UNIVERSITÉ CATHOLIQUE DE LOUVAIN
FACULTY OF SCIENCE, SCHOOL OF GEOGRAPHY
EARTH AND LIFE INSTITUTE
GEORGES LEMAÎTRE CENTRE FOR EARTH AND CLIMATE RESEARCH

EXPLAINING AGRICULTURE AND FOREST DYNAMICS: FROM ECONOMIC STRUCTURES TO AGENTS' EXPECTATIONS.

DOCTORAL DISSERTATION PRESENTED BY

VIRGINIA RODRÍGUEZ GARCÍA

IN FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR IN SCIENCES

THESIS COMMITTEE:

Prof. Patrick Meyfroidt (Supervisor)	Université Catholique de Louvain
Prof. Frédéric Gaspart (Co-supervisor)	Université Catholique de Louvain
Prof. Marie-Laurence De Keersmaecker (Chair)	Université Catholique de Louvain
Prof. Sophie Vanwambeke (Secretary)	Université Catholique de Louvain
Prof. Eric Lambin	Université Catholique de Louvain
Prof. Yann Le Polain De Waroux	McGill University
Prof. Peter H. Verburg	Vrije Universiteit Amsterdam

LOUVAIN-NEUVE

31st AUGUST 2021

“Toutes les grandes personnes ont d’abord été des enfants”

Le Petit Prince, Antoine de Saint Exupéry

First and foremost, I'd like to thank my supervisor Patrick Meyfroidt and my co-supervisor Frederic Gaspart. I would not have decided to start this dissertation without the encouragement and fruitful discussions during the writing of my masters thesis under the supervision of Frederic, so thanks for bringing me here. Likewise, I would never have finished this dissertation without the superb supervision of Patrick. I can say without any doubt that I could not have had a better supervisor. You've given me the best example I could have hoped for on how to carry out multidisciplinary research. Thank you so much for your guidance and understanding during all this time, especially during the hard moments and unexpected events. Thanks for bringing me here too.

I’d also like to thank one of the biggest pilar of my life, my parents and two siblings. Specially to Verónica. Gracias por estar siempre a mi lado, por tu amor e incansable paciencia, nada me hace más feliz como estar cerca de tí. Gracias a los cinco por recordarme que *“Toutes les grandes personnes ont d’abord été des enfants”.* To Melanie, Letizia and Mathilde, thank you so much for being such an important part of my live, and for all we shared together during my stay in Belgium. To Irene…no hay palabras que puedan describir la increíble persona que eres…te quiero tanto…gracias por todo. Last but not least, to Jesús, gracias por cruzarte en mi camino con un emocionante concierto, por ser tan especial, y por acompañarme hasta aquí.

A kind thanks to all members and peers at our institute, Pierre, Monique, and to all the members of this Jury. Thanks for guiding me here.

CONTENTS

ABSTRACT

In the context of growing societal demands for land-based products, pressure on forest and agricultural areas is increasing. This dissertation explains some of the complexities of agricultural and forest dynamics, namely, the indirect consequences of agricultural intensification and the interconnection with agricultural expansion, and the non-linear trajectories and feedbacks of agricultural and forest areas' trajectories. The following three research questions were explored: i) What are the causal relationships between changes in cropland intensity and expansion and contraction of cropland areas? ii) What are the causal relationships and indirect effects between changes in forest, agricultural areas and major socioeconomic drivers? iii) Which factors and conditions shape an agricultural frontier's trajectory? Combining panel data econometric techniques, causality analysis, theoretical modelling and qualitative data from experts' interviews, this dissertation shows the importance of distinguishing between short and long run dynamics. In addition, the results provide evidence of a rebound effect in low and middle-income countries over the short run for several commodities with high price elasticity of demand, as well as a land sparing over the long run for some stable cereals. Furthermore, this dissertation made advances on old-fashioned ways of testing the Environmental Kuznets Curve (EKC), and the results show a rejection of the EKC for most of the country groups and globally, with the exception of high-income countries in the long run and post-forest transition countries in the short run. Lastly, the forest scarcity and the state of forest policy were the most relevant forest transition pathways. When zooming in on the agricultural frontier regions, this dissertation shows the important role of both structural and human agency factors -namely, the agglomeration economies and the actors' expectations- on the frontier trajectory while also providing a framework to compare agricultural frontier areas according to their dependence on initial conditions or on both initial conditions and actors' expectations.

SUMMARY

Balancing the increase in agricultural production to satisfy current and future demands, along with the conservation of the biodiversity and ecosystem structures and the safeguard of poor smallholder's livelihoods is a major challenge for humanity. This challenge involves important complex environmental, social and economic tradeoffs which can be associated with different land trajectories. This dissertation aims to explore the complexity of land systems through the exploration of two kinds of complex interactions on those systems: i) the indirect consequences of agricultural intensification and the interconnection with agricultural expansion, and ii) the non-linear trajectories and feedbacks of agricultural and forest areas' trajectories through the exploration of forest transitions and agricultural frontiers. The following three research questions were explored: i) What are the causal relationships between changes in cropland intensity and expansion and contraction of cropland areas? ii) What are the causal relationships and indirect effects between changes in forest, agricultural areas and major socioeconomic drivers? iii) Which factors and conditions shape an agricultural frontier's trajectory?

Combining panel data econometric techniques and causality analysis, theoretical modelling and experts' interviews, this dissertation shows the importance of distinguishing between short and long run dynamics. These have often been overlooked, yet they represent an essential part of the understanding of forest and agricultural dynamics. This dissertation confirms the existence of a long-run dynamic equilibrium relationship between forest cover area, agricultural area, agricultural intensification (measured through both technological progress and increase of inputs use) , economic development and rural population density. Furthermore, our results provided evidence of some of the major theories explaining agricultural and forest dynamics. A rebound effect was present in low and middle-income countries over the short run for several commodities with high price elasticity of demand, and a land sparing over the long run for some stable cereals. Furthermore, intensification driven by technological progress was generally more likely to result in a rebound effect than intensification due to increases

in input use. In contrast, this dissertation shows that old-fashioned ways of testing the EKC result in biased estimates and erroneous conclusions and it makes advances on this by providing more robust results than previous assessments. Results show a rejection of the EKC for most of the country groups, with the exception of the EKC for high-income countries in the long run and post-forest transition countries in the short run. In addition, this dissertation showed no evidence for a global EKC (for all countries included). Lastly, the forest scarcity and the state of forest policy were the most relevant forest transition pathways. By zooming in on the agricultural frontier regions, this dissertation shows the important role of both structural and human agency factors - namely, the agglomeration economies and the farmers' expectations- on the frontier trajectory. This dissertation provides a framework to compare agricultural frontier areas according to their dependence on initial conditions, or on both initial conditions and actors' expectations. These results have potential normative implications too. Our findings show that agricultural intensification should be accompanied by appropriate policies and other interventions to control and orient cropland expansion. Furthermore, the development strategies of low and middle-income economies should not rely on previous experiences of high-income countries as they are experiencing different trajectories of relationships between economic development and forest cover. I also emphasize the importance of governance as a positive feedback mechanism for forest areas. Finally, the two types of frontier dynamics identified in this dissertation contribute to a better design of policies in the following way: If the trajectory appears mainly driven by initial conditions, policies to steer land use dynamics in one direction or another should focus on modifying the structural conditions including interest rates, agglomeration economies and adjustment costs. Instead, when frontiers are driven by the actors' expectations, this provides an additional policy lever. Policy-makers or other actors may seek to steer land use trajectories by fostering certain visions about the future of a region. Furthermore, the level of agglomeration economies and the changes in agents' expectations are identified as ways in which early warning signals will be transmitted. A good measurement of both will contribute to the detection of warning signals used to anticipate potential land regime shifts.

1. INTRODUCTION

A key challenge of the 21st century consists in satisfying the world's increasing demands for land-based products while not undermining the ecosystems on which we depend for life (Hannah et al., 2020). Global demand for agricultural crops is increasing and it may continue to do so as a result of any future increases in the global population, greater per capita incomes and the heightened use of biofuel crops (Tilman et al., 2011). At the same time, options to expand the area devoted to crop production are diminishing (Phalan et al., 2014). This results in complex socio-environmental trade-offs associated with different trajectories of land systems. Furthermore, the issue is not only about agricultural intensity and expansion but also the nutritional needs of the poorest areas.

On one hand, agricultural activities result in ecological detriments, including land degradation, tropical deforestation, biodiversity loss, greenhouse gas emissions, water and air pollution, climate change and human health (Tilman et al., 2011). On the other hand, agricultural expansion and intensification have served to meet humans' demands for food and other land-based products (Ramankutty et al., 2008). Moreover, it is increasingly recognized that future productivity gains in agriculture need to save not only land but natural resources as well (water quantity and quality, and soil quality, biodiversity, etc.) (Fuglie, 2015). Both, agricultural expansion and intensification could contribute to meet those demands through an increase in crop production and other land-based products. However, the tradeoffs and environmental impacts of those paths are still unclear (Tilman et al., 2011; Goldfray et al 2010). Balancing the increase in agricultural production to satisfy current and future demands, along with the conservation of the biodiversity and ecosystem structures, and the safeguard of poor smallholder's livelihoods is a major challenge for humanity (Foley et al., 2011; Hannah et al., 2020). This challenge involves important environmental, social and economic tradeoffs that need to be understood for a sustainable land system (Meyfroidt et al., 2018). This section introduces the agricultural and forest dynamics which are the base for many unavoidable tradeoffs.

1.1 Global agriculture dynamics

At a global level crop production has been historically driven[1] in large part by expanding the land area devolved mainly to three of the major cereals (rice, wheat and maize) (Grassini et al., 2013). The global spread of human populations and the increase of an intensive land use happened gradually over the past 12,000y. Although, a global acceleration occurred in the 19th century (Oberlack et al., 2019). By the twentieth century, large areas of farmland had been established around the world and there was a dramatic change in the way we farm in the early 1960s when the so-called Green Revolution increased the global per capita food supply (Tilman et al., 2011). As a result, the world's agriculture shifted from a long period of geographic expansion to a period of rapid industrial intensification of lands already in use. Some studies suggest that the period of relative constant total crop area came to an abrupt end in 2002 when land expansion due to the production of the major cereals and soybean was observed; most of it happening in South America, Asia and Africa (Grassini et al., 2013). According to the Food and Agriculture Organization of the United Nations agricultural land and pastures occupy about 38% (12% and 26% respectively) of the Earth's terrestrial surface, and most of this land is used to raise and feed animals (roughly 75%) (FAO, 2011). Altogether, agriculture and pastures represent the largest use of land on the planet (Ellis and Ramankutty, 2008; Ramankutty 2008). In line with this, there is a different way of classifying the current terrestrial land surface -the anthromes characterization- which indicates that 51% of the terrestrial biosphere is categorized by intensive anthromes, 30% by cultured anthromes and only 19% by wildlands. In addition, agriculture has major global environmental impacts: it causes land clearing and habitat fragmentation, one-quarter of global greenhouse emissions result from agriculture (namely from land clearing, crop production, and fertilization), it is the biggest user of freshwater and the major cause of

---

[1] Although, we need to go back enough on history as showed by Oberlack et al. (2019): "most of terrestrial nature (72.5%) was already inhabited by hunter-gatherer and/or early agricultural societies at the beginning of the current interglacial interval, transforming "wildlands into Cultured anthrome".

freshwater eutrophication, and it can harm human health (Foley et al., 2011; Laurance et al., 2014; Ramankutty & Coomes, 2016; Tilman et al., 2011). Beyond the direct (or in situ) environmental effects of land use conversion through agriculture, commodity crop intensification or expansion can lead to land displacement somewhere else, something that has been referred to as indirect land use change (iLUC) (Meyfroidt et al., 2014). If this displacement is the result of conservation restriction in existing production regions, then it is known as "leakage" (R.D. Garrett et al., 2018a). Both of these distant effects of land use changes, and the globalized nature of trade in land-based products play a decisive role in the area's land use trajectories and may lead to regional disparities in those trajectories (Eigenbrod et al., 2020). These factors are explicitly or implicitly included in this dissertation, but they are not the focus of it.

Nevertheless, the supply of agricultural products and ecosystem services are both essential to human existence and quality of life (Tilman et al., 2002). Therefore, solutions to satisfy both are needed. Agricultural expansion and intensification are two options to meet the demands that humans place on land for food production and other land-based goods, nature protection and ecosystem services, and they both have respective socio-environmental trade-off (Meyfroidt et al., 2018).

Changes in global agricultural intensity

Agricultural intensification can be defined as the increase of agricultural output per unit of land and can happen through increases in land productivity i.e., crop yields or in total factor productivity (TFP)[2]. Over the past 60 years, global agricultural production has increased substantially and food production more than tripled (Pretty, 2018), with

[2] Total factor productivity (TFP): a broad agricultural productivity metric that accounts for the contributions of all inputs to production, calculated as a ratio of total agricultural outputs (crop and animal products) to total inputs (land, labour, capital and resource inputs), where outputs and inputs are typically aggregated on a monetary unit basis. TFP growth occurs when total agricultural output grows faster than inputs. TFP can be contrasted with partial factor productivity which focuses on individual factors of production, such as land area. Source: Coomes et al., (2019).

the primary driver initially being input intensification rather than area expansion, which represented less than 20% of total agricultural output growth (Coomes et al., 2019; Fuglie, 2015; Hannah et al., 2020). The agricultural intensification in this period was attributable to improved seeds, more fertilizer, chemical inputs and water. Cereals represent a clear example of this; from 1961 to 2000, the global population more than doubled and per capita cereal consumption increased by 20% but harvested areas of cereals grew only 7% (Byerlee et al., 2014a). Thus, global rates of yields for most major cereal crops were linearly increasing since the start of the Green Revolution (Grassini et al., 2013). However, not all crops experience the same patterns and the prospect for yield increases comparable to those of the past 40 years are unclear (Tilman et al., 2002). Past rates of agricultural productivity growth will probably not be sustained in the 21st century (Fuglie, 2015); Most of the best quality farmland is already used for agriculture. Thus, further area expansion would occur on marginal land that is unlikely to sustain high yields and some crops are approaching a yield potential ceiling determined by biophysical limits (Grassini et al., 2013).

Since the 1990s, most growth in global agricultural output has come not from inputs intensification or extensification but from more efficient use of labor, land, capital and inputs which have boosted the total factor productivity (TFP). This shift toward TFP growth became evident in the industrialized countries in the 1970s and in some developing countries such Brazil and China, in the 1990s and 2000s agriculture (Coomes et al., 2019). In comparison to the period of the Green Revolution, a more efficient input use in the 1990s and 2000s contributed to about three-quarters to agricultural output growth in comparison to less than 10% from input intensification. These TFP gains corresponded to modern advances in genetic and genomic sciences and as well as the discovery of additional options for weed control through biotechnology. In addition, most of these underlying productivity gains appear to have occurred in agricultural sectors other than “grains”. In high-income countries, the increase in TFP has been strong enough since the 1980s to maintain increasing agricultural output while achieving considerable reductions in agricultural inputs. This trend suggests that TFP growth could play an important role in the sustainability of future global farming (Coomes et al., 2019).

Changes in global agricultural land extent

Notwithstanding the predominance of agricultural intensification in the past growth of agricultural production, a share of these production gains also comes from cropland expansion. Between 1985 and 2005, the world's croplands and pastures experienced a slow net increase (about 3%). However, this agricultural expansion had unequal distribution with tropical areas experiencing significant expansion while little changes or a decrease was experienced in other areas such the template zone (Foley et al., 2011). Currently, the rate of global yield gains per area is decreasing and the rate of agricultural expansion has been increasing since 2000, mostly in the tropics (Grassini et al., 2013). However, there is uncertainty about how much future expansion of cropland is likely to occur and where (Eigenbrod et al., 2020). The main direct cause of tropical deforestation to date remains agricultural expansion, especially commercial agriculture (Lambin & Meyfroidt, 2011a). Therefore, agricultural and forest dynamics are directly and indirectly linked.

1.2 Global forest cover dynamics

Forest dynamics and deforestation are one of the main processes of land change with multiple implications for the global environment and livelihoods (Meyfroidt & Lambin, 2011). Forest ecosystems harbor most of Earth's terrestrial biodiversity and they play an essential role in terms of storing carbon and mitigating climate change while contributing to the livelihood of millions of poor people living in those areas (FAO, 2019). Archaeological and ethnobotanical evidence suggest that human activities have influenced forest ecosystems and their biodiversity since ancient times (Ellis & Ramankutty, 2008). According to the Forest Resource Assessment (FRA) in 2020, forests currently cover 30.8 percent of the global land area but they are not equally distributed around the world as more than half of the world's forest are found in only five countries (Russian Federation, Brazil, China, Canada and the United States of America). Furthermore, forest area decreased form 32.5% between 1990 and 2020 and the naturally regenerating forest has decreased by 7% over the past 30 years. Nevertheless, average rates of net forest loss declined by roughly 40 percent between 1990-2000 and 2010-2020 (from 7.84 million hectares per year to 4.74 million hectares per year). As a result, forest losses were reduced in some countries while there was forest gain in others (FAO, 2019). Losses of forest are also not equally distributed; Africa had the highest net loss of forest area in 2010-2020 followed by South America. Instead, Asia showed the highest net gain in forest area during 2010-20, followed by Oceania and Europe. Complementary to the FRA, there are other sources of data on gradual global land changes from satellite imagery that have shown a global greening trend of the Earth (Hansen et al., 2013; Song et al., 2018; Zhu et al., 2016). The greatest tree cover area gain reported happened in Europe, including European Russia. This may be explained by the former Soviet Union and Eastern European socialist countries where a process of natural afforestation on abandoned agricultural land took place following collapse of the Soviet Union (Potapov et al., 2015). All of these sources of data agree on the decreasing trend in annual tropical forest which represents the largest area of net tree canopy loss in the world - estimated as a loss of 2101km2 per year by Hansen et al (2013)-. Most of the losses in tropical forest occurred in South American rainforest with the tropical dry forest of South America having the highest rate of tropical forest loss. Moreover, Brazil exhibited the largest decline in annual forest loss and Indonesia the largest increase in forest loss among other tropical countries between 2000-2012 (Hansen et al., 2013).

These negative changes in tree cover of the tropical rainforest (FAO, 2019) and the fact that tropical regions cover the largest part of the forest (45%) in the world, focused the attention on them. Tropical ecosystems harbor much of Earth’s biological diversity and they play a key role in the global carbon and hydrological cycles. In addition, much of the projected growth in global population will occur in tropical nations where living standards and per capita food consumption are also expected to vary (Gibbs et al., 2010). The main driver of deforestation and forest fragmentation is agricultural expansion (Curtis et al., 2018; FAO, 2019; Hosonuma et al., 2012) by either expansion of commercial agriculture (see ref in Stevenson et al 2013) or local subsistence agriculture to support rural livelihoods (FAO, 2019). Between 1980 and

2000, 83% of all new agricultural land in the tropics came from either intact forests (55%) or disturbed forests (28%) (Gibbs et al., 2010). The surviving forests are often modified to varying degrees and half of the tropical or subtropical forest that persists today has been substantially altered (Laurance et al., 2014). Therefore, changes in agricultural intensity or extension result in environmental detriment directly or indirectly. The relationship between agricultural and forest dynamics is thus evident and it deserves further exploration.

Last but not least, the complexity of interactions between the environment and the production of agricultural products is also linked to the global food production and food security issues. In terms of production, we are currently able to produce enough food to feed the current global needs, as measured in calories produced and compared with the FAO average kcal per capita per day (Meyfroidt et al,, 2018). However, this may not occur in the future and there are still a lot of people who suffer from food insecurity in urban and rural areas with food prices affecting food-insecure households directly or indirectly. Furthermore, a large share of the food-insecure households are still farmers. Since 2002, more than 25% of the increase in demand for stable food crops has come from increase in harvested crop area (Grassini et al., 2013). Nevertheless, changes in consumer tastes and diets can strongly influence future food demands and its environmental implications (Meyfroidt et al., 2018)-e.g. reducing meat consumption or its replacement by insects, cultured meat or imitation meat- if such food behavioral changes occur not only in high income countries but also within the incoming urban middle class in emerging countries. If current trends in population, diets and affluence continue, projections suggest that increasing demands for agricultural products will be met in the remaining areas of arable land in the tropics, mainly in South America and Sub-Saharan Africa bringing important environmental losses with them (Gibbs et al., 2010).

