (Tversky and Kahneman 1974) describe that individuals might be biased by events that are easily recallable, such as recent, frequent and salient events.

Instead of focusing on the impacts of irregular and catastrophic events, a concept closer to our measurement of household shocks is risk vulnerability. It describes that background risk (non-diversifiable and non-insurable risk) may make individuals less tolerant towards additional risks (Guiso and Paiella 2008; Gollier and Pratt 2016). Applications have shown that this effect holds true for specifications under EUT and RDU (Harrison et al., 2007). This is particularly the case for marginalized smallholders in developing economies, whose livelihood options are limited by the aggregation of frequent small and medium scale household shocks. This form of long term risk and vulnerability is a central feature of their lives with strong implications for their behavior under risk. A study that examines a broader set of household shocks is Gloede et al. (2013). The researchers correlate impacts of a broad range of household shocks with a survey item which measures a self-assessment of risk attitudes of rural dwellers in Thailand and Vietnam. The study documents that household shocks can increase risk aversion, depending on the local context. Menkhoff and Sakha (2016) explore the effects of macroeconomic conditions and microeconomic shocks on experimentally measured risk preferences with a panel data set in rural Thailand. They find that both, micro- and macroeconomic shocks lead to higher risk aversion among the sampled population. Nielsen et al. (2013) test six hypothetical risk elicitation methods and one experimental method to assess risk preferences in Vietnam. They examine determinants for risk preferences and include variables measuring monetary losses due to covariate and idiosyncratic shocks in their regression models. Idiosyncratic shocks are positively correlated and significant in four methods, but three further methods have a negative correlation to risk preferences. The impact of covariate shocks is only significant in one method.

3.2.2 Household shocks and loss aversion

One major drawback of the above cited literature is that indecisive results may be caused by an omission of important behavioral insights or in the characterization of risk preferences of poor populations. Specifically, not known are the impacts of cumulative household shocks on loss aversion, even though this aspect of risk aversion is credited to be fundamental to smallholders in developing countries (Abdellaoui et al. 2008; Nguyen and Leung 2009; Yesuf and Bluffstone 2009; Fafchamps 2010). The most common approach to estimate risk aversion, is to calculate one parameter that displays the concavity of the utility function under the EUT framework, or to estimate risk attitude with a hypothetical survey item (Dohmen et al., 2011). Due to violations of EUT, which are confirmed in lab and field studies, it is widely suggested to replace EUT with a framework which includes risk aversion in the loss domain, reference depended utility, and nonlinear weighting of probabilities of outcomes (Harless and Camerer 1994; Starmer 2000; Humphrey and Verschoor 2004a, 2004b). The most prominent theory which contains these behavioral insights is CPT developed by Tversky and Kahneman (1992). Other models which include behavioral insights are RDU developed by (Quiggin, 1982) and Disappointment Aversion developed by (Gul, 1991). See Starmer (2000) or Harless and Camerer (1994) for a thorough review for non-expected utility models.

Several studies estimate risk preferences in a developing country context and support the view that smallholders display a value function according to CPT (e.g. Tanaka et al. 2010; Liu 2013). Overall, there is emergent scientific proof that farmers in developing countries display loss aversion. Several recent studies underline this importance, by using aversion to losses as a predictor of technology adoption (Liu 2013; Liu and Huang 2013; Holden and Quiggin

2016; Shimamoto et al., 2017). One major drawback of this literature is that these studies assume only one parameter σ for value function curvature, instead of separating this parameter into its theoretical components laid out by Tversky and Kahnemann: parameter α for value function curvature in the gain domain and parameter β for value function curvature in the loss domain. Apart from accounting for loss aversion through parameter λ, a decrease in the marginal utility at a faster pace for α than for β can be accounted for as loss aversion. Thus, the picture of loss aversion in the above cited literature is not complete. Furthermore, there is no empirical evidence from this literature that $\lambda>1$ is derived from utility loss aversion as opposed to probability loss aversion (Harrison and Ross, 2017).

A particular advantage to include loss aversion in the estimation of risk preferences of smallholders is that farm management decisions include the possibilities of economic losses because adequate insurance mechanisms are not available. Household consumption might depend on farming, which puts an additional weight on farm management decisions. Indeed, there is evidence that individuals have target income levels, which serve as a reference point to distinguish between losses and gains (Camerer et al. 1997; Heath et al. 1999; Fehr and Goette 2007). Thus, households that are exposed to downside risks are forced to prevent further losses to protect their target income level. This mechanism is presented in Dercon and Christiaensen (2011), who show that low consumption outcomes caused by harvest failures discourage the application of fertilizer in Ethiopia. This is also exemplified in the work of Emerick et al. (2016), who show that reducing downside risk through tolerant varieties causes farmers to re-optimize production along several dimensions, which leads to a crowding in of modern inputs.

Thus, the decision to apply inputs is not dependent on probabilities and profits in the gain domain alone, but also in the loss domain. Kőszegi and Rabin (2007) develop a framework under which utility is derived from differences between consumption and expected consumption, where the utility function exhibits loss aversion. Therefore, we argue that household shocks can disrupt expected consumption and lead to an increase in loss aversion. A further burden for resource poor households comes from the fact that necessities and temptations are concave to income. Therefore, poor households are disproportionally taxed by household necessities and temptations (Mullainathan, 2007; Wicker, Hamman, Hagen, Reed, & Wiehe, 1995). This case is presented in Wicker et al. (1995), who show that loss aversion is greater when a larger proportion of household resources is designated for necessities. If a household loses income due to an adverse event, the proportion of household income which is absorbed to satisfy necessities is even greater, thus increasing loss aversion. Accordingly, household consumption and farm input purchase compete for resources, which deters smallholders in their ability to take advantage of profitable inputs. Thus, poverty traps can emerge through the mechanism of loss aversion.

3.3 Methodology, data collection and research area

3.3.1 Conceptual framework and estimation strategy

In this study we apply 3 estimations of utility functions: EUT, RDU and CPT (Tversky and Kahneman, 1992) to characterize individual risk preferences of farmers through observed choices in binary lotteries. The approach is adapted from appendix E of Harrison and Rutström (2008). The simplest specification is EUT. The estimation approach is laid out with EUT as an example and stays the same under RDU/CPT specifications, with the only changes in adding probability weighting and ranking in RDU and probability weighting, sign dependent utility and loss aversion in CPT.

Generally, we assume that utility over income is defined by

$$U\ (x) = \ x^{\alpha} \tag{3.1}$$

Where x is the lottery price and α is the parameter to be estimated. We assume that lottery outcomes are immediately integrated into the stream of consumption (Harrison and Rutström, 2009). The probabilities for each outcome i are $p(x_i)$. The expected utility is the probability weighted utility of each outcome *i* in lottery *j*:

$$EU^j = \sum_{i=1,I} (p(x_i) \cdot U(s + x_i)) \tag{3.2}$$

To elicit the risk preference parameters, structural estimation after Harrison and Rutström (2008) is applied. The utility function for participant k in decision task j under EUT can be written as:

$$EU_k^j \left(X_k; Z^j\right) + \varepsilon_k^j \tag{3.3 }$$

where $EU_k^j \left(X_k; Z^j\right)$ is the utility function under EUT. X_k are individual characteristics of the participant. Z^j are the characteristics of the decision task and ε_k^j is the error term. The present utility streams for each choice can be expressed as the difference between the utilities of lottery A and B, which can be expressed as the latent choice index ΔEU_k:

$$\Delta EU_k = (EU_k^A - EU_k^B)/\mu \tag{3.4}$$

where μ is a structural noise parameter introduced by Fechner and popularized by Hey and Orme (1994). The latent choice index ΔU_k is linked to the observed choices made by the participants in the experiment through a normal cumulative distribution function $\Phi(\Delta U_k)$. The conditional log-likelihood derived from each of 35 decision tasks is written as follows:

$$ln\, L_k(\alpha; X_k; Z^j; y_k^j) = \sum_{i=1,I} [\ln \Phi(\, \Delta EU_k) |\, y_k^j = A] + [ln(\, \Phi \Delta EU_k) |\, y_k^j = B)] \qquad (3.5)$$

The log-likelihood depends on the EUT parameter α and on observed individual and household characteristics of the participant $k(X_k)$, the characteristics of choice scenario $j(Z^j)$ and the decision made by participant k in choice scenario (y_k^j). The implementation was done in STATA through maximum likelihood estimation, with clustered standard errors by farmer and a Fechner error term.

3.3.2 Extending the specification with decision weights under RDU

To extend the above specification to RDU (Quiggin, 1982) we follow appendix E of Harrison and Rutström (2008). The decision weights of lottery prices have to be calculated and formula (2.2) is replaced by

$$RDU_j = \sum_{i=1,I} \, (w_i \cdot U(x_i)) \qquad (3.6)$$

where

$$w_i = \omega(p_i + \cdots + p_I) - \omega(p_{i+1} + \cdots + p_I) \tag{3.7}$$

for i=1,…,I-1, and

$$w_i = \omega(p_i) \tag{3.8}$$

for i ranking outcomes from worst to best and $\omega(.)$ being the probability weighting function. The function we apply here is the single parameter weighting function $w(p)$ according to Prelec (1998):

$$w(p) = \frac{1}{exp(ln\left(\frac{1}{p}\right))^{\alpha}} \tag{3.9}$$

The parameter α is a proxy for probability weighting (α>0). If α=1, probabilities are weighted linearly and the specification collapses to EUT.

3.3.3 Extending the specification of risk attitudes with loss aversion and sign dependent utility under CPT

Tversky and Kahneman (1992) assumed that the two part power function assigns different values for gains and losses:

$$CPT_j = \begin{cases} x^{\alpha}, & if\ x \geq 0 \\ -\lambda(-x)^{\beta}, & if\ x < 0 \end{cases} \tag{3.10}$$

The risk aversion parameters are α and β. In case α<β (if α<1), utility is concave for gains and convex for losses. The parameter λ is the coefficient of loss aversion. This amounts to the implicit scaling convention that u(1)=−u(−1)=1, implying utility loss aversion λ=(-U(-1))/(U(1)). Hence, it is the empirical strategy to evaluate estimates of α and β and then infer λ by evaluating the implied utility function at +-1. Estimates of all 3 parameters are used to evaluate each lottery, in combination with the decision weights (Harrison and Swarthout, 2016). Probabilities are weighted by the single parameter weighting function (3.9).

3.3.4 Data collection

The data was collected between May and July of 2016. In total we surveyed 184 households, 93 from Ratanakiri province, Cambodia and 91 from XK, Lao PDR. In each of the provinces, two districts with 10 villages each were purposefully selected along a gradient of more remote locations at the forest margin, where more traditional forms of semi-subsistence agriculture are common, to sites which are more specialized, intensified and market-oriented. The participants of the field experiment in each village were selected based on a nonprobability sample (Levy and Lemeshow, 2008). Since there are no comprehensive lists of farming households in the villages, we rely on the expert knowledge of the extension workers from the regional government offices in both countries to select participants haphazardly in collaboration with the respective village official. The workshops were publicly announced by the provincial department of agriculture in collaboration with the mayors of the villages. The only condition to participate in the workshop was that the respondent had to know about the income situation of the household and the production decisions of the family farm. The data was collected during workshops in the respective village centers. The workshops consisted of two parts, a household survey and lottery choice experiments (see Appendix B for details). The household survey was conducted with a comprehensive questionnaire which included socio economic information, farm characteristics, household income and information on experience of shocks.

3.3.5 Research area

The study was conducted in the province of Ratanakiri, Cambodia and in XK province, Lao PDR. Both provinces are located in mountainous regions and have undergone fast structural transformations, but are still coined with slow rural development (IFAD, 2016). The agricultural sector in both provinces can be characterized in economic terms with low specialization and commercialization. Most recently, local political focus aimed to develop the Greater Mekong Subregion. Investments materialized mainly in infrastructure projects, mining operations and hydropower facilities. The agricultural sector developed due to investments of large co-operations and concession based agriculture. Deforestation and planting of industrial crops (rubber, cashew, oil palm, cassava and maize) has brought considerable rural investment and development, but participation of local communities in the markets is still low or non-existent (Fox and Castella, 2013). Nevertheless, in both provinces, the agricultural sector accounts for a share of 48 percent and 54 percent of GDP, respectively (Anh, 2016). Both countries are exposed to many climate risks and disasters, with flood and drought being the most frequent events. SEA can also expect changes in the frequency and magnitude of extreme weather and climate events, such as heat waves and heavy precipitation (IPCC, 2012). These changes in climate pose a big constraint for the agricultural sector in Lao

PDR and Cambodia. Climate change scenarios predict losses ranging between 20 percent and 33 percent of crop net revenue (Mendelsohn, 2014).

3.4 Results and discussion

3.4.1 Risk preference parameters of EUT, RDU and CPT

In a first step, risk preferences parameters of three models (EUT, RDU and CPT) are estimated. In these models we leave out individual covariates. Table 3.1 presents the mean values of the parameters calculated with maximum likelihood estimation from observed choices made in the risk experiments. The total sample has 6440 single choice observations from 184 clusters (households). A z-test is carried out to test if the risk preference parameters are significantly different from zero. Additionally, we conduct tests on single parameters. In the EUT specification, we have an alfa of 0.55, which indicates concavity of the utility function. This result points to moderate risk aversion and resemble results from other developing countries. Harrison et al. (2010) for example report a risk aversion parameter of 0.536 in India Uganda and Ethiopia.

Under RDU, the risk preference parameter has the same parametric specification as EUT, so parameter alfa stays the same. The parameter gamma is γ=2.224. This indicates s-shaped probability weighting. Furthermore, we do an additional test on the constant of gamma. If γ=1, equation (7) collapses to $\omega(p) = p$ as it is the case under EUT. We take this test as a convenient equivalent of testing if preferences follow the RDU specification or the EUT specification. We can reject the H_0: γ=1 with a p-value<0.001, thus we infer that probability weighting is present in our sample which posits a deviation from the conventional EUT representation. Comparable results of an s-shaped probability weighting function are reported in three studies which were conducted in Mozambique (γ=1.370), as well as in India, Ethiopia and Uganda (γ=1.384) by de Brauw and Eozenou (2014) and Harrison et al. (2010), respectively.

Turning to the CPT case we have a probability weighting parameter γ=1.413, smaller compared to the RDU case. Again our test statistic shows that $\gamma \neq 1$. If $\gamma > 1$, individuals tend to underweight small probabilities relative to the objective and overweight large probabilities. The form of the probability weighting function is s-shaped. Further support for an s-shaped probability weighting function comes from Humphrey and Verschoor (2004b). The researchers use common consequence effects to investigate if rank depended utility describes lottery choice of rural Ethiopians, Indians and Ugandans.

The curvature of the utility function has two parameters, α=0.350 for curvature in the gain domain and β=0.638 for curvature in the loss domain. This implies concavity in the gain domain and convexity in the loss domain since $\alpha < \beta$. Comparable studies from developing countries are not existent, in most cases curvature of the utility function is not measured separately for the gain and loss domain. The assumption that $\alpha=\beta=\sigma$, with σ being the only parameter that expresses curvature of the utility function is very restrictive and has a dramatic effect on the loss aversion coefficient λ. The parameters indicate that farmers in our study show a higher risk aversion than cotton farmers in China (σ=0.480), farmers in Vietnam (σ=0.590), rice farmers in India (σ=0.772) and fishermen in Vietnam (σ=1.012), but a lower degree of risk aversion than herders in Mali and Burkina Faso (σ=0.133) (Liebenehm and Waibel 2014; Nguyen 2011; Tanaka et al. 2010; Ward and Singh 2015).

The loss aversion parameter lambda is 0.951, so under this estimation we have no evidence of loss aversion since loss aversion calls for a $\lambda > 1$. Our additional test on the constant of λ ($H_{0:}$ λ=1) has a p-value of 0.646 and we cannot reject the null. These results contrast most results from the literature. Tanaka et al. (2010) find λ=2.630 of farmers in Vietnam. An even higher degree of loss aversion (λ=4.464) is reported in Ward and Singh (2015) who use a sample of

Indian farmers. Nguyen (2011) reports a λ=3.255 of fishermen in Vietnam and Liu and Huang (2013) report a loss aversion parameter λ=3.470 among rice farmers in China. A similar result to our study comes from Liebenehm (2014) who report a loss aversion coefficient of West African herders to be 1.351, but fail to reject the H_0: λ=1. Anyhow, it has to be stated that these studies don't use the same specification, and the difference of loss aversion stems from the fact that we include curvature of the utility function in the gain and loss domain to derive the parameter for loss aversion.[1]

To discriminate between RDU and CPT we apply Clarke's test for non-nested models (Clarke, 2003). We use this test because the log likelihoods are non-normally distributed, with highly significant p-value (p<0.001) of the Shapiro-Wilk test. The distribution is displayed graphically in Figure B 2, supplementary materials. The graph shows that the distribution is more peaked than the normal distribution. The Clarkes test uses a non-parametric strategy to compare individual likelihoods for each observation. In our case, CPT outperforms RDU in 71 percent of the cases due to the Clarke's test so we infer that the CPT model better explains the data.

[1] In fact we also conducted an estimation without differentiating between utility curvature in the gain and loss domain, which is comparable to the cited literature. Strikingly, we find a coefficient λ=2.044 for loss aversion, which is significantly different from 0 (F-stat (H_0: λ=1) 116.53, p-value <0.001. Results are reported in Table B 20 in Appendix B.

Table 3.1: Risk preference parameters of EUT, RDU and CPT

	Parameter	Point Estimate	Standard Error	p-value	95% Confidence interval	
Expected Utility Theory						
	α	0.550	0.051	0.000***	0.450	0.649
Rank Dependent Utility Theory						
	γ	2.224	0.214	0.000***	1.805	2.643
	α	0.550	0.154	0.000***	0.248	0.853
Additional test statistics						
F-stat (H_0: γ=1)	32.78					
p-value	0.000***					
Cumulative Prospect Theory						
	γ	1.413	0.039	0.000***	1.337	1.489
	α	0.350	0.023	0.000***	0.305	0.396
	β	0.638	0.049	0.000***	0.543	0.734
	λ	0.951	0.107	0.000***	0.741	1.161
Additional test statistics						
F-stat (H_0: γ=1)	114.52					
p-value	0.000***					
F-stat (H_0: λ=1)	0.21					
p-value	0.646					
F-stat (H_0: α-β=0)	36.28					
p-value	0.000***					

N= 6440. Significance levels: * p<0.10, **p<0.05, *** p< 0.01.

3.4.2 CPT model with covariates

Since we identified that the CPT specification best explains our data, we take this model for further analysis here and include variables on socio-economic characteristics and household shocks. Three distinct models of CPT are calculated and displayed in Table 3.2; each includes one of three variables that measure the impacts of household shocks. One variable describes the total number of household shocks, a second variable distinguishes between households that reported no shocks and the third variable measures only exogenous household shocks due to flood, drought and price shocks. The control variables are carefully selected to reflect influencing factors on individual preference parameters (see Table B 13 in Appendix B).

The first three columns of Table 3.2 display the results of our full model including socioeconomic variables and household shocks and their correlation with probability weighting. Generally, increasing age lessens the overweighting of subjective probabilities. This might be due to the fact that people gain experience over time and can estimate probabilities of events better. Out of our three variables describing shocks, the variable accounting for total household shocks is statistically significantly correlated with our probability weighting parameter. Hence, adverse events distort subjective probability weighting directly. These results have strong implications in a farming context. For example, smallholders might underweight the probabilities of rainfall. As a consequence, they might delay planting, which in turn reduces yields. Furthermore, the beneficial effects of technologies are underestimated, which decreases adoption rates of useful, and potentially risk reducing technologies.

Columns four through nine of Table 3.2 present the results regarding the curvature of the value function of farmers. In the gain domain (coefficient alpha), we find consistent effects of age on value function curvature. According to our results, older people get more risk averse, a result that is consistent with the literature (de Brauw & Eozenou, 2014; Tanaka et al., 2006). Interestingly, we find that female respondents are less risk averse than their male counterparts. This effect is more pronounced in the loss domain than in the gain domain, where the coefficients are smaller and not significant. This result contests the common believe that female respondents are more risk averse and reflects results found in Vietnam (Nguyen, 2011). Our tentative explanation for this effect is that typically, women in Cambodia are managing the household finances. This gives them experience when handling monetary decisions which is what our lottery is. The variable no shocks is significantly correlated with utility curvature in the loss domain (coefficient beta). Thus, farmers who were not exposed to any shocks have a lower degree of risk aversion in the loss domain. Risk averse smallholders forgo profits by applying non-profit maximizing farming systems and strategies. Consequently, they miss out on opportunities which can increase the productive potential of their farm.

The last three columns of Table 3.2 present the results regarding the loss aversion parameter. We see that older people are less loss averse, though not statistically significant. Education also reduces loss aversion, but is only significant in one model. A more consistent and significant effect comes from the variable cattle owned, which significantly reduces loss aversion in all three models. This result is in line with our expectation that variables describing wealth hamper the effect of loss aversion. In the case of livestock holdings, it is well known that farmers in our research area keep cattle because it is a safe deposit. They can resort to this resource in times of need, thus decreasing the pressure on household consumption and relaxing loss aversion. The variable total household shocks is significantly increasing loss aversion in our sample. As a perfect mirror image, not having experienced any shocks decreases loss aversion significantly. Thus, household shocks can trigger loss aversion, controlling for a wide range of socio economic variables. A possible interpretation of this result is that farmers have a target income level. The experience of household shocks endangers this target income level which increases their aversion to losses. Therefore, household shocks not only cause economic damages, but alter the behavior of smallholders under uncertainty. Another possible mechanism that comes into play if smallholders are sensitive to losses, is the reflection effect described by Tversky and Kahneman (1992). Hereafter, individuals become risk seeking to prevent sure losses. In a farming context, this could mean that if a rice field is hit by a drought, smallholders exceedingly allocate resources to this field to prevent a crop loss, instead of abandoning a hopeless situation.

For completeness, the supplementary materials present robustness checks (Table B 16 and B 17) and results from the EUT (Table B 18) and RDU (Table B 19) specifications with variables describing single household shocks, but we exclude them from the main text for brevity. In short, the EUT specification suggests a strong relationship of risk preference parameter α with household shocks. The exercise to estimate several utility specifications helps us to compare our results to the literature, which predominantly applies utility according to EUT. Using a EUT estimation strategy, we find clear correlations with household shocks and utility curvature.

Table 3.2: Cumulative Prospect Theory parameters with covariates

	gamma			alpha			beta			lambda		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Age	-0.025	-0.047*	-0.060***	-0.018*	-0.023**	-0.022***	0.024	0.013	0.025	-0.011	-0.012	-0.035
(Years)	(0.022)	(0.025)	(0.023)	(0.009)	(0.010)	(0.007)	(0.034)	(0.030)	(0.021)	(0.020)	(0.014)	(0.027)
Gender	-0.161*	-0.065	-0.176	-0.026	-0.003	-0.025	-0.182*	-0.140	-0.123**	0.057	0.004	-0.037
(Male=1)	(0.094)	(0.077)	(0.120)	(0.026)	(0.027)	(0.026)	(0.110)	(0.098)	(0.057)	(0.048)	(0.030)	(0.055)
Education	-0.008	0.002	0.010	-0.005	-0.002	-0.002	0.011	0.008	0.007	-0.014*	-0.003	-0.002
(Years)	(0.012)	(0.012)	(0.011)	(0.006)	(0.006)	(0.004)	(0.025)	(0.020)	(0.009)	(0.007)	(0.003)	(0.005)
Dependent Household	-0.016	-0.025	-0.000	-0.007	0.000	0.003	0.061	0.067	0.077	-0.047**	-0.026	-0.032*
members (count)	(0.020)	(0.023)	(0.022)	(0.009)	(0.009)	(0.009)	(0.047)	(0.064)	(0.049)	(0.019)	(0.024)	(0.017)
Household income	-0.000	-0.001*	-0.000	0.000	-0.000	-0.000	-0.001*	-0.002**	-0.002**	0.002	0.001*	0.003*
(Per capita)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Cattle owned	0.005*	0.005	-0.002***	0.002**	0.002	0.000	0.005	0.007	0.000	-0.008**	-0.004**	-0.005***
(Heads)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.000)	(0.006)	(0.007)	(0.000)	(0.003)	(0.002)	(0.002)
Country	0.029	-0.064	-0.054	0.008	-0.060	-0.003	-0.780	-0.758	-0.046	0.741	0.620	-0.207
(Base level Cambodia)	(0.141)	(0.142)	(0.151)	(0.044)	(0.047)	(0.042)	(0.591)	(0.491)	(0.562)	(0.514)	(0.685)	(0.851)
Total shocks	0.060*			0.014			-0.008			0.061**		
	(0.034)			(0.013)			(0.025)			(0.030)		
No Shock		-0.063			0.060			0.266**			-0.186**	
		(0.100)			(0.047)			(0.111)			(0.077)	
Exogenous shocks			-0.018			-0.001			-0.006			0.023
			(0.050)			(0.018)			(0.020)			(0.052)
Constant	1.662***	1.851***	1.925***	0.446***	0.478***	0.476***	1.196**	1.060**	0.664**	0.450	0.679	1.160
	(0.182)	(0.202)	(0.192)	(0.080)	(0.080)	(0.069)	(0.526)	(0.490)	(0.313)	(0.423)	(0.471)	(0.792)

Summary statistics		Model 1	Model 2	Model 3
	Noise Parameter	0.149***	0.158***	0.151***
	Noise Standard Error	(0.034)	(0.025)	(0.025)
	Observations	6440	6440	6440
	Log Likelihood	6288	6176	6283
	BIC	-3037	-2981	-3034

Significance levels: * p<0.10, **p<0.05, *** p< 0.01. Standard errors in parentheses. Village Dummies are excluded from the table.