In this introductory chapter, I first explain the concept of complex land systems. I then discuss two core themes which are the focus of this dissertation: spillovers from land use intensification and land use transitions. Next, I clarify the motivation, novelty, and objectives of this dissertation, and outline the remaining structure.

1.3 Complex land systems

> *"Human interactions with ecosystems are inherently dynamic and complex; any categorization of these is a gross oversimplification. Yet there is little hope of understanding and modeling these interactions at a global scale without such simplification."* (Ellis & Ramankutty, 2008)

Land use and cover changes could be understood as being part of an open and complex human-environment system dominated by long distance flows, capital and people (Lambin & Meyfroidt, 2011a). The complexity of those systems requires going beyond one-way causal relations and linear dynamics, and it implies the integration of more complex processes such as distant causes (i.e., telecouplings), indirect consequences (i.e., spillover effects, displacement, leakage) and also feedbacks and non-linear trajectories (e.g., land use transitions). Here we focus on some of those complex interactions of land systems: i) the indirect consequences of agricultural intensification and the interconnection with agricultural expansion, and ii) the non-linear trajectories and feedbacks of agricultural and forest areas' trajectories through the exploration of forest transitions and agricultural frontiers; Both of them being a type of land transitions or regime shifts with different temporal and spatial scales.

1.3.1 Spillovers from land use intensification

Agricultural intensification it is often seen as a tool for sustainability and a key strategy to produce agricultural products while simultaneously using ecosystems in a sustainable way (Egli et al., 2018a; Meyfroidt et al., 2018; Rasmussen et al., 2018). Nevertheless, and beyond the in situ or direct environmental impacts of agricultural intensification, the land use impacts of agricultural intensification are also distant or indirect (also referred as spillovers). Various forms of land use spillovers have been distinguished such as leakage, indirect land use change (iLUC) and rebound effects among others (Meyfroidt et al., 2018). Here we focus on the rebound effect which is a form of spillover where the adoption of intensifying practices stimulates land use expansion and its alternative hypotheses, the land sparing effect.

Land sparing and rebound effect

The relationship between agricultural intensification and expansion is also complex (Meyfroidt et al., 2014). The idea that agricultural intensification may reduce pressure on land by fulfilling a given demand for land-based products with a smaller land base that could be released for other uses such as nature preservation is described by the so-called Borlaug hypothesis of land sparing (Meyfroidt et al., 2018). A key assumption of this theory is that intensification spares land and that this land is returned to nature. However, agricultural intensification can also lead to a rebound effect. This rebound effect, also known as the Jevons paradox, suggests that an increase in the productivity of one factor (e.i. cropland) leads to its increased utilization (Angelsen & Kaimowitz, 1999a; Byerlee et al., 2014a; Hertel et al., 2014a; Lambin & Meyfroidt, 2011a; Patrick Meyfroidt et al., 2018). According to this theory, the increase in productivity of a land system increases its returns and acts as an incentive to expand the land. Therefore, intensification can also lead to a form of spillover that stimulates land use expansion. In addition, the land sparing can be absolute, i.e., resulting in net cropland contraction of the area of study (Rudel et al., 2009) or relative, i.e., reducing per-capita rate of land expansion compared to the counterfactual scenario without the intensification, although the net agricultural area is still increased (Stevenson et al., 2013). Globally, intensification of staple crops through the Green Revolution has resulted in relative land sparing (Rudel et al., 2009; Stevenson et al., 2013; Meyfroidt et al., 2018). However, the potential impacts of land sparing on the environment and livelihoods are debated (Meyfroidt et al., 2018).

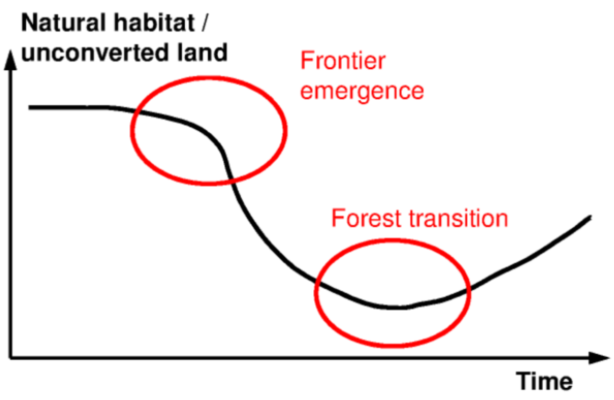
Both hypotheses have been tested with different methodological approaches. From an empirical point of view, the evidence supporting these alternative hypotheses is mixed. Modelling and statistical approaches have found empirical evidence of the land sparing hypothesis under specific subset of crops such the staple crops (Ewers, 2006; Rudel et al., 2009; Stevenson et al., 2013). In some cases, strong land use policies were needed (Phalan et al., 2014; Rudel et al., 2009) or the land sparing effect was of smaller magnitude than the theory proposed (Stevenson et al., 2013). In other cases, only a specific subset of countries was studied (Hertel et al., 2014). In contrast, and using the same modelling and statistical methods, other works found evidence of the Jevons Paradox (Ceddia et al., 2014a). Thanks to these and other works, the conditions under which intensification can spare land or be more likely to lead to a rebound effect have been identified (Meyfroidt et al., 2018). However, the social-ecological outcomes associated with them remain poorly documented (Rasmussen et al., 2018). Multiscale and cross-regional studies are needed because differences across countries, e.g., the rates of land, machinery and suppliers have increased at very different rates and with large differences between continents and periods of time (Peregrini and Fernandez, 2018); and many times, projections based on historical trends of agricultural intensification of expansion do not take into account the cross-national and temporal nuances. Therefore, there is a need for a robust statistical analysis framework across countries (Grassini et al., 2013) and a stronger focus on causal explanation (Rasmussen et al., 2018).

1.3.2 Land use transitions or land regime shifts

Human societies constantly coevolve with their environment through change, instability and mutual adaptation (Lambin & Meyfroidt, 2010). This results on land dynamics characterized by feedbacks and nonlinear trajectories of land which are associated with societal and biophysical system changes through a series of transitions (Lambin & Meyfroidt, 2010). A transition can be defined as a process of system change in which the structural character of the system transforms (Martens and Rotmans, 2002) and the land systems shift from one to another dynamic equilibrium- and a regime shifts has occurred- (Lambin & Meyfroidt 2010; Muller et al 2014). The regime shift concept in complex systems theory is equivalent to land use transitions (Scheffer et al., 2001; Biggs et al., 2012; Filatova et al., 2016; Kull et al., 2017). Thus, the use of the terms “land use transitions” and “land system regime shifts” depends on the context and it is merely a matter of vocabulary. The field of ecology has mainly used the term regime shift (Filatova et al., 2016) but its use has been proven to be a good analytical and communicative tool in the land use context (Kull et al., 2017). Different methodologies have been applied to study regime shifts. The most common modeling approaches to study regime shifts are statistical (e.g., Daskalov et al., 2007), system dynamics (Pacheco et al., 2010), equilibrium (Eboli et al., 2010), and agent-based models (Castella et al., 2005).

Land use transitions or regime shifts are characterized by abrupt changes which are persistent, difficult or even impossible to reverse (Kull et al., 2017) and those processes can occur at different temporal and spatial scales. Scientists and decision makers are challenged by those transitions because the outcome of abrupt sudden shifts is difficult to anticipate (Filatova et al., 2016). On the contrary, when the system only experiences gradual changes, predictions of future states are simpler, and they can be done relying on historic trends and historic data (Filatova et al., 2016). Furthermore, land use transitions or regime shifts have received little attention in land science (Ramankutty&Coomes 2016). Here, we focus on two types of land use transitions or regime shifts occurring at the interplay between agricultural and forest dynamics, these are: i) The frontier emergence as a shift from stable or gradually changing land systems towards rapid expansion of some land use (e.g., from subsistence agriculture to commercial agriculture). ii) Forest transitions, as structural shifts from net deforestation to net reforestation at national or regional scale. Figure 1 represents both types of land transitions although, the scale of time may be different for each of them - forest transitions are usually expected to span over a longer period compared to frontier emergence.

Figure 1: Types of regime shifts in land systems



Source: Meyfroidt, 2015

1.3.2.1 Forest transition

Forest change is not always linear in the long run and a forest transition has occurred in some places over the years. The notion of forest transition was introduced by Mather (1992) and it describes a structural shift from net forest loss to net forest gain through natural regeneration or planted forest. Therefore, forest transition represents an empirical regularity that constitutes a well-known example of a non-linear land use transition (Meyfroidt & Lambin, 2011). Forest transitions in the 19th and early 20th centuries occurred mainly in industrial countries from Europe and North America. More recently, other countries such as China, India, Vietnam, Costa Rica and Puerto Rico among others have experienced a forest transition and similar patterns of forest cover change are taking place in some tropical countries with developing economies (Meyfroidt & Lambin, 2011).

The causes of these transitions are different (Lambin & Meyfroidt, 2010). Nevertheless, three kinds of land use dynamics have been identified that explain how land is made available for restoration of natural ecosystems and reforestation, apart from a decrease in demand, these are: agricultural or forestry intensification (Green et al., 2005), a better match of the spatial redistribution of land use in increasingly integrated markets (Mather and Needle., 1998; Nanni and Grau, 2017), and international trade in land-based products which may facilitate forest recovery in one place by displacing pressure elsewhere (Meyfroidt et al., 2010). Processes of forest transition have been described through interacting pathways combining these three land use dynamics (Jinlong Liu et al., 2017; Patrick Meyfroidt & Lambin, 2011; Thomas K. Rudel et al., 2005): the economic development path, the globalization path, the forest scarcity path, the state forest policy path and the smallholder's pathway (Meyfroidt et al., 2018). In addition, other theories such as the Environmental Kuznets Curve (EKC) that are related to the forest transition have been explored in the literature (Cropper & Griffiths, 2021; Panayotou, T., 1993; Shafik, N., & Bandyopadhyay, S., 1992). Altogether, this literature has brought important insights into our current understanding of forest transitions.

In particular, they have allowed us to identify a wide set of relevant variables, and they provide short-run empirical evidence of some hypotheses and theories about forest dynamics. However, these works had little emphasis on the causal analysis of forest dynamics, and they do not build on an explicit causal modelling linking the variables under study and their indirect effects. Comparative works that could test those theories on a unified conceptual causal model are desirable. In addition, works that are based on statistical designs and time series data are often affected by problems of spurious regression that could be overcome by improved econometric techniques. Such works are relevant under the context of international commitments trying to halt deforestation (such as the commitments to deforestation free supply chains or the UN Climate Summit). These initiatives aim for a rapid global forest transition which has been supported by some works (Waggoner&Ausubel, 2001). However, a recent study suggested that a global forest transition may be more challenging than achieving local or regional forest transitions. In addition, the net effect of these forest transitions on global forest areas remains an open question due to the lack of comprehensive data on deforestation drivers (Pendrill et al., 2019).

1.3.2.2 Agricultural frontier

Frontier areas could be defined as places with abundant land and natural resources compared to a relative lack of capital and/or labor to exploit it, and rapid land use changes (Eigenbrod et al., 2020; Meyfroidt, 2018). Those areas are often related to the notion of "opportunity" and they imply the human agents rush for the frontier resource appropriation. Frontier areas often labelled based on the land uses they harbor (e.g., agricultural frontier, mining frontier, forest frontier, conservation frontier or reforestation frontier) or the actors that operate on them (e.g., smallholder frontier, commercial frontier) (Kronenburg García et al., 2021). This dissertation focuses on the agricultural frontier areas which are often inhabited by people who use the land for subsistence purposes and local markets. On these frontiers, the expansion of commodity crops can occur as these local actors switch to new land use, but oftentimes they are displaced to more marginal lands or lose access to important resources by more powerful foreign actors or local elites

(Kronenburg García et al., 2021). Thus, frontier areas are characterized by their complexity with multiple processes interacting at different scales, their uncertainty and the diversity of actors together with their perceptions and interactions (Junquera et al., 2020; Patrick Meyfroidt, 2015). Those actors interact to create a dynamic land use system that is space and time dependent, where feedbacks between human activities, land use change and ecological dynamics produce nonlinearities (Rindfuss et al., 2007). In addition, those areas are influenced by factors at different levels; micro level factors such as individual decisions of heterogeneous actors, meso-level institutions and organizations and macro-level factors such as international policies or geopolitics.

Several theories have been proposed to explain transformative processes occurring in frontier areas. The classic pattern of frontier development explains the frontier expansion process as subsequent steps moving from a populist or pioneer state, dominated by smallholders, toward a capitalized or consolidated state, where powerful actors consolidate land into large holdings (Pacheco, 2005; le Polain et al., 2018). This pattern took place in North America from the 18^{th} to early 20^{th} centuries and was compared with the development of other frontier areas in the same era (Barbier, 2012). However, contemporary frontiers have deviated significantly form the classic pattern. An example, it is the dualistic pattern described by Barbier (2012) in which modern and profitable commercial economic activities coexist with more traditional and relatively poor agricultural activities. Furthermore, in addition to the structural factors that influence the frontier development, there is an important role played by the heterogeneous actors' agency on these frontiers and the different responses of these actors to changes to land rents (le Polain et al. 2018; Jepson et al., 2006).

All of these factors have been used on works have that have contributed to the understanding of land change processes on specific frontier areas (Dwyer & Vongvisouk, 2019; R.D. Garrett et al., 2018a; Jepson, 2006; Junquera & Grêt-Regamey, 2019; Kronenburg García et al., 2021; le Polain de Waroux, 2019; R. Müller et al., 2012; Ramankutty & Coomes, 2016; Rindfuss et al., 2007). However, they have mainly focused on case studies and current explanations are limited to ex post facto explanations (Ramankutty & Commes, 2016). Thus, the literature has identified the need for: i) developing better theoretical frameworks for studying regime shifts (Ramankutty & Commes, 2016), ii) methodologies that will permit and foster hypothesis testing and generalizability across frontier areas (Rindfuss et al., 2007) and a iii) compilation of data from frontier case studies. On this last point, recent initiatives have arisen such as the database created by the Stockholm Resilience Centre (https://www.regimeshifts.org). However, it is important to understand the extent to which frontier regions can be understood based on past events (Dearing et al., 2010) and in which cases these approaches are not sufficient. The same applies to predictions for land use changes, if the system only experiences gradual changes, predictions of future states could be done relying on historic trends and historic data (Filatova et al., 2016) but the same cannot be expected for frontier regions where anticipating the timing and nature of land regime shifts is extremely challenging (Müller et al., 2014). This dissertation contributes to this problematic by characterizing agricultural frontier trajectories based on either past and current levels of agglomeration economics or the farmer's expectations about the future.

1.4 Causality

A major contemporary scientific challenge is to develop ways to resolve causality, not just correlation, in complex systems (Larse et al. 2014). Many works on the land systems science (LSS) are concerned with understanding the causes and consequences of land changes (Geist et al., 2006). However, the terminology they use in this field for causal analysis remains imprecise with studies often relying on terms like "drivers", "driving forces", "determinants", "factors" which hint at causal relations without clearly articulating them (Meyfroidt, 2016). Thus, the concept of cause remains elusive and there are different notions of causality (Meyfroidt, 2016). This dissertation goes beyond the state of the art of those previous studies by explicitly tackling the issue of causality through the definition of two notions of causality (long and short run Granger causality and theoretical causality) and a methodological approach that allows to quantify them (namely, the error correction model and the characterization of trajectory path to a stable equilibrium of the frontier).

First, in Chapters 2 and 3, I used the Granger causality which is characteristic of time series data. This approach considers that a causal relation exists when "past values of one series (x_t) are useful for predicting future values of another series (y_t), after past values of y_t have been controlled for" (Wooldridge, 2009). Under the premise that causes must occur before the outcome they produce, this causality can be operationalized through time series analysis or through panel data analysis as it has been the case in those chapters. As part of the methodology in Chapters 2 and 3 (a cointegration model, see Section 1.5.3), I used the Granger causality on the short and long run. This short run causality is linear and it allows to identify some of the causal effects, while the long run causality is systemic and it does not allow to estimate those mechanisms, only the net effect of many of those causal mechanisms while the structural relations remain hidden. Building on these types of causalities (the long run systemic one and the linear one in the short run) and the use of structural relations in the short run, we are able to explore the complex causal pathways of the processes studied in this dissertation. These complex causal pathways include: variables that are both causes and outcomes of each other (e.g. cropland intensification and cropland expansion), situations when the causal relationship is cause by a third variable or situations where the same initial cause (e.g., agricultural intensification) triggers several mechanisms with counteracting effects (e.g., the land sparing and rebound effects).

Second, in chapter 4, I have used a different notion of causality. In this chapter, the causal mechanism is a representative agent model. This agent optimizes an objective function (the discounted sum of future profits) under some constrains which reflect her environment, and this decision makes all variables connected -some variables are decided by the representative agent and some others are parameters describing the environment around her-.

1.5 General framework

1.5.1 Motivation and objectives

The main aim of this dissertation is to explore the complexity of gradual and abrupt changes in agriculture and forest areas trajectories. Gradual changes in agricultural land and forest have been observed and analyzed through the use of satellite or statistical data and modelling or econometrics techniques. However, the agricultural and forest dynamics are complex in themselves and in relation with each other. I carried out a global cross-country study that explores the main hypothesis explaining agricultural and forest gradual changes using rigorous econometric techniques. In addition, abrupt changes have received comparatively little attention in land system science and mainly focusing on forest transitions rather than agricultural transitions. The present dissertation contributes to fill this gap through the exploration of the following three research questions:

(i) What are the causal relationships between changes in cropland intensity and expansion and contraction of cropland area?

(ii) What are the causal relationships and indirect effects between changes in forest, agricultural area and major socioeconomic drivers?

(iii) Which factors and conditions shape an agricultural frontier's trajectory?

1.5.2 Conceptual framework

Most studies focus on either the agricultural, forest or frontier dynamics but they rarely take an approach that integrates all of them. This dissertation benefits from an approach that combines an exhaustive analysis of the agricultural and forest dynamics on their own and also, the study of the areas in which both land uses interplay directly-frontier regions. Figure 2 represents the conceptual framework of this dissertation. This framework is composed by a set of macro-level variables affecting forest and agricultural land (see periphery of Figure 2), and a core space where the agricultural area and forest interplay directly. These macro level variables located at the periphery of the figure represent:

i) the economic development of a country, this is a broad concept which is usually measured with increases or decreases of the Gross Domestic Product (GDP). At certain levels of GDP per capita and capital stock depending on each country, the process of economic development may drive the urbanization and industrialization of the country which in turn drives the labor force away from rural areas at the same time that agricultural intensification happens on the most suitable land. Economic development is therefore connected to all other variables in this framework;

ii) the government quality which represents a country's performance in different areas associated with governance such as the control of corruption, the government's effectiveness, the regulatory quality, the rule of law and the voice and accountability of government institutions in the country. Governance is also a transversal variable that impacts everything else in our framework which at the same time may feedback on the governance of the country;

iii) the countries' rural population density as an important demographic characteristic of a country;

iv) the trade of agricultural and forest products with other countries, not only measure as quantity or area of forest and agricultural product traded, but also the destination of this trade (if the trade partners have higher, lower or similar income than the country of origin). Both elements of trade impacts changes on agricultural and forest cover areas but they also influence the economic development of the country;

v) the producer prices which are an average of prices received by farmers at the first point of sale for their forest and agricultural products nationally. Prices are usually identified as mayor drivers of land changes and they also influence the economic development of a country;

vi) the agricultural productivity which is a measure of the quantity of agricultural outputs produced by a given quantity of inputs. It can be measured through yields (land productivity) or through the total factor productivity (TFP). TFP growth usually reflects changes in efficiency and technological progress. Agricultural productivity is directly, and indirectly linking with agricultural and forest areas as well as to rural population density and economic development of the country; and

vii) climate change which is a global issue affecting agricultural productivity mostly through the rise in temperature and changes in the frequency and intensity of precipitation. Climate change may also have an effect on all the other variables. This framework only includes the effect of those variables on the agriculture and forest areas (the core zone of the figure). However, the different chapters of this dissertation tackle all the described effects.

Lastly, the core zone of the figure represents the place where the forest and agricultural area directly interplay, which are to a large extent agricultural frontier or forest frontier. This area is characterized by non-linear dynamics (see curve between the forest and agricultural area boxes) and the influence of the different actor's cognitive process in land use decision such for example the farmers' expectations.