3.5 Suggestions for future research

In our experimental design, participants of field experiments receive a show up fee. Since it would be unethical to receive payments from farmers, the losses that could potentially occur to them never exceed this fee. It is a possibility that the respondents of our field experiments integrate the participation fee into their utility which results in the fact that we never play for real losses. To circumvent this, one possible avenue for future research would be to pay the participation fee ahead of the experiments, so that the show up fee is already incorporated into the stream of household consumption. If the lottery is played weeks after this, we could rule out the house money effect as a potential source of biased estimates (Thaler & Johnson, 1990).

The one parameter probability weighting function we apply here is not very flexible. The two parameter weighting function (Prelec, 1998), where one parameter controls the curvature of the function and one parameter for the elevation is more flexible. Furthermore, the function can allow for probability weighting in the gain and loss domain separately. This is important because different decision weights for gains and losses could induce the same behavior as in utility loss aversion (Harrison & Swarthout, 2016). Anyhow, these restrictions are forced due to the design of lottery choices that were employed in our experiment. In our design, only 7 out of 35 binary choices are in the mixed frame, with equal probabilities over prospects, and there is a lack of choice in the pure loss domain. This imposes a restriction to apply more complex functions of probability weighting. Also, identification of probabilistic loss aversion is not possible due to the fact that we have only 50-50 choices in the mixed frame (Harrison & Swarthout 2016). Probabilistic loss aversion can occur when the utility curves for both loss and gains are linear and there is no evidence of conventional utility loss aversion. In this case, differences in the decision weights alone could induce behavior similar to loss aversion as noted by Wakker, (2010). As Harrison and Ross (2017) state: Most of the apparently loss-averse choice behavior results from probability weighting rather than from direct disutility experienced when an outcome is framed as a loss against an idiosyncratic reference point. Therefore, a future avenue of research is to estimate a model that applies the more flexible Prelec two parameter probability weighting function (Prelec, 1998) and allow the parameters of the function to vary over gains and losses.

One more obvious weakness stems from the fact that we use cross sectional data to measure the impact of shocks on risk preference parameters. The studies that employ panel data on this topic also use more restrictive functional forms of behavior. In future research, the strengths of both approaches should be combined.

3.6 Conclusion

Behavior under risk is a central feature of vicious cycles of poverty. The occurrence of household shocks can further aggravate poverty traps by altering behavior of poor smallholders under uncertainty. We apply three different specifications (EUT, RDU, CPT) of utility to measure the effects of household shocks on risk aversion, probability weighting and loss aversion. First, we find that CPT is a good representation of smallholders' behavior under uncertainty. One striking result of our study is that if we allow for two parameters to measure curvature of the utility function, we find no evidence for loss aversion in the traditional sense, but rather convex utility in the loss domain. Second, we find that household shocks can increase loss aversion and they have effects on the curvature of the utility function in the loss domain. The underlying mechanism is that household shocks can disrupt expected consumption or target income levels of smallholders and lead to an increase in loss aversion. Since many farm management decisions in developing countries entail the possibility of net losses, risk aversion in the loss domain deters smallholders from profit maximizing investment decisions. Thus, the aversion to losses is tightly linked to the reinforcing nature of poverty and is important to address this in experiments with smallholders in developing economies.

An important policy recommendation from this study is that new technologies and farming systems might be rejected by smallholders due to loss aversion. The result suggests that it is important for extension agencies to introduce new technologies in a way that losses are excluded for farmers. Initial risk free demonstration and on-farm experimentation has to be provided for farmers until they can identify expected returns of such new technologies. Farmers will be more likely to invest in new technologies if they can evaluate risks and probabilities associated with the technology. Regarding loss aversion, insurance and financing mechanisms should be applied in a way that they protect stable household consumption. One potential avenue is to offer flexible loan payback schedules for agricultural technologies. For example, loan payback schedules can be linked to yields achieved by farmers. This has the benefit that agricultural shocks can be buffered by decreasing amortization rates when yields are low. This protects target income levels and household consumption against shocks. A future avenue of research is to identify reference points and target income levels of smallholders and their influence on risk aversion in the gain and loss domain.

Conducting this research in remote mountain areas of Cambodia and Lao PDR might have contributed to the fact that the observed effects are significant. In areas of limited infrastructure and an undeveloped insurance sector, the impacts of shocks are severe enough to be recorded by the instruments. Nevertheless, shocks seem to be case specific and further research in other regions and contexts should be carried out to verify the results.

References

Abdellaoui, M., Bleichrodt, H., & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty*, 36(3), 245–266.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Lost in state space: Are preferences stable? *International Economic Review*, 49(3), 1091–1112.

Anh, N. T. (2016). Effects of natural disasters on agricultural production activities in the Cambodia-Laos- Vietnam development triangle area: Case studies of Ratanakiri (Cambodia), Attapeu (Laos) and Kon Tum (Vietnam) provinces. ASEAN-Canada Research Partnership Working Paper Series, 2.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2), 407–441.

Cameron, L., & Shah, M. (2015). Risk-taking behavior in the wake of natural disasters. *Journal of Human Resources*, 50(2), 484–515.

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: Empirical and policy implications. *Journal of Development Studies*, 42(2), 178–199.

Clarke, K.A.(2003). Nonparametric model discrimination in international relations. *Journal of Conflict Resolution*, 47(1), 72–93.

de Brauw, A., & Eozenou, P. (2014). Measuring risk attitudes among Mozambican farmers. *Journal of Development Economics*, 111, 61–74.

Dercon, S. (2004). Growth and shocks: Evidence from rural Ethiopia. *Journal of Development Economics*, 74(2), 309–329.

Dercon, S., & Christiaensen, L. (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics*, 96(2), 159–173.

Dohmen, T., Falk, A., Huffmann, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3), 371–589.

Eckel, C. C., El-gamal, M. A., & Wilson, R. K. (2009). Risk loving after the storm : A Bayesian-network study of Hurricane Katrina evacuees. *Journal of Economic Behavior & Organization*, 69,110–124.

Emerick, K., De Janvry, A., Sadoulet, E., & Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, 106(6), 1537–1561.

Fafchamps, M. (2010). Vulnerability, risk management and agricultural development. *African Journal of Agricultural and Resource Economics*, 5(1), 243–260.

FAO. (2011). The state of the world's land and water resources for food and agriculture. Managing systems at risk. Food and Agriculture Organization. 978-1-84971-326-9 doi:10.1111/j.1467-8489.2010.00527.x

Fehr, E., & Goette, L. (2007). Work more if wages are high? Evidence from a randomized field experiment. *The American Economic Review*, 97(1), 298–317.

Fox, J., & Castella, J.-C. (2013). Expansion of rubber (Hevea brasiliensis) in mainland Southeast Asia: What are the prospects for smallholders? *Journal of Peasant Studies*, 40(1), 155–170.

Gloede, O., Menkhoff, L., & Waibel, H. (2013). Shocks, individual risk attitude, and vulnerability to poverty among rural households in Thailand and Vietnam. *World Development*, 71,54–78.

Gollier, C., & Pratt, J. W. (1996). Risk vulnerability and the tempering effect of background risk. *Econometrica*, 64(5), 1109–1123.

Guiso, L., & Paiella, M. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic Association*, 6 (6), 1109–1150.

Gul, F. (1991). A theory of disappointment aversion. Econometrica, 59(3), 667–686. doi:10.2307/2938223 Harless, D.W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6), 1251–1289.

Harrison, G.W.,& Rutström, E.(2008). Risk aversion in the laboratory. In J. C. Cox & G.W. Harrison (Eds.), Risk Aversion in Experiments (Vol. 12, pp. 41–196). Bingley, UK: Emerald, Research in Experimental Economics. doi:10.1016/S0193-2306(08)00003-3

Harrison, G. W., Humphrey, S. J., & Verschoor, A. (2010). Choice under uncertainty: Evidence from Ethiopia, India and Uganda. *The Economic Journal*, 120(543), 80–104.

Harrison, G. W., List, J. A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2), 433–458.

Harrison, G. W., & Ross, D. (2017). The empirical adequacy of cumulative prospect theory and its implications for normative assessment. *Journal of Economic Methodology*, 24(2), 150–165.

Harrison, G. W., & Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, 12(2), 133–158.

Harrison, G. W., & Swarthout, J. T. (2016). Cumulative prospect theory in the laboratory: A Reconsideration (Experimental Economics Center Working Paper Series No. 2016–4). Retrieved from https://cear.gsu.edu/category/working-papers/wp-2016/.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. Cognitive Psychology, 38(1), 79–109. Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6), 1291–1326.

Holden, S. T., & Quiggin, J. (2016). Climate risk and state-contingent technology adoption: Shocks, drought tolerance and preferences. *European Review of Agricultural Economics*, 1–24. doi:10.1093/erae/jbw016

Humphrey, S. J., & Verschoor, A. (2004a). Decision-making under risk among small farmers in East Uganda. *Journal of African Economies*, 13(1), 44–101.

Humphrey, S. J., & Verschoor, A. (2004b). The probability weighting function: Experimental evidence from Uganda, India and Ethiopia. *Economics Letters*, 84(3), 419–425.

IFAD. (2016). Rural development report 2016- fostering inclusive rural transformation. Rome: International Fund for Agricultural Development.

IPCC. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. Special report of the Intergovernmental panel on climate change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)] (pp. 582). Cambridge, UK: Cambridge University Press. doi:10.1596/978-0-8213-8845-7

Kaufman, B. E. (1999). Emotional arousal as a source of bounded rationality. *Journal of Economic Behavior & Organization*, 38(2), 135–144.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4), 1047–1073.

Levy, P. S., & Lemeshow, S. (2008). Sampling of populations : Methods and applications (4th ed.). Hoboken, NJ: Wiley.

Li, J. Z., Li, S., Wang, W. Z., Rao, L. L., & Liu, H. (2011). Are people always more risk averse after disasters? Surveys after a heavy snow-hit and a major earthquake in China in 2008. *Applied Cognitive Psychology*, 25(1), 104–111.

Liebenehm, S., & Waibel, H. (2014). Simultaneous estimation of risk and time preferences among small-scale cattle farmers in West Africa. *American Journal of Agricultural Economics*, 96(5), 1420–1438.

Liu, E. M. (2013). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in china. *The Review of Economics and Statistics*, 95(4), 1386–1403.

Liu, E. M., & Huang, J. (2013). Risk preferences and pesticide use by cotton farmers in China. *Journal of Development Economics*, 103(1), 202–215.

Malmendier, U., & Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1), 373–416.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcome of risky options. *Psychological Science*, 8, 6.

Mendelsohn, R. (2014). The impact of climate change on agriculture in Asia. *Journal of Integrative Agriculture*, 13(4), 660–665.

Menkhoff, L., &Sakha, S. (2016). Determinants of risk aversion over time: Experimental evidence from rural Thailand. *DIW Berlin Discussion Paper*, No. 1582. doi:10.2139/ssrn.2785467

Mosley, P., & Verschoor, A. (2005). Risk attitudes and the “vicious circle of Poverty”. *The European Journal of Development Research*, 17(1), 59–88.

Mullainathan, S. (2006). Development economics through the lens of psychology. Proceedings of the Annual Bank Conference on Development Economics. Retrieved from: https://sendhil.org/wp-content/uploads/2019/08/Publication-45.pdf

Nguyen, Q. (2011). Does nurture matter: Theory and experimental investigation on the effect of working environment on risk and time preferences. *Journal of Risk and Uncertainty*, 43(3), 245–270.

Nguyen, Q., & Leung, P. (2009). Do fishermen have different attitudes toward risk? An application of prospect theory to the study of Vietnamese fishermen. *Journal of Agricultural and Resource Economics*, 34(3), 518–538.

Nielsen, T., Keil, A., & Zeller, M. (2013). Assessing farmers’ risk preferences and their determinants in a marginal upland area of Vietnam: A comparison of multiple elicitation techniques. *Agricultural Economics*, 44(3), 255–273.

Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3), 497–527.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3, 323–343.

Rosenzweig, M. R., & Binswanger, H. P. (1993). Wealth, weather risk and the composition and profitability of agricultural investments. *The Economic Journal*, 103(416), 56–78.

Said, F., Afzal, U., & Turner, G. (2015). Risk taking and risk learning after a rare event: Evidence from a field experiment in Pakistan. *Journal of Economic Behavior and Organization*, 118, 167–183.

Sandmo, A. (1971). On the theory of the competitive firm under price uncertainty. *The American Economic Review*, 61(1), 65–73.

Shimamoto, D., Yamada, H., & Wakano, A. (2017). The different effects of risk preferences on the adoption of agricultural technology: Evidence from a rural area in Cambodia. *The Journal of Development Studies*,1–19. doi:10.1080/ 00220388.2017.1329527

Starmer, C. (2000). Developments in non-expected utility theory : The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, XXXVIII (June), 332–382.

Stigler, G. J., & Becker, G. S. (1977). De Gustibus Non Est Disputandum. *American Economic Review*, 67,76–90.

Tanaka, T., Camerer, C. F., & Nguyen, Q. (2006). Poverty, politics, and preferences: Field experiments and survey data from Vietnam (Working Paper), (October 2005) (pp. 1–47). Retrieved from http://authors.library.caltech.edu/22469/

Tanaka, T., Camerer, C. F., & Nguyen, Q. (2010). Risk and time preferences: Experimental and household survey data from Vietnam. *American Economic Review*, 100(1), 557–571.

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, 36(6), 643–660.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory - Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.

Van Den Berg, M., Fort, R., & Burger, K. (2009, August 16–22). Natural hazards and risk aversion: Experimental evidence from Latin America. In Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference (pp. 1–35). Beijing, China.

Voors, M. J., Nillesen, E. E. M., Verwimp, P., Bulte, E. H., Lensink, R., & Van Soest, D. P. (2012). Violent conflict and behavior: A field experiment in Burundi. *American Economic Review*, 102(452), 941–964.

Ward, P. S., & Singh, V. (2015). Using field experiments to elicit risk and ambiguity preferences: Behavioural factors and the adoption of new agricultural technologies in rural India. *The Journal of Development Studies*, 51(6), 707–724.

Wicker, F. W., Hamman, D., Hagen, A. S., Reed, J. L., &Wiehe, J. A. (1995). Studies of loss aversion and perceived necessity. *The Journal of Psychology*, 129(1), 75–89.

Yesuf, M., & Bluffstone, R. A. (2009). Poverty, risk aversion, and path dependence in low-income countries: Experimental evidence from Ethiopia. *American Journal of Agricultural Economics*, 91(4), 1022–1037.

4 The effect of poor vision on economic farm performance: Evidence from rural Cambodia

Abstract

Roughly one-fifth of the global population is affected by poor visual acuity. Despite the fact that inhabitants of rural areas in low-income countries are most distressed by this, no prior research has studied the impact of poor visual acuity on the economic performance of farms. We conduct a standardized eye test with 288 farm managers in rural Cambodia and find that around 30 percent of our sample suffers from poor visual acuity in terms of nearsightedness (myopia). Our analyses indicate a statistically significant and economically meaningful association of poor visual acuity with economic farm performance. Our results show that gross margins for cropping activities per year could be, on average, around 630 USD higher if farm managers were able to correct for poor vision. Our results suggest that poor visual acuity impairs farm managers from tapping the full potential of their business, which in turn decreases their chance to break the vicious cycle of poverty.

This chapter is coauthored by Frederik Sagemüller (FS), Selina Bruns (SB) and Oliver Mußhoff (OM). The contributions to the article are: FS and SB conceptualized the idea, SB collected the data, FS analyzed the data, and wrote the article. OM and SB reviewed the article. This chapter is published in the journal *PLoS ONE* with the following citation: Sagemüller, Bruns & Mußhoff O (2022). The effect of poor vision on economic farm performance: Evidence from rural Cambodia. *PLoS ONE* 17(9): e0274048. https://doi.org/10.1371/journal.pone.0274048

4.1 Introduction

An estimated 596 million people are affected by mild, moderate or severe distance vision impairments and blindness (Burton et al., 2019). To put the problem into an economic perspective, the total worldwide financial cost of visual impairments was estimated at three trillion USD (Gordois et al., 2012) and the economic burden of uncorrected distance refractive error alone was estimated to be 202 billion USD per annum (Holden et al., 2016).

A growing body of literature is investigating occurrence, cause and consequence of visual impairments. With regards to occurrence, most disease burden is carried by low income economies of South Asia, East Asia and SEA (Bourne et al., 2017). Due to population growth and demographic change, a substantial increase in the prevalence of visual impairments is expected in the future, e.g. the global incidence of population with myopia will increase to roughly 50 percent by 2050 (Holden et al., 2016). In terms of cause, cataract and uncorrected refractive error contributed to 55 percent of blindness and 77 percent of vision impairment in adults aged 50 years and older (Flaxman et al., 2017). Uncorrected refractive errors like myopia and presbyopia are known to be among the largest causes for moderate and severe vision impairments, as well as blindness (Flaxman et al., 2017). Looking at the consequences of poor visual acuity, studies show that uncorrected refractive error leads to a decrease in quality of life (Kandel et al., 2018; Tahhan, Papas, Fricke, Frick, & Holden, 2013). This result is confirmed through a randomized control trial carried out in Bangladesh, Kenya and the Philippines which shows that cataract surgery to relieve blindness improves the quality of life of participants (Kuper et al., 2010). Moreover, people who underwent the surgery were more likely to participate in productive activities.

The losses in public welfare and human capital due to visual impairments are preventable; unaided visual impairments almost exclusively occur in the context of poverty. The relationship between poverty and eye health is laid out in a literature review in which the authors suggest that visual impairments and poverty seem to be intertwined in a vicious cycle (Jaggernath et al., 2014). A recent study shows that improving eye health contributes directly and indirectly to achieving many SDGs, including reducing poverty and hunger, improving work productivity and educational equity. The researchers suggest that eye health needs to be reframed as an enabling, cross-cutting issue within the sustainable development framework (Zhang et al., 2022). A combination of a lack of access to affordable public eye care in low-income countries, and the large percentage of the population in these countries whose livelihoods depend on agriculture, results in a greater burden of visual impairments occurring among inhabitants of agrarian communities. Thus, there are strong connections between agriculture, poverty and visual impairments.

From an agricultural economics perspective, we can hypothesize that poor vision is a cause for productivity loss of farms. Undoubtedly, the literature suggests that better public health contributes to higher agricultural productivity in terms of total factor productivity and efficiency (Ajani & Ugwu, 2008; Antle & Capalbo, 1994; Audibert & Etard, 2003; Loureiro, 2009; Sabasi & Shumway, 2018), yet there are no case studies which describe effects of poor visual acuity on the profitability of farming. This research gap is astounding taking into account that raising agricultural productivity and profitability are central themes in the fight to end poverty (Christiaensen, Demery, & Kuhl, 2011; De Janvry & Sadoulet, 2009; Klasen & Reimers, 2017).

The linkages between agriculture and visual impairments have only been directly addressed in one previous study, which shows that providing glasses to correct for age-related farsightedness (presbyopia) improves work performance of tea pickers in India (Reddy et al., 2018). The effect of myopia on the profitability of family-owned farms remains unknown. This aspect is important since agricultural production in developing countries relies mostly on family-owned farms (Lowder et al., 2016) and the effect of vision impairments on economic performance could be biased if studied only among wage earners. Also unknown is the effect of myopia on agricultural profitability, as compared to the effect of presbyopia on work performance. Age-related presbyopia occurs in persons older than 40 years and therefore it is difficult to generalize the findings from their study.

We contribute to the understanding of the relationship between health and agriculture by adding a case study with farm managers in rural Cambodia on poor visual acuity in terms of myopia and its association with farm profitability. We address this topic by answering two research questions: 1) What is the prevalence of myopia among rural smallholders in Cambodia? 2) Is myopia associated with a loss in agricultural profitability of family-owned farms?

To illustrate the relationship between agricultural profitability and poor visual acuity, and to give first estimates on potential effect sizes, we carry out a household survey combined with a standardized eye test with 288 Cambodian smallholders. Cambodia was chosen for our empirical application for three reasons: Firstly, the anecdote that the leaders of the Communist Party of Kampuchea (CPK) ‘decided to kill anyone who wore glasses’, is commonly shared when relaying some of the horrors associated with life in Cambodia during the revolutionary period (BBC, 2018). It is plausible that cultural stigmas and stereotypes are playing parts in the low uptake of glasses in rural Cambodia. Secondly, the rural areas of Northeast Cambodia are coined by slow rural development. Most income is generated by small scale agriculture and most of the economically active population is employed in and depends on agriculture. Thirdly, the predicted increase in the number of people with avoidable vision impairment to 2050 is mainly occurring in South Asia and SEA (Flaxman et al., 2017).

This paper is addressed to development practitioners and researchers that work in agriculture and global health. There is a growing recognition that opportunities exist for agriculture to contribute to better health, and for health to contribute to agricultural profitability. We argue that joint action in agriculture and health could unlock synergies that substantially reduce poverty. From a research perspective, our study gives first results on the association between poor visual acuity and agricultural profitability and we want to motivate research at the intersection of public health topics and agriculture. The implications of our results are critical due to the magnitude of people who suffer from poor visual acuity and have limited access to modern optometric services. The issue becomes more pressing when we consider that the number of people affected by uncorrected poor visual acuity will continue to rise in the future (Holden et al., 2016).

The following paper is structured as follows: Section 4.2 reviews the literature on the global losses of productivity due to visual impairments with a special attention to the agricultural sector. Section 4.3 lays out a very simple conceptual framework to describe how visual acuity affects agricultural profitability. Section 4.4 gives an overview on the data collection process, the raw data, the important variables we include in our model and the empirical strategy. Section 4.5 presents the results and discussion. In Section 4.6 we draw a conclusion.

4.2 Literature review

Vision is often considered to be the sense that is most valued (Enoch et al., 2019). Vision requires structural and physiological integrity of the eyes, brain, and their connections. Disruption of any part of this pathway causes vision impairment. The most common causes of vision impairment in adults are uncorrected refractive error, cataract, glaucoma, age-related macular degeneration, diabetic retinopathy, corneal scarring, and trachoma. Vision-driven activities of daily living can be captured using quality of life tools and vision function-related tasks. The most common measure of visual function is distance visual acuity, which tests the ability to discern letters or characters of high contrast at decreasing size using the central retina (Bennett et al., 2019; Burton et al., 2021).