This framework was built on different steps. The first step corresponded to an extensive literature review about the complexity of agricultural and forest land dynamics. This literature review allowed me to identify the most relevant macroeconomic variables. Figure 2 contains the conceptual and aggregate form of those variables. The second step was to explore the causal relationships between the variables pointed by the different theories and hypothesis on the literature. The third step consisted on the analysis of frontier regions, their main characteristics and the factors influencing land use decision on them. These steps are related to Chapter 2, 3 and 4 of this dissertation on the following way: Steps one and two through the identification the macro economic variables and relationship between them, served as a starting base for Chapters 2 and 3- Chapter 2 focused on the relationship between gradual changes on agricultural land extension and changes in agricultural productivity; instead, Chapter 3 set the forest area as its focus, as well as the different hypothesis about changes in forest area-. Step three, by which I focus on the frontier regions, corresponds to Chapter 4. This Chapter focused on abrupt changes on agriculture land. It also took into account the macro economic variables, but it approached agricultural frontier regions from a micro level perspective in which different actors influence the land changes on the area. A detailed description of these chapter is provided on the next section.

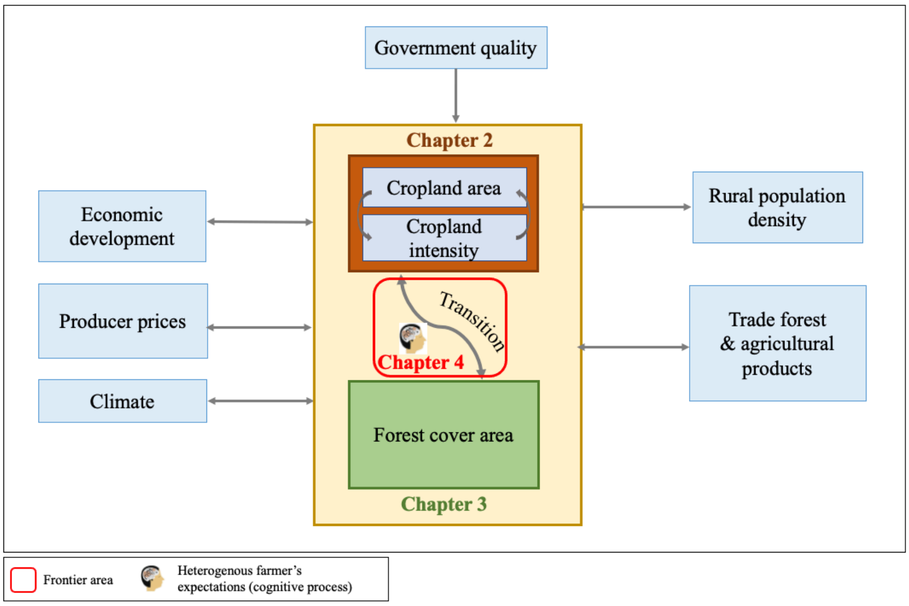


Figure 2: Conceptual framework of this thesis. Land changes on agricultural and forest cover areas are influenced by different macro level variables. Frontier areas harbors the interplay between forest and agricultural cover areas, and they are dominated by different actors whose cognitive process influence land use decision on those areas.

1.5.3 Main methodology

In this dissertation I have used two distinct methodological approaches. In Chapters 2 and 3, I used panel data econometrics to analyze the causal relationship between agriculture area and intensity, and forest area. In Chapter 4, I took a theoretical dynamic modelling approach with quantitative and qualitative data from local experts for the exploration of agricultural frontiers.

For the time series and panel techniques, I used the cointegration approach developed by Granger (1981) which is a solution for the problem of spurious regression (Granger & Newbold, 2001; Wooldridge, 2009). This problem can appear when the distribution (average and standard deviation) of the dependent variable and at least one independent variable change over time – i.e., they are non-stationary. When this happens, the relations observed among non-stationary variables with ordinary least squares (OLS) regressions may not reflect a causal relationship but be actually caused by a third variable, possibly time – known as “spurious” relation (Baltagi, 2005; Wooldridge 2009). There are different techniques that can be applied to solve this problem such the use of first differences’ method or the cointegration model. In this dissertation, I chose to use the cointegration method which consist on the existence of time series variables that are non-stationary individually but there exists a linear combination of them which is stationary. In such a case the variables are said to be cointegrated (Engel and Granger, 1987). This method solves the problem of spurious regression by which correlation is confused with causality, because effects identified in a cointegration regressions were shown to be non-spurious. And it also describes the short and long run causality relations between the variables. Chapters 2 and 3 of this dissertation show the existence of a long run equilibrium relationship between agricultural area, forest and some of the variables included on the general framework (Figure 2)- namely agricultural yield, TFP, GDP per capital and rural population density. Assuming that there is only one cointegration relationship, this implies that those variables have evolved together in an equilibrium relationship during the time period covered by our dataset. Consequently, shocks in the short run do not alter the long run equilibrium relationship. Due to the nature of our variables, cointegrated of order one, there is no possibility of having more than one long run equilibrium relationship.

For the exploration of frontier dynamics in Chapter 4, I used a theoretical model based on economic theory. This model allowed me to characterize the agricultural frontier’s trajectories towards a stable equilibrium. For this, I used two equations to describe a situation around an unstable point (the agricultural frontier). Followed by this, I used qualitative data from local experts to calibrate some values of the parameters that describe the behavior of the model at a specific point of time. This calibration is crucial to describe the behavior of the dynamic system and to answer the research questions of this chapters.

1.5.4. Outline

This section details the outline of the three chapters that constitute the core of this dissertation and their contribution to answer the three main research questions exposed above.

Chapter 2 investigated the empirical relationships between changes in cropland intensity and extent at country scale. For this, I assessed the major hypotheses of the relations between these variables - land sparing, rebound effect and induced intensification- which I performed using a global cross-country panel dataset over 55 years. In this chapter, I tackle the following specific objectives:

i) To explore the effects of changes in cropland area on agricultural intensification, measured through yields and total factor productivity?

ii) To explore the effects of change in cropland intensity on the expansion and contraction of cropland area?

To achieve these objectives, an econometric cointegration approach with additional tests to disentangle long- and short-run causal relations between variables was used with data on cropland area and productivity, and key control variables. In addition, previous methodological issues from econometric analysis were addressed in this work through the use of: two distinct measures of intensification were used as well as lagged variables and first difference estimators to address possible endogeneity; a cointegration to disentangle long- and short-run relationships and identify correlations due to third variables—i.e. spurious relations; and an approach to test for the direction of the causal effects in our long-run estimated coefficients. The cointegration approach is composed by a panel data analysis with an error correction model which allowed obtaining unbiased short-run effects and exploring the long-run equilibrium and causal relationships between these variables. This national- level focus allowed exploring the broad patterns across different regions and different types of crops, using long-term data on units large enough to internalize some of the spillover effects.

Chapter 3 formalized into testable hypotheses some of the most prominent theories explaining forest dynamics on the literature, namely the environmental Kuznets curve (EKC), the forest transition and the ecologically unequal exchange. These hypotheses were tested using a panel dataset covering 111 countries during the period the period 1992–2015 and relying on cointegration techniques to assess both long- and short-run dynamics in forest change, avoiding possible spurious results. Moreover, I attempted to disentangle direct and indirect effects of our independent variables to uncover the mechanisms that underly forest change dynamics. The specific objectives of this Chapter are:

i) To investigate the long- and short-run causal relationships between changes in forest, agricultural cover areas, and socio-economic drivers, going beyond potentially spurious correlations.

ii) To explore the direct and indirect effects of these variables and their interactions on forest cover, by building on a causal model informed by major theories about structural forest cover changes.

To achieve these objectives, I followed several steps. First, the narratives of well-known theories were translated into testable hypothesis and combined into a unified conceptual causal model. Secondly, the long- and short-term mutual relationships and causal pathways between key variables and forest cover were assessed using a cointegration methodology. Finally, for the short run analysis, a comparison between the direct and total effects of the independent variables on forest was performed.

Chapter 4 explores theoretical properties of frontiers dynamics, and it compares different frontiers areas by formalizing the role of agglomeration economies and people's expectations about future market conditions in these frontier dynamics. The model represents the frontier area as an unstable equilibrium between two land systems regimes which are: a pre-frontier situation dominated by semi-subsistence farming and a consolidated frontier dominated by commercial agriculture. The model also considers the presence of three types of actors in those areas. The specific goals of this chapter are:

i) To explore the conditions under which land users will perceive one or the other land regimes as being optimal and will shift their land use accordingly, thereby driving the frontier's dynamics to one or the alternative land regimes.

ii) To understand how agglomeration economies and land users' expectations shape a frontier's trajectory.

iii) To explore the conditions under which agglomeration economies and land users' expectations lead to the dominance of semi-subsistence versus commodity land systems at the frontier region?

iv) To increase generalization around agricultural frontiers.

To achieve these goals, a complex system dynamics perspective was taken, and I translated and explored a theoretical model from economic geography into the context of land use dynamics. A frontier area was the complex system of study, and the model represented the aggregate of all land decision makers at the frontier region. The model allowed to explore theoretical properties of frontiers dynamics as regime shifts and to compare different frontiers areas around the world by formalizing the role of agglomeration economies and people's expectations about future market conditions. The model was calibrated empirically with a set of five contemporary tropical deforestation and land use expansion frontiers cases- Niassa (Mozambique), Bolivian lowlands, Mato Grosso (Brazil), West Kalimantan (Indonesia) and Luang Namtha (Laos)-. The data for the calibration was taken from interviews to experts on each frontier area carried out in 2020. Altogether, this work brought some light into the frontiers' complexity by analyzing when frontier dynamics can be understood essentially based on its initial conditions, or as depending also on the expectations of the land users.

2. AGRICULTURAL INTENSIFICATION AND LAND USE CHANGE

This chapter is based on the following paper:

Rodríguez García, V., Gaspart, F., Kastner, T., & Meyfroidt, P. (2020). Agricultural intensification and land use change: assessing country-level induced intensification, land sparing and rebound effect. *Environmental Research Letters*, *15*(8), 085007.

Abstract

In the context of growing societal demands for land-based products, crop production can be increased through expanding cropland or intensifying production on cultivated land. Intensification can allow sparing land for nature, but it can also drive further expansion of cropland, i.e., a rebound effect. Conversely, constraints on cropland expansion may induce intensification. We tested these hypotheses by investigating the bidirectional relationships between changes in cropland area and intensity, using a global cross-country panel dataset over 55 years, from 1961-2016. We used a cointegration approach with additional tests to disentangle long- and short-run causal relations between variables, and total factor productivity and yields as two measures of intensification. Over the long run we found support for the induced intensification thesis for low-income countries. In the short run, intensification resulted in a rebound effect in middle-income countries, which include many key agricultural producers strongly competitive in global agricultural commodity markets. This rebound effect manifested for commodities with high price-elasticity of demand, including rubber, flex crops (sugarcane, oil palm and soybean), and tropical fruits. Over the long run, strong rebound effects remained for key commodities such as flex crops and rubber. The intensification of staple cereals such as wheat and rice resulted in significant land sparing. Intensification in low-income countries, driven by increases in total factor productivity, was associated with a stronger rebound effect than yields increases. Agglomeration economies may drive yield increases for key tropical commodity crops. Our study design enables the analysis of other complex long- and short-run causal dynamics in land and social-ecological systems.

2.1 Introduction

As populations grow in size and affluence, their demands for land-based products and services increase. Crop production can be augmented by allocating more land to crop production (cropland expansion) or by producing more crops per unit of cropland (cropland intensification) (Meyfroidt, et al., 2018; Thomson et al., 2019; Yao et al., 2018a). Both options create socio-environmental tradeoffs, including the loss of grassland and forest (Curtis et al., 2018). Agricultural intensification, through increases in land productivity – i.e., crop yields – or in total factor productivity (TFP) (Coomes et al., 2019), may reduce pressure on land through the so-called Borlaug hypothesis of land sparing. This hypothesis postulates that intensification, by fulfilling a given demand for land-based products with a smaller land base, releases land for other uses including nature preservation (Meyfroidt et al., 2018). Modeling studies suggested that the intensification of staple crops globally during the Green Revolution from 1965 to 2004 resulted in relative land sparing – i.e., reducing per-capita land demand or the rate of agricultural expansion compared to the counterfactual scenario without intensification (Stevenson et al. 2013).

However, the relationship between agricultural intensification, demand for agricultural products and land use is complex. An alternative hypothesis builds on the “Jevon’s paradox” or “rebound effect” (Ruttan, 1997; Ruttan & Hayami, 1984). The paradox occurs if an increase in the productivity of one factor (here cropland) leads to its increased utilization, in a form of spillover where adoption of intensifying practices increases agricultural profitability and stimulates land-use expansion (Angelsen & Kaimowitz, 2001; Hertel 2012, Byerlee et al., 2014; Lambin & Meyfroidt, 2011, Meyfroidt et al., 2018).

Three main factors have been hypothesized to condition whether agricultural intensification in one country may lead to land sparing or a rebound effect domestically (Byerlee et al., 2014; Hertel et al., 2014; Lambin & Meyfroidt, 2011; Rudel et al., 2009; Villoria et al., 2014, Meyfroidt et al. 2018): (i) land-supply elasticity, (ii) price elasticity of demand of agricultural products and (iii) competitiveness of the intensifying region. First, land-supply elasticity refers to biophysical, technical (e.g., lack of accessibility in some regions) or institutional (land-use policies, tenure or access rules) constraints on cropland expansion, which are expected to promote land sparing. Weak constraints would support a rebound effect. Over 1970-2005, paired increases in yields and declines in cropland for ten major crops were mostly observed in countries with strong land use policies (Rudel et al., 2009). Second, a rebound effect is more likely when intensification results from a technological change shifting the supply curve, and the demand is elastic to prices. This is more likely for products such as meat, luxury or leisure crops (e.g., cocoa, coffee), feed or bioenergy crops, or when markets are well integrated and the intensifying region is large enough to influence prices. With inelastic demand to prices, which is particularly the case for staple crops and when markets are closed, price-reducing intensification will trigger less reactions on the demand and may result in land sparing. Statistical approaches have found evidence supporting this hypothesis for staple crops in developing countries, but non-staple crops have expanded onto lands spared due to the intensification of staple crops (Ewers et al., 2009). Third, when intensification increases the ability and performance of a set of producers to sell a product in a given market compared to competitors – i.e., competitiveness – , a local rebound effect is likely either because consumers substitute other products by the more competitive one, or because the region gains market shares against competitors abroad. In the latter case, increased competitiveness may spare land in other regions especially if the product demand is inelastic (Villoria 2019 a,b).

Conversely, changes in cropland area are also expected to influence intensification. Induced intensification theories, building on Boserup's (1965) insights, suggest that intensification arises as a response to growing demand – due to increases in population density, local consumption patterns, and connections with distant markets – along with scarcity of productive and accessible land for agricultural expansion (Turner & Ali, 1996). This demand is moderated and mediated by availability of technologies and institutional constraints on land-use expansion among others. Over the long term, technological innovation can endogenously respond to changing scarcity of production factors, as agents invest in innovations that enhance labor or land productivity, determined by the scarcity of either (Ruttan, 1997;

Ruttan & Hayami, 1984). Theoretically, induced intensification is compatible with both land sparing and rebound effect, although in practice, intensification induced by land scarcity and population pressure is unlikely to cause a strong rebound effect.

Our objective is to test the three hypotheses of land sparing, rebound effect and induced intensification by exploring the empirical relationships between changes in cropland intensity and extent, at country scale (Figure 2.1). We assess: (i) What are the effects of changes in cropland area on agricultural intensification, measured through yields and total factor productivity? and (ii) What are the effects of change in cropland intensity on the expansion and contraction of cropland area?

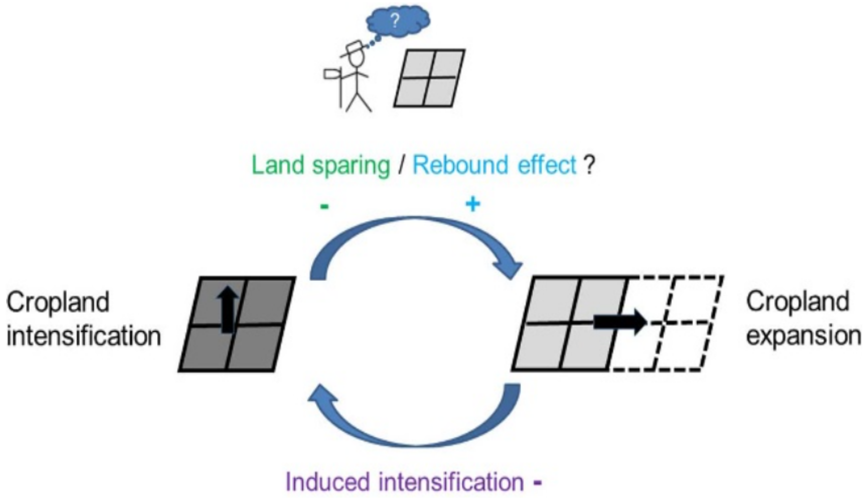


Figure 2.1: Mutual causation between changes in cropland area and intensity, as hypothesized by the theories of induced intensification, land sparing and rebound effect.

We used an econometric approach with a dataset covering 161 countries over the 1961-2016 period – i.e., a panel –, with data on cropland area and productivity, and key control variables (Table S1 (available online at stacks.iop. org/ERL/15/085007/mmedia)). Previous works have explored these questions using general and partial equilibrium models to simulate episodes of agricultural intensification and their effects on cropland (Stevenson et al., 2013; Hertel et al., 2014). Others works used econometric analysis of country-level panels (Ewers et al. 2009; Rudel et al. 2009; Ceddia et al. 2013, 2014). Here we make progress towards addressing methodological issues in this literature: (i) The use of yield growth as a single measure of intensification, (ii) endogeneity between technological change, factor substitution and land use on the long run, possibly related to simultaneity bias, and (iii) corroborating casual inference via causality tests (Byerlee et al., 2014; Villoria et al. 2014). Recent efforts explored the role of TFP and used instrumental variables and ordinary least squares (OLS) to explore the potential effects of simultaneity bias and other sources of endogeneity between TFP and cropland area changes (Villoria 2019 a,b).

Our study addresses previous methodological issues by using (i) TFP – i.e., the efficiency of the overall mix of production factors (land, labor, and capital) due to improved technologies, farmer's skills, and knowledge (Coelli & Rao, 2005; Coomes et al., 2019; Fuglie, 2015) – along with yield, as two distinct measures of intensification, (ii) lagged variables and first difference estimators to address possible endogeneity, (iii) cointegration to disentangle long- and short-run relationships and identify correlations due to third variables – i.e., spurious relations (Granger, 1981), and (iv) an approach designed by Canning & Pedroni (2008a) to test for the direction of the causal effects in our long-run estimated coefficients. This allowed obtaining unbiased short-run effects and exploring the long-run equilibrium and relationships between these variables. Our national-level focus allows exploring the broad patterns across different regions and different types of crops, using long-term data on units large enough to internalize some of the spillover effects.

2.2. Materials and methods

2.2.1 Data and variables

We assembled an unbalanced panel dataset of country-level variables covering the 1961-2016 period (Table S1). Our key variables were retrieved from the FAOSTAT database. This remains the only long-term, cross-country dataset for exploring our research questions, although it has limitations. Although subnational data is not available, FAOSTAT's national-level data provides some insight into the institutional and technological nuances within and between countries. Crop-specific analyses rely on harvested areas, thereby accounting increasing harvest frequency as area expansion, which may have blurred the results for some crops (Portmann et al., 2010), and not adequately capturing intensification through shortening of fallows. Planted area data, which are unavailable in FAOSTAT, would allow avoiding the counting for failed harvest, but would still have other issues including multi-cropping.

We excluded very small countries and those with very little cropland or potential for cropland expansion (Supplementary Information text at the Appendix). We assumed cross-section independence in each group, and parameter homogeneity. We expressed all variables in natural logarithm to interpret the results as percentage change and elasticities, except for TFP which is constructed as a growth rate and directly interpreted as percentage change, and the annual average temperature and rainfall. We first performed the analyses for all crops together, separating the countries in three income groups (SI text, Tables S7, S9, S11, S13 at the Appendix). We then extracted subpanels covering specific crops for the main producing countries for these crops (Tables S2, S3), selected based on the global importance of these crops and the relation with our hypotheses (SI text, Tables S7-S14).

2.2.2 Hypotheses tested

We tested the following hypotheses (Table 2.1):

Hypothesis 1: Following the induced intensification theory, cropland intensity will increase if local pressure on land increases, here represented by an increasing rural population density or lower rates of cropland expansion. We do not explicitly account for changes in demand from distant consumers. As we focus on the land-use impacts of intensification, here we used only yields as the dependent variable representing intensification, not TFP.

Hypothesis 2: With intensification, staple crops with low price-elasticity of demand will exhibit land sparing. In contrast, intensification of flex crops – crops that can be used for multiple purposes including food, feed and fuel, such as oilseeds – and leisure crops, such as coffee and cocoa, is more likely to result in a rebound effect as these crops have higher price-elasticity of demand.

Hypothesis 3: Intensification occurring through increase in TFP, which directly influences competitiveness, is more likely to result in a rebound effect compared to intensification resulting from increasing yield, which has an ambiguous relationship with competitiveness.

Land sparing and rebound effects can occur within the region that experiences intensification or remotely (Meyfroidt et al., 2018). Here we focus on country-level intensification and potential land sparing and rebound effect within the intensifying country, rather than assessing potential spillovers in other countries. In the energy domain, rebound effects include any net decrease in or dampening of the reduction in overall resource consumption stemming from efficiency gain. In land use, the debate around rebound effect has focused on the most extreme form of rebound effect (>100% of rebound effect), called "backfire", when resource consumption actually increases with efficiency improvements (Hertwich, 2005; Sorrell, 2007). In our analyses, a positive effect of intensification on cropland area indicates such a backfire effect. A negative effect indicates that, on net, intensification is land sparing, but low absolute values indicate a partial rebound effect, compared to a coefficient with larger absolute value.

Table 2.1. Hypotheses tested

Hypotheses	Dependent variable	Independent variable	Sign	Expected conditions for which the hypotheses should hold	Selected crops and regions
1	Yield	Cropland	Negative	Staple crops	Wheat, rice, maize, staple

					crops in Sub-Saharan Africa
		Rural population density	Positive	Staple crops, labor-intensive crops	The same as above, plus cocoa, coffee, banana, tropical fruits
2, 3	Cropland area	TFP	Positive	Countries with low competitive advantage, for crops that they import. Countries with land available for cropland expansion. Crops with high price elasticity of demand.	Banana, cocoa, coffee, sugarcane, oil palm, rubber (overall and in Southeast Asia), soybean (overall and in South America), tropical fruits in tropical countries, flex crops, high-elasticity crops
			Negative	Staple crops, crops with low price elasticity of demand. Countries with physical or institutional	Wheat, rice, maize, staple crops in Sub-Saharan Africa

				restrictions on cropland expansion.	
		Yield	Negative	Staple crops, crops with low price elasticity of demand. Countries with physical or institutional restrictions on cropland expansion.	Wheat, rice, maize, staple crops in Sub-Saharan Africa
			Positive	Countries with low competitive advantage, for crops that they import. Countries with high competitive advantage, for crops that they export. Countries with land available for cropland expansion. Crops with high price elasticity of demand.	Banana, cocoa, coffee, sugarcane, oil palm, rubber (overall and in Southeast Asia), soybean (overall and in South America), tropical fruits in tropical countries, flex crops, high-elasticity crops

2.2.3 Time series properties of the data

Many time series and panel techniques assume that the series under investigation are stationary. However, time series analyses are often affected by the problem of spurious regression (Granger & Newbold, 2001; Wooldridge, 2009). This problem can appear when the distribution (average and standard deviation) of the dependent variable and at least one independent variable change over time – i.e., they are non-stationary. This corresponds to time series that are integrated of order 1, I(1), or higher (Wooldridge, 2009a). Relations observed among non-stationary variables with ordinary least squares (OLS) regressions may not reflect a causal relationship but be actually caused by a third variable, possibly time – known as "spurious" relation (Baltagi, 2005; Wooldridge 2009). The notion of cointegration was developed to address the spurious regression problem and make unbiased conclusions on the short- and long-run dynamics (Granger (1981). If two time series are non-stationary individually but there exists a linear combination of them which is stationary, those variables are cointegrated (Engle & Granger, 1987a). Such a cointegration regression is not spurious and describes the long-run equilibrium relationship between the variables, in which there must be causation in at least one direction (Granger, 1988).

We used Stata software for all analyses (version 15.1). We assessed the time series properties of our data (see SI text Appendix) by first testing the stationarity of our variables using the Fisher unit root test (StataCorp, 2017), performed by income groups. The null hypothesis was that all panels contained a unit roots, and are thus non-stationary. The alternative was that at least one panel was stationary for each variable. We included time trend and one lag. Specifications with different lag structures up to four lags introduced collinearity between the lag variables and the other variables. When the variables were non-stationary, we regressed them and tested for the stationarity of the error term of that regression, which indicates cointegration. We tested for panel cointegration using the Kao test, using the Modified Dickey-Fuller t statistics (StataCorp, 2017). As all the long-run equations contained the same set of variables, we performed this test once. Rejection of the null hypothesis indicated that the error term was stationary and that the variables were cointegrated. In some cases, this test was not significant but the error correction term in the short-run equation was significant, we then also concluded that the variables were cointegrated. For a few short-run equations, the error correction terms (ECT) were not significant, suggesting an absence of cointegration. In these cases, we relied directly on the lagged first-difference model, although these results are more uncertain than with cointegration.

2.2.4 Specification of the dynamic error correction model

Our time series were individually non-stationary and cointegrated, and could thus be represented with a dynamic error correction model (ECM) (Engle & Granger, 1987a). We estimated this model using the Engle and Granger two-step procedure. First, we estimated the long-run cointegrating relationship (Eq. 1-2). Second, we specified the short-run equations (Eq.3-4) by introducing the residuals of the long-run regression ($\mu_{1,i,t}$), called the error correction term (ECT). The ECT indicates the degree to which the last period's deviation from a long-run equilibrium influences short-run dynamics. The ECT's coefficients (γ_7, γ'_7) are comprised between [-1; 0] and adjusts the variables towards the equilibrium keeping the long-run relationship intact, with a larger value indicating a more rapid adjustment (Canning & Pedroni, 2008a). The ECM allows obtaining unbiased estimates of the short-run effects of the explanatory variables.

We explored different model specifications including a range of explanatory variables (Table S6), and selected a reduced set of variables, with GDP per capita, rural population density, and average temperature and precipitation as control variables, allowing to include a wide range of countries and a long period (Eq. 1-4). Yield and cropland area are the dependent variables measuring cropland intensification and expansion respectively. To check the model's robustness, we replicated our analyses without TFP, producing similar results for most variables and subsets (Tables S11-S14).

In these equations, *i* represents the country (cross sectional component), *t* the time series and *c* the specific crop or crop group. We used a model with country fixed effects for both long- and short-run equations – controlling for the unobserved heterogeneity between countries – and dummy variables for each year in the long-run equations to control for time fixed effects such as global inputs and outputs prices fluctuations or other uncontrolled weather conditions. As domestic prices mediate the relationship between changes in cropland area and intensity, we did not include producer price indices as control variables (further explanation in SI Text).

We specified short-run models by using first-differences and lagging the variables by one period to avoid endogeneity, such as between TFP and yields (Eq. 3-4).

2.2.4.1 Long-run model specifications:

Eq. 1:

$$LogYield_{i,c,t} = \propto_{1i} + \beta_1 LogCropland_{i,c,t} + \beta_2 TFP_{i,t} + \beta_3 LogGDPcapita_{i,t} + \beta_4 LogRPDensity_{i,t} + \beta_5 AvgRain_{i,t} + \beta_6 AvgTemperature_{i,t} + \beta_7 Timedummy_t + \mu_{1,i,t}$$

Eq. 2:

$$LogCropland_{i,c,t} = \propto_{2i} + \beta'_1 LogYield_{i,c,t} + \beta'_2 TFP_{i,t} + \beta'_3 LogGDPcapita_{i,t} + \beta'_4 LogRPDensity_{i,t} + \beta'_5 AvgRain_{i,t} + \beta'_6 AvgTemperature_{i,t} + \beta'_7 Timedummy_t + \mu_{2,i,t}$$

For all equations:

$$i\epsilon\{1,2,\dots,N\}, t\epsilon\{1,2,\dots,T_i\}, c\epsilon\{1,2,\dots,C_i\}, \mu_{i,c,t}(0,\sigma_\mu^2)$$

2.2.4.2 Short-run model specifications:

Eq. 3:

$$\Delta LogYield_{i,c,t} = \gamma_1 \Delta LogCropland_{i,c,t-1} + \gamma_2 \Delta TFP_{i,t-1} + \gamma_3 \Delta LogGDPcapita_{i,t-1} + \gamma_4 \Delta LogRPDensity_{i,t-1} + \gamma_5 \Delta AvgRain_{i,t-1} + \gamma_6 \Delta AvgTemperature_{i,t-1} + \gamma_7 \hat{\mu}_{i,t-1} + \varepsilon_{i,t}$$

Eq. 4:

$$\Delta LogCropland_{i,c,t} = \gamma'_1 \Delta LogYield_{i,c,t-1} + \gamma'_2 \Delta TFP_{i,t-1} + \gamma'_3 \Delta LogGDPcapita_{i,t-1} + \gamma'_4 \Delta LogRPDensity_{i,t-1} +$$

$$\gamma'_5 \Delta AvgRain_{i,t-1} + \gamma'_6 \Delta AvgTemperature_{i,t-1} +$$
$$\gamma'_7 \hat{\mu}_{i,t-1} + \varepsilon_{i,t}$$

For all equations:

$$i\epsilon\{1,2, \dots ., N\}, t\epsilon\{1,2, \dots ., T_i\}, c\epsilon\{1,2, \dots ., C_i\}, \varepsilon_{i,c,t}(0, \sigma_\varepsilon^2)$$

2.2.5 Long-run causality test

The presence of cointegration is sufficient to prove the existence of at least one non-spurious long-run causal relationship between the variables in at least one direction (Engle and Granger, 1987), but it does not indicate the direction of the causality between the variables. This is because the long-run estimated coefficients represent the net effect of a system of relationships and cannot be directly interpreted as unilateral causal effects (Granger causality) but rather as a systemic causality. These long-run coefficients are unbiased but their standard errors are large and unreliable, and thus not robust for statistical inference (Canning & Pedroni, 2008; Stock, 1987). However, Canning and Pedroni (2008) proposed a way to test for the direction of long-run estimated coefficients in panel data avoiding this long-run inference problem, by carrying out standard hypothesis tests on the short-run coefficients estimated, which are superconsistent. We extended their two-variables design to a seven-variables model (Eq. S1-S6). This test identifies the sign of the ratio of error correction coefficients under a transformed specification, and thus allows us to identify the sign of the possible long-run causal effects between variables, noting that mutual causation can occur.

2.3 Results

2.3.1 Panel unit root tests and cointegration

For all crops aggregated, the statistical properties of most variables and income groups varied through time, as indicated by the results of our Fisher unit root tests (Table S4). The test indicated stationarity in at least one panel of the yield, temperature and rainfall data across all income groups, and rural population density in low-income countries. Yet, visual inspection of these data showed temporal trends in most variables, indicating that most panels were non-stationary (Figures S1-S6). The panel cointegration test results for the aggregated analysis indicated that the variables in our long-run equations were panel cointegrated in low and middle-income countries and thus showed a long-run equilibrium relationship (Table S5). In order to treat the different sub-groups with a consistent cointegration approach, we used the results of the Fisher unit root test for the variables that are not crop-specific and the visual inspection of the different yield and cropland variables, which are crop-specific (Figures S7-S12). The error correction terms (ECT) were significant in most short-run equations, justifying the cointegration approach. For a few regressions which had non-significant ECTs, i.e. different groups of oil palm and soybean producers with cropland area as dependent variable, our results may be less reliable.

2.3.2 All crops, by income groups (Figures 2.2, 2.3, Tables S7, S9)

The coefficients of the long-run regressions are unbiased and thus can be interpreted for their respective samples, but the test for the sign of the causality was significant for only some of those long-run coefficients. TFP had a positive and strong impact on yields for middle and high-income countries in the long run and for middle-income countries in the short run (Figure 2.2, Table S7). Cropland area had a negative effect on yields over the long run within our sample, which covers most countries of the globe over a meaningful time period, but the causality test failed to show statistically significant result beyond the sample. For all countries together, TFP had a positive effect on yields and a negative one on GDP. Over the long run, TFP had a positive effect on cropland area for all countries together, and for low-income countries, TFP had a positive effect and rural population density a negative effect (Figure 2.3, Table S9).

All short-run regressions confirmed a long-run equilibrium between the variables, with the variable ‘residuals’, corresponding to the ECT, being significant and between 0 and -1. Farmers adjusted yields towards the long-run equilibrium generally more rapidly than cropland area, as indicated by the higher ECTs for yield as dependent variable compared to cropland area. Cropland area had a negative effect on yields in high-income countries and positive effect for middle-income countries (Figure 2.2, Table S7). TFP had a positive impact in low, high-income countries and all countries together, while rural population density had a negative effect in middle-income countries and all countries together. For cropland area (Figure 2.3, Table S9), yield had a positive but weak impact in middle-income and all countries together. TFP also had a positive and stronger impact in low- and middle-income countries, as well as all countries together. Rural population density and GDP had positive impacts on cropland area in low, middle-income and all countries together, and low-income countries and all countries together, respectively.

2.3.3 Crop- and region-specific results on yields (Figure 2.2, Tables S7, S8)

Over the long run, cropland area had a negative effect on yields, and thus an induced intensification effect, only for wheat. Yet, supporting the induced intensification thesis, in the short run cropland area had a negative effect on yields for wheat and maize, as well as the three main cereals (wheat, rice and maize) together. The same effect occurred for oil palm in Africa, where this constitutes a main staple crop. Rural population density had a positive impact on yields for several crops, i.e., soybean over the long run, and cocoa, coffee, oil palm in Africa, and soybean over the short run. Cocoa and coffee are labor-intensive crops, and may thus be particularly responsive to increases in labor availability. In contrast, several crops showed a positive effect of cropland area change on yields changes. In the long run, this was the case for cocoa, sugarcane, oil palm, rubber, high elasticity crops and flex crops. In the short run, this was the case for banana, soybean, and soybean in South American frontier countries and staple crops in sub-Saharan Africa.

2.3.4 Crop and region-specific results on cropland area (Figure 2.3, Tables S9, S10)

The intensification variables had a negative effect on cropland area in several cases, indicating a net land sparing effect. For yields, this was the case for rice and the group of the three main cereals in the long run, and for wheat in the short run. TFP had a negative effect for wheat and all countries together in the short and long run, as well as on rubber and the group of rubber and oil palm in Southeast Asian frontier countries, in the short run. In contrast, in other cases we observed a positive effect of intensification variables on cropland area, indicating a strong rebound effect. For yields, this was the case for rubber and flex crops in the long and short run, as well as for banana, cocoa, sugarcane, tropical fruits, high elasticity crops and soybean in South American frontiers in the short run. TFP had a positive long-run effect on harvested area for rice, maize, and all countries together, as well as a short-run effect for rice, banana, coffee, flex crops, staple crops in Sub-Saharan Africa, and oil palm in Africa. The positive effect of TFP increase on rice harvested area likely reflects increases in harvest frequency (e.g. double and sometimes triple cropping), rather than actual rice cropland expansion.

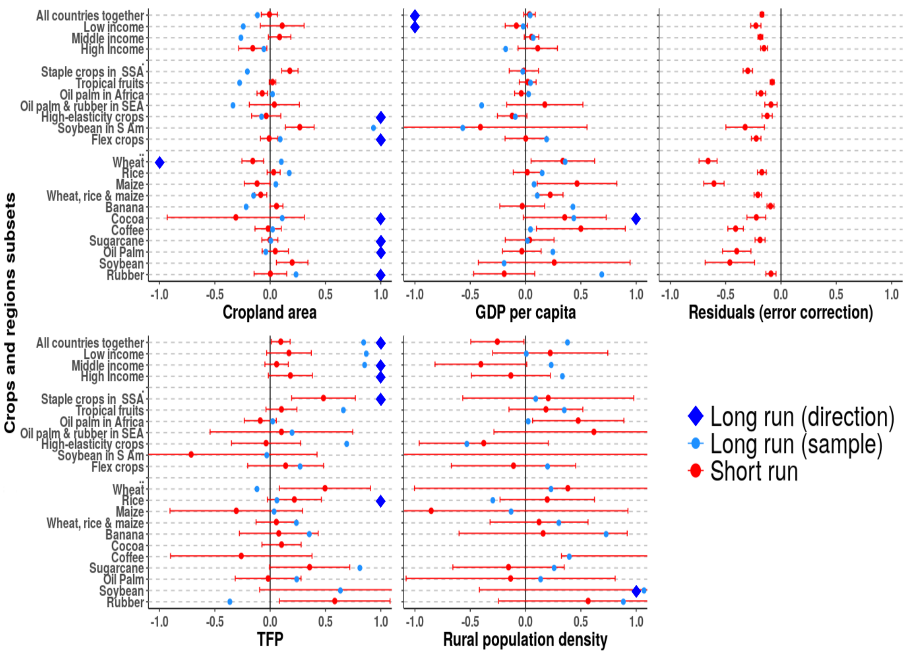


Figure 2.2: Regression results for yields as dependent variable. Results are shown for only the main explanatory variables related to our hypotheses, results for additional control variables (temperature and rainfall) are in tables S7 and S8. For the long-run regressions, the results show the regression coefficient for the long-run sample, the direction of the long-run coefficient as statistically estimated using the modified test from (Canning and Pedroni, 2008), either positive or negative, arbitrarily indicated as 1 or −1. For the short-run regressions, the coefficient and 95% confidence interval (1.96 × standard error) are shown.

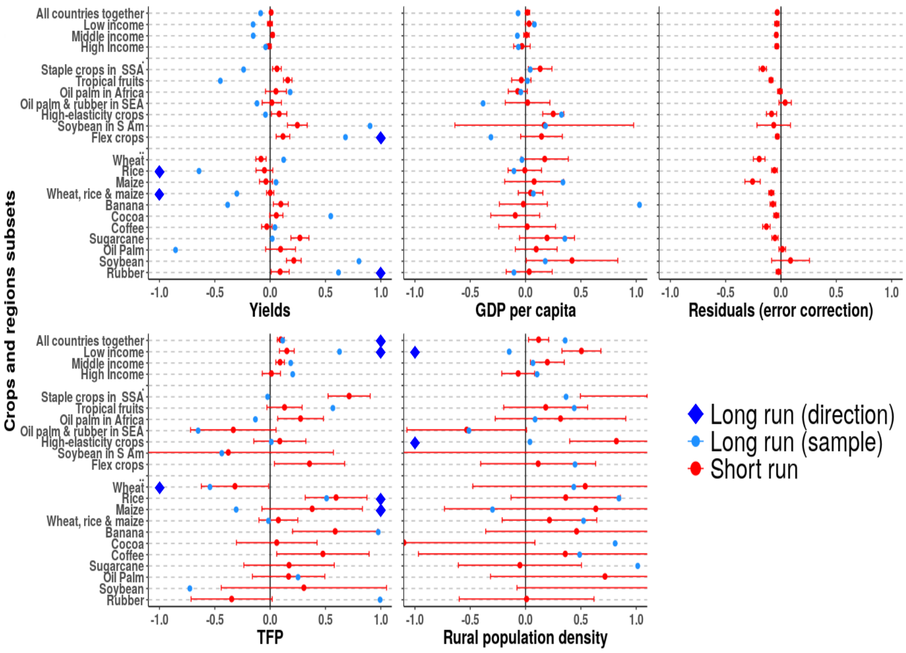


Figure 2.3: Regression results for Cropland as dependent variable. Results are shown for only the main explanatory variables related to our hypotheses, results for additional control variables (temperature and rainfall) are in tables S7 and S8. For the long-run regressions, the results show the regression coefficient for the long-run sample, the direction of the long-run coefficient as statistically estimated using the modified test from (Canning and Pedroni, 2008), either positive or negative, arbitrarily indicated as 1 or −1. For the short-run regressions, the coefficient and 95% confidence interval (1.96 × standard error) are shown.