In 2020, an estimated 510 million people worldwide, of whom most live in low and middle income countries, had uncorrected near vision impairment, and a further 596.2 million people have distance vision impairment (Burton et al., 2021). Due to population growth and demographic change, a substantial increase in the prevalence of visual impairments is expected in the future, e.g. the global incidence of population with myopia will increase to roughly 50 percent by 2050 (Holden et al., 2016). An important gap in this literature is the lack of data from low income regions, including SEA. Due to this lack of data, most prevalence based studies extrapolate estimates across regions. Anyhow, there are studies that report regional estimates for SEA countries, though most of them report data on the sub-national level, within specific age groups and on a variety of different indicators for visual impairments. For example, a study from rural Myanmar looks at the prevalence of refractive errors in a population cohort of 40 years and older. The study reports a prevalence of refractive errors of 42.7 percent (Gupta et al., 2008). A national survey from Thailand reports an incidence of refractive errors via self-assessment, which is reported to be 28 percent (Yiengprugsawan et al., 2011). A study from an urban area in Lao PDR conducts comprehensive ophthalmic examinations and finds that the incidence of bilateral visual impairments of the population was 22.4 percent (Tan et al., 2022). A survey from Cambodia reports a prevalence of low vision in adults 50 years and older to be 21.1 percent (Morchen et al., 2015). The prevalence of vision impairment in school children between the ages 12–15 in Vietnam is reported to be 19.4 percent (Paudel et al., 2014). Despite the methodological differences and the differences of the cohort populations in age and in rural and urban locations, these studies show that the prevalence of low vision in the region ranges around a fifth to a quarter of the population.

From a macroeconomic perspective, there is evidence that vision impairments have a large economic impact worldwide. The scientific literature heavily relies on studies that calculate welfare costs of vision impairments by using visual acuity prevalence reported in national and global datasets, together with data for relative reduction in employment and reduction in wages due to visual impairments. Anyhow, the underlying data, methods and measurements of visual impairments have undergone drastic changes in recent decades. In 1996, the first global estimate of the worldwide productivity cost of blindness was estimated at 168 billion USD using 1993 data on visual impairment prevalence rates, GDP and world population data (Smith & Smith, 1996). The weakness of this study is that it only accounts for blindness as a visual impairment, and that the researchers assumed zero productivity for the blind and 100 percent productivity for the non-blind. A further study used data from the year 2000 to identify the potential effect on the global economic productivity of interventions that were planned as part of the "VISION 2020- right to sight" initiative (Frick & Foster, 2003). The economic gain of the interventions was estimated at 102 billion USD. These results were rather conservative estimates, as admitted by the researchers, since it was assumed that only working individuals at working age (15-64 years) produce goods and services valued at GDP per capita. Another important weakness of these studies is that they only account for best corrected visual acuity. However, using best corrected visual acuity obscures that, especially in settings of low income economies, people may not own spectacles, and so live with vision impairment from uncorrected refractive error. This underestimation is possibly large, the total number of persons with visual impairment worldwide including uncorrected refractive error was estimated to be 61 percent higher than the commonly quoted estimates which exclude uncorrected refractive error (Dandona & Dandona, 2006). Another estimate of the loss in productivity for 7 world regions was published in 2012. The study reports losses to be 168.3 billion USD, with projections for the year 2020 to be 177.5 billion USD (Gordois et al., 2012). Nevertheless, since losses to the economy were only accounted for in high income regions, these estimates are also underestimating the global costs due to losses in productivity. A recent study estimates the annual potential productivity losses associated with reduced employment due to blindness, moderate and severe vision impairment at the regional and global level. In this study, it is estimated that globally, 160.7 million people with moderate or severe vision impairment or blindness were within the working age. The relative reduction in employment by people with vision loss was 30.2 percent which result in a global potential productivity loss due to vision impairments of 410.70 billion USD purchasing power parity (Marques et al., 2021). However, the study captures only a limited amount of productivity loss components. Components not included in the analysis, because reliable data at country and regional level remain scarce, were absenteeism and presenteeism (reduced productivity in the working place), premature mortality due to visual impairments, productivity losses of people older than 64 years, productivity losses of caregivers, and value of time lost from unpaid or informal labor activities.

Thus, what remains unknown from these estimates, are the losses that occur in the informal agricultural sector due to visual impairments. Therefore, even though these estimates are the best guess of global productivity losses to date, they most likely present a conservative estimate. Regarding the underlying mechanisms that would lead to losses in agricultural productivity, the existing scientific literature provides consistent evidence for an association of visual impairments with reduced quality of life (Assi et al., 2021), reduced educational outcomes (Glewwe et al., 2016), reduced social status and reduced economic activity (Finger et al., 2012) and cognitive impairment, cognitive decline, and dementia (Burton et al., 2021). Other health domains and their association with agricultural productivity received more attention in the literature compared to visual impairments. For example, it is documented that health care access has a positive impact of total factor productivity for aggregate U.S. agricultural production (Sabasi & Shumway, 2018) and that the general health status of Filipino, Malian, Nigerian and Norwegian farmers increases production efficiency (Ajani & Ugwu, 2008; Antle & Pingali, 1994; Loureiro, 2009) and labor productivity (Audibert &

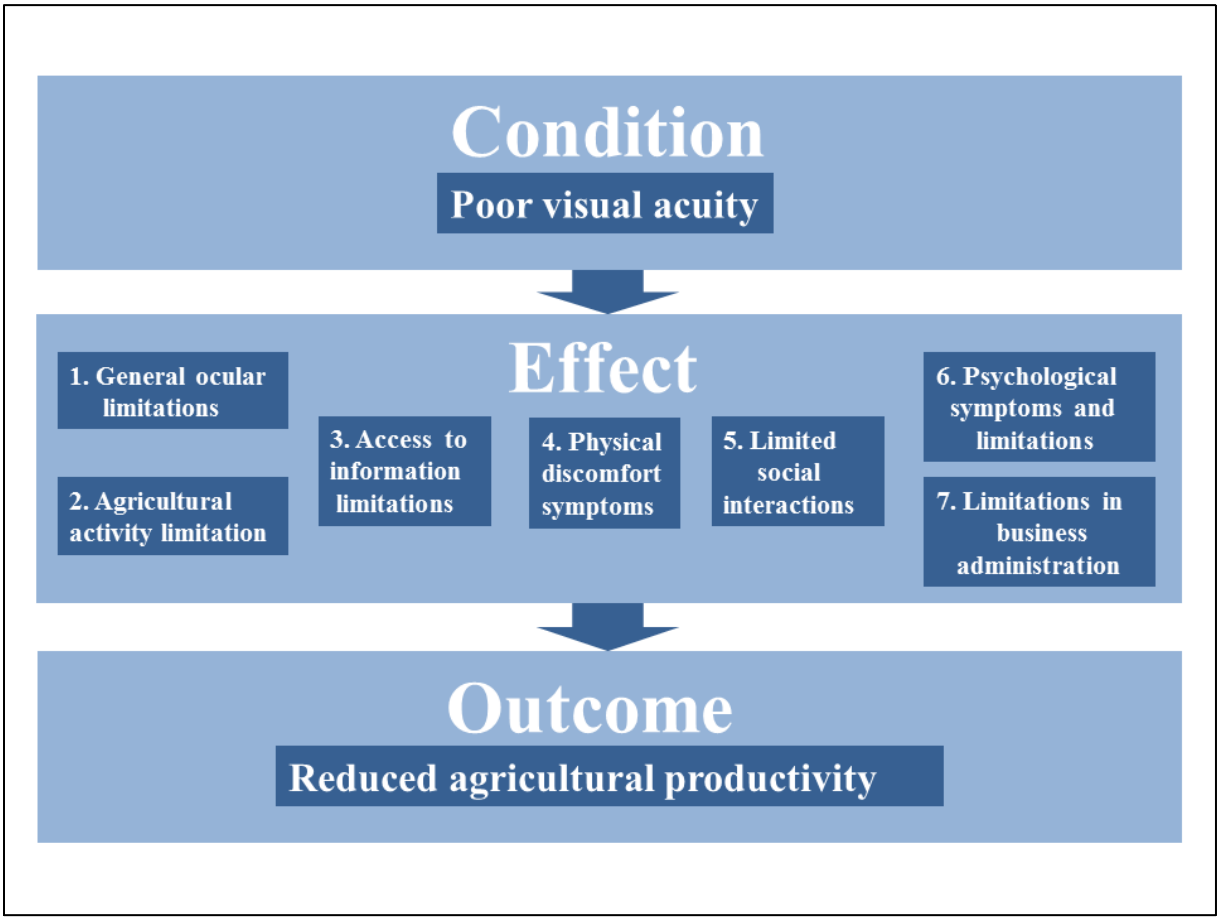
Etard, 2003). The linkages between agriculture and visual impairments have only been directly addressed in one previous study, which shows that providing glasses to correct for age-related farsightedness (presbyopia) improves work performance of tea pickers in India (Reddy et al., 2018). The effect of myopia on the profitability of family-owned farms remains unknown. This aspect is important since agricultural production in developing countries relies mostly on family-owned farms (Lowder et al., 2016) and the effect of vision impairments on economic performance could be biased if studied only among wage earners. Also unknown is the effect of myopia on agricultural profitability, as inference cannot be drawn from the effect of presbyopia on work performance. Age-related presbyopia occurs in persons older than 40 years and therefore it is difficult to generalize the findings from their study.

4.3 Conceptual framework

Our conceptual framework is related to the work on general impacts of health on economic outcomes, which describe how healthier populations tend to have higher labor productivity, because their workers are physically more energetic and mentally more robust (Bloom & Canning, 2000). Healthier children learn and perform better at school which leads to greater productivity and higher incomes. Furthermore, good health promotes school attendance and enhances cognitive function. We take a qualitative report from rural dwellers in Nepal as a starting point to map out the potential effects of poor vision on agricultural profitability (Kandel et al., 2018). The report explores the impact of corrected and uncorrected refractive error on Nepalese people’s quality of life. We sort the qualitative statements from their report and group them by categories that affect the outcome we aim to study. Figure 4.1 shows that we expect strictly negative impacts of poor visual acuity on agricultural profitability. To describe the pathways from visual acuity to agricultural profitability, we order the qualitative statements into the following categories: 1. General ocular limitations. This category includes blurred vision and vision problems in general, like sensitivity to bright or dim light and limitations regarding reading and writing as well as riding a motorcycle. 2. Agricultural activity limitation. This category is specific to limitations in field work, like problems in seeing small insects, harvesting or using hand tools. 3. Limitation in access to information. This category entails limitations like reading newspapers, using a computer, reading calendars and clocks, as well as using a phone. 4. Physical discomfort symptoms are grouped and entail examples like squinting, a loss of balance or a pain in the eyes. 5. Limited social interactions. This entails examples of how people avoid crowded spaces, meeting people, attending social functions or recognizing faces. 6. Psychological symptoms and limitations. This refers to feelings of worry and depression, as well as nervousness and fears. 7. Limitations in business administration. Examples include making bank transfers, signing documents and recharging credit on the mobile phone. A detailed list of all limitations that are mentioned in the original study is given in Appendix C. This paper does not examine the effects of poor vision on all individual categories, but tries to estimate their aggregate effect.

Regarding the condition of poor visual acuity in Figure 4.1, poor visual acuity can be assessed by testing the performance of different components of the visual system. There are visual function tests that assess factors such as visual acuity, contrast sensitivity, color, and depth and motion perception. These properties each represent an aspect of visual function and impact an individual's level of functional vision. Visual acuity is the main component tested to assess the performance of the visual system and it is arguably the most crucial component of the visual system when it comes to working in agriculture (Bennett et al., 2019). Therefore, our study focuses on visual acuity as a proxy for visual function.

Figure 4.1: Poor visual acuity and its negative effects on agricultural profitability



The effects are derived from Kandel et al. (6), themes 1, 3, 4, 5, 6 and 7. Theme 2 was excluded because it deals exclusively with the negative effects of wearing glasses, contact lenses and corrective surgery. Source: Own depiction.

4.4 Methods

4.4.1 Ethics statement

When we conducted the fieldwork for this project in 2018, the University of Göttingen only had an internal review board system in place for clinical trials. Since our research does not qualify as a clinical trial, we were not eligible for an internal review. Anyhow, in close collaboration with our partners at The International Center for Tropical Agriculture (CIAT) and the Royal University of Agriculture (RUA) in Phnom Penh, we designed this research project according the principles of ethic responsibility in research involving human subjects and the national legislation of Cambodia. Since 2020, the University of Göttingen has an ethics committee that reviews and approves research that involves human subjects. We submitted our research protocol to them and obtained a retrospective approval.

Before visiting the villages, we met the commune leader and the government extension officers of the respective sub-province. The extension officer then accompanied us to each village and introduced us to the village chief, whom we presented our research endeavor in order to get his/her consent for data collection in his/her village. Only after receiving verbal consent from a) the commune chief, b) the extension officer, and c) the village chief were we able to undertake the data collection in the respective village. Before the survey, we explained to each participant what the research is about, what their participation in the project entails and that participation is voluntary. After this was understood we gathered written consent from each participant to be included in the research project. As a compensation for their lost time, the participants where paid equal to half a day of paid labor. We employed students from RUA who spoke English and Khmer to enable us to communicate with the farmers as well as other local stakeholders. We carried out intensive training sessions on survey methods which included the importance of explaining our research and obtaining informed written consent. One researcher was always present during data collection and she checked every questionnaire and if the protocol for obtaining consent was followed. The data was recorded by paper based surveys. The finished questionnaires were transcribed to an excel table by the researchers with codes for each observation so that re-identification without the paper based survey is not possible. The paper based surveys are archived at the University of Göttingen and only the authors of this article have access to them.

4.4.2 Data and descriptive statistics

We collect and explore cross-sectional primary data from household surveys and standardized eye examinations with 288 smallholder farmers from 16 villages throughout Ratanakiri province, Cambodia. The data was collected between August and October 2018 by a team of student enumerators from RUA in Phnom Penh, the University of Göttingen as well as staff from the provincial department of agriculture. We trained the enumerators and accompanied them during the data collecting process and conducted data quality checks. Ratanakiri province is remotely situated in northeastern Cambodia. This multi-ethnic province is categorized as one of the poorest areas in Cambodia (ADB, 2014). Of its 150,000 citizens, 88 percent live in rural areas and depend predominantly on income from agriculture. Rice is typically cultivated for household consumption whereas cassava, cashew, and rubber are the main cash crops (Burra et al., 2021; Paul et al., 2022; Ritzema et al., 2019). The target villages were selected by the managers of a greater project on sustainable farming practices in the region. Since there are no comprehensive lists of farming households in the villages, we relied on the expert knowledge of the extension workers from the regional government offices and the respective village officials to select participants based on a nonprobability sample (Levy & Lemeshow, 2008). The household surveys recorded detailed data on crop production for the growing season of 2017-2018. Observations with missing values were dropped from the data set. The final data set contains 260 observations. Table 4.1 displays the variables that are used in the estimations and their precise measurements.

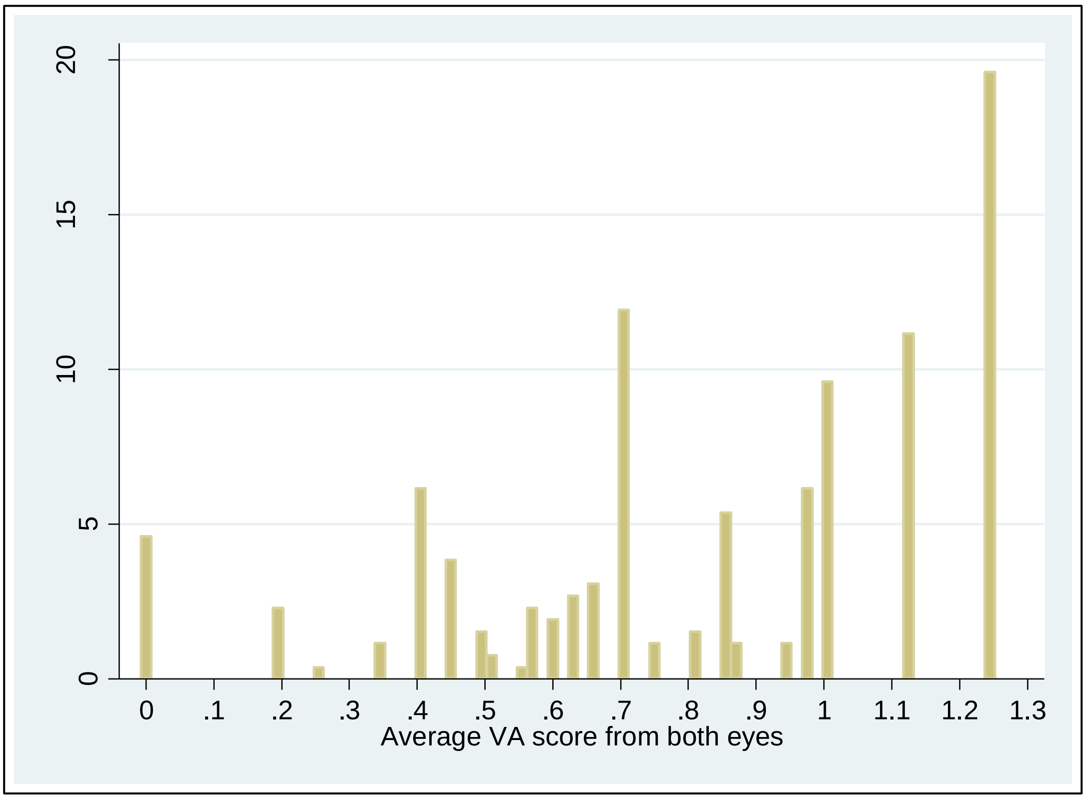
Table 4.1: Data description for selected variables

Variable	Description
Gross margin	All produce valued at average product prices minus cost for seeds, fertilizer, insecticides, fungicides, herbicides, machine hours, land and costs for hired labor for all cropping activities (transplanting, weeding, application of agrochemicals, harvesting and irrigation). Relates to the growing season 2017-2018. All values are transformed to USD/year
Single factor productivity	Calculates revenues per farm and year, divided by the area under cultivation for the growing season 2017-2018. All values are transformed to USD/ha/year
Eyesight	Calculates the results from Landolt C-Test, classifying respondents into "poor vision" and "good vision". The threshold is an average visus on both eyes≥0.7
Eyesight: Upper bound comparison	We shift the threshold of assignment to the "good vision" group to a visus≥0.75
Eyesight: Lower bound comparison	We shift the threshold of assignment to the "good vision" group down to a visus≥0.45
Age	Age in years
Area of cultivation	Total area of cultivation in hectares for all plots that belong to the farm
Education	Years in school
Household size	Number of people living in the household

Source: Own data.

To assess visual acuity of participants, we carry out a standardized eye examination, the Landolt–C test. This test is particularly useful to test young, illiterate or non-English speaking populations for visual acuity (Reich & Ekabutr, 2002). The participants of the test go through six lines of a vision chart, with each line corresponding to a value of visual acuity. The Landolt rings get smaller in each line so it becomes increasingly difficult to identify the rings and the test is stopped when the participant cannot recognize the rings anymore and the last ring that was identified is recorded as the result. Following the standard test procedure, there are 22 rings and 7 visual acuity groups: VA=0, VA=0.3, VA=0.4, VA=0.63, VA=0.7, VA=1, VA=1.25. The test is carried out first with the left eye, while blocking the right eye. Afterwards the test is repeated with the right eye while blocking the left eye. Also, the test is carried out without vision aids like glasses or contact lenses. In the following we use average VA score for both eyes. We also carry out calculations by taking the VA score for the weaker and stronger eye separately. The results do not differ from the results presented here. Figure 4.2 shows the raw results from the eye examination. Almost 5 percent of participants didn't identify a single ring and almost 20 percent of participants identified all 22 rings.

Figure 4.2: Raw results from the standardized eye examination



Number of observations=260. Displayed are the average values from both eyes. Source: Own depiction.

We use the visual acuity threshold of 0.7 in the decimal notation (visus) to classify participants into two groups: “poor vision” and “good vision”. We use this threshold because it is a widely accepted indicator for the performance of the visual system, i.e. it is used to verify if a person’s visual function is well enough to safely operate a vehicle (Bennett et al., 2019). In the context of our study, it is a suitable indicator because it applies a measure of visual acuity to a visual function. Thus, our indicator corresponds to visual functionality in everyday tasks, which connects to the idea of disadvantages in farm management activities for people who belong to the poor vision group. Another reason for selecting the two categories is that in practice, these categories resemble a real-world treatment. If we would give glasses to a person from our experiment, they could (theoretically) switch instantly from the poor vision group to the good vision group. Figure 4.3 displays the difference in test scores between the two groups. The poor vision group has an average VA score of 0.41 and the good vision group has an average VA score of 1.01.

Figure 4.3: Results from the standardized vision test by visual acuity group

Number of observations =260. Source: Own depiction.

From the household survey we collect production data on the growing season of 2017/2018. We record data on all arable crops of the household farm. In total, we recorded data from 543 individual plots on 260 farms. To aggregate the production of all commodities into a single measure, all produce is valued at average farm gate price in USD per year. Revenues here are not farm income, because not all produce is sold in the market; a large quantity of fruits, vegetables and rice is consumed by the household.

For each farm, we calculate the gross margin (USD/year) for the growing season 2017-2018:

Gross Margin (USD/ year) = Revenues – (Labor costs + Input costs + Land rent)

Table 4.2 displays the gross margins. Input costs include expenditure for fertilizer, pesticides, fungicides and insecticides, seeds, planting materials and hired labor for all cropping activities and rent for land, where we apply average land prices per hectare. The overview statistics are displayed in Table 4.2 for the good vision group and the poor vision group.

Table 4.2: Calculations of contribution margins

	Good vision				Poor vision			
	Mean	S.E.	95% CI		Mean	S.E.	95% CI	
Revenues[1]	2,265	314	1,648	2,883	2,058	218	1,629	2,486
Labor cost	129	23	83	175	199	41	118	281
Input cost	164	23	120	209	255	45	166	344
Land rent	514	33	450	578	674	51	574	774
Gross margin	1,466	296	883	2,050	934	163	613	1,255

[1]All calculations are based on cost benefit analysis of 260 farms for the growing season 2017/2018. Calculations are based on cropping activities on 543 single plots. Crops are cashew, cassava, fruits, maize, rice (upland), rubber, soybean and vegetables. All output is valued at average product price at farm gate in USD/year. Source: Calculated by the authors.

The mean value of revenues for the good vision group is 2,265 USD per year and 2,058 USD per year for the poor vision group. In terms of input costs, the poor vision group has higher costs of hired labor, higher input costs and higher land rent costs on average. The resulting contribution margins are higher for the good vision group with 1,466 USD per year on average, compared to the poor vision group with 934 USD per year on average.

Table 4.3 shows a mean comparison of the independent variables for the two groups. The most important point here is the variation in age between the two groups. In the good vision group, participants are on average 33 years old, compared to the poor vision group with an average age of 49 years. This negative relationship of age with visual acuity is to be expected, because visual acuity is fully developed at about 12 months of age and decreases over time (Pitts, 1982). We carry out a t-test and see that this difference is statistically significant. Another variable with a statistically significant difference between the groups is area of cultivation, where the good vision group has on average 3.17 ha of arable land, compared to the poor vision group with 4.20 ha.

Table 4.3: Mean comparison of key variables for the good vision and poor vision groups

Variable	Mean		T-test	
	Good vision	Poor vision	T-value	p>|t|
Age (years)	33	49	-10.63	<0.01***
Area of cultivation (ha)	3.17	4.20	-3.18	<0.01***
Education (years)	3.05	2.56	1.44	0.15
Gender (female=1)	0.56	0.61	-1.00	0.32
Household size (people)	5.33	5.17	0.62	0.53

Significance levels: * p<0.10, ** p<0.05, *** p<0.01, N=260. Source: Calculated by the authors.