2.4 Discussion

The induced intensification theory supporting our *Hypothesis 1* has a solid body of local evidence (Meyfroidt et al., 2018; Turner & Ali, 1996). In our results, aggregating all crops, we found significant evidence for intensification induced by restrictions on cropland expansion, be they biophysical, technological, economic or institutional, only in the short-run in high-income countries. Over the long run, though, our unbiased coefficient supports the hypothesis for all our income groups for our sample of countries and years, which represent a crucial period of intensification such as through the Green Revolution (Pellegrini & Fernández, 2018; James R. Stevenson et al., 2013). Increased population pressure caused cropland expansion in the short run in low and middle-income countries, but the opposite over the long-run in low-income countries. This suggests that although the short-term response of increasing population pressure is generally to expand cropland, over the long-run induced intensification may reduce cropland expansion in low-income countries. We also found support for our *Hypothesis 1* for staple crops such as cereals and oil palm in Africa, as well as for labor-intensive crops such as cocoa and coffee in the short run. Staple crops respond more directly to local demand than non-staples such as high value crops, and thus are more likely to see their production increasing through intensification if expansion is constrained.

In contrast, for several key commodity crops, cropland expansion induced an increase in yields over the long run (cocoa, sugarcane, oil palm, rubber, the group of high-elasticity crops and flex crops) or the short run (whole cropland area in middle-income countries, staple crops in SSA, soybean, banana). This could suggest the presence of *agglomeration economies* for these crops. When a country progressively expands the production of these crops, sometimes by specializing into them, producers may benefit from positive externalities such as increased knowledge transmission, skilled labor force, access to cheaper inputs, policy support, and others, which allow for yield improvements, as shown for soybean in Brazil (Garrett et al., 2018; Richards, 2018) and banana and pineapple in Costa Rica (Jadin et al., 2016; Shaver et al., 2015). Induced intensification and agglomeration economies are not mutually exclusive intensification pathways, but our results suggest that in some circumstances one dominates. Despite the lack of a strong induced intensification effect in our dataset for some key commodities, local studies provide evidence that restrictions on land use expansion can indeed induce intensification even for high-demand commodities, as shown in the Brazilian Amazon for soy and beef (Garrett et al., 2018c; Koch et al., 2017).

We tested conditions under which intensification results in land sparing or rebound effect. Our *Hypothesis 2* differentiated crops depending on their price-elasticity of demand. At the crop-specific level, we found evidence of land sparing over the long run for some staple cereals: With increasing yields for rice as well as for wheat, rice and maize together, and increasing TFP for wheat, and over the short term for wheat with TFP and yields. Among major crops, wheat and rice were shown to have lower growing-area elasticities to price than corn and soybeans (Iqbal & Babcock, 2018). Land sparing with intensification of these crops is thus expected (Phalan et al. 2014; Stevenson et al. 2013). Yet, even for these crops, low absolute values in the short-term effects of yields increase indicates a partial rebound effect that reduces the land sparing.

Other subsets showed evidence of a rebound effect, mostly for low and middle-income countries. For all crops aggregated, yields had a rebound effect in middle-income countries over the short run. Many countries in this group, such as in Latin America, have export-oriented agriculture strongly engaged in global commodity markets, and land available for cropland expansion. Intensification increased their competitive advantage and supported increasing exports as well as growing domestic consumption with economic growth (Yao et al., 2018). Crop-specific results showed a rebound effect of yields or TFP increases for several commodities with high price-elasticity of demand in which middle-income countries are dominant players, including bananas, coffee, sugarcane, rubber, flex crops, tropical fruits, palm oil in Africa, soybean in South American frontier countries, and the group of high-elasticity crops in tropical countries in the short run. Villoria (2019a) also found a rebound effect for countries with large commodity exporting sectors. Over the long run, the evidence for rebound effects is less strong, still existing for some crops such as rubber and flex crops, but disappearing for most other crops and, aggregating all crops, remaining mainly for intensification through technological progress, and only significant in low-income countries. Over the long run, and with economic development, rebound effects may decline as demand for many products saturate, or with stronger constraints on cropland expansion.

Regarding our *Hypothesis 3*, for low-income countries, TFP improvements induced a rebound-effect over both the long and the short run, corresponding to theoretical expectations for countries with poorly competitive agriculture, in which increasing TFP and competitiveness may decrease reliance on imports (Hertel et al., 2014). This result was supported by the short-run rebound effect of yield increases on staple crops area in Sub-Saharan Africa. In contrast, the lack of effect of TFP on cropland area in high-income countries could be related to the much stronger constraints on land use expansion in these countries (Rudel et al., 2009). There, technological progress mostly affects the efficiency of other inputs, in particular labor, rather than land. In the short run, neither the land sparing or rebound effect hypotheses were validated for high-income countries, indicating that other forces such as land use zoning and agricultural policies including subsidies may have stronger effects on cropland area changes. Increases in TFP had a positive impact on yields in many subsets, confirming the importance of continuing agricultural research as a strategy to increase yields (Coomes et al. 2019).

Our analyses measure rebound-effect and land sparing as the effect of changes in one variable on another one – i.e., in relative terms –, but not directly indicating absolute cropland expansion or contraction. Our aggregated analyses summing all crops together allow investigating land sparing and rebound through substitution among crops, but do not account for substitution with other land uses (pasture, forestry). In our crop-specific analyses, the rebound and land sparing effect may represent effects on the total cropland area or substitution for one crop to another. Area expansion of a particular crop can come from land cultivated with other crops, from other land uses such as pastures, or from new conversion of land to agriculture. Thus, we cannot directly draw conclusions regarding net effects on deforestation or other natural habitat clearing. A rebound effect of cropland expansion occurring on pastures may still result in net contraction and land sparing on total agricultural lands, if crops produced intensively substitute for grass as animal feed (Yao et al., 2018b). Further, we only examine the rebound or land sparing within the region that experiences intensification, and thus cannot measure if land use changes in one country have land sparing or rebound effects elsewhere (Villoria 2019 a,b). Finally, the crop-specific analyses used harvested areas of these crops as a proxy for cropland area. Increasing harvest frequency was thus accounted for as area expansion, which likely blurred the results. This is particularly relevant for rice, the main crop for which several harvests per year are grown over significant areas (Portmann et al. 2010), and to a lesser extent for soy which is increasingly double cropped (Spera et al., 2014). Thereby we also could not adequately capture intensification occurring through shortening of fallows. We kept all variables together in one long-run relationship. Further works could also explore these questions by separating the long-run relationships among variables into several equations, with a different estimation of the ECM (Johansen, 1988).

2.5 Conclusions

Human societies face trade-offs between production-oriented land uses and ecological outcomes (Phalan et al. 2014; Phalan, 2018; Meyfroidt et al., 2018). Globalized food systems can contribute to a more efficient use of agriculture globally (Egli et al. 2018; Heck et al. 2018) but can also trigger increased consumption of natural resources (Bruckner et al., 2012; Tramberend et al., 2019). Agricultural intensification is often seen as a key tool for sustainability, to lessen competition for productive land and mitigate associated trades-offs, but the dynamics and spillover effects of intensification remain insufficiently understood (Meyfroidt et al., 2018).

Here, we developed and applied a methodological design that addresses key issues from previous statistical approaches, by analyzing different dimensions of intensification, and assessing long- and short-run mutual causal relationships between cropland area and intensity. We provided empirical evidence that the presence of a land-sparing versus rebound effect of intensification, as well as of induced intensification, varied across subsets of crops with distinct price elasticity of demand and labor intensity, and across regions with different conditions in terms of demand elasticity, constraints on cropland expansion, competitiveness and openness to markets. Our results support the induced intensification hypothesis for low-income countries over the long run to some extent, as well as for staple crops and oil palm in Sub-Saharan Africa, and labor-intensive crops such as cocoa and coffee in the short run. Intensification resulted in a rebound effect over the short run in low and middle-income countries, the latter including many key agricultural producers with strong competitiveness in global agricultural commodity markets, and land available for cropland expansion. This rebound effect manifested over the short run for several commodities with high price-elasticity of demand, including bananas, sugarcane, rubber, flex crops, tropical fruits, palm oil in Africa and soybean in Latin America frontier countries. Over the long run, "backfire" rebound effects remained for some key high-elasticity commodities such as flex crops and rubber for their main producer countries, and soybean within our sample. Intensification of staple cereals such as wheat and rice, as well as total cropland area in middle-income countries, resulted in significant land sparing. Lastly, intensification driven by technological progress was generally more likely to result in a rebound effect than intensification due to yields increases driven by increased inputs use.

Our approach allows distinguishing short-term from long-term effects, which is key as crucial socio-environmental impacts occur when gross natural habitat conversion occurs. These insights can contribute to improve land-use policies targeted to specific contexts and crops, confirming that apart for staple crops over the long run, intensification is unlikely to result in actual land sparing for nature unless accompanied by appropriate policies and other interventions to control and orient cropland expansion (Lambin et al. 2014). Further analyses could replicate our study with distinct sets of crops, countries and control variables, as well as with subnational-level data, to test for additional conditions and further refine theories explored. Our study design, building on cointegration, error-correction models and tests for long-run causality direction could be used for addressing other complex long- and short-run causal dynamics in land and social-ecological systems.

3. LONG- AND SHORT-RUN FOREST DYNAMICS

This chapter is based on the following paper:

Rodríguez García, V., Caravaggio, N., Gaspart, F., & Meyfroidt, P. (2021). Long-and Short-Run Forest Dynamics: An Empirical Assessment of Forest Transition, Environmental Kuznets Curve and Ecologically Unequal Exchange Theories. *Forests*, *12*(4), 431.

Abstract

Forest dynamics are changing at a local and global level, with multiple social and environmental implications. The current literature points to different theories and hypotheses to explain these forest dynamics. In this paper, we formalized some of those theories, the environmental Kuznets curve (EKC), the forest transition and the ecologically unequal exchange, into hypotheses tested with a panel dataset covering 111 countries during the period the period 1992–2015. Considering the nature of our data, we relied on cointegration techniques to assess both long- and short-run dynamics in forest change, avoiding possible spurious results. Moreover, we attempted to disentangle direct and indirect effects of our independent variables to uncover the mechanisms that underly forest change dynamics. The results show that there is a long-run dynamic equilibrium relationship between forest cover area, economic development, agricultural area and rural population density. Furthermore, our results confirmed an EKC for high-income countries and post-forest transition countries, while low- and middle-income economies are experiencing different paths. We showed the importance of government quality as a positive feedback mechanism for previous periods of deforestation when tested for all countries together as well as for pre-transition and middle- income economies. Moreover, in low-income economies, economic development affects forest mainly indirectly through the agricultural area.

3.1. Introduction

Under the current global context of rapid land use changes, deforestation is one of the main processes at stake (Patrick Meyfroidt & Lambin, 2011). The multiple implications of changes in forest cover for climate change and livelihoods have made deforestation a global problem of public concern (Duan & Tan, 2019). As a result, several international sustainability agendas have emphasized the role of forest in transition to sustainability, with special attention to tropical and subtropical regions (Oldekop et al., 2020).

Deforestation, as other land use and cover changes, take place at the plot level over short time periods but those changes act as elementary building blocks of complex and structural processes that take place over broader extents and longer time scales (Meyfroidt et al., 2018). Following this perspective, we focus on theories that explain forest changes from such a structural point of view. The literature exploring these dynamics is extensive and builds on different data and methods, as well as on their spatial and temporal coverage. Here, we focus on cross-country studies using different statistical techniques (Caravaggio, 2020; Culas, 2012; Duan & Tan, 2019; Joshi & Beck, 2017; Leblois et al., 2017; Jinlong Liu et al., 2017; Ogundari et al., 2017). Most of these studies used the well-known [11] forest cover datasets—in particular, the Forest Resource Assessment (FRA) database or older FAO sources such as the Statistical Yearbook, but some have also been using more recent forest cover datasets with high spatial resolution. Altogether, this literature brought important insights to our current understanding of global forest dynamics—in particular, they allowed us to identify a wide set of variables relevant to explaining forest dynamics; they provided empirical evidence for some hypotheses and theories on the short run at different regions of the world; finally, those results were used to give recommendations for forest policies on those regions. However, these works have some limitations. They rely on time series data, which can be affected by problems of spurious regression due to non-stationarity data (Culas, 2012; Ogundari et al., 2017). Yet, few of these works use statistical designs that can correct for these potential spurious relations and disentangle long- and short-run relations between forest cover and socio-economic variables. As a consequence, there is little emphasis on the causal analysis of forest dynamics (Oldekop, et al., 2020). Finally, several works explore the effects of a range of variables whose selection is informed by theories, but do not build on an explicit causal model linking these variables together. As a consequence, these studies rarely explore interactions between independent variables and indirect effects, even though these indirect effects are mentioned as something that should be further assessed on forest dynamics (Crespo Cuaresma et al., 2017; Culas, 2012). Nevertheless, econometrics techniques do exist to overcome these limitations, and they have recently been applied to explain land system issues by Rodríguez García et al., (2020).

In this paper, we aim to contribute to fill those gaps by i) investigating the long- and short-run causal relationships between changes in forest, agricultural cover areas, and socio-economic drivers, going beyond potentially spurious correlations; ii) investigating the direct and indirect effects of these variables and their interactions on forest cover, by building on a causal model informed by major theories about structural forest cover changes, i.e., the environmental Kuznets curve (EKC), forest transition and unequal eco- logical exchange theories. For this purpose, we used an econometric approach with a dataset covering 111 countries over the period 1992–2015. First, we translated the narratives of well-known theories into testable hypotheses, and we combined them in a unified conceptual causal model. Secondly, we assessed the long- and short-term mutual relation- ships and causal pathways between key variables and forest cover. We used forest cover data from a recent land cover database from the Climate Change Initiative (CCI) (ESA, 2017), and a cointegration methodology that solves the main statistical problems on previous literature (Granger, 1981). This methodology is composed by a panel data analysis with an error correction model in order to disentangle the long- and short-run dynamics and causal relationships of forest cover area with the other variables. Finally, for the short-run analysis, we com- pared the direct and total effects of the independent variables on forest by using simple algebra on the coefficients obtained by the least squared method.

The body of this paper is organized as follows: Section 2 presents the conceptual framework of this paper and the hypotheses we tested. Section 3 contains an explanation of the data and methods used for our analysis. Section 4 explains the results obtained. Section 5 discusses the results, while Section 6 draws some conclusions from our results.

3.2 Causal framework and hypotheses tested

This section introduces the theories from which we drew out the hypotheses assessed in this work. Our hypotheses mainly build on the forest transition theories and their sub- sequent paths. The notion of forest transition was introduced by Mather (Mather, 1992) and describes a turning point when the forest cover in a region or country reaches its minimum and stops decreasing. Afterwards, a recovery occurs through the conservation of remaining primary forest, plantations and reforestation (Barbier & Tesfaw, 2015). Therefore, forest transition represents an empirical regularity that constitutes an example of a non-linear land use transition (Meyfroidt & Lambin, 2011). Two main forest transition pathways have been identified (Rudel et al., 2005): (i) the economic development pathway and (ii) the forest scarcity pathway. Later on, recent case studies led to the identification of three more paths that are a contemporary version of the previous ones: (iii) the state forest policy pathway, (iv) the globalization pathway, and (v) the small- holder, tree-based land use intensification pathway, which takes place at a smaller geo- graphical scale (Meyfroidt et al., 2018; Lambin & Meyfroidt, 2010).

Taking into account these forest transition pathways and other environmental eco- nomics and environmental sociology theories, we have grouped our hypotheses in the following three groups: (1) the economic development path and the EKC; (2) the forest scarcity and forest policy pathways; (3) the globalization pathway together with the unequal ecological exchange theory. We formalized these hypotheses in a causal diagram (Figure 1), with their operationalization and expected signs explained in Table 3.1.

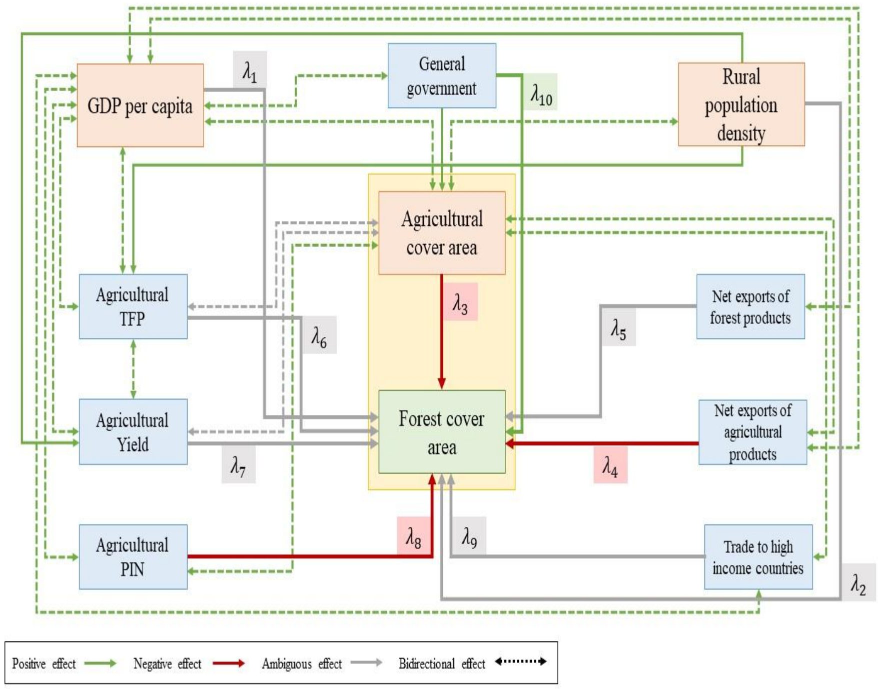


Figure 3.1. Theoretical causal framework. Note: Green, red, and grey arrows identify a positive, negative, or ambiguous effect from one variable to another, respectively. Dotted arrows are double headed, meaning a bidirectional effect between variables. Forest cover area represents the dependent variable of the model. Variables with an orange shape are those implemented in our long-run model while both orange and blue shape variables have been implemented in the short-run model. The $\lambda_{1...10}$ represent variables' coefficients of our models.

This framework (Figure 3.1) is linked with the next section on the hypothesis tested and also with the methods section in the following way: The arrows containing a λ coefficient represent the (direct) effects of those variables on forest, also captured by our econometric models (see coefficients in Equations (1) and (2)). The rest of the arrows represent the indirect effects of the variables on forest (captured by our procedure in Section 3.3.3). We attributed different expected signs to the λ coefficients based on theoretical knowledge about each correspondent theory. We tested these hypotheses by comparing the sign of the coefficients in our model with the expected sign in Table 3.1.

Table 3.1. Hypothesis tested

Hypothesis	**Dependent variable**	**Independent variable** (level or first differences)	**Expected Signs**
Hypothesis 1	Forest cover (% Land)	GDP per capita GDP per capita^2	Negative Positive
	ΔForest cover (% Land)	ΔGDP per capita ΔGDP per capita^2	Negative Positive
Hypothesis 2	ΔForest cover (% Land)	Δagricultural employment	Negative
Hypothesis 3	ΔForest cover (% Land)	ΔTFP Δagricultural yields	Positive Positive
Hypothesis 4	ΔForest cover (% Land)	Δ(Government quality/Forest t-1)	Positive
Hypothesis 5	ΔForest cover (% Land)	Error correction term	Negative
Hypothesis 6	ΔForest cover (% Land)	ΔNet Exports of roundwood ΔNet Exports of agricultural products	Negative
Hypothesis 7	ΔForest cover (% Land)	ΔTrade to high income countries	Negative

3.2.1. Environmental Kuznets Curve and the "Economic Development" Pathway to Forest Transition

The EKC brought its name from Simon Kuznets who first explored the inverse U- shaped relationship between income per capita and income inequalities (Kuznets, 1955). Later on, this was extended to the relationship between indicators of environmental degradation and economic growth (Panayotou, 1993; Shafik, & Bandyopadhyay, 1992). Such EKC was applied to deforestation, by exploring a U- shaped relationship between rate of deforestation or forest cover and the income per capita (Cropper & Griffiths, 2021; Panayotou, 1993). The theory asserts that at low levels of development, the environmental degradation is limited to the subsistence economic activity on natural resources. During the course of economic growth, with increasing consumption, and thus agricultural expansion, income increases (in this paper, we use interchangeably the term income and gross domestic product (GDP)) are first positively related with deforestation (and thus a de- crease in forest cover area). The shift to more industrialized economies spurred the impact of human activities on the environment with further exploitation of natural resources, such as primary forest products boosted by an increasing demand. However, after a certain level of economic development, further economic growth will be associated with a slowdown in deforestation till reaching a transition from forest losses to gains. Multiple mechanisms are hypothesized to allow this trajectory to happen—in particular, changes in economic and industrial structure with the advent of tertiary sectors, technological changes, and the creation of reforestation projects together with policy and societal changes and increased demand for environmental amenities (Hyde, 2012; Panayotou, 1993). The EKC literature for deforestation shows conflictual results concerning the effective existence of a U-shaped relationship between deforestation and economic growth with a recent work which results seems to confirm a possible existence of it (Caravaggio, 2020).