4.4.3 Empirical strategy

According to the conceptual framework in Figure 4.1, this paper explores the effect of poor vision of farm managers on the profitability of agriculture. An important problem of causal inference is how to estimate treatment effects in observational studies, where (like an experiment) a group of units is exposed to a well-defined treatment, but (unlike an experiment) no systematic methods of experimental design are used to maintain a control group (Dehejia & Wahba, 2002). To circumvent this problem, we apply Mahalanobis Distance Matching (MDM) and Propensity Score Matching (PSM). Matching involves pairing treatment and comparison units that are similar in terms of their observable characteristics. These matching methods have become popular in impact evaluations and are used in a variety of fields, including to assess impacts related to agricultural production (Ahmed, Dompreh, & Gasparatos, 2019; Costedoat et al., 2015; Lawin & Tamini, 2019; Mishra, Kumar, Joshi, & D'Souza, 2018; Nakano, Tanaka, & Otsuka, 2018). Admittedly, the ideal data to answer our research questions would come from a controlled experiment, where agricultural profitability is quantified before and after treating poor vision, for example by giving out glasses to the participants. In our case, an adequately powered randomized control trial is not feasible due to reasons such as ethic concerns, time and cost. Instead, we apply a less expensive strategy to explore observational data that is naturally occurring in the field. This way, we apply a cost-efficient analysis to generate first results on the topic to motivate further research. Two conditions about remote areas in SEA enable this strategy: 1) a high and near-random incidence of "poor vision" among the target population and 2) the symptoms of poor vision are not treated by vision aids such as spectacles or contact lenses, which allows for a clear identification of treatment. Thus, we apply an estimation strategy that resembles a natural field experiment with regards to the near-random assignment to the group of poor vision and good vision. The strongest confounders in our data are observable, and we control for them in our estimations. This approach is restricted to the rural areas of most low-income countries, or anywhere the incidence of poor vision is high and health care infrastructure low (Mills, 2014).

We cannot measure the effect of belonging to the good vision group on agricultural profitability for each individual because we can only observe one outcome for each individual. Therefore, the focus of our analysis is on the average or population treatment effects, by using a potential outcome approach (Rubin, 1974). In our case it makes sense to investigate the average treatment effect on the treated (ATT), to explicitly evaluate the effects on the population for which the intervention is intended. This way we can estimate the realized gross gain that individuals get from having at least intact visual functions in terms of visual acuity. Put simply, our treatment effect resembles a possible real-world scenario, which would be a benchmark for the realized gross gain of an intervention to remedy poor visual acuity by prescribing glasses or contact lenses. The average treatment effect on the treated τ_{ATT} of our population is defined as follows:

$$\tau_{ATT} = E(\tau|D_i=1) = E(Y_{1i}\,|D_i=1) - E(Y_{0i}|D_i=1), \qquad (4.1)$$

where D_i is a binary variable equal to one if participant i passes the threshold of the vision test, zero otherwise, and Y_{1i} and Y_{0i} are the outcomes of the individuals with good vision and poor vision, respectively. The last term on the right-hand side of the equation is not observed, since it describes the hypothetical case of the outcome without treatment for the treatment group. To continue our estimation and find the true parameter for τ_{ATT} we replace the last term in formula (4.1) $E(Y_{0i}|D_i=1)$ with $E(Y_{0i}|D_i=0)$ so that $E(Y_{0i}|D_i=1)$ - $E(Y_{0i}|D_i=0)=0$.

To do this, we apply MDM and PSM We apply multiple matching methods, because it is recommended to use of several matching methods in combination with diagnostic checks to make a sample robust to the failures of individual methods (Rosenbaum, 2020). We create the missing counterfactual from the pool of observations in the poor vision group by observable characteristics x_i, which is highly dimensional. To reduce the problem of multidimensionality in matching, we match on a single index, the propensity score (Rosenbaum & Rubin, 1983). Matches are constructed on the basis of observed characteristics x_i of the poor vision group and the probability to belong to that group $Pr(D_i=1|x_i)=P(x_i)$ (Rosenbaum & Rubin, 1983). In the case of the MDM we calculate the inverse of the covariance matrix for all the covariates. Now, we can express the τ_{ATT} as:

$$\tau_{ATT}\ P(x_i)=E[Y_{1i}|D=1, P(x_i)] - E[Y_{0i}|D_i=0, P(x_i)]. \qquad (4.2)$$

We use two different PSM estimators to obtain the results, the first being Kernel Based Matching (KBM), the second being Nearest Neighbor Matching (NNM). KBM averages over multiple individuals in the poor vision group for each individual in the good vision group, with weights defined by their distance (Imbens, 2004), NNM is a one-to-one matching method where observations from the good vision group are assigned their closest match from the poor vision group. A major advantage of KBM is a lower variance, which is achieved because information from more observations is used. A drawback of the KBM method is that observations are possibly used that are bad matches (Caliendo & Kopeinig, 2008).

To assess matching quality, a balancing test is required. The algorithm we apply splits a sample into equally spaced intervals of propensity scores and then tests whether average propensity scores between treated and control units are different (Becker & Ichino, 2002). Tests continue until the average propensity scores of the good vision group and propensity scores of the poor vision group do not differ in each interval. If the means of each characteristic between the good vision group and the poor vision group for the same propensity score do not differ, the balancing test is satisfied. We restricted the algorithm to test in the area of common support (the area belonging to the intersection of the propensity score of good vision and poor vision), as this condition enhances the quality of matches in ATT estimation.

In terms of sensitivity analysis, we apply a method that can reveal robust baseline results by comparing our results with a model that includes a binary variable that is a proxy for a potential unobserved confounder (Ichino et al., 2008). This potential confounder can be simulated in the data and used as an additional covariate in combination with the preferred matching estimator. The comparison of the estimates obtained with and without matching on the simulated confounder show to what extent the baseline results are robust to specific sources of failure of the conditional independence assumption.

4.5 Results and discussion

4.5.1 Main results

In Table 4.4, the results from PSM are displayed. As expected, the effect of the variable *Age* is statistically significant and negatively correlated with our binary treatment variable. We restrict the model to the region of common support and 255 of 260 observations are in the region of common support (176 observations from the good vision group and 79 from the poor vision group). The model has a high degree of sensitivity (92.18 percent) and specificity (60.49 percent). For both, negative and positive predicted values, the model correctly classifies 82.31 percent of observations.

Table 4.4: Estimates from the PSM (treatment=good vision) with a probit model

Variable	Coefficient	Standard error	Z-value	P>Z
Age (years)	-0.06	0.01	-7.39	<0.01***
Area of cultivation (ha)	-0.38	0.03	-1.18	0.23
Education (years)	-0.04	0.03	-1.16	0.24
Household size (number of people)	0.08	0.04	1.85	0.06
Number of observations	260			
Sensitivity (%)	92.18			
Specificity (%)	60.49			
Positive predictive value (%)	83.76			
Negative predictive value (%)	77.78			
Correctly classified (%)	82.31			
LR chi2(7)	-18.40			
Pseudo R^2	0.27			
Observations on support (treatment)	176			
Observations on support (control)	79			

Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Sensitivity is the ratio of predicted positives/ actual positives and specificity is the ratio of predicted negatives/ actual negatives. Source: Calculated by the authors.

In Table 4.5 we compare the matched and unmatched samples. We use several parameters to assess matching quality. Apart from the mean values, we calculate the standardized differences of the original and matched samples. We can see that for the variable *Age*, matching reduces the standardized differences between the two samples to 18.00 percent, down from -133.80 percent in the original sample. The next highest standardized difference in the matched sample is reported for the variable *Education* with 13.70 percent. The test statistics show that the Rubin's B value, which is the absolute standardized difference of the means of the propensity score between the two groups, is 26.00 in the matched sample compared to 130.10 in the unmatched sample. This is slightly higher than the value of 25.00, which is an indicator for good matching quality (Rubin, 2001). Additionally, Rubin's R value gives the ratio of the treated to control variances of the propensity scores. The value of our matched sample is 1.64, which is in the satisfactory range between 0.5 and 2.0 (Rubin, 2001) and the value for the unmatched sample is 0.72. In summary, the results in Table 4.5 provide evidence of the reliability of the model that we selected and that matching significantly improved covariate balance.

Table 4.5: Covariate balance for the good vision and poor vision groups before and after MDM

Variables	Unmatched sample				Matched sample			
	T	C	T-value	Stand. Diff. %	T	C	T-value	Stand. Diff. %
Age	34.06	50.00	-9.58	-133.80	34.06	36.26	-0.39	-18.00
Area of cultivation	3.20	4.37	-3.08	-40.50	3.20	3.41	-0.78	-7.20
Education	3.10	2.34	1.72	24.50	3.10	2.70	1.26	13.70
Household size	5.29	5.34	-0.18	-2.50	5.29	5.19	0.49	4.60

Test statistics	Unmatched	Matched
Propensity score R^2	0.24	0.01
LR chi^2	74.73	6.06
$P>chi^2$	<0.01	0.19
Mean Bias	48.60	10.90
Rubin’s B	130.10	26.00
Rubin’s R	0.72	1.64

Mean values for the good vision group (T) and the poor vision group (C). Standardized differences are in percent. Rubin’s B is the absolute standardized difference of the means of the propensity score in the good vision and poor vision groups (unmatched and matched). Rubin’s R is the ratio of the good vision to poor vision variances of the propensity scores. Rubin’s B is good if < 25 and Rubin’s R is good if >0.5 and <2.0.N matched sample=255. N unmatched sample=260. Source: calculated by the authors.

Table 4.6 displays the main results from MDM, KBM and NNM. The ATT shows that farmers in the good vision group have net contribution margins that are 632.58 USD higher on average when compared to the poor vision group. The results from KBM are very similar with 627.20 USD/year, though with a lower statistical significance level. The NNM method gives us an ATT of 589.38 USD/year, which is slightly lower than for the other two methods. As opposed to KBM and MDM, NNM does not match on all controls which reduces sample size and inflates the Standard Error. In summary, the results obtained by all three methods are quite close to each other, and taken together give evidence of a positive ATT in the range of 589-632 USD/year associated with having at least intact visual functions in regards to everyday tasks.

Table 4.6: ATT comparison between good vision and poor vision groups with PSM and MDM

	Treatment	Control	ATT	SE	T-value
Mahalanobis Distance Matching	179	76	632.58	287.98	2.20
Kernel Density Matching	179	76	627.20	329.10	1.91
Nearest Neighbor Matching	179	40	589.38	445.96	1.32
Sensitivity analysis					
E-value: 1.94					
E-value CI: 1.24					
Critical level hidden bias: 1.25					

Total observations are 260, 5 drop out when enforcing area of common support. Source: Calculated by the authors.

4.5.2 Robustness checks

To assess the robustness of our results we apply two slightly different thresholds of belonging to the good vision group. For the lower bound comparison group, we lower the threshold of belonging to the good vision group to all average VA scores that are bigger or equal to 0.45. For the upper bound comparison group we raise this threshold to an average VA score of both eyes bigger or equal to 0.75. The results are displayed in Table 4.7. For the lower bound comparison group, the results from MDM are robust to the main estimations with an ATT of 679.00 USD/year. For the NNM in the lower bound comparison group, the control group is reduced to only 28 observations which inflate the standard errors and the ATT is not statistically significant. The upper bound comparison group is more balanced in terms of observations in the control and treatment groups and the calculations yield robust results when compared to the main results.

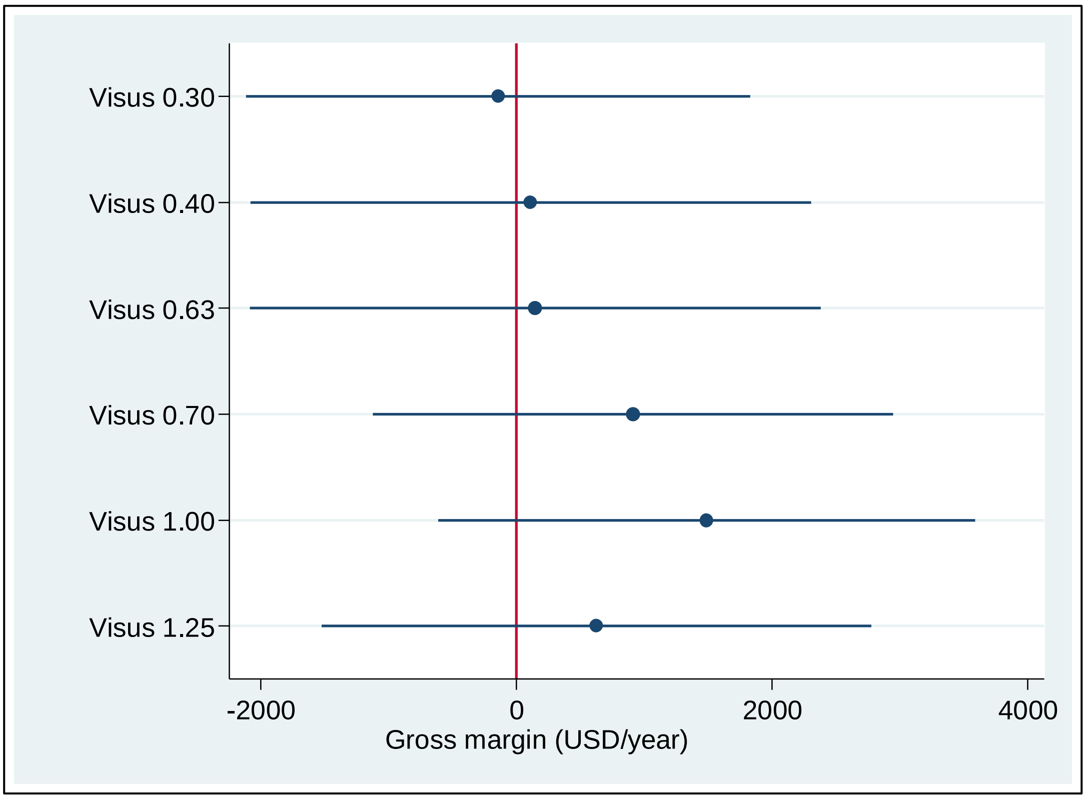
Table 4.7: ATT's for MDM, KBM and NNM with upper and lower bound comparison groups

	Treated	Control	ATT	SE	T-value
Lower bound comparison group (visus≥0.45)					
Mahalanobis Distance Matching	222	38	679.00	287.26	2.36
Kernel Density Matching	222	38	273.58	518.70	0.53
Nearest Neighbor Matching	222	28	135.34	707.29	0.19
Upper bound comparison group (visus≥0.75)					
Mahalanobis Distance Matching	148	108	617.16	365.44	1.69
Kernel Density Matching	148	108	682.90	337.04	2.03
Nearest Neighbor Matching	148	53	750.36	395.55	1.90

Total observations are 260, 5 drop out when enforcing area of common support. Source: Calculated by the authors.

To look deeper into the effects of the treatment variable, we code visual acuity as a categorical variable, where each line in the vision chart corresponds to one of the following seven visual acuity groups: VA=0.00; VA=0.30, VA=0.40, VA=0.63, VA=0.70, VA=1.00, VA=1.25. We conduct a regression with this factorial variable, holding all other control variables constant. Figure 4.4 displays the regression results where the group with a visus of 0 is the reference group. We can observe a stepwise increase in farm profitability up to the visual acuity group with a visus 1.00. The highest increase in farm profitability over the 0 visus group is observed in the group with a visus of 1.00. This advantage over the 0 visus group drops slightly in the group with the highest visus of 1.25. In general, we can observe a near linear increase of farm productivity with visual acuity, which confirms our results from the propensity score matching. Also, the difference between the first three groups is not very strong. But with a visual acuity score of 0.7 we see a sharp increase in contribution margins. Since our results are robust at the upper and lower bound levels, we can conclude that the VA score of 0.70 is a good indicator to assess the impacts of poor vision on farm profitability.

Figure 4.4: Results from OLS regression on gross margins, with visual acuity groups as independent factorial variable



Total observations =260. Reference group is visual acuity group with a visus of 0. Control variables are not displayed. Source: Own depiction.

As an additional robustness check we recalculate the outcome variable gross margins. As an alternative indicator we use single factor productivity in which our only input is cropping area per farm, an approach which can be easily interpreted, understood and calculated (FAO, 2018). More specifically, we multiply produce per plot by average product price in our sample divided by total crop area per farm for the growing season of 2017/2018 for all arable crops of the farm. Table 4.8 displays the results for the three visual acuity groups and matching algorithms. The ATT for the MDM is 243.17, which means that on average, a farmer in the good vision group earns 243.17 USD more per hectare. The results across groups and matching algorithms remain robust, with the lowest estimation of 182.06 USD/hectare and the highest estimation of 343.83 USD/hectare. Overall, the results obtained by all three methods and visual acuity thresholds taken together show a gain in single factor productivity in the range of 182.06-343.83 USD/hectare/year associated with having at least intact visual functions in regards to everyday tasks.

Table 4.8: Treatment effects of visual acuity on single factor productivity

	Treatment	T	C	ATT	SE	T-value
Mahalanobis Distance Matching	visus≥0.70	179	76	243.17	74.74	3.25
Kernel Density Matching	visus≥0.70	179	76	260.98	67.45	3.87
Nearest Neighbor Matching	visus≥0.70	179	40	237.75	109.00	2.18
Mahalanobis Distance Matching	visus≥0.45	220	24	320.39	69.12	4.63
Kernel Density Matching	visus≥0.45	220	35	320.39	67.11	4.77
Nearest Neighbor Matching	visus≥0.45	220	24	343.83	95.07	3.62
Mahalanobis Distance Matching	visus≥0.75	148	107	194.28	73.23	2.65
Kernel Density Matching	visus≥0.75	148	107	190.77	71.05	2.68
Nearest Neighbor Matching	visus≥0.75	148	54	182.06	96.40	1.89

The outcome variable single factor productivity is expressed in USD per hectare land and year for all plots of the farm. T=Treatment, C= Control. Source: Calculated by the authors.

The sensitivity analysis was carried out and we calculate a critical value for the Rosenbaum bounds of 1.25 (Table 4.6). For example, for the impact of good vision on gross margins, the sensitivity analysis suggests that at a level of 1.25, there is no hidden bias due to an unobserved confounder. In other words, if the odds of an individual belonging to the good vision group are 1.25 times higher because of the unobserved covariate, despite being identical on the matched (observed) covariate, there may be a change in inference. We can compare this number to the main observable confounder in our data. The variable *Age* alone explains roughly 25 of the variability in treatment status (See Table C21 in Appendix C). Thus, the unobserved confounder needs to have a bigger influence on treatment selection than the variable *Age*. For example, if we had an unobservable confounder like a genetic prevalence that influences treatment selection, this unobservable confounder needs to be unrealistically high. Thus, we assume that our estimates are robust to such unobservable confounders and that possible departures from randomization in our data are not big enough to explain away the pattern that poor eyesight leads to lower farm profitability. The E-value reported in Table 4.6 supports this result. The observed treatment effect could be explained away by an unmeasured confounder that was associated with both the treatment and the outcome by an effect of 2-fold each, above and beyond the measured confounders. The results have to be taken with caution since we cannot exclude the possibility of potential confounders which influence vision status and gross margins. This can lead to an upward-biased estimate, but we believe that the breadth and depth of our analyses show a clear association between economic farm performance and visual acuity.

4.6 Conclusion

Our study presents first results on the impacts and pathways of visual acuity on economic farm performance in rural Cambodia. We aim to present estimations on the maximum achievable treatment effect, i.e. to estimate how much profit is forgone because farmers are disadvantaged in managing their farms because of reduced visual functions.

Almost 30 percent of farmers in our sample suffer from poor vision. Furthermore, if a farm manager moves from poor vision to good vision, her gross margins would increase on average by around 630 USD per year. This effect is particularly outstanding considering the Cambodian gross national income per capita (GNI) is 1,380 USD (World Bank, 2018). The result is simple, as is the cure for the problem: glasses. With the help of glasses, a farm manager can potentially switch from poor vision to good vision instantly. According to our data, the economic benefits from this simple intervention can be enormous.

It is questionable if a real-world intervention would deliver a treatment effect of this size, because behavioral aspects would most likely reduce the treatment effect. For example, if a participant would be prescribed glasses, she perhaps wouldn’t wear glasses for all activities or couldn't use them for all activities equally. Wearing glasses in field work under direct sun could be practiced less if irritations like fogging and blurred vision due to sweat and dust outweigh the advantages of wearing glasses. Thus, we present estimations on the maximum achievable effect, against which real-world interventions can be measured.

It is clear though, that access to modern optometric services generate high returns on human capital with long lasting effects on educational attainment for example. We provide a framework that shows a variety of effects of poor vision on agricultural profitability. Despite the magnitude of the problem and its relatively cheap solution, which is modern eye care, the relationship between myopia and economic farm performance has received extremely little attention from development actors and researchers alike. Farm managers in the Global South are continuously challenged with many technology adoption issues. To make sound management decisions they require an intact visual system. Our results show that there are important linkages between agriculture and public health and that there is a need for more collaboration across the agricultural and public health sectors to address the negative impacts of ill-health on agricultural profitability.

In future research, a better identification of the causal relationships between myopia and farm profitability can be established by collecting longitudinal data. Future research should investigate (1) which entrepreneurial activities are most affected by poor vision and (2) which steps need to be taken to drive the usage of glasses. A repeated measure within-subjects design, e.g. a controlled experiment that applies pre- and post-measurement in relation to the treatment of glasses or contact lenses would be optimal for determining the causal effect of visual acuity on the economic performance of farms of smallholders.

References

Ahmed A, Dompreh E, Gasparatos A. (2019). Human wellbeing outcomes of involvement in industrial crop production: Evidence from sugarcane, oil palm and jatropha sites in Ghana. *PLoS ONE*, Vol. 14. 1– 33 p. https://doi.org/10.1371/journal.pone.0215433 PMID: 31022186

Ajani OIY, Ugwu PC. (2008). Impact of Adverse Health on Agricultural Productivity of farmers in Kainji Basin North-Central Nigeria Using a Stochastic Production Frontier Approach. *Trends in Agricultural Economics*, 1 (1):1–7.

Antle JM, Capalbo SM. (1994). Pesticides, Productivity, and Farmer Health: Implications for Regulatory Policy and Agricultural Research. *American Journal of Agricultural Economics*, 76(3):598–602.

Antle JM, Pingali PL. (1994). Pesticides, Productivity, and Farmer Health: A Philippine Case Study. *American Journal of Agricultural Economics*, 76(3):418–30.

Audibert M, Etard JF. (2003). Productive benefits after investment in health in Mali. *Economic Development and Cultural Change*, 51(3):769–82.

Assi L, Chamseddine F, Ibrahim P, Sabbagh H, Rosman L, Congdon N, et al. (2021). A global assessment of eye health and quality of life a systematic review of systematic reviews. *JAMA Ophthalmology*,139 (5):526–41. https://doi.org/10.1001/jamaophthalmol.2021.0146 PMID: 33576772

BBC (2018). Khmer Rouge: Cambodia's years of brutality [Internet]. BBC News. https://www.bbc.com/ news/world-asia-pacific-10684399. Date accessed: 29.06.2022

Becker SO, Ichino A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal*, 2(4):358–77.

Bennett CR, Bex PJ, Bauer CM, Merabet LB. (2019). The Assessment of Visual Function and Functional Vision. *Seminars in Pediatric Neurology*, 2019; 31:30–40. https://doi.org/10.1016/j.spen.2019.05.006 PMID: 31548022

Bloom DE, Canning D. (2000) The health (and wealth) of nations. *Science*, 287:1207–9. 39.

ADB. (2014). Cambodia: Country Poverty Analysis 2014. Asian Development Bank, Mandaluyong City, Philippines. Retrieved from: https://www.adb.org/sites/default/files/institutional-document/151706/cambodia-country-poverty-analysis-2014.pdf

Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli M V., Das A, Jonas JB, et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Global Health*, 5(9):e888–97.

Burra DD, Parker L, Than NT, Phengsavanh P, Long CTM, Ritzema RS, et al. (2021). Drivers of land use complexity along an agricultural transition gradient in Southeast Asia. *Ecological Indicators*, 124.

Burton MJ, Faal HB, Ramke J, Ravilla T, Holland P, Wang N, et al. (2019). Announcing the Lancet Global Health Commission on Global Eye Health. *Lancet Global Health*, 7(12):e1612–3. https://doi.org/10. 1016/S2214-109X(19)30450-4 PMID: 31606327

Burton MJ, Ramke J, Marques AP, Bourne RRA, Congdon N, Jones I, et al. (2020). The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. Supplementary Appendix 1. *Lancet Global Health*, 9(4):e489–551.

Caliendo M, Kopeinig S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.

Costedoat S, Corbera E, Ezzine-de-Blas D, Honey-Rose J, Baylis K, Castillo-Santiago MA. (2015). How effective are biodiversity conservation payments in Mexico? *PLoS One*, 10(3):1–20. https://doi.org/10. 1371/journal.pone.0119881 PMID: 25807118

Christiaensen L, Demery L, Kuhl J. (2011). The (evolving) role of agriculture in poverty reduction- An empirical perspective. *Journal of Development Economics*, 96(2):239–54.

Dandona L, Dandona R. (2006). What is the global burden of visual impairment? *BMC Medicine*, 4.

Dehejia RH, Wahba S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, 84(1):151–61.

De Janvry A, Sadoulet E. (2009). Agricultural Growth and Poverty Reduction: Additional Evidence. *World Bank Research Observer,* 25(1):25:1–20.

Enoch J, McDonald L, Jones L, Jones PR, Crabb DP. (2019). Evaluating Whether Sight Is the Most Valued Sense. *JAMA Ophthalmology*, 137(11):1317–20. https://doi.org/10.1001/jamaophthalmol.2019.3537 PMID: 31580383

FAO. (2018). Guidelines for the measurement of productivity and efficiency in agriculture. Food and Agriculture Organization, Rome. Retrieved from: https://www.fao.org/3/ca6428en/ca6428en.pdf

Finger RP, Kupitz DG, Fenwick E, Balasubramaniam B, Ramani R V., Holz FG, et al. (2012). The Impact of Successful Cataract Surgery on Quality of Life, Household Income and Social Status in South India. *PLoS One*, 7(8):1–7. https://doi.org/10.1371/journal.pone.0044268 PMID: 22952945

Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli M V., et al. (2017). Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Global Health*, 5(12):e1221–34. https://doi.org/10.1016/S2214-109X(17)30393-5 PMID: 29032195

Frick KD, Foster A. (2003). The magnitude and cost of global blindness: An increasing problem that can be alleviated. *American Journal of Ophthalmology*, 135(4):471–6. https://doi.org/10.1016/s0002-9394(02)02110-4 PMID: 12654362

Glewwe P, Park A, Zhao M. (2016). A better vision for development: Eyeglasses and academic performance in rural primary schools in China. *Journal of Development Economics*, 122:170–82.

Gordois A, Cutler H, Pezzullo L, Gordon K, Cruess A, Winyard S, et al. (2012). An estimation of the worldwide economic and health burden of visual impairment. *Global Public Health*, 7(5):465–81. https://doi. org/10.1080/17441692.2011.634815 PMID: 22136197

Gupta A, Casson RJ, Newland HS, Muecke J, Landers J, Selva D, et al. (2008). Prevalence of Refractive Error in Rural Myanmar. The Meiktila Eye Study. *Ophthalmology*, 115(1):26–34.

Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. (2016). Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. *Ophthalmology*, 123 (5):1036–42. https://doi.org/10.1016/j.ophtha.2016.01.006 PMID: 26875007

Jaggernath J, Øverland L, Ramson P, Kovai V, Chan VF, Naidoo KS. (2014). Poverty and Eye Health. *Health*, 06(14):1849–60.

Ichino A, Mealli F, Nanncini T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Economics*, 23:305–27.

Imbens GW. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Kandel H, Khadka J, Shrestha MK, Sharma S, Neupane Kandel S, Dhungana P, et al. (2018). Uncorrected and corrected refractive error experiences of Nepalese adults: a qualitative study. *Ophthalmic Epidemiology*, 25(2):147–61. https://doi.org/10.1080/09286586.2017.1376338 PMID: 28985110

Klasen S, Reimers M. (2017). Looking at Pro-Poor Growth from an Agricultural Perspective. *World Development*, 90:147–68.

Kuper H, Polack S, Mathenge W, Eusebio C, Wadud Z, Rashid M, et al. (2010). Does cataract surgery alleviate poverty? Evidence from a multi-centre intervention study conducted in Kenya, the Philippines and Bangladesh. *PLoS One*, 5(11).

Lawin KG, Tamini LD. (2019). Tenure Security and Farm Efficiency Analysis Correcting for Biases from Observed and Unobserved Variables: Evidence from Benin. *Journal of Agricultural Economics,* 70(1):116–34.

Levy PS, Lemeshow S. (2008). Sampling of populations: methods and applications. 4th ed. Hoboken, N.J.: Wiley. 576 p.

Loureiro ML. (2009). Farmers' health and agricultural productivity. *Agricultural Economics,* 40(4):381–8.

Lowder SK, Skoet J, Raney T. (2016). The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development,* 87:16–29.

Marques AP, Ramke J, Cairns J, Butt T, Zhang JH, Muirhead D, et al. (2021). Global economic productivity losses from vision impairment and blindness. *EClinical Medicine,* 35:100852. https://doi.org/10. 1016/j.eclinm.2021.100852 PMID: 33997744

Mills A. (2014). Health Care Systems in Low- and Middle-Income Countries. *New England Journal of Medicine*, 370(6):552– 7. https://doi.org/10.1056/NEJMra1110897 PMID: 24499213

Mishra AK, Kumar A, Joshi PK, D'Souza A. (2018). Impact of contract farming on yield, costs and profitability in low-value crop: evidence from a low-income country. *Australian Journal of Agricultural and Resource Economics,* 62(4):589–607.

Morchen M, Langdon T, Ormsby GM, Meng N, Seiha D, Piseth K, et al. (2015). Prevalence of blindness and cataract surgical outcomes in Takeo province, Cambodia. *Asia-Pacific Journal of Ophthalmology,* 4(1):25–31. https://doi.org/10.1097/APO.0000000000000061 PMID: 26068610

Nakano Y, Tanaka Y, Otsuka K. (2018). Impact of training on the intensification of rice farming: evidence from rainfed areas in Tanzania. *Agricultural Economics*, 49(2):193–202.

Paudel P, Ramson P, Naduvilath T, Wilson D, Phuong HT, Ho SM, et al. (2014). Prevalence of vision impairment and refractive error in school children in Ba Ria-Vung Tau province, Vietnam. *Clinical and Experimental Ophthalmology*, 42(3):217–26. https://doi.org/10.1111/ceo.12273 PMID: 24299145

Paul BK, Epper CA, Tschopp DJ, Long CTM, Tungani V, Burra D, et al. (2022). Crop-livestock integration pro- vides opportunities to mitigate environmental trade-offs in transitioning smallholder agricultural systems of the Greater Mekong Subregion. *Agricultural Systems*, 195:103285.

Pitts DG. (1982). Visual acuity as a function of age. *Journal of the American Optometric Association*, 53(2):117–24. PMID: 7069103

Reddy PA, Congdon N, MacKenzie G, Gogate P, Wen Q, Jan C, et al. (2018). Effect of providing near glasses on productivity among rural Indian tea workers with presbyopia (PROSPER): a randomised trial. *Lancet Global Health*, 6(9):e1019–27. https://doi.org/10.1016/S2214-109X(18)30329-2 PMID: 30049615

Reich LN, Ekabutr M. (2002). The effects of optical defocus on the legibility of the Tumbling-E and Landolt-C. *Optometry and Vision Science,* 79(6):389–93. https://doi.org/10.1097/00006324-200206000-00013 PMID: 12086306

Ritzema R, Douxchamps S, Fraval S, Bolliger A, Hok L, Phengsavanh P, et al. (2019). Household-level drivers of dietary diversity in transitioning agricultural systems: Evidence from the Greater Mekong Subregion. *Agricultural Systems*, 38(4):467–86

Rosenbaum PR. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application* , 7:2.1–2.34.

Rosenbaum PR, Rubin DB. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–5.

Rubin D.B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2:169–88.

Rubin D. B. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Sabasi D, Shumway CR. (2018). Climate change, health care access and regional influence on components of U.S. agricultural productivity. *Applied Economics*, 50(57):6149–64.

Smith AF, Smith JG. (1996). The economic burden of global blindness: A price too high! *British Journal of Ophthalmology*, 80(4):276–7. https://doi.org/10.1136/bjo.80.4.276 PMID: 8703872

Tahhan N, Papas E, Fricke TR, Frick KD, Holden BA. (2013). Utility and uncorrected refractive error. *Ophthalmology*, 120(9):1736–44. https://doi.org/10.1016/j.ophtha.2013.02.014 PMID: 23664469

Tan Y, Guo B, Nygaard S, Carillo C, Pham H-D, Nouansavanh KO, et al. (2022). Prevalence and causes of visual impairment and blindness in Lao People's Democratic Republic: the Vientiane Eye Study. *British Journal of Ophthalmology*, https://doi.org/10.1136/bjophthalmol-2021-320127 PMID: 35354562

Yiengprugsawan V, Seubsman S-A, Sleigh A. (2011). Epidemiological associations of vision impairment and health among a national cohort of 87,134 adults in Thailand. *Journal of Epidemiology and Community Health*, 65 (Suppl 1):A373–A373.

Zhang JH, Ramke J, Jan C, Bascaran C, Mwangi N, Furtado JM, et al. (2022). Advancing the Sustainable Development Goals through improving eye health: a scoping review. *Lancet Planetary Health*, 6(3): e270–80. https://doi.org/10.1016/S2542-5196(21)00351-X PMID: 35219448

5 Where can the crow make friends? Sci-Hub's activities in the library of development studies and its implications for the field

Abstract

This study examines data on the worldwide use of the shadow library website Sci-Hub. It focuses particularly on the discipline of development studies, taking a critical look at current practices in scientific publishing and their implications for scientific conduct in this field. In the context of discussions about open science, the data demonstrate that Sci-Hub represents an existing network of open access literature. The study first describes the extent and geographic distribution of download requests from Sci-Hub and then identifies the underlying socio-economic drivers (i.e. GDP and population). The authors find that Sci-Hub is used the most by researchers from the Global South, primarily from middle-income countries, whereas researchers from the poorest countries in the data set use Sci-Hub the least. This approach allows some conclusions to be drawn about factors that have an effect on a functioning open access network.

This chapter is coauthored by Frederik Sagemüller (FS), Luise Meißner (LM) and Oliver Mußhoff (OM). The contributions to the article are: FS conceptualized the idea, analyzed the data, and wrote the article. LM unpacked the download logs. OM and LM reviewed the article. This chapter is published in the Journal *Development and Change* with the following citation: Sagemüller, Meißner & Mußhoff (2021). Where can the crow make friends? Sci-Hub's activities in the library of development studies and its implications for the field. *Development and Change*, https://doi.org/10.1111/dech.12638

5.1 Introduction

The scientific community is engaged in a heated debate over the ethical implications of scientific publishing. The major argument made by the proponents of open access to science is that scientific publishing under the present system actually harms science. The gist behind this argument is that the research process is facilitated by ensuring rapid and widespread access to research findings, such that all communities have the opportunity to build upon them and to participate in scholarly conversations (Tennant et al., 2016). Thus, if scientists are denied access to high-quality scientific content, they are hindered from producing high-quality content themselves.

Recent trends in the market for journal subscriptions have made access to scientific journals increasingly expensive, meaning that poorer institutions and scientists are more likely to be excluded. Data from the Association of Research Libraries from 1986 to 2005 reveal an average annual price increase of 7.6 percent for all serials, with total expenditures for journal subscriptions increasing by 302 percent (McGuigan and Russel, 2008). At the same time, funds available to libraries of public institutions have remained static or even declined when viewed in real terms (Guarria and Wang, 2011). As a backdrop to these trends, the rise of tertiary education around the world, widespread access to broadband internet and the growth of social media, have all fed into globalization of free science and education.[2]

In 2011, Sci-Hub, a project managed by a Kazakhstani computer programmer, Alexandra Elbakyan, started to provide free access to a vast amount of journal articles. Known as a shadow library or pirate website, Sci-Hub does not restrict itself to openly licensed content but instead it retrieves and distributes scholarly literature without regard to copyright. Sci-Hub has been growing rapidly since its creation, and by March 2017, its database contained 85

[2] See Our World in Data 'Tertiary Education' (https://ourworldindata.org/tertiary-education), 'How many internet users does each country have?' (https://ourworldindata.org/how-many-internet-users-does-each-country-have) and 'The rise of social media' (https://ourworldindata.org/rise-of-social-media), accessed on 15.12.2020.

percent of articles published in toll-access journals (Himmelstein et al., 2018). In 2016, Elbakyan published data showing Sci-Hub traffic over a six-month period (see Bohannon and Elbakyan, 2017). Bohannon (2016) has analysed these log data from Sci-Hub servers; his study shows that researchers from all continents download articles via Sci-Hub. Most surprisingly, perhaps, articles are also downloaded by prestigious US and European universities that actually grant access to most journals. Bohannon (2016) concludes that not only need, but also convenience play a role in why researchers use Sci-Hub.

The analysis of aggregated data hides interesting patterns within specific disciplines. The field of development studies presents a special case in this discussion, with researchers raising the question of how complicit they are in upholding and reproducing racial hierarchies that underpin development studies (Pailey, 2020). Publications in highly ranked development journals use data which is collected 'in the field', which very often means in rural areas of low-income countries. The results and conclusions generated by this data, however, remain inaccessible for scientists from those locations, unless they are in the privileged position of working at an institute that has subscription access to these articles, or have enough money to pay on a per-article basis (Tennant et al., 2016). Thus, a separate analysis for the discipline of development studies is warranted, to revisit the long-standing hypothesis that piracy is a means for researchers from developing economies to bypass unjust hierarchies and access results from studies which are based on data from their home countries. Additionally, certain salient facts are not widely known, including the size and geographic distribution of Sci-Hub's network in the discipline of development studies and its socio-economic drivers.

It is the aim of this study to fill this research gap by describing Sci-Hub's activities in the field of development studies. We achieve this by screening through 28 million download logs in Sci-Hub for articles in the field of development studies; we identify the geographic location of download requests; and we use country metadata (i.e. GDP and population) to reveal factors that influence the use of Sci-Hub. To the best of our knowledge, our contribution is the first from the social sciences to cover a comprehensive list of journals for a specific subject category and to enrich the Sci-Hub data with country metadata to draw conclusions on influencing factors of Sci-Hub's use. It is our hope that our results can help to inform future applications of open access systems.

5.2 Literature review

5.2.1 Sci-Hub

Informal systems of copyright infringement often have been a topic of interest for researchers in academia. As a researcher, it is a common practice to share journal articles from toll-access journals with students and colleagues (Bodó, 2016). These favour-based systems of copyright infringement never really bothered publishers, nor did they do enough economic damage to provoke publishing houses to act against them. From the publishers' perspective, academic journals represent one of the most profitable investments, with profit margins reaching 36 percent and global revenues exceeding USD 24 billion in 2011 (Buranyi, 2017). The explosion of piracy via internet libraries, however, has been a game changer (ibid.).

The institution that spearheaded this online movement is Sci-Hub. In 2011, Sci-Hub began to provide free access to a vast amount of journal articles in a clear violation of copyright law. Researchers from all over the world have downloaded millions of copies of paywalled journal articles. In an interview with the magazine The Verge in 2018, Elbakyan states that she was participating in informal sharing networks in online communities as a graduate student, because she could not afford to purchase scientific content from journals (Graber-Stiehl, 2018). At this time, she also began responding to requests from other researchers by pirating journal articles. From her point of view, pirating paywalled journal articles constitutes a legitimate form of civil disobedience (Banks, 2016). She argues that a lack of universal access to scientific content violates Article 27 of the United Nations Universal Declaration of Human Rights, which states that 'Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits' (UN General Assembly, 1948). In this context, Elbakyan developed Sci-Hub as an automated process of informal sharing for the whole research community. Sci-Hub's contribution is to make the process instantly available to everyone with a very high degree of reliability.

According to Elbakyan, she obtains login credentials from researchers who are frustrated with the status quo of scientific publishing (Bohannon, 2016). These credentials enable Sci-Hub to use institutional networks as proxies and gain access to journals. Elbakyan refuses to disclose the sources of such login credentials but states that they have been given willingly by people from around the world (Banks, 2016). There also have been reports of phishing and hacking, although Elbakyan (2017) denies that Sci-Hub would directly engage in such practices. As Banks (2016) states, the most plausible conclusion is that Sci-Hub has obtained credentials through a combination of willing donations and more nefarious means. Elbakyan (2017) insists that the credentials are only used to download journal articles.

In 2013–14, Sci-Hub relied on the Library Genesis (LibGen) Scimag repository to store articles (Himmelstein et al., 2018). LibGen has unique features; its mission is to provide access to the collection by being radically open. LibGen's main focus is the distribution of its own library infrastructure, including source code, catalogue and terabyte sized collection to anyone who wants to start their own library. This openness has led to the creation of a lively ecosystem of shadow libraries via mirror sites. The mirror sites deliver the LibGen collection to the public and, at the same time, increase the likelihood of LibGen's long-term survival (Bodó, 2018). The mirror sites also serve as lightning rods for lawsuits: if one site is shut down, another mirror site surfaces to replace it. Thus, there is little reason to believe that law enforcement agencies will be capable of shutting down Sci-Hub in the near future (Elbakyan, 2016; Hoy, 2017).

Clearly, the demand for open access websites like Sci-Hub is there and the idea is brilliantly simple. To download a journal article, one needs only to go to the website, check the availability of a mirror site, and follow the procedure to retrieve a copy of a paywalled article. In the formal market, this copy would cost USD 30 or more (Hoy, 2017), which represents a substantial barrier for many researchers. However, the price incentive is not the only thing driving researchers to download pirated journal articles; the attraction of convenience is another draw. As Bohannon (2016) shows, researchers from US institutions who have access through their official library systems often use Sci-Hub instead. Further evidence for this is the availability of open access articles on Sci-Hub (Babutsidze, 2016). This is because Sci-Hub is simpler in its use, with three specific advantages. First, Sci-Hub requires only one identifier like the DOI to be directly referred to an article; a library system might need the names of the authors, the journal name, the article name, volume, issue etc. Second, in Sci-Hub the access to the desired article is almost certain, whereas in the library system is still a chance that the library does not have full access to the volume the researcher needs (Himmelstein et al., 2018). Third, articles can be downloaded without a VPN-client or a connection to institutional servers.

It is not the aim of this article to advocate for or against Sci-Hub. Sci-Hub's activities represent a clear violation of copyright law, and the use of Sci-Hub constitutes copyright infringement in many jurisdictions. In fact, the United States District Court of the Southern District of New York has ruled against Sci-Hub, citing as violations of US copyright law the unlawful access to, use, reproduction, and distribution of Elsevier's copyright works.[3] In the hallways of universities, however, Sci-Hub is still dubbed 'the Robin Hood of science'

[3] See the United States District Court, SD of NY, 2015. Case 1:15-cv-04282-RWS.

(Oxenham, 2016), and its activities are seen as the result of a moral imperative (Swartz, 2008).

5.2.2 Problems of access to scientific publications and implications for low income countries

As described in literature, current practices in scientific publishing can actually harm science, with far-reaching consequences for society. The research capacity of a society is known to have a profound effect on its economic development and its ability to address problems in such areas as public health, infectious diseases, agriculture, environmental management, or industrial progress (Kirsop and Chan, 2005). Publications in highly ranked development studies journals use data that is collected ‘in the field’, which often refers to rural areas of low-income countries. The results and conclusions generated by this data, however, often remain inaccessible not only for scientists but also for policy makers and their technical advisors from those locations. Thus, important policy recommendations made in scientific publications remain unheard.

In an open letter to the editor of the Indian Journal of Community Medicine, Deshpande and Naik (2012) point out that payments to a high impact factor journal, either by researcher or reader, obstruct the flow of scientific information. In their view, economically constrained researchers are forced to publish in lower-ranked journals that charge the reader rather than the author, which in turn reduces the reach of the article. This contributes to a performance gap between researchers in low- and high-income countries. Gibbs (1995) has shown that both north–south and south–north knowledge gaps exist. The resulting invisibility of research from developing countries contributes to an incomplete picture of a scientific discipline. According to Gibbs (1995), just 2 percent of participation in international scientific discourse accounts for 80 percent of global scientific output (ibid.).

In short, the ability to pay still determines how scientific contributions are disseminated, with the consequence that researchers in developing economies have limited access to the scientific literature they need (Kirsop and Chan, 2005). For many scholars and scientists, accessing literature on a day-to-day basis is onerous and confronts them with a moral dilemma. Bendezú-Quispe et al. (2016) describe a case that reflects the challenges faced by many researchers from low-income countries, and raise the question: is it ethical for economically marginalized researchers to use an illegal medium (such as Sci-Hub) to access information for the greater good? Recent data suggest that many would adopt a utilitarian approach in answering this question: for example, a study on the sources for scientific investigation across six Latin American countries shows that about 62 percent of medical students use Sci-Hub (Mejia et al., 2017).

It is important to stress that although Sci-Hub opens access to paywalled journal articles, it does not change the publishing system itself. In fact, Harrison et al. (2018) argue that Sci-Hub does the opposite, because it enables access to journal articles and only serves to reinforce the notion that these final, peer-reviewed articles are de facto the currency of science. This perversely enhances the status of prestige publications, rather than working towards a system of open science which is based on transparency, collaboration and open source. As Priego (2016) states, there is no real cultural change: because digital copies are reproducible ad infinitum at negligible cost, commercial publishers profit from the consequences of citations, rankings, reputation and legacy. Thus, it is not clear whether or to what extent Sci-Hub improves the ability of developing-country scholars to publish, especially when putting Sci-Hub in the context of the open science movement. Most notably, a number of key attributes of open science that are discussed in the literature are not addressed by Sci-Hub: (1) eliminating the use of journal-based metrics, such as journal impact factors; (2) changing the peer review system towards an ongoing post-publication process of transparent peer review and rating of articles; (3) making underlying data, research methods and research documentation available to the public to enhance reproducibility; and (4) opening up licensing agreements to improve re-use of outputs through data mining. Thus, Sci-Hub does not ultimately solve the problem for researchers who are disadvantaged by non-transparent hierarchies and practices within the current system of scientific conduct, but rather eases the symptom of immediate exclusion by paywalls.

5.2.3 Related research on Sci-Hub use around the world

The data provided by Bohannon and Elbakyan (2017) on Sci-Hub traffic opens up the opportunity to look at how the scientific community makes use of Sci-Hub, and to test the hypothesis that piracy is a means for researchers from low-income countries to bypass unjust hierarchies. Bohannon (2016) shows that researchers from all continents download articles via Sci-Hub. Greshake (2016) analyses the same data set and finds that per capita GDP is positively correlated with downloads. Most surprising to these researchers was the fact that articles are also downloaded by researchers at prestigious US and European universities that grant full access to most journals. A calculation of IP ranges shows that about 10 percent of Sci-Hub downloads were made from university campuses, with most traffic occurring during working hours and less traffic over the weekend (Greshake, 2016). Thus, the received wisdom that piracy primarily serves researchers from low-income countries to bypass unjust hierarchies cannot be confirmed from these studies. However, the aggregated data are skewed towards specific disciplines: Himmelstein et al. (2018) show that physical sciences and engineering together with life sciences achieve the greatest coverage within Sci-Hub’s library, and that downloads favour newer publications.

Since very few studies have looked into the data for specific sub-disciplines, the aggregated data set might hide interesting patterns within certain disciplines. Timuș and Babutsidze (2016) use the Sci-Hub download logs to examine patterns of Sci-Hub use in European studies research. To do so, they screen the data for the six most prestigious journals in the field and identify 2,310 downloads on 1,537 distinct manuscripts. They find that the non-European countries Brazil, China and the USA belong to the top 15 countries that are active in illegal downloads. Babutsidze (2016) examines the same data with a focus on economics, analysing downloads from the top five economics journals. The study identifies 2,147 downloads of economics articles from these journals, which is only 0.009 percent of the whole data set. To test the poor-country enabler status of Sci-Hub, they add the geographic composition of downloads and find that most downloads are from low-income countries. Similar to the aggregated data presented by Bohannon (2016), OECD countries including France, Germany and the USA show up in the list, as well as the BRIC states Brazil, China and Russia. Indonesia, Iran, Malaysia and Pakistan all make it into the top 10 of downloading countries.

The one study that analyses a whole catalogue for a specific discipline is Androcec (2017). This study screens the data for a comprehensive list of 86,000 journal volumes, conference proceedings and monographs in the discipline of computer science. In total, download data were retrieved on 607,023 computer science texts. The five countries with the most downloads were China, India, Indonesia, Iran and the USA.