Similar to the EKC, the economic development path of forest transition theory links deforestation to the level of income. This path highlights the processes of urbanization and industrialization which drive labor force away from rural areas at the same time that agricultural intensification happens on the most suitable land. This switch from an agriculture-based economy to an industry or service-based one happens at a certain level of income per capital and capital stock which is different for each country. The sequence of events described by this path are the following: farm workers leave the land for better wages in non-farm sectors; the loss of laborers raises the salaries of the remaining workers and makes agriculture less profitable in areas that can hardly mechanize or with unfavorable agro-environmental conditions (Rudel et al., 2005). Thus, farmers in remote and less productive fields abandon their activity and these lands may potentially experience forest regrowth (Meyfroidt & Lambin, 2011; Meyfroidt et al., 2018). Improved markets and transport networks as well as infrastructures allow food's redistribution from the most productive regions to the least productive ones, reinforcing the abandonment of agriculture in marginal regions and possibly their specialization in forestry. This leads to a U-shaped relationship between time and forest cover which is consistent with the EKC hypothesis applied to forest cover change assuming a continuous economic growth over time. Operationally, we thus used the quadratic form of GDP per capita as an aggregate measure of all structural changes coming from economic development and growth to test the general pattern posited in both theories. Yet, to move further, and because as identified above, multiple mechanisms are bundled in GDP measures, we included additional variables that allow us to test some of these main mechanisms. For this, we operationalized these narratives into the following testable hypotheses:

- *Hypothesis 1*: There is an inverted U-shaped relationship between deforestation and GDP per capita at an aggregate level. We test this hypothesis using the quadratic form of GDP per capita as an aggregate measure of economic growth.
- *Hypothesis 2*: A reduction in agricultural employment results in lower deforestation rates.
- *Hypothesis 3*: Increases in agricultural intensification, measured by either total factor productivity (TFP) and agricultural yields, lead to higher levels of forest area. In addition, we introduced an interaction term between agricultural intensification and GDP per capita on our model, hypothesizing that there may be

a synergistic relation between the effects of these two dynamics on forest cover.

2.2. “Forest Scarcity” and “State Forest Policy” Forest Transition Pathways

Deforestation in a country, due to agricultural expansion, wood extraction or mining activities, may create a scarcity of forest products and a decrease in the ecosystem services that forest provides. In addition, this process could be reinforced by the increase in de- mand for wood products. Such perception of scarcity and degradation may, in turn, drive forest protection efforts, forestry intensification and tree-planting projects (“forest scar- city” pathway) including specific political responses (“state forest policy” pathway). Those feedback responses to low levels of forest cover could contribute to preserve the remaining forest and revert the decreasing trend of forest [1]. This has also been explored by others works such as (Ewers, 2006). Here we operationalized these feedback mechanisms through hypotheses 4 and 5:

- *Hypothesis 4*: Countries with higher government ability to draft and enforce stricter environmental and forest policies have a stronger effect of past decreasing forest area on present forest area. We operationalize this through an interaction between a proxy of environmental governance capacity and lagged forest cover. As a proxy for the government ability to draft and implement environmental legislation, we use an aggregate measure of the following indices: control of corruption, government effectiveness, regulatory quality, rule of law, voice and accountability (see variable “government quality” on Table A1 on Appendix A).

Building on our statistical design using cointegration to investigate the long- and short-term dynamics, another way of testing this effect is through the error correction term (ECT) of the model. This term represents the forest’s speed of adjustment to the long-run forest equilibrium relationship with other variables (here GDP, agricultural area and rural population density). A larger, and more significant value of this parameter indicates that forest cover tends to adjust more rapidly to the long-term equilibrium relation between forest cover and these other variables. A larger value would thus suggest that changes in forest cover produce a relatively rapid feedback on land users' decisions that then affects the forest cover itself. A smaller value would suggest that when GDP and other variables change, and with them the long-term equilibrium level of forest cover, the forest cover takes a longer time to adjust to this new equilibrium. However, it is not possible to specify how this response varies with either positive deviation of forest cover (abundance of forest) or negative deviation (forest scarcity).

- *Hypothesis 5*: There is a national feedback response from forest scarcity or forest abundance to the forest cover under the long equilibrium relationship. We test this hypothesis using the error correction term (ECT) variable: a significant ECT indicates that there is a feedback from forest abundance or scarcity on further forest cover levels, and thus the forest scarcity mechanism appears to hold, with larger values of ECT indicating stronger response.

3.2.3. "Globalization" Pathway of Forest Transition, and Ecologically Unequal Exchange Theory

These two theories point to trade with other countries as a mechanism that drives changes on national forest areas. With the increasing integration of the world via international trade, forest dynamics can no longer simply be explained by national dynamics such the ones explained on the previous forest transition paths (Pendrill et al., 2019). Thus, the globalization pathway represents a modern version of the economic development pathway in which countries are integrate into global markets (Lambin & Meyfroidt, 2010). This integration influences the national-scale forest dynamics and also the national labor, capital, tourism and ideologies. For this work, we focus on the agricultural and roundwood international trade as a process that alters the national forest area. International trade could displace deforestation from countries that undergo a forest transition to regions with higher levels of forest area (Kastner et al., 2011; Lambin & Meyfroidt, 2010; Meyfroidt et al., 2010; Pendrill et al., 2019). We tested the effect of net exports of agricultural products and roundwood on the exporter country through the following two hypotheses:

- *Hypothesis 6a:* National net exports of agricultural products have a negative effect on the national forest cover of the exporter country.
- *Hypothesis 6b:* National net exports of roundwood have an ambigous effect on national forest cover of the exporter country. On the one hand, roundwood exports may con- tribute to forest exploitation. On the other hand, these may incentivize tree plantation and improved management of forest resources (Barbier et al., 2010).

The ecologically unequal exchange theory also identifies trade with other countries as a mechanism that drives changes on national forest areas. This theory points to trade between non-equivalent economically countries as a mechanism for cross-national disparities over access to environmental space (Rice, 2007). As an effort to empirically test this theory, a recent work analyzed transfers of biophysical resources such as materials, energy, land and labor embodied on trade of commodities and services (Dorninger et al., 2021). Other work from Jorgenson (Jorgenson, 2006) applied this hypothesis to deforestation as follows: less-developed countries with higher levels of exports sent to more-developed countries experience greater rates of deforestation. We built our hypothesis on this work (Jorgenson, 2006), and focused on agricultural products as the ones that are more likely to operate through this mechanism. Based on his work, we formulated the following hypothesis:

- *Hypothesis 7:* Exports of agricultural products to high-income countries have a negative effect on the national forest cover of the exporter country. We tested this hypothesis for low-, middle-, and high-income country groups, the latter being for consistency purposes as the theory is not supposed to apply to trade between high-income countries.

3.3 Materials and methods

3.3.1. Data and Variables

We assembled an unbalanced panel dataset of country level variables covering the period 1992–2015. The dataset considered 111 countries, excluding those with less than 1 million hectares of forest cover in 2000. Those countries classified geographically are shown in Table 3.2 but we used different classification for our analysis (Appendix B). Our data were retrieved from different sources. The dependent variable, forest cover, came from the CCI database (ESA, 2017). The CCI database is a land cover dataset of the European Space Agency (ESA) derived from remote sensing. The category of forest tree coverage therein accounts for both natural forests as well as forest plantations. Many studies have used the Food and Agriculture Organization's FRA data on forest area. Even though the CCI database cover shorter time span than the FRA's data, we decided to use this dataset because of its focus on land cover (forested areas as measured by remote sensing with a spatial resolution of 300 m) instead of land use (land designated for forest use, as in the FAO data). We acknowledge that there are other relevant sources such as the widely used Global Forest Watch (GFW) dataset of Hansen et al. (2013). Nevertheless, we decided to rely on the CCI database for two main reasons: first, the time coverage, which is larger for CCI data, 1992–2015, compared to GFW, 2000–2019; second, GFW data consider only annual gross loss of forests, with less temporal details in measuring forest gains such as from natural regeneration or forest restoration. Because our analysis aims in assessing forest transition's hypotheses, we needed a reliable source able to account for both forest losses and gains. Eventually, although some other land cover sources are characterized by a larger time span (Jinlong Liu et al., 2017; Song et al., 2018), several of our explanatory variables have shorter time span, which match with the CCI dataset. The list of all variables selected in our analysis are reported in Table A1 on Appendix A and, a reduced form of the table is presented below (Table 3.3).

Table 3.2. Countries included in the analysis.

Africa	America	East and South Asia and Pacific	Europe and Central Asia
Angola	Argentina	Afghanistan	Austria
Benin	Belize	Australia	Azerbaijan
Burkina Faso	Bolivia	Bangladesh	Belarus
Cameroon	Brazil	Bhutan	Bosnia and Herzegovina
Central African Republic	Canada	Cambodia	Bulgaria
Congo, Dem. Rep.	Chile	China	Croatia
Congo, Rep.	Colombia	Fiji	Czech Republic
Cote d'Ivoire	Costa Rica	India	Estonia
Equatorial Guinea	Cuba	Indonesia	Finland
Ethiopia	Dominican Republic	Japan	France
Gabon	Ecuador	Korea, Rep.	Georgia
Ghana	El Salvador	Lao PDR	Germany
Guinea	Guatemala	Malaysia	Greece
Guinea-Bissau	Guyana	Myanmar	Hungary
Kenya	Honduras	Nepal	Iran, Islamic Rep.
Liberia	Mexico	New Zealand	Italy
Madagascar	Nicaragua	Pakistan	Kazakhstan
Malawi	Panama	Papua New Guinea	Kyrgyzstan
Morocco	Paraguay	Philippines	Latvia
Mozambique	Peru	Sri Lanka	Lithuania
Namibia	Suriname	Thailand	Norway
Niger	United States	Vietnam	Poland
Nigeria	Uruguay		Portugal
Senegal	Venezuela, RB		Romania
Sierra Leone			Russian Federation
South Africa			Slovak Republic

Sudan	Slovenia
Tanzania	Spain
Togo	Sweden
Uganda	Switzerland
Zambia	Turkey
Zimbabwe	Ukraine
	United Kingdom

Table 3.3 Summary of variables.

Variable's name	Definition	Source
Forest (%)	Coefficient between: i) Forest cover (ha) (the sum of tree-covered areas and mangroves categories) and ii) land area (ha) (country area excluding area under inland waters and coastal waters).	Land Cover CCI Product User Guide Version 2.0 (2017).
Agr land (%)	Coefficient between: i) Agricultural cover area (sum of herbaceous crops, wood crops, and grassland) and ii) land area.	Land Cover CCI Product User Guide Version 2.0 (2017).
Rural pop density	Population living in rural areas over the land area of the country	World Bank
GDP cap; GDP2 cap	Gross domestic product divided by midyear population.	World Bank
TFP[3]	The ratio of an output index (total amount of crop and livestock	United States Department of Agriculture.

[3] To reduce potential index number bias in TFP growth estimates, cost shares are varied by decade whenever such information is available. For outputs, base year prices are fixed (the base period for output prices is 2004-06). Source: https://www.ers.usda.gov/data-products/international-agricultural-productivity/documentation-and-methods/

	output) to an index of land and nonland inputs (all land, labor, capital and material resources employed in farm production).	
Yield	Aggregate of all crops' harvested production / harvested area for all crops.	FAOSTAT (FAO, 2018)
Agr (PIN)	Producer Price Index (2004-2016=100). It measures the average annual change over time in the selling prices received by farmers (prices at the farm gate or at the first point of sale).	FAOSTAT (FAO, 2018).
Agri empl (%)	Agricultural employment: Employment in agriculture (% of total employment) (modeled ILO estimate).	The World Bank,
Agr prod net ex	Area of land embodied on exports of agricultural products (ha) minus imports agricultural products (ha).	Own calculations using the data from [27].
For prod net ex	Exports of roundwood (m3) minus imports of roundwood (m3)	FAOSTAT (FAO, 2018)
Government quality	An average of the following five index: Control of corruption, Government Effectiveness,	Worldwide Governance Indicators

	Regulatory quality, Rule of Law, Voice and Accountability.	
Trade high	Exports of agricultural products send to high income countries divided by the total exports of agricultural products.	Own calculations using data from [27].
Avg Temperature	Average temperature per year (computed from monthly data)	World Bank,
Avg Rain	Average temperature per year (computed from monthly data)	World Bank

We assumed cross-section independence in each group, and parameter homogeneity. We first performed the analyses separating the countries in three income groups following the World Bank's classification (WB, 2018) (Appendix B). Thereafter, we classified countries according to their forest transition path following the classification suggested by Pendrill et al., (Pendrill et al., 2019) (Appendix B). These classifications were only made once for the whole analysis period (based on year 2018 for income groups, and on the classification directly retrieved from the work of Pendrill et al., (2019), which is constructed over the period 2001–2015). Thus, when showing scatterplots for variables that change over time, the range of values of these different groups may partly overlap.

3.3.2. Time Series Properties of the Data

Many time series and panel techniques applied to test theories about forest cover and area assume that the series under investigation are stationary (a stochastic time series whose probability distributions are stable over time; for a formal definition of stationary, see Wooldridge (p. 378) (2009b) (Leblois et al., 2017; Perman & Stern, 2003). However, this assumption may lead to a misspecification as time series analyses are often affected by the problem of spurious regression (Granger & Newbold, 2001; Wooldridge, 2009b). This problem can

appear when the dependent variable and at least one independent variable are non-stationary, i.e., their distribution (average and standard deviation) change over time. With non-stationary variables, the central limit theorem does not hold, and relations observed among these variables with ordinary least squares (OLS) regressions may be spurious, i.e., not reflecting a causal relationship among the two variables but being actually caused by a third variable, possibly time (Baltagi B H, 2005; Wooldridge, 2009b). The notion of cointegration was developed by Granger (Granger, 1981a) to address the spurious regression problem and make conclusions on the short- and long-run dynamics. If two time series are non-stationary individually but there exists a linear combination of them which is stationary, those variables are said to be cointegrated (Engle & Granger, 1987b). When such a cointegration relation can be identified, the regression is not spurious and describes the long-run equilibrium relationship between the variables, in which there must be causation in at least one direction (Granger, 1981b). The cointegration approach has been used in the energy domain (Narayan & Smyth, 2005; Zoundi, 2017) and by the land use and land cover change scientific community thanks to recent studies such the one from Rodríguez García et al. (2020). This study adopted the cointegration methodology to study dynamic interactions between cropland extent and intensity and it showed how this methodology can be successful on improving the understanding of land use dynamics.
We used Stata software for all analyses (version 15.1). We assessed the time series properties of our data by first testing the stationarity of our variables and assuming non- structural breaks. We used the Fisher unit root test to assess the stationary character of our variables, which allows for gaps in the panel series (StataCorp, 2017). This test performs the Augmented Dickey–Fuller unit root tests on each panel to verify whether a variable has a unit root or, equivalently, that the variable follows a random walk, i.e., a process in which the dependent variable y at time t is obtained by starting at its previous value, $yt-1$, and adding a zero mean random variable that is independent of $yt-1$. This test performs a unit-root test on each panel (made of each variable) separately and combines the p-values to obtain an overall test (StataCorp, 2017). We performed this test by income groups and clusters of forest transition phases by also including time trend and one lag for each test. The null hypothesis was that all panels contained a

unit root, i.e., they were integrated of order 1, I(1) (the order of integration, denoted I(d), reports the minimum number of differences (d) required to obtain a stationary series, namely I(0)), or higher. The alternative was that at least one panel was stationary for each variable. We explored specifications with different lag structures up to four lags, but this introduced collinearity between the lag variables and the other variables of the model. Thus, we decided to use only one lag. When the variables were non-stationary, we continued by regressing them and testing for cointegration, i.e., for the stationarity of the error term of that regression. We tested for panel cointegration using the Pedroni's test (Pedroni, 1999) and we choose to report the Modified Phillips–Perron t statistics for the sake of simplicity. Further results and other cointegration tests, namely (Kao, 1999; Westerlund, 2005), are available upon request. They show mixed results, but as a general pattern they suggest the possible presence of cointegration among the considered variables. The null hypothesis is 'no cointegration' against the alternative of 'all panels are cointegrated.' If the test rejected the null hypothesis of no cointegration, this indicated that the error term was stationary and that the variables were cointegrated.

3.3.3 Specification of the Dynamic Error Correction Model

Our time series were individually non-stationary and cointegrated, thus, they can be represented in the form of a dynamic error correction model (Engle & Granger, 1987). We estimated this model by using the Engle and Granger two-step procedure (Engle & Granger, 1987). First, we estimated the long-run cointegrating relationship (Equations (1) and (2). Second, we specified the short- run equations (Equations (3) and (4) by introducing in them the residuals of the long-run regressions (*μi,*), namely ECT. The ECT indicates the degree to which the equilibrium behavior drives short-run dynamics. The last period's deviation from a long-run equilibrium influences its short-run dynamics. The ECT's coefficient (λ15 in Equation (2)) should be statistically significant and comprised between -1 and 0 to adjust the variables towards the equilibrium keeping the long-run relationship intact, with a larger value indicating a more rapid adjustment (Canning & Pedroni, 2008b). This concept could be formalized through the idea that a proportion of the disequilibrium from one period is corrected in the next period (Engle &

Granger, 1987). The ECM also allows to obtain unbiased estimates of the short-run explanatory variables' effects.

In these equations, *i* represents the cross-sectional component of the data and *t* the time series. We used a model with country fixed effects for both long- and short-run equations— controlling for the unobserved heterogeneity between countries such as climatic and other geographic conditions—and dummy variables for each year in the long-run equations to control for time fixed effects such as global inputs and outputs prices fluctuations or other uncontrolled weather conditions. Due to the lack of data at the national level we did not include timber prices in our analysis. Nonetheless, variations in timber prices on global markets are potentially captured by fixed effects. We also consider including the annual average temperature and rainfall as control variables on the long run to account for broader changes in environmental conditions, but they showed to be stationary in our time series, which is relatively short to capture remarkable changes in climate conditions, hence we decided not to include them in our long-run model. Further, we want to keep the long equation parsimonious and we assume that there is only one long term relationship for simplicity purposes. Thus, we only included gross domestic product per capita (hereafter as GDP), GDP^2, agricultural area, and rural population density on the long-run equation. We also estimated our long-run model with the inclusion of variables on: i) the net agricultural exports, and ii) the net forest product exports. However, these variables had a very small impact on forest dynamics (see Appendix C, Table A2, Table A3, Table A4, Table A5, Table A6, Table A7, Table A8, Table A9). Therefore, we decided to keep the long-run model without these additional variables. In order to characterize indirect impacts, we specified two long-run equations, with and without the agricultural area (Equations (1) and (3). Changes in agricultural area are expected to be a key mediator through which other underlying causes such as GDP and rural population density affect forest cover (Figure 3.1). Thus, using models with and without agricultural area allowed us to identify the net effects of these underlying variables as well as to assess how much of this effect is mediated by agricultural area changes. The presence of cointegration is sufficient to prove the existence of at least one non-spurious long-run causal relationship between the variables. Yet, the

estimated coefficients represent the net effect of a system of relationships and statistical inference is not straightforward (see Appendix D for further explanation). Thus, we use the long run estimators only for our sample.
In the short-run models, we used first differences and we lag the variables by one period to avoid endogeneity between the variables (Equations (3) and (4). As explained above, we limited the number of variables that we included in the long-run regression for the error correction model. However, there is no corresponding limit for the short-run model; thus, we included additional variables to test our hypotheses as long as other control variables. We performed different robustness checks by excluding the variables agricultural production price index or, GDP^2 from our short-run model. For both variables, their absence did not influence much the results; however, we opted for maintaining them in the presented results as they constitute important control variables. We also tested the interaction terms of GDP with TFP and GDP with agricultural yield in the short run. However, we did not include those interaction terms in the final results as they were not significant, and their inclusion would have reduced the degree of freedom or our models.