5.3 Data and methods

The data set from Bohannon and Elbakyan (2017), which can be downloaded from the open access data portal Dryad, was used to analyse download requests from development journals per country. The data set contains all Sci-Hub downloads from September 2015 to February 2016, which amounts to about 28 million requests. To isolate the field of development studies within the data, we referred to the SCImago Journal and Country Rank. These data are publicly available and include the journals and country indicators developed from information contained in the Scopus database. In total, there were 152,918 documents. Within this database, journals are grouped into 27 major thematic areas: one of these subject categories is the discipline of development studies. We downloaded this list, which contains 230 journals, proceedings and book chapters. We identified the DOI number for 190 of these journals. Observations where the variable 'country' was not available were deleted. This left us with a dataset with all geo-referenced Sci-Hub requests from 190 development journals, containing 43,909 observations. These data were than grouped by country. We enriched the data with country-level socio-economic indicators retrieved from the World Bank: Population and GDP per capita. [4]

5.4 Results and discussion

We first want to show the supply side of academic publishing in the field of development studies. The first panel of Table 5.1 shows the documents which are produced in this field by country. The only non-OECD countries that make it into this list are China, India and South Africa. The USA is in first place with 34,234 documents, followed by the United Kingdom and Australia with 19,650 and 7,151 documents, respectively. If we normalize the data by population size, China and India drop out of the top 10. If we group the data by quartiles of GDP, we see that the upper quartile provides the vast majority of documents to the discipline, with a sharp drop in the third quartile. Thus, the richest nations in the world provide 70 percent of total contributions to the discipline of development studies. This confirms the assumption that research from developing countries in the current system is largely invisible. These numbers also give a point of reference for the size of the catalogue from which the 43,909 Sci-Hub download requests stem.

[4] See World Bank Development Indicators (https://databank.worldbank.org/source/world-development-indicators), accessed 23.10.2019.

Table 5.1: Country statistics for documents provided to the catalogue of development studies, and Sci-Hub downloads taken from it

Documents per country			**Downloads per country**		
Rank	Country	Documents	Rank	Country	Downloads
1.	United States	34,234	1.	Iran	5,280
2.	UK	19,650	2.	Indonesia	4,576
3.	Australia	7,151	3.	China	2,687
4.	India	5,796	4.	Malaysia	2,598
5.	Canada	5,709	5.	India	2,487
6.	China	5,410	6.	Pakistan	2,402
7.	Germany	5,000	7.	United States	1,997
8.	France	4,418	8.	Brazil	1,641
9.	Netherlands	4,215	9.	Morocco	1,329
10.	South Africa	3,884	10.	Tunisia	1,152

Documents per capita			**Downloads per capita**		
Rank	Country	Docs./ 1,000 capita	Rank	Country	Downl./ 1,000 capita
1.	Norway	.320	1.	Tunisia	.102
2.	UK	.300	2.	Malaysia	.085
3.	Australia	.296	3.	Iran	.066
4.	Netherlands	.247	4.	Portugal	.051
5.	Hong Kong	.206	5.	Morocco	.038
6.	Sweden	.199	6.	Netherlands	.029
7.	Switzerland	.196	7.	Chile	.022
8.	Canada	.158	8.	Indonesia	.017
9.	Belgium	.134	9.	Germany	.012
10.	United States	.106	10.	Pakistan	.011

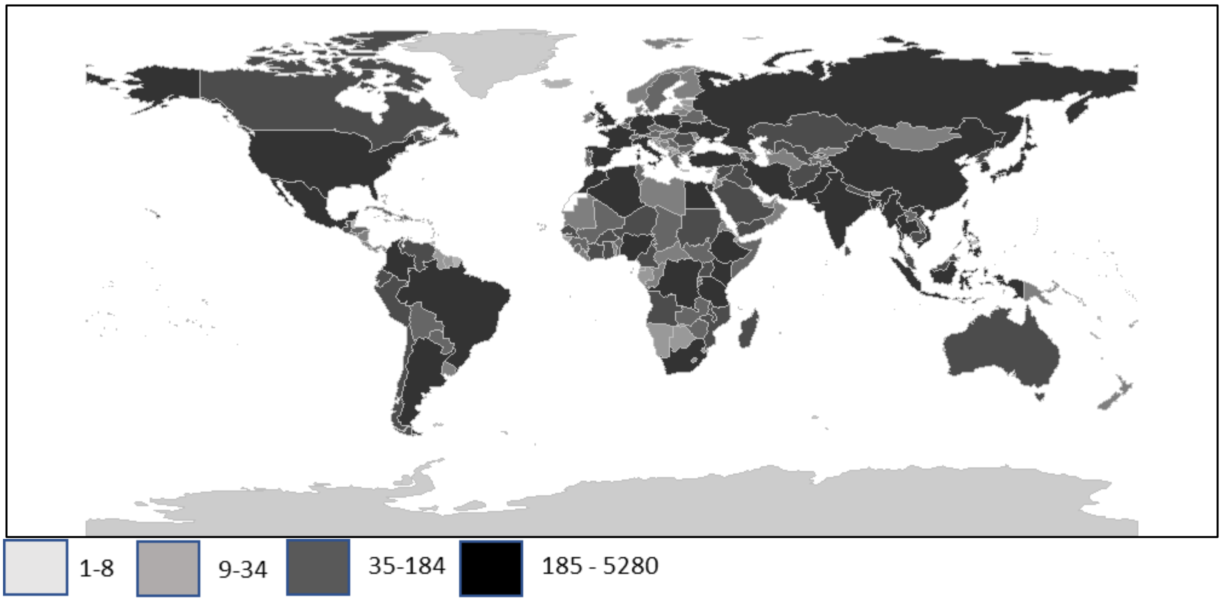
Documents by quartile GDP			**Downloads by Quartile GDP**		
Quartile	Total documents	Documents/ 1,000 capita (mean)	Quartile	Total Downloads	Downloads/ 1,000 capita (mean)
1st Quartile	15,441	.002	1st Quartile	6,876	.008
2nd Quartile	9,659	.012	2nd Quartile	16,457	.015
3rd Quartile	18,772	.015	3rd Quartile	12,209	.046
4th Quartile	106,781	.013	4th Quartile	8,135	.125

The data on documents produced by country is taken from SCImago country rankings in the subcategory development. In total N= 152,918 documents are listed. Download statistics show the number of Sci-Hub downloads within this catalogue, total N= 43,909.

Looking at the Sci-Hub downloads per country, we rank the top 10 countries from which articles are downloaded. In this list, the only high-income country is the USA. Iran leads the ranking with 5,280 downloads, followed by Indonesia and China. The only countries listed from the African continent are Morocco and Tunisia; the majority of countries come from the Asian continent. If we again normalize the data for population, the high-income countries Germany, The Netherlands and Portugal move into the top 10, and the USA drops out. In this ranking Chile is the only country from South America. If we now look at the data by quartiles of GDP, we can see that the lowest-income countries downloaded 6,876 documents, which is the lowest number by quartile. The second lowest number of downloads (8,135) came from the quartile with the highest income. The highest number of downloads fall in between these extremes, with 16,457 downloads for the second quartile and 12,209 downloads for the third quartile.

There are two main points that we can derive from these results. First, countries from the Global South are the main users of Sci-Hub in the discipline of development studies. This result stands in contrast to the aggregated data and results from other disciplines (Androcec, 2017; Babutsidze, 2016; Bohannon, 2016; Timuș and Babutsidze, 2016). Second, looking at GDP, we can see that the poorest countries in our dataset use Sci-Hub the least, especially when normalizing the data by population. The biggest downloading countries in this cohort are India, Kenya, Nigeria and Pakistan. It is countries from the second and third quartiles of GDP that are using this service the most, namely Indonesia, Iran, Tunisia and Vietnam from the second quartile, and Brazil, China and Malaysia from the third quartile. We also compare our data with results from Babutsidze (2016) and Timuș and Babutsidze (2016). This shows that countries with the lowest GDP in our dataset downloaded far higher numbers of development articles (6,876 downloads) than the whole-world downloads articles from the leading economics (2,147 downloads) and European studies journals (2,310 downloads).

Figure 5.1: World map of Sci-Hub download requests by country



The number of download requests is displayed by quantiles. The darker the shade, the more downloads are requested.

The map in Figure 5.1 makes those trends visible. Countries with higher numbers of downloads have a darker shade, countries with less downloads a lighter shade. The map shows that Central and SEA countries have large numbers of downloads, with the exceptions of Lao PDR and Cambodia in SEA. In Latin America, Argentina and Brazil have higher download rates, whereas less developed economies like Bolivia and Paraguay have the least downloads. This result is robust when normalized by population. On the African continent, we have two hotspots of download use: one is in the east, with the four neighbouring countries of DRC, Ethiopia, Kenya and Tanzania; the other is in the north, with Algeria, Egypt and Morocco. Large parts of West Africa and sub-Saharan Africa have very little download traffic.

In Figure 5.2, we plot GDP with download requests. The hyperbolic relationship of download frequency and GDP per capita that we expect from the results in Table 5.1 becomes visible. The countries with lower GDP also have a low download frequency. As GDP rises, we have countries that show higher frequency of download requests; this is the case for Brazil, China, Indonesia, Iran, Malaysia and Vietnam. As GDP rises further, download rates decrease. Thus, we have to be careful when making statements about the correlation between GDP and Sci-Hub downloads, since the countries with the lowest and the highest GDP show lower download rates than the countries in-between. This might be due to the fact that the least developed countries do not have the research infrastructure that is critical for scientific discourse, leading to less demand for scientific literature. Our data show that Sci-Hub is used mainly by researchers from middle-income countries to bypass paywalls. The hypothesis that is commonly found in the literature — that Sci-Hub is a means for researchers from poor countries to access scientific literature — must therefore be rejected.

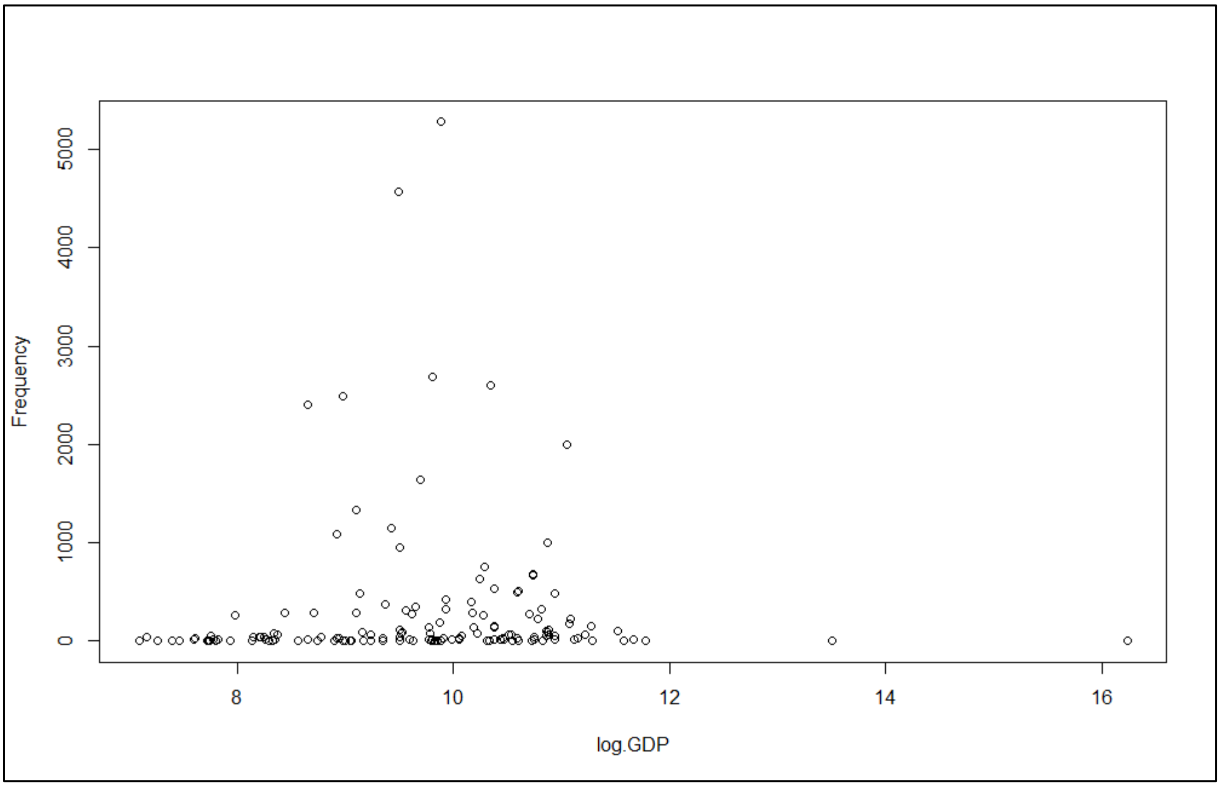
Table 5.2 presents a simple regression framework on total Sci-Hub downloads per country, with two explanatory variables: population and GDP. In general, the model has a good fit with an R-squared of 0.277. Of course, the variable population is positively correlated with Sci-Hub downloads. GDP is also positively correlated in this framework, though at a lower statistical significance level. Thus, we find that GDP per capita has a positive influence on Sci-Hub use. This result confirms our expectations and the results that are found in the literature (Greshake, 2016).

Table 5.2: Regression framework of total downloads from Sci-Hub by country

Variable	Estimate	Standard Error	t-value	Pr(>|t|)	Significance
Population	243.18	32.54	7.474	7.88e-12	***
GDP per capita	93.65	42.99	2.178	0.031	*
Summary statistics					
Residual standard error	638.5				
Df	139				
Adjusted R-squared	0.277				

Significance levels: *** <0.001, **<0.01, *<0.01

Figure 5.2:Scatterplot of GDP per capita and frequency of downloads by country



The plot shows download frequency and the natural log of GDP per capita.

5.5 Conclusion

A small but growing community of researchers is advocating support for and the use of Sci-Hub (Bendezú-Quispe et al., 2016; Bodó, 2018; Deshpande, 2019; Deshpande and Naik, 2012; Elbakyan, 2016; Greshake, 2016). Our analysis proves that especially researchers from the Global South follow this call. For them, Sci-Hub provides access to all literature they need. Sci-Hub can thus be described as an already existing network of open access literature in the field of development studies. However, our analysis also shows that even in the Sci-Hub world of open access, the poorest communities are excluded. We believe that analysing Sci-Hub data, which implies learning lessons from Sci-Hub, can uncover socio-economic drivers and their effects on a functioning open access network. These lessons can then be used in future applications of open access to provide evidence-based policy recommendations.

References

Androcec D. (2017). Analysis of Sci-Hub Downloads of Computer Science Papers. *Acta Universitatis Sapientiae, Informatica* 9(1): 83–96.

Babutsidze Z. (2016). Pirated Economics. Munich: Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/72621/

Banks M. (2016). Sci-Hub: What it is and Why it Matters. The Essentials on an Open Access Controversy, *American Libraries Magazine*. Date accessed: 11.01.2023. https://americanlibrariesmagazine.org/ 2016/05/31/why-sci-hub-matters/

Bendezú-Quispe, G., W. Nieto-Gutiérrez, J. Pacheco-Mendoza and A. Taype-Rondan (2016). Sci-Hub and Medical Practice: An Ethical Dilemma in Peru. *The Lancet Global Health*, 4(9): e608. https://doi.org/10.1016/S2214-109X(16)30188-7

Bodó, B. (2016). Pirates in the Library: An Inquiry into the Guerilla Open Access Movement. Paper prepared for the 8th Annual Workshop of the International Society for the History and Theory of Intellectual Property, CREATe, University of Glasgow (6–8 July).

Bodó, B. (2018) The Genesis of Library Genesis: The Birth of a Global Scholarly Shadow Library. Cambridge, MA: MIT Press. https://www.ivir.nl/publicaties/download/library_genesis.pdf

Bohannon, J. (2016). Who's Downloading Pirated Papers? Everyone. *Science*, 352(6285): 508– 12.

Bohannon, J. and A. Elbakyan (2017). Data from: Who's Downloading Pirated Papers? Everyone. *Dryad*, dataset. https://doi.org/10.5061/dryad.q447c

Buranyi, S. (2017). Is the Staggeringly Profitable Business of Scientific Publishing Bad for Science? *The Guardian*. Date accessed: 11.01.2023. www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

Deshpande, P.R. (2019). Why Should Sci-Hub Be Supported? *International Journal of Health and Allied Sciences* 8(3): 210–12.

Deshpande, P.R. and A.N. Naik (2012). Chargeless/Free Availability of Medical Literature: The Ethical Need for Development of Global Healthcare, *Indian Journal of Community Medicine* 37(4): 264. https://doi.org/10.4103/0970-0218.103478

Elbakyan, A. (2016). Sci-Hub Is a Goal, Changing the System Is a Method. *Engineering.* Date accessed: 11.01.2023. https://engineuring.wordpress.com/2016/03/11/sci-hub-is-a-goal-changing-the-system-is-a-method/

Elbakyan, A. (2017). Some Facts on Sci-Hub that Wikipedia Gets Wrong, *Engineuring*. Date accessed: 11.01.2023. https://engineuring.wordpress.com/2017/07/02/some-facts-on-sci-hub-that-wikipedia-gets-wrong/

Gibbs, W.W. (1995). Lost Science in the Third World. *Scientific American*. Date accessed 11.01.2023. www.scientificamerican.com/article/lost-science-in-the-third-world/

Graber-Stiehl, I. (2018). Science's Pirate Queen: Alexandra Elbakyan is Plundering the Academic Publishing Establishment. *The Verge*. Date accessed 11.02.2023. www.theverge.com/2018/2/8/16985666/alexandra-elbakyan-sci-hub-open-access-science-papers-lawsuit

Greshake, B. (2016). Correlating the Sci-Hub Data with World Bank Indicators and Identifying Academic Use. *The Winnower.* Date accessed: 11.01.2023. https://doi.org/10.15200/winn.146485.57797

Guarria, C.I. and Z. Wang (2011). The Economic Crisis and its Effect on Libraries. *New Library World* 112(5/6): 199–214.

Harrison, R., Y. Nobis and C. Oppenheim (2018). A Librarian Perspective on Sci-Hub: The True Solution to the Scholarly Communication Crisis is in the Hands of the Academic Community, not Librarians', *LSE Impact of Social Sciences Blog*. Date accessed: 11.01.2023.

https://blogs.lse.ac.uk/impactofsocialsciences/2018/11/09/a-librarian-perspective-on-sci-hub-the-true-solution-to-the-scholarly-communication-crisis-is-in-the-hands-of-the- academic-community-not-librarians/

Himmelstein, D.S. et al. (2018). Sci-Hub Provides Access to Nearly All Scholarly Literature. *eLife*, 7: e32822. https://doi.org/10.7554/eLife.32822

Hoy, M.B. (2017). Sci-Hub: What Librarians Should Know and Do about Article Piracy. *Medical Reference Services Quarterly*, 36(1): 73–78.

Kirsop, B. and L. Chan (2005). Transforming Access to Research Literature for Developing Countries. *Serials Review*, 31(4): 246–55.

McGuigan, G.S. and R.D. Russel (2008). The Business of Academic Publishing: A Strategic Analysis of the Academic Journal Publishing Industry and its Impact on the Future of Scholarly Publishing. *Electronic Journal of Academic and Special Librarianship*, 9(3).

Mejia, C.R. et al. (2017). Use, Knowledge, and Perception of the Scientific Contribution of Sci-Hub in Medical Students: Study in Six Countries in Latin America. *PLoS One*, 12(10): e0185673. https://doi.org/10.1371/journal.pone.0185673

Oxenham, S. (2016). Meet the Robin Hood of Science, Alexandra Elbakyan. *Big Think*. Date accessed 11.01.2023. https://bigthink.com/neurobonkers/a-pirate-bay-for-science

Pailey, R.N. (2020). De-centring the “White Gaze” of Development. *Development and Change*, 51(3): 729–45.

Priego, E. (2016). Signal, not Solution: Notes on Why Sci-Hub Is not Opening Access. *The Winnower*. Date accessed: 11.01.2023. https://doi.org/10.15200/winn.145624.49417

Swartz, A. (2008). Guerilla Open Access Manifesto. Date accessed: 11.01.2023. https://ia800605.us.archive.org/15/items/GuerillaOpenAccessManifesto/Goamjuly2008.pdf

Tennant, J.P. et al. (2016). The Academic, Economic and Societal Impacts of Open Access: An Evidence-based Review. *F1000Research*, 5: 632. https://doi.org/10.12688/ f1000research.8460.1

Timuș, N. and Z. Babutsidze (2016). Pirating European Studies. *Journal of Contemporary European Research*, 12: 783–91.

UN General Assembly (1948). Universal Declaration of Human Rights (217 [III] A)’. New York: United Nations.

6 General conclusion

Overall, this dissertation aims to contribute to the understanding of smallholder farmers in SEA. Chapter 2 discerns the linkages between land use change and plot-level and household-level characteristics and processes by contrasting these developments in a subsistence-oriented site in Lao PDR and a market-oriented site in Vietnam. Chapter 3 analyzes risk preferences of smallholders in Cambodia and Lao PDR with an incentivized lottery design under the framework of EUT, RDU and CPT and tests the effect of household shocks on these risk preferences. Chapter 4 examines the prevalence of poor vision among rural smallholders in Cambodia and investigates if poor vision is associated with a loss in agricultural profitability of family-owned farms. Chapter 5 screens through 28 million download logs for papers in the field of development studies and lays out trends for Sci-Hub use in this discipline, including geographic location of downloads and socioeconomic drivers. Furthermore, it discusses ethical considerations relevant to scientific conduct in the field of development studies.

6.1 Main findings

In chapter 2, we find that land use dynamics varied strongly between the subsistence-oriented site in Lao PDR and the market-oriented site in Vietnam. In the Lao PDR site, 66 percent of the land use types were completely replaced by others during the past 10 years. In the Vietnam site, only 15 percent of the land use types were replaced. The associated key drivers of land use change also differed significantly: while market orientation of agricultural products was the main driver behind land use changes in the Vietnam site, mostly agronomic challenges like slope, soil tillage, and agrochemical input use are associated to land use change in the Lao PDR site. Likewise, land use complexity does not exhibit the same relationship with farm performance in the two sites. In the Vietnam site, households with higher food availability are half as likely to transition, whereas in the Lao PDR site, land use complexity was significantly correlated with the PPI, with better-off farms being associated with a higher likelihood of transition.

In chapter 3, we apply three different specifications (EUT, RDU and CPT) of utility to measure the effects of household shocks on risk aversion, probability weighting and loss aversion. First, we find that CPT is a good representation of smallholders' behavior under uncertainty and that the CPT model outcompetes the other models in this regard. Second, we find that household shocks can increase loss aversion and that they have effects on the curvature of the utility function in the loss domain. In our sample, the variable *total household shocks* significantly increases loss aversion. As a perfect mirror image, not having experienced any shocks decreases loss aversion significantly. Thus, household shocks can trigger loss aversion, controlling for a wide range of socioeconomic variables. A possible interpretation of this result is that farmers have a target income level. The experience of household shocks endangers this target income level, which increases their aversion to losses. Therefore, household shocks not only cause economic damages, but alter the behavior of smallholders under uncertainty.

In chapter 4, we find that almost 30 percent of farmers in our sample suffer from poor vision. To generate robust results on the impacts of poor vision on farm profits, we apply various methods to test if farmers with poor vision lose farm profits in comparison to farmers with good vision. In summary, the results obtained by all estimation procedures are robust, and taken together give evidence of forgone profits in the range of 589–632 USD/year associated with having, at least, intact visual functions in regard to everyday tasks. This effect is particularly outstanding considering the Cambodian gross national income per capita (GNI) is 1,380 USD.

In chapter 5 we find that Sci-Hub is used the most by researchers from the Global South, primarily from middle-income countries; whereas researchers from the poorest countries in the data set use Sci-Hub the least. There are two main points that we can derive from these results. First, researchers from the Global South are the main users of Sci-Hub in the discipline of development studies. Second, looking at GDP, we can see that users from the poorest countries in our data set use Sci-Hub the least, especially when normalizing the data by population. The most downloads from this cohort come from India, Kenya, Nigeria and Pakistan. Researchers from countries in the second and third quartiles of GDP are using this service the most, namely: Indonesia, Iran, Tunisia and Vietnam from the second quartile, and Brazil, China and Malaysia from the third quartile. Large parts of West Africa and Sub-Saharan Africa have very little download traffic. Thus, our analysis also shows that in terms of Sci-Hub downloads, the poorest communities are underrepresented.