3.3.1. Long-Run Model Specifications (with and without Agricultural Cover Area)

$$LogForest_{i,t} = \propto_{1i} + \beta_1 LogGDPcap_{i,t} + \beta_2 LogGDPcap^2_{i,t} + \beta_3 logAgri_{i,t} + \beta_4 LogRuralPopDensity_{i,t} + \beta_5 timedummy_t + \mu_{i,t} \quad (1)$$

$$LogForest_{i,t} = \propto_{1i} + \beta_1 LogGDPcap_{i,t} + \beta_2 LogGDPcap^2_{i,t} + \beta_3 LogRuralPopDensity_{i,t} + \beta_4 timedummy_t + \mu_{i,t} \quad (2)$$

For all equations:

$$i \epsilon \{1,2, \ldots, N\}, t \epsilon \{1,2, \ldots, T_i\}, \mu_{i,t} \left(0, \sigma^2_\mu\right)$$

It has to be noted that these long-run coefficients are superconsistent, i.e., the distribution of the estimated coefficients converges to the real value as the sample size tends to infinity (Wooldridge, 2009b). However, their standard errors are large and unreliable, and thus not robust, so statistical inference is not straightforward (Oldekop et al., 2020; Stock, 1987). Thus, we have only used those estimates for our sample without performing statistical inference (see Appendix C for further explanation).

3.3.2. Short-Run Model Specifications (with and without Agricultural Cover Area)

$$\Delta LogForest_{i,t} = \lambda_1 \Delta LogGDPcap_{i,t-1} + \lambda_2 \Delta LogGDPcap^2_{i,t-1} + \lambda_3 \Delta LogRuralPopDensity_{i,t-1} + \lambda_4 \Delta LogAgri_{i,t-1} + \lambda_5 \Delta NetExportsAgri_{i,t-1} + \lambda_6 \Delta NetExportForest_{i,t-1} + \lambda_7 \Delta TFP_{i,t-1} + \lambda_8 \Delta LogYield_{i,t-1} + \lambda_9 \Delta LogAgri(PIN)_{i,t-1} + \lambda_{10} \Delta TradeHigh_{i,t-1} + \lambda_{11} \Delta Goverment_{i,t-1} + \lambda_{12} \Delta AgriEmployment + \lambda_{13} \Delta Temperature(avg) + \lambda_{14} \Delta Rainfall(avg) + \lambda_{15} \mu_{i,t-1} + \varepsilon_{i,t} \quad (3)$$

$$\Delta LogForest_{i,t} = \lambda_1 \Delta LogGDPcap_{i,t-1} + \lambda_2 \Delta LogGDPcap^2_{i,t-1} + \lambda_3 \Delta LogRuralPopDensity_{i,t-1} + \lambda_4 \Delta NetExportsAgri_{i,t-1} + \lambda_5 \Delta NetExportForest_{i,t-1} + \lambda_6 \Delta TFP_{i,t-1} + \lambda_7 \Delta LogYield_{i,t-1} + \lambda_8 \Delta LogAgri(PIN)_{i,t-1} + \lambda_9 \Delta TradeHigh_{i,t-1} + \lambda_{10} \Delta Goverment_{i,t-1} + \lambda_{11} \Delta AgriEmployment + \lambda_{12} \Delta Temperature(avg) + \lambda_{13} \Delta Rainfall(avg) + \lambda_{14} \mu_{i,t-1} + \varepsilon_{i,t} \quad (4)$$

3.3.4. Assessment of Indirect Effects in the Short Run

Based on our theoretical causal framework (Figure 3.1), we further explored the direct and indirect short-run relations between the variables. The short-run coefficients in Equation (2) capture the direct effect of each independent variable on the dependent variable, as well

as the indirect effects through variables that are not included in the model. However, the effect of a given variable $X1$ that is mediated by other variables present in the model is not captured in the coefficient of this variable $X1$. Those effects refer to the correlation between independent variables, also known as multicollinearity. Our goal is not to correct for multicollinearity, as it does not bring any bias to our estimators unless it is perfect collinearity (Wooldridge, 2009b), but to assess the indirect effects between independent variables. If we take into account these indirect effects, the matrix information of the regression would not change but the significance and the absolute size of the regression may change (Baron & Kenny, 1986). To assess the influence of these indirect effects of our independent variables on the out- come variable for the short-run analysis, we carried out two approaches. In the first one, we performed our model without the agricultural area variable, as this variable is a key mediator for many other drivers of forest cover change. In the second approach, instead, we used simple algebra on the coefficients obtained by the least squared method. In this procedure, we decided to exclude the variables of average temperature and average rainfall; we assumed that those variables impact all other independent variables and we used them essentially as control variables, not as variables that are part of our focus on the causal framework (Figure 3.1). This simple way of assessing indirect effects does not bring any bias.
The second approach to assess the influence of indirect effect is characterized by the following four steps (see Appendix E for further exploration of this method, Figure A2, Figure A3, Figure A4, Figure A5, Figure A6, Figure A7, Figure A8, Figure A9, Figure A10, Figure A11):

- Step 1. Short-run regression: direct effects.

This step corresponds to Equation (5) in the previous section but excluding the variables annual average rain and temperature as well as the interaction term between government quality and forest, due to simplicity purposes. From this equation, we obtained the direct effects ($\lambda 1$ to $\lambda 15$) of each independent variable on the dependent variable.

$$\Delta LogForest_{i,t} = \lambda_1 \Delta LogGDPcap_{i,t-1} + \lambda_2 \Delta LogGDPcap^2_{i,t-1} + \lambda_3 \Delta LogRuralPopDensity_{i,t-1} + \lambda_4 \Delta LogAgri_{i,t-1} + \lambda_5 \Delta NetExportsAgri_{i,t-1} + \lambda_6 \Delta NetExportForest_{i,t-1} + \lambda_7 \Delta TFP_{i,t-1} + \lambda_8 \Delta LogYield_{i,t-1} + \lambda_9 \Delta LogAgri(PIN)_{i,t-1} + \lambda_{10} \Delta TradeHigh_{i,t-1} + \lambda_{11} \Delta Goverment_{i,t-1} + \lambda_{12} \Delta AgriEmployment + \lambda_{13} \Delta Temperature(avg) + \lambda_{14} \Delta Rainfall(avg) + \lambda_{15} \mu_{i,t-1} + \varepsilon_{i,t} \quad (5)$$

• Step 2. Orthogonalization of independent variables: indirect effects.

The orthogonalization of variables is a transformation by which variables that were correlated become perfectly uncorrelated, i.e., orthogonal (here we use orthogonal and non-correlated as synonyms; see (Papoulis & Pillai, 2002) for a formal definition of orthogonality). This trans- formation can be achieved without altering the disturbance term of the regression model in which the variables are embedded. The transformation we made consisted on breaking our multivariate short-run model (Equation (3)) into simpler bivariate models. We estimated a set of auxiliary regressions for each targeted independent variable (henceforth named as “$x1$”). We built one auxiliary regression for each variable to which $x1$ has an indirect impact to, and this $x1$ was introduced as a regressor in each of these auxiliary regressions. Afterwards, we estimated the residuals of these auxiliary regressions. They capture the part of the dependent variable’s auxiliary regression that is orthogonal to, or thus not explained, by the variable $x1$. If those residuals are equal to zero, there is perfect multicollinearity between the variable $x1$ and the auxiliary variable—in that case, we would originally not have included both variables in our model. In the example below, we apply those steps for $x1$ being equal to the variable $NetExportsAgri_{i,t}$.

Auxiliary Equation:

$$\Delta \mathrm{LogGDP}_{i,t-1} = \gamma 1 \Delta NetExportsAgri_{i,t-1} + residualgdp \quad (6)$$

Auxiliary Equation:

$\Delta LogGDPcap^2_{i,t-1} = \gamma 2 \Delta NetExportsAgri_{i,t-1} + residualgdp2$ (7)

Auxiliary Equation:

$\Delta LogAgri_{i,t-1} = \gamma 3 \Delta NetExportsAgri_{i,t-1} + residual\text{Agri}$ (8)

• Step 3. Netting out indirect and direct effects: total effect.

Thirdly, we netted out the direct and indirect effects in order to estimate the total or net effect of each targeted variable ($x1$). For that, we ran a modified version of Step 1′s regression which consists in replacing the corresponding original variables on Equation (3) ($\Delta LogAgri_i$and $\Delta LogGDP, \Delta LogGDP^2{}_i$ in our example) with the corresponding residuals obtained from Step 2′s auxiliary regressions (*residualAgri*, *residualgdp*, *residualgdp*2). The total effect of variable $x1$ on forest cover area is the sum of two different effects: the direct effect ($\lambda 5$ on Step 1) and the indirect effects of this variable through the auxiliary variables ($\lambda'_1\gamma_1 + \lambda'_2\gamma_2 + \lambda'_4\gamma_3$). We obtained one modified regression for each variable of focus. Those modified regressions are linear combinations of Equation (3).

$$\begin{aligned}\Delta LogForest_{i,t} &\\ &= \lambda'_1 \Delta residual_{gdp_{i,t-1}} + \lambda'_2 \Delta residual_{gdp2_{i,t-1}} \\ &+ \lambda'_3 \Delta LogRuralPopDensity_{i,t-1} \\ &+ \lambda'_4 \Delta residual_{agri_{i,t-1}} \\ &+ (\lambda'_5 + \lambda'_1\gamma_1 + \lambda'_2\gamma_2 \\ &+ \lambda'_4\gamma_3) \Delta NetExportsAgri_{i,t-1} \\ &+ \lambda'_6 \Delta NetExportForest_{i,t-1} + \lambda'_7 \Delta TFP_{i,t-1} \\ &+ \lambda'_8 \Delta LogYield_{i,t-1} + \lambda'_9 \Delta LogAgri(PIN)_{i,t-1} \\ &+ \lambda'_{10} \Delta TradeHigh_{i,t-1} + \lambda'_{11} \Delta Goverment_{i,t-1} \\ &+ \lambda_{12} \Delta AgriEmployment + \lambda'_{13} Temperature(avg) \\ &+ \lambda'_{14} Rainfall(avg) + \lambda'_{15} \mu_{i,t-1} + u_{i,t}\end{aligned} \quad (7)$$

• Step 4. Total vs. direct effect.

Eventually, we compared the direct effect (λ_5) and total effect $\left(\lambda'_5 + \lambda'_1\gamma_1 + \lambda'_2\gamma_2 + \lambda'_4\gamma_3\right)$ of the independent variable of interest— net exports of agricultural products in this example— and discussed the relative differences.

3.4. Results

3.4.1. Panel Unit Root Test and Cointegration Test Results

The stationarity tests of the variables forest cover, GDP, GDP^2, agricultural area, and rural population density for income and forest transition clusters, respectively (Tables 3.2 and 3.3), showed that the variables are non-stationary of order 1, I(1), for GDP, GDP^2, and agricultural area. Concerning rural population density, the results suggest an I(0) variable but the visual inspection of the data suggests that it can be considered as a non-stationary variable (Appendix F, Figure A12, Figure A13, Figure A14). Furthermore, other panel data studies which used population as control variable in their analysis conclude I(1) as a result (Al Mamun et al., 2014; Atasoy, 2017). For the sake of simplicity, only one test (inverse Chi-squared) was reported (additional results are available upon request). Assuming that all our variables are I(1), we tested the presence for panel cointegration through the Pedroni test (Pedroni, 1999). The results indicated the presence of cointegration among the considered variables for income and for forest transition clusters, respectively (Appendix G).

3.4.2. Results by Income Groups

In line with the results from the cointegration test, we found a long-run equilibrium relationship between the variables forest cover, GDP, GDP^2, agricultural area, and rural population density (Tables 3.4 and 3.5). This is corroborated by the results on the variable 'residual', which corresponds to the ECT, and which is always significant and between 0 and −1. This result indicates that the variables from the long-run equations evolve together along a dynamic path that has a long-run equilibrium. We observed differences between income groups on the speed of adjustment towards the level of forest at the long- run equilibrium relationship. The ECTs are −0.326 and −0.369 for low-income countries; −0.119 and −0.120 for middle-income countries; −0.084 and −0.109 for high-income countries with and without agricultural area for each income group, respectively (Table 3.5). This indicates that for low-income countries, the speed of adjustment towards the level of forest that is in a long-run equilibrium with any level of GDP per capita, agricultural area, and rural population density is much faster than for middle- and high-income countries.

Table 3.4. Long-run equations, income clusters.

	(1)	(2)	(3)	(4)	(5)	(6)
	Low	Low	Middle	Middle	High	High
GDP cap (log)	0.168	-0.357	-0.071	-0.030	-0.406	0.072
GDP2 cap (log)	-0.013	0.029	0.005	0.002	0.024	-0.004
Rural pop (log)	-0.025	-0.108	0.020	-0.031	-0.073	-0.072
Agr Land (log)	-1.357		-0.258		-0.515	
Const.	-3.461	-0.352	-1.081	-0.937	-0.014	-1.474
Obs.	470	470	1455	1455	703	703
R-squared	0.489	0.042	0.187	0.027	0.360	0.126
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Table Note: The estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. Therefore, we interpreted those results only for our sample, which it is the reason why standard errors are not reported.

Over the long run and only for our sample (we do not perform statistical inference over the long run), the results including agricultural area (columns (1), (3), (5) of Table 3.4) showed a negative effect of GDP and a positive effect of GDP^2 on forest for high-income countries. Additionally, a negative effect of agricultural area on forest cover for all income groups, with important differences on the coefficients' size: −1.357 for low-income countries, −0.258 for middle-income countries and −0.515 for high-income countries. The graphical display of these results for the variables forest and GDP (Figure 3.2a) shows a somewhat flat parabola between forest cover and GDP per capita for high-income countries, and almost a flat line for low- and middle-income countries. When not including the agricultural cover area in the long-run model (columns (2), (4), (6) of Table 3.4, and Figure 3.2b), results changed slightly. Over the long run, for low-income countries, the GDP had a negative effect on forest, while GDP^2 a positive effect. Instead, for high-income countries, GDP showed a positive effect on forest and GDP^2 had a negative one. For middle-income countries, rural population density became positive. In addition, in the absence of agricultural area, the coefficients' sizes of the other variables are affected, with larger coefficients of all variables for low-income countries and slightly smaller coefficients of GDP and GDP^2 variables for middle- and high-income countries.

Table 3.5. Short-run equations, income clusters.

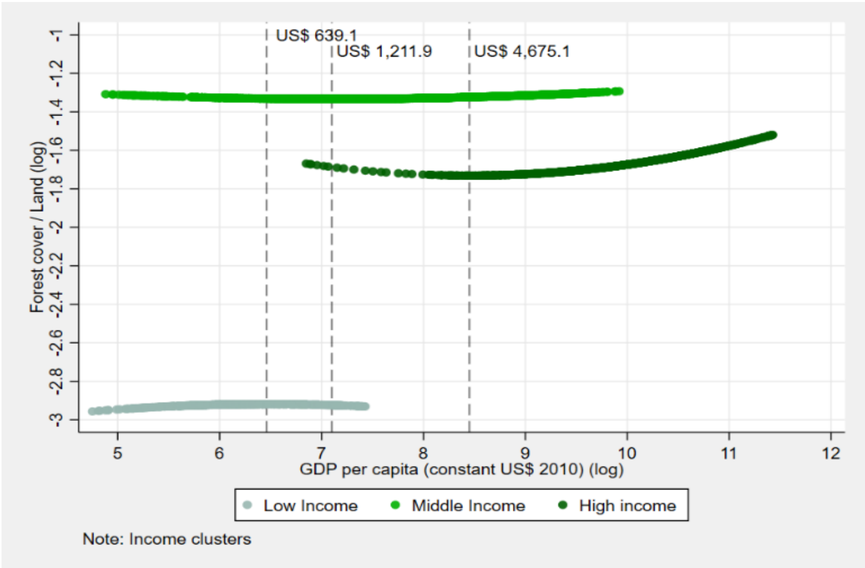
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Low	Low	Middle	Middle	High	High
LD.GDP cap (log)	0.097	0.244*	-0.009	0.013	0.004	0.043
	(0.145)	(0.126)	(0.039)	(0.039)	(0.106)	(0.101)
LD.GDP 2 cap (log)	-0.008	-0.020*	0.001	-0.001	0.000	-0.001
	(0.012)	(0.011)	(0.002)	(0.002)	(0.005)	(0.005)
LD.Rural pop (log)	0.043	0.181	-0.040	-0.048*	0.007	0.001
	(0.174)	(0.149)	(0.030)	(0.029)	(0.027)	(0.025)

LD.Agr Land (log)	-0.45***		-0.013		-0.111**	
	(0.104)		(0.013)		(0.046)	
LD.Agr empl	0.000	-0.001	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
LD.Agr prod netex	0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LD.For prod netex	0.000	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LD.TFP	-0.170	3.296	-0.103	-0.494	0.143	0.032
	(2.713)	(2.350)	(0.481)	(0.474)	(0.692)	(0.661)
LD.Agr (PIN) (log)	-0.014	-0.004	0.006*	0.006*	0.002	0.001
	(0.016)	(0.014)	(0.003)	(0.003)	(0.006)	(0.006)
LD.Yield (log)	-0.006	-0.015	-0.001	-0.001	0.005	0.005
	(0.012)	(0.011)	(0.002)	(0.002)	(0.003)	(0.003)
LD.Trade high	-0.000	-0.001	0.000*	0.000**	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LD.Government	-0.018	-0.013	-0.002	-0.005	-0.005	-0.008
	(0.019)	(0.016)	(0.003)	(0.003)	(0.008)	(0.008)
LD.Temperature	-0.003	-0.002	-0.001*	-0.001*	-0.000	-0.000
	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
LD.Rainfall	0.000	0.000**	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

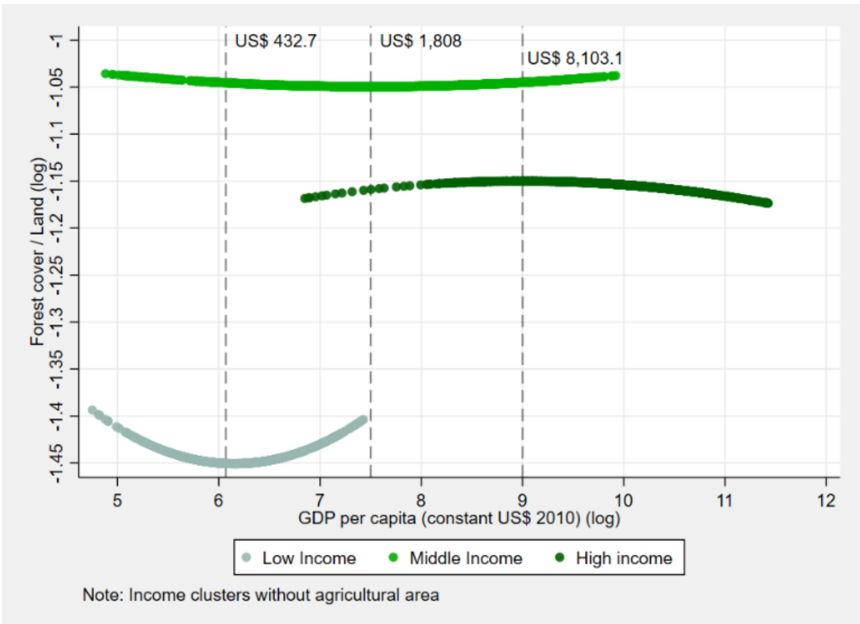
ECT	-0.326***	-0.369***	-0.122***	-0.120***	-0.083***	-0.109***
	(0.036)	(0.024)	(0.010)	(0.009)	(0.016)	(0.013)
DL(Gov/L.For)	-0.006	-0.004	0.001*	-0.000	-0.002	-0.003
	(0.009)	(0.008)	(0.000)	(0.000)	(0.005)	(0.005)
Const.	0.034***	0.013***	-0.003***	-0.003***	-0.006***	-0.007***
	(0.006)	(0.003)	(0.000)	(0.000)	(0.001)	(0.001)
Obs.	244	244	828	828	390	390
R-squared	0.406	0.558	0.186	0.209	0.141	0.203

Table Note: Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

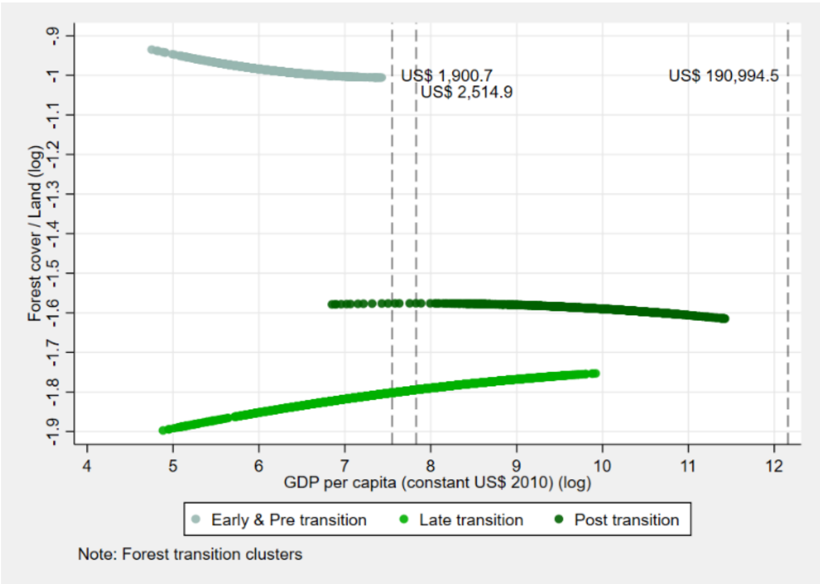
Over the short run, when agricultural area is included (columns (1), (3), (5) of Table 3.5), this variable showed a negative effect for low- and high-income groups while the agricultural producer price index had a positive and small impact for middle-income countries. Government quality on its own had no impact on forest for any of the three income groups. However, the interaction term of government quality and forest in previous period was significant and positive for middle-income countries. Thus, when forest in previous period becomes lower, improved government quality has a positive effect on forest. Finally, average temperature had a negative and small impact for middle-income countries. When removing agricultural area (columns (2), (4), (6) of Table 3.5), GDP and GDP^2 became significant for low-income countries, and rural population density became significant and negative for middle-income countries.