6.2 Limitations and further research

There are two methodological limitations in the study presented in chapter 2 that are worth noting. First, the CI integrates seasonal information without consideration for the novelty of the crops: rice-fallow rotations are treated equivalently to switching from one crop to a completely new one, from coffee to sugarcane, for example. Seasonal crop rotations have no consequence on the complexity of the system, and care must be taken not to translate this complexity as a measure of diversity. Second, performance indicators and household characteristics are defined at one point in time (in the year 2017) whereas the CI is the aggregate result over a recall period of 10 years. This assumes that there has been no major alteration in the households and that the present situation comprehensively summarizes all happenings on the farm during the last ten years. This was deemed to be a reasonable assumption for the purposes of this study, but future studies could provide more time variant perspectives into household and plot level characteristics via panel data.

There are two limitations in the study presented in chapter 3. First, the respondents were paid a fee to participate in the workshop, where they answered the household questionnaire and took part in the risk experiment. It is possible that the respondents of our field experiments integrate the participation fee into their utility which results in the fact that we never play for real losses. To circumvent this, one possible avenue for future research would be to pay the participation fee ahead of the experiment, so that the participation fee is already incorporated into the stream of household consumption. If the risk experiment is played weeks after this, we could rule out the so-called house money effect as a potential source of biased estimates. The second weakness stems from the fact that we use cross-sectional data to measure the impact of shocks on risk preference parameters. The studies that employ panel data on this topic also use more restrictive functional forms of behavior. In future research, the strengths of both approaches should be combined. A future avenue of research is to identify reference points and target income levels of smallholders and their influence on risk aversion in the gain and loss domain.

In chapter 4 we report results on the average treatment effect on the treated, which are generated by statistical matching methods. It has to be stated that it is questionable if a real-world intervention, like providing glasses to a subset of farmers, would deliver a treatment effect of this size. Behavioral aspects would most likely reduce the treatment effect. For example, if a participant would be prescribed glasses, she perhaps would not wear glasses for all activities, or could not use them for all activities equally. Glasses may be removed in field work under direct sun, or if irritations like fogging and blurred vision due to sweat and dust outweigh the advantages of wearing glasses. Thus, we rather present estimations on the maximum achievable effect, against which real-world interventions can be measured. In future research, a better identification of the causal relationships between myopia and farm profitability can be established by collecting longitudinal data. Future research should investigate (1) which entrepreneurial activities are most affected by poor vision and (2) which steps need to be taken to drive the usage of glasses. A repeated measure within-subjects design, e.g. a controlled experiment that applies pre- and post-measurement in relation to the treatment of glasses or contact lenses, would be optimal for determining the causal effect of visual acuity on the economic performance of smallholders.

6.3 Policy recommendations

The research in this thesis highlights the transitions of smallholder farming systems in marginal mountainous areas of SEA and addresses challenges regarding farmer risk preferences, household shocks and farmer health. The goal is to identify policies that support smallholders in creating viable livelihoods, be it through market-based approaches or public programs. Successful policies in this corridor depend on a deep understanding of farmers themselves. Since I want to put smallholders at the heart of this dissertation, the following recommendations are focused on practical implications, in the hope that they find their way into the ears of development practitioners.

One overarching factor is that agricultural intensification must be a key component of the development strategy, particularly over the next few decades because the number of people living in the highlands of SEA continues to grow and further encroachment into forest areas is not a sustainable option. Because of poor infrastructure, low yield potential, fragile soils and high climate risk, the strategy will need to be different from the approach adopted in irrigated and high-potential rainfed areas during the so-called green revolution. With the lessons on land use transitions from chapter 2 in mind, innovative land use options for intensification have higher chances for adoption when they take into consideration the diverse set of livelihood strategies that smallholders follow. For example, if plots under consideration for adoption of innovative technologies are devoted to staple crop production, there is a lower chance of adoption, unless the intervention complements staple production rather than replacing it. Similarly, initial experimentation on intensified systems should not be located on plots with steep slopes. Innovative land use options would be successful on slopes only if complemented by appropriate soil and water conservation measures that are acceptable from a capital and labor investment point of view. Furthermore, land area devoted to low-input, low-risk cash crops, such as wild tea or forages, can offer viable livelihood strategies to smallholders. This is consistent with the results from chapter 3 which highlight the risk averse behavior of SEA upland smallholders. A policy recommendation from chapter 3 is that new technologies and farming systems might be rejected by smallholders due to loss aversion. Our result suggests that it is important for extension agencies to introduce new technologies in a way that losses are excluded for farmers. Initial risk-free demonstration and on-farm experimentation has to be provided for farmers until they can identify the expected returns of such new technologies. Farmers will be more likely to invest in new technologies if they can evaluate risks and probabilities associated with the technology. Furthermore, insurance and financing mechanisms should be applied in a way that they protect stable household consumption. One potential avenue is to offer flexible loan payback schedules for agricultural technologies. For example, loan payback schedules can be linked to yields achieved by farmers. This has the benefit that agricultural shocks can be buffered by decreasing amortization rates when yields are low. This protects target income levels and household consumption against shocks. Another possible mechanism that comes into play if smallholders are sensitive to losses is the reflection effect. Hereafter, individuals become risk seeking to prevent sure losses. In a farming context, this could mean that if a rice field is hit by a drought, smallholders exceedingly allocate resources to this field to prevent a crop loss, instead of abandoning a hopeless situation. Active decision support of extension workers via the calculation of damage thresholds can alleviate this problem.

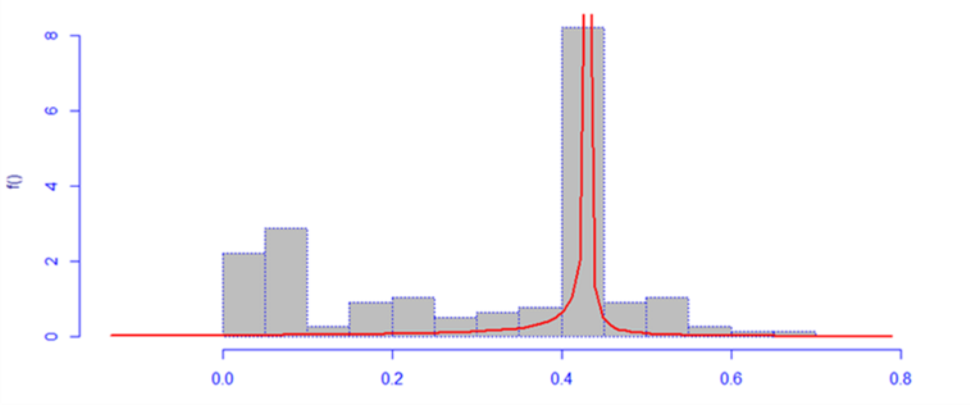
In chapter 4, we discuss that the basis to make sound management decisions by smallholder farmers is that they require an intact visual system. Our results show that there are important linkages between agriculture and public health and that there is a need for more collaboration across the agricultural and public health sectors to address the negative impacts of poor health on agricultural profitability. Any form of agricultural extension has to take into account that farmers might be visually impaired, which is why the selection of delivery methods is crucial. Visually impaired farmers might not be able to comprehend presentations at a far distance, as is common in the farmer field day approaches.

Lastly, we want to touch upon the issues of scientific publishing and open access, which were raised in chapter 5. Especially for researchers in the field of development studies, it is important to guarantee that the research outputs we create are widely available. It is a sad fact that many scholars rely on pirate websites like Sci-Hub. Many other legal avenues exist, be it through open access journals, scientific archives or informal networks. It is our responsibility to close the north/south gap in research outputs; opening up the library is a fundamental step to a more just way of doing science. In the end, most researchers are doing their work for the greater good, and it would be unfortunate if the body of knowledge that we create through mutual collaboration ends up behind a paywall.

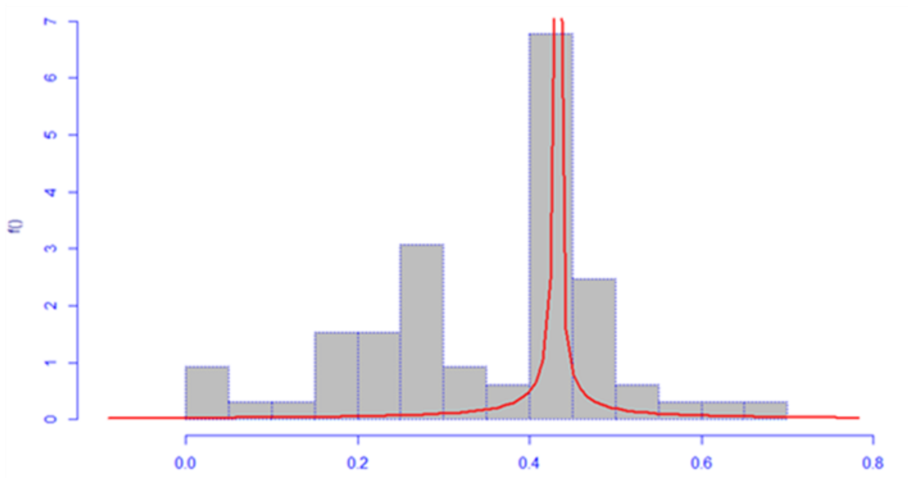
Appendix A. Drivers of land use complexity along an agricultural transition gradient in Southeast Asia

Figure A 1: Histogram of CI (A) and Agg-CI (B) for site XK

A



B



Histogram of (A) CI (B) Agg-CI for Site XK, overlaid with the selected distribution function, sinh-arcsinh (in red) obtained as the distribution function with the lowest Global Akaike Information Criterion (GAIC) using the fitDist function in GAMLSS.

Table A 1: Residual distribution-based model diagnostics for models selected and described for plot level, household level and performance level characteristics based regression analysis for site XK

Residual based model diagnostic parameters	**Plot level**	**Household level**	**Performance**
Mean	0.06	0.18	0.03
Variance	0.86	1.01	1.09
Coefficient of Skewness	-0.35	-0.65	-0.13
Coefficient of Kurtosis	2.67	2.93	2.05

Table A 2: Descriptive analysis of variables disaggregated by plot level, used for plot level regression analysis in site CH

Plot-level CI	**Variable**	**Category**	**Total number**
No transition	Soil tillage	No	11
No transition	Soil tillage	Yes	104
Transition	Soil tillage	No	10
Transition	Soil tillage	Yes	72
No transition	Agrochemical inputs	Fertilizers	3
No transition	Agrochemical inputs	Pesticides and fertilizers	112
Transition	Agrochemical inputs	Fertilizers	1
Transition	Agrochemical inputs	Pesticides and fertilizers	81
No transition	Property	Family	112
No transition	Property	Rented	3
Transition	Property	Collective	1
Transition	Property	Family	80
Transition	Property	Rented	1
No transition	Irrigation	Canal	6
No transition	Irrigation	Pump	87
No transition	Irrigation	Rainfed	22
Transition	Irrigation	Canal	1
Transition	Irrigation	Pump	65
Transition	Irrigation	Rainfed	16
No transition	Slope	Flat	63
No transition	Slope	Modest	38
No transition	Slope	Strong	14
Transition	Slope	Flat	37
Transition	Slope	Modest	33
Transition	Slope	Strong	12
No transition	Source of planting material	Bought	110
No transition	Source of planting material	Mix	5
Transition	Source of planting material	Bought	76
Transition	Source of planting material	Mix	6
No transition	Final use	Home consumption	23
No transition	Final use	Sale	92
Transition	Final use	Home consumption	4
Transition	Final use	Sale	78

Variable refers to the variable used for regression analysis, category refers to the responses for each variable, and performance level characteristics based regression analysis for site XK.

Table A 3: Descriptive analysis of (a) categorical and (b) continuous variables disaggregated by Agg-CI in the CH

(a)

Aggregated CI	**Variable**	**Category**	**Total number**
No transition	Ethnic group	Kinh	20
No transition	Ethnic group	Mnong	3
No transition	Ethnic group	Tay	1
Transition	Ethnic group	Ede	1
Transition	Ethnic group	Kinh	39
Transition	Ethnic group	Mnong	1
Transition	Ethnic group	Tay	1
Transition	Ethnic group	Thai	1
No transition	Origin	No	21
No transition	Origin	Yes	3
Transition	Origin	No	40
Transition	Origin	Yes	3
No transition	Off farm employment	No	17
No transition	Off farm employment	Yes	7
Transition	Off farm employment	No	30
Transition	Off farm employment	Yes	13
No transition	Labor	Family	11
No transition	Labor	Family contracted	9
No transition	Labor	Family contracted mutual help	3
No transition	Labor	Family mutual help	1
Transition	Labor	Contracted	2
Transition	Labor	Family	10
Transition	Labor	Family contracted	22
Transition	Labor	Family contracted mutual help	6
Transition	Labor	Family mutual help	3
No transition	Education	Post secondary	5
No transition	Education	Primary	4
No transition	Education	Secondary	15
Transition	Education	Illiterate	1
Transition	Education	Literate	1
Transition	Education	Post secondary	8
Transition	Education	Primary	8
Transition	Education	Secondary	25

Table continues on next page

Table continuation from previous page

(b)

Aggregated CI	Variable	Mean	Standard Deviation
No transition	Household size	2.87	1.07
Transition	Household size	3.52	1.26
No transition	Land cultivated	2.29	2,18
Transition	Land cultivated	2.45	1.54
No transition	Fertilization	87.08	82.79
Transition	Fertilization	96.43	70.28
No transition	Crop diversity	4.5	1.35
Transition	Crop diversity	4.3	1.52
No transition	Livestock diversity	1.91	1.21
Transition	Livestock diversity	2.13	1.10
No transition	Total number of fields	2.04	0.95
Transition	Total number of fields	2.79	1.66

Variable refers to the variables used for regression analysis, category refers to the responses for each variable and total number refers to the number of households belonging to corresponding combination of AggCI, variable and category. In Table B, Mean and Standard Deviation represent descriptive statistics of the corresponding variable

Table A 4: Descriptive analysis of variables representing household level performance characteristics

Aggregated CI	Variable	Mean	Standard Deviation
No transition	Progress out of poverty index	7.52e+07	1.73e+07
Transition	Progress out of poverty index	6.95e+07	1.72e+07
No transition	Total income	6.65e+09	6.65e+09
Transition	Total income	5.15e+09	4.69e+09
No transition	Food availability	9.21e+10	7.95e+10
Transition	Food availability	5.88e+10	4.91e+10
No transition	Food self sufficiency	4.69e+09	5.38e+09
Transition	Food self sufficiency	4.14e+09	5.39e+09
No transition	Total energy available	9.22e+13	8.76e+13
Transition	Total energy available	7.23e+13	6.21e+13
No transition	Food insecurity	4.58e+06	4.18e+06
Transition	Food insecurity	5.30e+06	4.49e+06

Characteristics are disaggregated by AggCI identified in site CH. Variable refers to the variables used for regression analysis, mean and standard deviation represent descriptive statistics of the corresponding variable.

Table A 5: Descriptive analysis of plot level CI in site XK

Variable	**Category**	**Mean of plot-level CI**	**Standard deviation of plot-level CI**
Soil tillage	No	0.20	0.16
Soil tillage	Yes	0.40	0.12
Agrochemical inputs	Fertilizers	0.45	0.04
Agrochemical inputs	No chemical	0.30	0.18
Agrochemical inputs	Pesticides	0.29	0.14
Property	Collective	0.20	0.19
Property	Family	0.31	0.17
Property	Rented	0.43	0
Irrigation	Canal	0.42	0.11
Irrigation	Rainfed	0.29	0.17
Slope	Flat	0.37	0.14
Slope	Modest	0.32	0.18
Slope	Strong	0.24	0.17
Planting material source	Bought	0.34	0.13
Planting material source	Combination	0.30	0.19
Planting material source	Exchanged	0.21	0.30
Planting material source	Gift	0.48	
Planting material source	Own	0.30	0.17
Planting material source	Subsidized	0.25	0.36
Final use	Home consumption	0.30	0.18
Final use	Home consumption and sale	0.32	0.17
Final use	Sale	0.28	0.15

Table A 6: Descriptive analysis and model diagnostics

(a)

Variable	**Category**	**Mean of aggregated CI**	**Standard deviation of aggregated CI**
Ethnic group	Hmong	0.31	0.15
Ethnic group	Lao	0.38	0.13
Origin	No	0.31	0.12
Origin	Yes	0.33	0.16
Off farm employment	No	0.33	0.12
Off farm employment	Yes	0.33	0.18
Labor	Family	0.31	0.15
Labor	Family contracted	0.24	0.17
Labor	Family contracted mutual help	0.36	0.14
Labor	Family mutual help	0.36	0.14
Education	Illiterate	0.24	0.19
Education	Literate	0.34	0.12
Education	Post-secondary	0.48	0.08
Education	Primary	0.33	0.14
Education	Secondary	0.34	0.16

(b)

Variable	**Pearson correlation coefficient with aggregated CI**
Household size	-0.02
Land cultivated	-0.06
Fertilization	0.004
Crop diversity	0.19
Livestock diversity	0.28
Total number of fields	-0.18

(a) Descriptive analysis, i.e. mean and standard deviation of agg-CI score of categorical variables (b) Pearson correlation coefficient score between log transformed values of agg-CI score and log transformed values of household level variables, that are continuous in nature, identified across site XK.

Table A 7: Pearson correlation coefficient score between log transformed values of agg-CI and log transformed values of performance indicators identified across site XK

Variable	**Pearson correlation coefficient with aggregated CI**
Progress out of poverty index	0.25
Total income	0.07
Food availability	0.16
Food self sufficiency	0.04
Total energy available	0.16
Food insecurity	-0.05

Table A 8: Single term deletion analysis of variables used in plot level regression for site CH

Plot - level variables	**Df**	**Deviance**	**AIC**	**LRT**	**Pr(>Chi)**
Soil tillage	1	251.18	271.18	0.05	0.82
Agrochemical inputs	1	251.85	271.85	0.72	0.39
Property	2	254.85	272.85	3.72	0.15
Irrigation	2	251.81	269.81	0.68	0.70
Slope	2	251.14	269.14	0.01	0.99
Source of planting material	1	251.92	271.92	0.79	0.37
Final use	1	260.33	280.33	9.20	0.002

Single term deletion analysis of variables used in plot level regression for Site CH showing significant impact of only the final use plot level variable on the outcome variable (i.e. plots with and without transition). The final use variable is responsible for explaining the largest deviance and was responsible for the largest drop in Akaike Information Criterion (AIC), upon removal from the analysis, in comparison to the other variables.

Table A 9: Single term deletion analysis of household level variables used in regression for site CH

Variable	Df	Deviance	AIC	LRT	Pr(>Chi)
Ethnic group	3	69.16	103.16	4.71	0.19
Origin	1	65.57	103.57	1.12	0.29
Off farm employment	1	64.52	102.52	0.07	0.79
Labor	4	66.79	98.79	2.33	0.67
Household size	1	70.04	108.04	5.59	0.01
Education	3	64.96	98.96	0.52	0.91
Land cultivated	1	64.45	102.45	0.0014	0.96
Fertilization	1	64.68	102.68	0.23	0.63
Crop diversity	1	66.55	104.55	2.10	0.14
Livestock diversity	1	65.10	103.10	0.65	0.42
Total number of fields	1	67.31	105.31	2.87	0.09

Single term deletion analysis of household level variables used in regression for site CH showing significant impact of household size and total number of fields on the outcome variable (i.e. plots with and without transition). Household size and total number of fields variable are responsible for explaining the largest deviance and were responsible for the largest drop in Akaike Information Criterion (AIC), upon removal from the analysis, in comparison to the other variables.

Table A 10: Single term deletion analysis of performance variables used in regression for site CH

Variable	Df	Deviance	AIC	LRT	Pr(>Chi)
Progress out of poverty index	1	80.217	90.21	1.02	0.31
Total income	1	79.314	89.31	0.12	0.72
Food availability	1	83.801	93.80	4.61	0.03
Food self sufficiency	1	79.472	89.47	0.28	0.59
Total energy available	1	79.257	89.25	0.06	0.79
Food insecurity	1	80.725	90.72	1.53	0.21

Single term deletion analysis of performance variables used in regression for site CH showing significant impact of food availability on the outcome variable (i.e. plots with and without transition). Food availability variable is responsible for explaining the largest deviance and was responsible for the largest drop in Akaike Information Criterion (AIC), upon removal from the analysis, in comparison to the other variables.

Appendix B. Effects of household shocks on risk preferences and loss aversion: Evidence from upland smallholders of Southeast Asia

The risk experiment

To elucidate risk preferences of smallholders we apply an incentivized lottery choice experiment using the method from Tanaka et al. (2010), which follows the multiple price list format. The experiments are adapted to be played in a locally appropriate way. Because of the low educational level of participants, we do not use price lists but rather played out the experiments with tennis balls and bags, simulating each choice scenario in front of the participants. The participants of the experiments have to make 35 pairwise choices between two options. The lottery is divided into three separate series, where the first two series have 14 choices each with only positive payouts; the third series has seven choices which include negative payouts to estimate loss aversion. Tables B 11 and B12 display the full set of lottery choices in all three series. Option A represents a safer choice and option B a riskier choice, therefore people who are more risk averse choose option A more often than risk loving participants. In series one and two, the payouts for the riskier option B are increasing, and the expected value of option B exceeds option A in the seventh row. Each participant is told that they are only allowed to switch once from option A to B. It is also allowed to always choose option A or option B. Both options A and B were set up in front of the participants, who can chose between option A and B in each row. Furthermore, we add a choice at the beginning of the series where both outcomes for option B are dominated by option A. This way we can check for comprehension of the lottery and this choice is not used in the analysis.

The participants receive a show up fee of 2.5 USD, which was clearly stated before the experiment started. Each participant first plays each of 35 choices. After this, one choice set is randomly chosen and played with real incentives which are added or subtracted from the show up fee. Because it would be unethical to receive payments from the participants, the show up fee is higher than the maximum lost in the experiment. The average payout for Cambodia and Lao PDR was 3.5 USD and 3.8 USD, respectively.

Table B 11: Design of the risk experiment in Cambodia

		Option A		Option B	
		Probability		Probability	
Series	Row	0.3	0.7	0.1	0.9
1	0	7,200	1,800	7,200	900
	1	7,200	1,800	12,300	900
	2	7,200	1,800	13,500	900
	3	7,200	1,800	15,000	900
	4	7,200	1,800	16,800	900
	5	7,200	1,800	19,200	900
	6	7,200	1,800	22,600	900
	7	7,200	1,800	27,100	900
	8	7,200	1,800	33,400	900
	9	7,200	1,800	39,800	900
	10	7,200	1,800	54,200	900
	11	7,200	1,800	72,300	900
	12	7,200	1,800	108,400	900
	13	7,200	1,800	180,700	900
	14	7,200	1,800	307,200	900
		Option A		Option B	
Series	Row	0.9	0.1	0.7	0.3
2	0	7,200	5,400	7,200	900
	15	7,200	5,400	9,800	900
	16	7,200	5,400	10,100	900
	17	7,200	5,400	10,500	900
	18	7,200	5,400	10,800	900
	19	7,200	5,400	11,200	900
	20	7,200	5,400	11,700	900
	21	7,200	5,400	12,300	900
	22	7,200	5,400	13,000	900
	23	7,200	5,400	13,900	900
	24	7,200	5,400	15,000	900
	25	7,200	5,400	16,300	900
	26	7,200	5,400	18,100	900
	27	7,200	5,400	19,900	900
	28	7,200	5,400	23,500	900
		Option A		Option B	
Series	Row	0.5	0.5	0.5	0.5
3	0	4,500	-700	4,500	-3,800
	29	4,500	-700	5,400	-3,800
	30	700	-700	5,400	-3,800
	31	200	-700	5,400	-3,800
	32	200	-700	5,400	-2,900
	33	200	-1,400	5,400	-2,900
	34	200	-1,400	5,400	-2,500
	35	200	-1,400	5,400	-2,000

Source: Own survey. All payouts are displayed in Khmer Riel (KHR).

Table B 12: Design of the risk experiment in Lao PDR

		Option A		Option B	
		Probability		Probability	
Series	Row	0.3	0.7	0.1	0.9
1	0	15000	4000	15000	2000
	1	15000	4000	24500	2000
	2	7200	1800	27000	900
	3	7200	1800	30000	900
	4	7200	1800	33500	900
	5	7200	1800	38500	900
	6	7200	1800	45500	900
	7	7200	1800	54500	900
	8	7200	1800	67000	900
	9	7200	1800	79500	900
	10	7200	1800	108500	900
	11	7200	1800	145000	900
	12	7200	1800	217500	900
	13	7200	1800	362500	900
	14	7200	1800	616000	900
		Option A		Option B	
Series	Row	0.9	0.1	0.7	0.3
2	0	15000	11000	15000	2000
	15	15000	11000	19500	2000
	16	7200	5400	20500	900
	17	7200	5400	21000	900
	18	7200	5400	21500	900
	19	7200	5400	22500	900
	20	7200	5400	23500	900
	21	7200	5400	24500	900
	22	7200	5400	26000	900
	23	7200	5400	28000	900
	24	7200	5400	30000	900
	25	7200	5400	32500	900
	26	7200	5400	36000	900
	27	7200	5400	40000	900
	28	7200	5400	47000	900
		Option A		Option B	
Series	Row	0.5	0.5	0.5	0.5
3	0	9000	1500	9000	7500
	29	9000	-1500	11000	-7500
	30	1500	-1500	11000	-7500
	31	500	-1500	11000	-7500
	32	500	-1500	11000	-6000
	33	500	-3000	11000	-6000
	34	500	-3000	11000	-5000
	35	500	-3000	11000	-4000

Source: Own survey. All Payouts are displayed in Lao Kip (LAK).

Household descriptive statistics

Table B 13 displays descriptive statistics of the sampled population. A Mann-Whitney test is carried out in order to detect differences between samples. The last column of Table B 13 displays the results of the Mann-Whitney test and the significance levels of the respective p-values.

Respondents in Cambodia are on average 38 years old. This is significantly younger compared to the sample from Lao PDR, which is 47 years of age on average. In both samples, there is a higher percentage of male participants, with 63 percent in Cambodia and 67 percent in Lao PDR. Participants in Lao PDR have spent on average 6 years in school. This is significantly more than the 3 years of average schooling in the Cambodian sample. In both samples, a household has on average two dependent household members. Landholdings are larger in Cambodia with 6.22 hectare on average compared to Lao PDR, where landholdings are 1.59 hectares on average. In Lao PDR, the average household owns 10 heads of cattle. In Cambodia, the average household owns 6 heads of cattle. The average per capita income of a working age household member is 691 USD/year in Cambodia and 436 USD/year in Lao PDR. In Cambodia, the largest contribution to household income comes from cropping activities with 2.186 USD/year, followed by off-farm income with 944 USD/year. In contrast to Lao PDR, the largest contribution to household income comes from off-farm activities, with 2.274 USD/year on average. In Lao PDR, the most important farm activity is livestock, with 740 USD/year, followed by income from crops with 447 USD/year. Income from perennials is rather marginal with 71 USD/year. In Cambodia, income from perennials and livestock are similar, with 395 USD/year and 377 USD/year, respectively.

Table B 13: Summary statistics of respondent characteristics

Variables	Cambodia		Lao PDR		Mann-Whitney *z*
	Mean	SD	Mean	SD	
Age (years)	38	14.25	47	12.79	-4.623***
Gender (1=male)	0.63	-	0.67	-	-5.100
Education (years)	3.18	2.85	6.07	2.95	-5.973***
Dependent household members (<16, > 65 years)	2	1.40	2	1.49	1.690*
Cattle owned (heads)	6	27.37	10	10.48	-5.805***
Total land size (ha)	6.22	7.38	1.59	1.36	8.907***
Income from crops (USD/year)	2186	5478	447	566	6.097***
Income from perennials (USD/year)	395	1101	71	296	3.535***
Income from livestock (USD/year)	377	1093	740	1308	-1.526
Income from off-farm activities (USD/year)	945	1768	2275	5682	1.709*
Household income (per working capita, USD/year)	691	845	436	903	3.814***

N= 184 (total sample), 93 (Cambodia), and 91 (Lao PDR). SD= standard deviation. Two tail significance levels of the Mann-Whitney test, * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Household shocks

Table B 14 shows that in Cambodia the most common household shock is drought, which is experienced at least once by 44 percent of all households. In Lao PDR, this number is much smaller, only 5 percent of all households experience a drought. The most common household shock in Lao PDR is disease or death of livestock, which is reported by 26 percent of all households. In Cambodia, this shock is only reported by 3 percent of households. The second most common shock in Cambodia is flooding, which is experienced by 34 percent of all households. This number is much lower in Lao PDR, with only 11 percent of all households experiencing a flood. The third most common shock, which is experienced by 33 percent of all households in Cambodia, is low market prices for produce. In Lao PDR, this shock is not reported by any household, which indicates that farms in Lao PDR are less market oriented. Furthermore, 9 percent of households in Cambodia report that no household shocks occurred in the previous three years. In contrast, 44 percent of all households in Lao PDR experienced no household shocks. Table B 14 displays the percentages of households that experienced household shocks in the years 2013-2015 by shock type. In Cambodia, 1.78 shocks occurred per household on average. This is higher than the sample from Lao PDR, which reports 0.84 household shocks on average.

Table B 14: Household shocks reported by smallholders

		Total sample	Cambodia	Laos
Agricultural shocks				
	Crop failure	5%	5%	4%
	Livestock disease/death	15%	3%	26%
	Drought	25%	44%	5%
	Flood	23%	34%	11%
Demographic shocks				
	Accident	4%	5%	3%
	Fatality	16%	14%	19%
	Disease	16%	26%	7%
Price shocks				
	High input prices	3%	6%	0%
	Low output prices	17%	33%	0%
Other shocks				
	Social shocks, conflicts, theft	2%	4%	0%
No shocks				
	No shocks	26%	9%	44%
Average number of shocks per household		1.31	1.78	0.84

Percentages of households that experienced shock between 2013-2015. N= 184 (total sample), 93 (Cambodia), 91 (Lao PDR)

Table B 15: Definitions of independent variables

Variable name	**Definition**
Age	Age of the respondent in years
Gender	1= Male, 0= Female
Education	Years of education in school or higher education
Dependent household members	Household members younger than 16 and older than 65 years
Household income	Household income. Sum of earnings from crops, perennial crops, livestock, jobs, own businesses, forest dependent activities and cash transfers minus input costs, rent and cash transfers. Calculated per working-age household member (between 16 and 65 years of age)
Cattle owned	Total heads of cattle (bull, cow and heifer) owned
Total number of household shocks	Total number of household shocks experienced 2013-2016
Exogenous household shocks	Total number of shocks from drought, flood and input/ output price shocks
No household shocks	Dummy variable= 1 if no shocks are reported by the household

Table B 16: CPT parameters and single shocks

		Accidents	Crop failure	Fatality in household	Disease in household	Drought	Flood	Unexpected high input prices	Livestock died/severe disease	Unexpected low price of produce	Other
Gamma	constant	1.546***	1.523***	1.528***	1.514***	1.445***	1.533***	1.540***	1.531***	1.513***	1.570***
	Standard Errors	(0.060)	(0.062)	(0.063)	(0.066)	(0.072)	(0.067)	(0.061)	(0.057)	(0.068)	(0.058)
	Coefficient of Covariate	0.026	0.142	0.110	0.160	0.230***	0.010	0.017	-0.036	0.125	-0.342
	Standard Errors	(0.183)	(0.158)	(0.094)	(0.103)	(0.078)	(0.085)	(0.169)	(0.110)	(0.102)	(0.234)
Alpha	Alpha constant	0.375***	0.362***	0.371***	0.374***	0.349***	0.370***	0.372***	0.369***	0.367***	0.386***
	Standard Errors	(0.027)	(0.030)	(0.031)	(0.029)	(0.033)	(0.030)	(0.029)	(0.027)	(0.029)	(0.027)
	Coefficient of Covariate	0.044	0.110**	0.028	0.025	0.060***	0.010	0.031	-0.067	0.040	-0.143*
	Standard Errors	(0.056)	(0.046)	(0.032)	(0.034)	(0.021)	(0.032)	(0.040)	(0.052)	(0.032)	(0.073)
Beta	Beta constant	0.800***	0.815***	0.752***	0.788***	0.869***	0.906***	0.792***	0.781***	0.820***	0.800***
	Standard Errors	(0.075)	(0.069)	(0.083)	(0.081)	(0.131)	(0.079)	(0.072)	(0.088)	(0.089)	(0.071)
	Coefficient of Covariate	0.433**	3.680	0.252***	0.153**	-0.177	-0.046	0.095	-0.326*	-0.063	-0.020
	Standard Errors	(0.201)	(3.823)	(0.084)	(0.064)	(0.199)	(0.031)	(1.118)	(0.168)	(0.143)	(0.177)
Lambda	Lambda constant	1.047***	0.945***	1.091***	1.085***	0.853***	0.745***	1.012***	1.061***	0.962***	1.045***
	Standard Errors	(0.152)	(0.133)	(0.165)	(0.164)	(0.226)	(0.129)	(0.147)	(0.183)	(0.162)	(0.148)
	Coefficient of Covariate	-0.619***	-0.923***	-0.205***	-0.301**	0.517	0.511***	0.545	0.066	0.265	-0.100
	Standard Errors	(0.217)	(0.169)	(0.071)	(0.149)	(0.479)	(0.165)	(3.125)	(0.247)	(0.255)	(0.114)
	Noise constant	0.209***	0.200***	0.207***	0.209***	0.204***	0.203***	0.206***	0.202***	0.209***	0.207***
	Standard Errors	(0.022)	(0.023)	(0.023)	(0.022)	(0.023)	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)
	N	6440	6440	6440	6440	6440	6440	6440	6440	6440	6440
	AIC	6576.977	6558.704	6571.199	6562.934	6558.761	6543.880	6573.401	6559.595	6583.549	6561.669
	ll	-3275.488	-3266.352	-3272.600	-3268.467	-3266.380	-3258.940	-3273.700	-3266.797	-3278.774	-3268.835

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Country and village Dummies not depicted

Table B 17: CPT specification with covariates restricting one parameter at a time

	Gamma			Alpha			Beta			Lambda		
	Total shocks	No shocks	Exogenous shocks	Total shocks	No shocks	Exogenous shocks	Total shocks	No shocks	Exogenous shocks	Total shocks	No shocks	Exogenous shocks
Age	-0.016	-0.014	-0.028	-0.005	-0.005	-0.005	0.001	0.000	0.005	0.005	0.004	0.008
	(0.037)	(0.032)	(0.047)	(0.008)	(0.008)	(0.008)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)
Gender	-0.129	-0.151	-0.121	0.018	0.019	0.018	-0.002	-0.006	0.000	-0.009	-0.013	-0.007
	(0.106)	(0.107)	(0.103)	(0.025)	(0.024)	(0.025)	(0.059)	(0.059)	(0.058)	(0.058)	(0.058)	(0.056)
Education	0.007	0.006	0.007	-0.003	-0.003	-0.003	0.010	0.011	0.009	0.014	0.014	0.013
	(0.013)	(0.013)	(0.012)	(0.004)	(0.003)	(0.003)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
Dependents	-0.001	-0.008	-0.004	-0.006	-0.004	-0.005	0.007	0.005	0.007	0.015	0.014	0.014
	(0.027)	(0.029)	(0.026)	(0.008)	(0.008)	(0.007)	(0.016)	(0.017)	(0.016)	(0.018)	(0.018)	(0.018)
HH income	-0.001*	-0.001*	-0.001*	0.000	0.000	0.000	0.001	0.001*	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Cattle owned	0.001	0.001	0.001	0.000	0.000	0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cambodia	0.024	0.107	-0.043	0.015	-0.012	0.022	-0.217*	-0.204*	-0.182	-0.280*	-0.260*	-0.232
	(0.148)	(0.170)	(0.138)	(0.037)	(0.043)	(0.034)	(0.116)	(0.121)	(0.117)	(0.143)	(0.146)	(0.143)
Shocks variable	0.073	-0.284**	0.018	-0.015	0.071***	-0.011	0.027	-0.095	0.054*	0.022	-0.081	0.063*
	(0.048)	(0.141)	(0.035)	(0.010)	(0.027)	(0.008)	(0.026)	(0.064)	(0.032)	(0.027)	(0.057)	(0.037)
Gamma constant	**1.619*****	**1.759*****	**1.792*****	1.404***	1.397***	1.397***	1.377***	1.377***	1.376***	1.378***	1.378***	1.377***
	(0.280)	**(0.224)**	**(0.341)**	(0.047)	(0.046)	(0.046)	(0.036)	(0.036)	(0.036)	(0.037)	(0.037)	(0.037)
Alpha constant	0.388***	0.381***	0.385***	**0.376*****	**0.341*****	**0.354*****	0.331***	0.331***	0.330***	0.331***	0.331***	0.331***
	(0.027)	(0.024)	(0.028)	**(0.070)**	**(0.055)**	**(0.061)**	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Beta constant	0.624***	0.622***	0.624***	0.640***	0.638***	0.639***	**0.845*****	**0.907*****	**0.798*****	0.895***	0.893***	0.903***
	(0.050)	(0.050)	(0.050)	(0.054)	(0.055)	(0.054)	**(0.171)**	**(0.163)**	**(0.171)**	(0.070)	(0.070)	(0.071)
Lambda constant	1.074***	1.052***	1.061***	0.954***	0.941***	0.940***	0.732***	0.724***	0.745***	**0.674*****	**0.728*****	**0.611*****
	(0.127)	(0.119)	(0.123)	(0.122)	(0.120)	(0.121)	(0.123)	(0.122)	(0.127)	**(0.189)**	**(0.179)**	**(0.187)**

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Village Dummies not depicted. Covariates only included in one parameter at a time (in bold), the others stay unrestricted

Table B 18: EUT specification and single shocks

Covariates	Coefficient
Accidents	-0.750*
	(0.412)
Crop failure	-0.735***
	(0.176)
Fatality	-0.492
	(0.318)
Disease	-0.568***
	(0.173)
Drought	-0.391**
	(0.170)
Flood	0.198**
	(0.094)
High input prices	-0.515*
	(0.263)
Livestock died/severe disease	0.534***
	(0.169)
Low output prices	-0.379
	(0.259)
Other	-0.749**
	(0.313)
Constant	0.749***
	(0.250)

Standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Country and village Dummies not depicted. N=6440.

Table B 19: RDU specification and single shocks

Variable	alpha	r
Cambodia	1.112*	-0.045
	(0.584)	(0.080)
Accidents	1.917***	-0.172
	(0.596)	(0.114)
Crop failure	0.254	0.067
	(0.196)	(0.066)
Fatality in household	0.749*	-0.099
	(0.406)	(0.068)
Disease	1.583***	-0.107
	(0.538)	(0.076)
Drought	0.912***	-0.018
	(0.327)	(0.030)
Flood	-0.125	0.003
	(0.142)	(0.030)
High input prices	0.179	0.120
	(0.626)	(0.103)
Livestock died/disease	-0.533***	-0.073
	(0.195)	(0.055)
Low output price	0.903	-0.080
	(0.599)	(0.071)
Other	3.825	-0.135
	(4.643)	(0.099)
Constant	1.877***	0.308
	(0.296)	(0.224)
Noise		0.032
		(0.024)

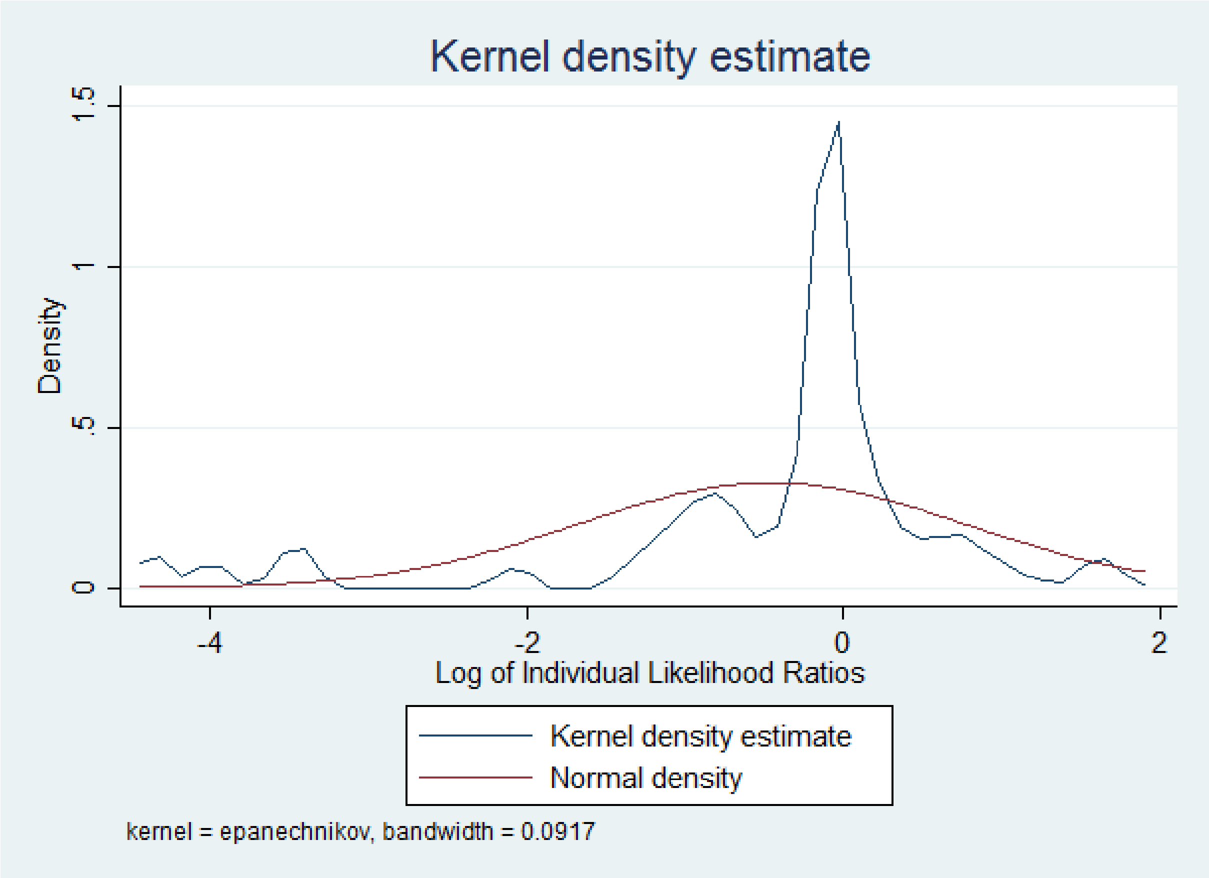
Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Country and village Dummies not depicted. N=6440.

Table B 20: CPT estimation with α=β

CPT Parameter	Coefficient
Gamma	1.416***
	(0.039)
Alpha	0.353***
	(0.023)
Lambda	2.044***
	(0.093)
Noise	0.514***
	(0.075)
F-stat ($H_{0:}$ γ=1)	126.55
p-value	0.000***
F-stat ($H_{0:}$ λ=1)	116.53
p-value	0.000***
N	6440

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Figure B 2: Kernel density plot of distribution of log likelihoods



N=6440

Appendix C. The effect of poor vision on economic farm performance: Evidence from rural Cambodia

Symptoms and limitations imposed on rural dwellers by poor visual acuity. Adapted from Kandel et al. (2018).

1. General ocular limitations (41 mentions): Blurred vision, Cloudy vision, Dazzled vision, Sensation like having a layer or net over the eyes, Distorted vision, Blurred vision at distance, Blurred vision at near, Recognizing objects at far, Reading at far, Seeing things in a relative motion, Headaches, Giddiness or dizziness, Nausea or vomiting, Using my other senses, Floaters in my vision, Flashes of light from within my eyes, Glare from lights, Sensitivity to light, Haloes around lights, Starbursts, Adapting to changes in light, Plain/even roads look like uneven, Prolonged reading, Blurred vision in rain, Poor vision in dim light, Double vision, Seeing objects with ghosts or shadows around them, Difficulty distinguishing colors, Difficulty focusing my eyes, Reading in dim light, Working in dim light, Writing, Riding a motorcycle, Riding, Riding a motorcycle in unfamiliar areas, Riding motorcycle/moped, Walking, Walk, Walking in dawn or dusk, Walking on uneven ground and negotiating bumps or cracks in my path, Going uphill or downhill

2. Agricultural Activity Limitation (19 mentions): Seeing small insect pests in vegetables, Seeing insects while doing agricultural works, Finding insect, Bending down, Looking after domestic animals or pets, Cutting grass, Carrying loads, Harvesting grains, Winnowing rice, Weeding, Seeing small insect pests in vegetables, Doing agricultural works like working in the fields, Finding something when it is surrounded by a lot of other things, Using

microscope, Reading medicine bottle, Bright sunlight, Cutting or chopping food, Using hand tools like screwdriver and hammer, Avoiding some tasks

3. Limitations in access to information (18 mentions):

Reading PowerPoint projected slides, Reading small print, Reading things written on a whiteboard, Telling the time from a clock, Reading the newspaper, Watching television, Reading a watch, Reading store names, Reading hoarding boards, Reading cookbooks, Reading magazines, Reading a book, Reading the phone book, Using a mobile phone, Reading large print, Using the computer, Reading my posts, Reading numbers or letters on the front of a bus or a motorcycle

4. Physical discomfort symptoms and limitations (18 mentions):

Squinting or squeezing my eyes, Feeling ill, Feeling like loss of balance, Discomfort in eyes, Dry eyes, Burning in eyes, Watery eyes, Grittiness in eyes, Red eyes, Stinging in eyes, Itchy eyes, Discharge in eyes, Loss of peripheral vision, Swelling of eyelids, Tired eyes, Heavy eyes, Pain in eyes, Poor vision in only one eye

5. Limited social interactions (18 mentions):

A crowded environment, Using public transport, Attending social functions, Participating in social activities at night, Meeting friends or family socially, Meeting people for the first time, Getting help and support, Maintaining usual social activities, Maintaining my roles and responsibilities in the family, Meeting someone for the first time, Group activities, People not understanding my eye condition, Taking part in recreational activities, Recognizing someone across the street, Recognizing faces and objects on a photograph, Avoiding classroom, conference hall

6. Limitations in business administration (4 mentions):

Reloading money on a mobile phone using a recharge card, Difficulty reading a wall-mounted calendar, Signing/putting on a signature, Writing on a cheque

7. Psychological symptoms and limitations (12 mentions):

Feel worried, Feel disabled, My eyesight getting worse, Going blind, My prescription (strength of glasses) getting worse, Not knowing what’s going to happen in the future, Fear of falling, Fear of tripping, Fear of getting lost, Feel afraid, Feel nervous, Feel depressed

Table C 21: Logit model of eyesight with age as only predictor

Eyesight	Coef.	SE	z	P>z	95% CI	
Age	-0.10	0.01	-7.52	<0.01	-0.13	-0.07
Obs.	260					
Mc Fadden’s R^2	0.26					
Mc Fadden’s Adjusted R^2	0.25					

N= 260.

Appendix D. Where can the crow make friends? Sci-Hub's activities in the library of development studies and its implications for the field

Figure D 3: The crow holding a key is the logo of Sci-Hub

Picture is taken from https://www.insidehighered.com/news/2016/08/08/letter-publishers-group-adds-debate-over-sci-hub-and-librarians-who-study-it

Erklärungen

1. Hiermit erkläre ich, daß diese Arbeit weder in gleicher noch in ähnlicher Form bereits anderen Prüfungsbehören vorgelegen hat.

Weiter erkläre ich, daß ich mich an keiner anderen Hochschule um einen Doktorgrad beworben habe.

Göttingen, den ....

....

(Unterschrift)

2. Hiermit erkläre ich eidesstattlich, daß diese Dissertation selbständig und ohne unerlaubte Hilfe angefertigt wurde.

Göttingen, den ....

....

(Unterschrift)