US$ 639.1
US$ 1,211.9
US$ 4,675.1
Forest cover / Land (log)
-1
-1.2
-1.4
-1.6
-1.8
-2
-2.2
-2.4
-2.6
-2.8
-3
5
6
7
8
9
10
11
12
GDP per capita (constant US$ 2010) (log)
Low Income
Middle Income
High income
Note: Income clusters

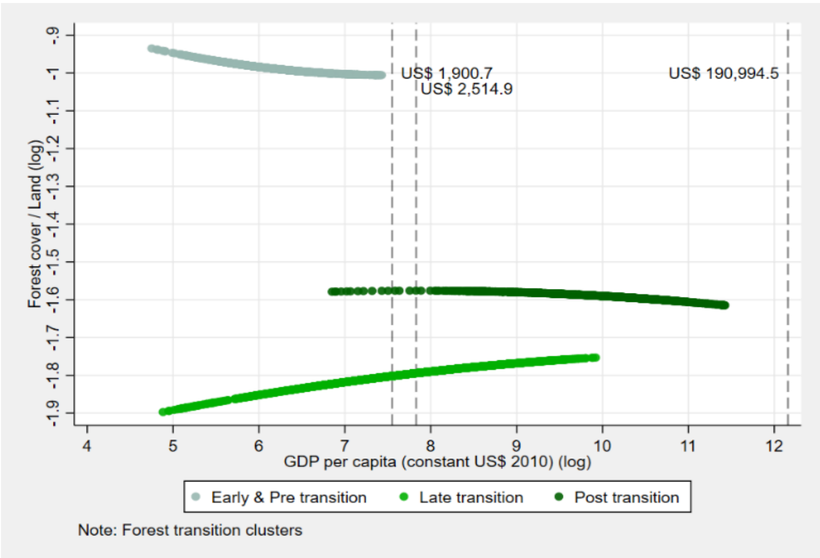
(a)


US$ 432.7
US$ 1,808
US$ 8,103.1
Forest cover / Land (log)
-1
-1.05
-1.1
-1.15
-1.2
-1.25
-1.3
-1.35
-1.4
-1.45
5
6
7
8
9
10
11
12
GDP per capita (constant US$ 2010) (log)
Low Income
Middle Income
High income
Note: Income clusters without agricultural area

(b)


US$ 1,900.7
US$ 2,514.9
US$ 190,994.5
Forest cover / Land (log)
-.9
-1
-1.1
-1.2
-1.3
-1.4
-1.5
-1.6
-1.7
-1.8
-1.9
4
5
6
7
8
9
10
11
12
GDP per capita (constant US$ 2010) (log)
Early & Pre transition
Late transition
Post transition
Note: Forest transition clusters

(c)


US$ 1,900.7
US$ 2,514.9
US$ 190,994.5
Forest cover / Land (log)
-.9
-1
-1.1
-1.2
-1.3
-1.4
-1.5
-1.6
-1.7
-1.8
-1.9
4
5
6
7
8
9
10
11
12
GDP per capita (constant US$ 2010) (log)
Early & Pre transition
Late transition
Post transition
Note: Forest transition clusters

(d)

Figure 3.2. Graphical representation of the long run results (four our sample). Note: Figure **(a)** shows the result for the long run equation for income groups in Table 4 columns 1, 3, and 5. The turning points are the following: US$ 639.1 for low income, US$ 1,211.9 for middle income, and US$ 4,675.1 for high income. Figure **(b)** shows the result for the long run equation in Table 4 columns 2, 4, and 6 (model without agricultural area). The turning points are the following: US$ 432.7 for low income, US$ 1,808 for middle income, and US$ 8,103.1 for high income. Figure **(c)** shows the results for the long run equation for the forest transition phases classification, in Table 6, columns 1, 3, and 5. The turning points are the following: US$ 1,900.7 for early and pre transition, US$ 190,994.5 for late transition, and US$ 2,514.9 for post transition. Figure **(d)** shows the result for the long run equation in Table 6, columns 2, 4, and 6 (model without agricultural area). The turning points are the following: US$ 1,408.1 for early and pre transition, US$ 2,724.4 for late transition, and US$ 1,669 for post transition.

3.4.3. Results by Forest Transition Groups

As previously, the variable "residual", corresponding to the ECT, is significant and between 0 and −1 for all the three groups, confirming the presence of a long-run relationship between the variables included in the long-run model (Tables 3.6 and 3.7). The differences in ECT between the different clusters are smaller than for income clusters, with the pre-transition countries having the largest ECT (−0253).

Table 3.6. Long-run equations, forest transition clusters.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Pre	Pre	Late	Late	Post	Post
GDP cap (log)	-0.136	-0.087	0.073	0.095	0.047	0.089

GDP2 cap (log)	0.009	0.006	-0.003	-0.006	-0.003	-0.006
Rural pop (log)	-0.019	-0.024	0.108	0.051	0.014	-0.015
Agr Land (log)	-0.244		-0.475		-0.472	
Const.	-0.492	-0.207	-2.182	-1.728	-1.760	-1.459
Obs.	599	599	493	493	1536	1536
R-squared	0.283	0.181	0.382	0.062	0.168	0.019
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Table Note: The estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. Therefore, we interpreted those results only for our sample, which it is the reason why standard errors are not reported.

Table 3.7. Short-run equations, forest transition clusters.

	(1)	(2)	(3)	(4)	(5)	(6)
	Pre	Pre	Late	Late	Post	Post
LD.GDP cap (log)	0.061	0.074	-0.070	-0.039	-0.120***	-0.131***
	(0.054)	(0.054)	(0.053)	(0.049)	(0.034)	(0.034)
LD.GDP 2 cap (log)	-0.003	-0.004	0.004	0.002	0.007***	0.008***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
LD.Rural pop (log)	-0.125	-0.102	0.018	0.016	0.009	-0.004
	(0.154)	(0.154)	(0.045)	(0.041)	(0.024)	(0.024)
LD.Agr Land (log)	-0.066**		-0.050		-0.083**	
	(0.030)		(0.035)		(0.041)	

LD.Agr empl	0.000	-0.000	-0.001*	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
LD.Agr prod net ex	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LD.For prod net ex	0.000	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LD.TFP	1.686	0.304	-0.225	-0.101	-0.055	-0.165
	(1.564)	(1.550)	(1.075)	(0.970)	(0.518)	(0.509)
LD.Agr (PIN) (log)	0.008	0.007	0.002	0.004	0.004	0.005
	(0.011)	(0.011)	(0.006)	(0.006)	(0.004)	(0.004)
LD.Yield (log)	-0.009	-0.010	-0.004	-0.004	0.001	0.000
	(0.009)	(0.009)	(0.004)	(0.004)	(0.002)	(0.002)
LD.Trade high	0.000	0.000	0.000	0.000	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LD.Government	-0.018*	-0.025**	0.003	0.003	-0.002	-0.004
	(0.010)	(0.011)	(0.009)	(0.008)	(0.004)	(0.004)
LD.Temperature	-0.003	-0.002	-0.002*	-0.002*	-0.000	-0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
LD.Rainfall	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ECT	-0.290***	-0.278***	-0.196***	-0.195***	-0.165***	-0.161***
	(0.027)	(0.025)	(0.021)	(0.015)	(0.009)	(0.009)
DL(Gov/L.For)	0.002*	-0.001	0.001	0.001	0.000	-0.001
	(0.001)	(0.001)	(0.006)	(0.005)	(0.001)	(0.001)
Const.	-0.007***	-0.010***	0.020***	0.022***	-0.006***	-0.007***

	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
Obs.	330	330	274	274	858	858
R-squared	0.321	0.318	0.310	0.429	0.302	0.323

Table Note: Standard errors are in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Over the long run, and only for our sample (we do not perform statistical inference over the long run), results when agricultural area is included (Table 3.6 and Figure 3.2c) showed that GDP had a negative effect and GDP^2 a positive one for the pre-transition countries of our sample. For the cases of late- and post-transition countries, GDP showed a positive effect and GDP^2 a negative one. Rural population density showed a positive effect on forest cover for late- and post-forest transition countries, and a negative effect for pre-transition countries. Agricultural area showed a negative effect on forest cover for all forest transition clusters.

In the short run, when agricultural area is included in the model (columns (1), (3), (5) of Table 3.7), agricultural area had negative and significant effect on forest cover for pre- and post-forest transition countries. Average temperature had a small and negative effect on late forest transition countries. Government quality has a negative effect for pre-transition countries. However, the interaction term between forest and government quality had a positive effect on those countries. Agricultural employment had negative effect on late-transition countries.

When the agricultural area is not present (columns (2), (4), (6) of Table 3.7, and Figure 3.2d), results were substantially the same. However, the variable of agricultural employment in the short run became not significant for the group of late forest transition.

3.4.4. Results for All Countries Together

When we considered the whole sample of countries without subclusters, the ECT was still negative and significant, showing the presence of a long-run relationship among variables (Tables 3.8 and 3.9). Moreover, on the short run, agricultural area had a negative impact on forest, and the interaction term of governance and forest had a positive impact on forest. In the absence of agricultural area, governance on its own had a negative impact on forest as well as the variable average annual temperature.

Table 3.8. Long-run equations, all countries.

	(1)	(2)
	All countries	All countries
GDP cap (log)	-.007	.02
GDP2 cap (log)	.001	-.001
Rural pop (log)	.011	-.029
Agr Land (log)	-.35	
Const.	-1.475	-1.164
Observations	2628	2628
R-squared	.189	.017
Year dummies	Yes	Yes

Note: The estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. Therefore, we interpreted those results only for our sample, which it is the reason why standard errors are not reported.

Table 3.9. Short run equations, all countries.

	(1)	**(2)**
	With Agri Land	**Without Agri Land**
LD.GDP cap (log)	-.006	.004
	(.021)	(.02)
LD.GDP2 cap (log)	.001	0
	(.001)	(.001)
LD.Rural pop (log)	.007	.006
	(.027)	(.027)
LD.Agr Land (log)	-.064***	
	(.016)	
LD.Agr empl	0	0
	(0)	(0)
LD.Agr prod net ex	0	0
	(0)	(0)
LD.For prod net ex	0	0
	(0)	(0)

LD.TFP	.309	-.333
	(.523)	(.512)
LD.Agr (PIN) (log)	.004	.004
	(.003)	(.003)
LD.Yield (log)	-.001	-.002
	(.002)	(.002)
LD.Trade high	0	0
	(0)	(0)
LD.Government	-.002	-.008**
	(.003)	(.003)
LD.Temperature	-.001	-.001*
	(0)	(0)
LD.Rainfall	0*	0
	(0)	(0)
ECT	-.205***	-.202***
	(.009)	(.008)
DL(Gov/L.For)	.002***	0
	(.001)	(.001)
Const.	-.002***	-.003***
	(0)	(0)
Observations	1462	1462
R-squared	.285	.31
Standard errors are in parentheses		
*** p<.01, ** p<.05, * p<.1		

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

3.4.5. Assessment of Indirect Effects on the Short Run

The results from disentangling direct and indirect effects are shown in Tables 3.10 and 3.11 for income and forest transition countries, respectively, with the inclusion of agricultural area. Concerning income clusters, we found very small changes between the direct and total effects of each short-run independent variables. The most significant change was observed for the variable GDP per capita, which was non-significant for high-income countries but became significant and positive after accounting for the indirect effects. In addition, the direct effect of the variable “yield” for high-income countries was not significant, but the total effect was positive and significant. For the forest transition clusters, there was one change on the post-transition countries: GDP^2 became non-significant when indirect effects were included. For the rest of the results, there were only small changes in the coefficient's size between the direct and total effects, with slightly larger significant coefficients for the total effects than for the direct ones (such in the case of the governance quality and its interaction term).

Table 3.10. Assessment of indirect effects, income clusters.

Effects	**Low**		**Middle**		**High**	
Variables	**Direct**	**Total**	**Direct**	**Total**	**Direct**	**Total**
GDP cap (log)	Not sig	Not sig	Not sig	Not sig	Not sig	0.000886**
GDP2 cap (log)	Not sig	Not sig	Not sig	Not sig	(+)0.00747***	(+)0.00747***
Rural pop (log)	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Government	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Government/L.Forest	Not sig	Not sig	0.000856*	0.000973**	Not sig	Not sig
Agr Land (log)	-0.453***	-0.453***	Not sig	Not sig	-0.111**	-0.0971**
Agr prod net ex	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
For prod net ex	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
TFP	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Yield (log)	Not sig	Not sig	Not sig	Not sig	Not sig	0.00551*
Agr (PIN) (log)	Not sig	Not sig	0.00593*	0.00603*	Not sig	Not sig
Trade high	Not sig	Not sig	0.0000829*	0.0000827*	Not sig	Not sig

Note: “Not sig” is the abbreviation for: not significant variable. For the cases where the variables were significant, we reported the results.

Table 3.11. Assessment of indirect effects, forest transition clusters.

Effects Variables	Pre		Late			
	Direct	**Total**	**Direct**	**Total**	**Direct**	**Total**
GDP cap (log)	Not sig	Not sig	Not sig	Not sig	-0.120***	-0.117***
GDP2 cap (log)	Not sig	Not sig	Not sig	Not sig	0.00725***	Not sig
Rural pop (log)	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Government	-0.0178*	-0.0241**	Not sig	Not sig	Not sig	Not sig
Government/L.Forest	0.00185*	0.00256***	Not sig	Not sig	Not sig	Not sig
Agr Land (log)	-0.0657**	-0.0634**	Not sig	Not sig	-0.0834**	-0.0858**
Agr prod net ex	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
For prod net ex	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
TFP	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Yield (log)	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Agr (PIN) (log)	Not sig	Not sig	Not sig	Not sig	Not sig	Not sig
Trade high	Not sig	Not sig	Not sig	Not sig	-0.0000686**	-0.0000686**

Note: “Not sig” is the abbreviation for: not significant variable. For the cases where the variables were significant, we reported the results.

3.5. Discussion

The results reveal the existence of a long-run equilibrium relationship between the variables forest cover area, GDP, GDP^2, agricultural area, and rural population density, for all the different country groups. Thus, they move together in a dynamic equilibrium relationship in the long run (Tables 3.4, 3.6 and 3.8). This is an important result since the long-run processes of forest are a key element for a good understanding of forest dynamics (Barbier et al., 2017). Short-run shocks propagate to these variables, but they tend to converge to these long-term equilibrium dynamic relations, as reflected by the error correction terms that correct a proportion of the deviation from the equilibrium each year. Low-income countries and those at pre-forest transition stages have faster adjustments of their forest cover area to the long-term dynamics. This suggests than land dynamics in these two groups may be quicker and would thus require a more adaptive design of policies and a more cautious monitoring of them.

Furthermore, our results confirm an inverse relationship between agricultural and forest cover area, which appears in all the country groups of our sample (including all of them together) over the long run. This inverse relationship is larger than a unit for our low-income country group and lower in magnitude for middle- and high-income countries. For low-income countries, every hectare of new additional agricultural land induces more than one hectare of deforestation. In contrast, agricultural area expansion in middle- and high-income countries may to some extent occur on already cleared but not previously used land or on other types of land. Low-income countries may lack such land resources or be unable to access them because of technical or biophysical constraints. According to the forest transition classification, we also found an inverse relationship between agriculture and forest cover. For late- and post-transition countries, the negative coefficient is larger than for pre-transition countries. This could be due to the fact that pre-transition countries have higher forest area where the agricultural expansion could happen, while the late- and post-transition countries expand not only in forest area but also in other types of land. As for short-run dynamics, we show that for low- and high-income countries, and pre-transition countries, the main direct factor influencing forest cover was changes in agricultural area. In contrast, short-run forest dynamics in middle-income countries seem more complex, with a wider range of variables (agricultural producer price, government quality and annual average temperature) having small effects on forest cover area. The middle-income countries' group gather a heterogeneous sample of countries; some studies with a smaller sample of middle-income countries have also identified a diversity of social and political variables that are important factors for land dynamics in those countries (Ceddia et al., 2014c; Imai et al., 2018).

Our results also brought important conclusions about our hypotheses described in Table 3.1:

Hypothesis 1, the EKC, was validated when agricultural cover land is present for high-income countries and for pre-transition countries in the long run (Tables 3.4 and 3.6), and for the post-transition countries in the short run (Table 3.7). This suggests that current low- and middle-income countries are experiencing different trajectories of relationships between economic development and forest cover compared to high-income countries.

Hypothesis 2, on the negative effect of agricultural employment on forest, was only validated for late forest transition countries in the short run (Table 3.7).

Hypothesis 3, on the positive effect of agricultural intensification on forest, was only validated when we take into account the indirect effects in our analysis, and only for high-income countries (see "yield" variable on Table 3.11). Thus, for high-income countries, agricultural yield had a positive impact on forest when we take into account the indirect effects of other variables on forest through agricultural yields. We also tested for the joint effect of economic growth and agricultural intensification on forest through an interaction term of the two, but we did not find any evidence of those synergies.

We explored the total (direct plus indirect) effects of other variables that are indirectly affecting forest through agricultural area. The results were particularly relevant for low- income countries on the short run, where economic development affects forest indirectly through the agricultural area (as shown when comparing columns (1) and (2) of Table 3.5). Henceforth, development policies should take this into account: any intervention on those countries' development will also cause indirect impacts on forest through the agricultural area, and both impacts could have opposite or similar effects on forest. Instead, the long run EKC in high-income countries disappeared when we included the indirect effects of development on forest through agricultural area (Table 3.10). This suggests that the direct and indirect effects of economic development on forest have opposite directions for high- income countries and this results in an insignificant net effect on forest cover for those countries. Lastly, the removal of agricultural area from our model revealed the negative effect of rural population density on forest for middle-income countries in the short run (as shown when comparing columns (3) and (4) of Table 3.5). Thus, rural population density affects forest area through agriculture in middle-income countries. We found similar effect in late forest transition countries for which agricultural employment affects forest indirectly through agriculture (as shown when comparing columns (3) and (4) of Table 3.7). Both results reveal two important mechanisms by which agriculture affects forest in those countries: rural population pressure and agricultural employment.

• *Hypothesis 4,* on the positive effect of government quality of forest when forest cover in previous period is low, was validated for middle-income, pre-transition countries and also for all countries together (as shown by the negative effect of the interaction term on Table 3.5, Table 3.7, and Table 3.9). In those countries the feedback mechanism of governance compensates for previous periods of deforestation. This suggests that the quality of governance plays a crucial role when countries face scarcity of forest. This goes along with other works that have also highlighted the role of governance on deforestation (Barbier & Tesfaw, 2015; Imai et al., 2018; Wolfersberger et al., 2015). On the same line, *Hypothesis 5*, about the existence of a feedback response from forest scarcity or forest abundance that adjusts forest over the long run to the dynamic equilibrium relationship with the other variables, was validated for all country's groups (see "ECT" on Table 3.5, Table 3.7, and Table 3.9). Several main mechanisms have been suggested to operate this feedback, including perception of scarcity of forest services and prices of forest products (Ewers, 2006; Satake & Rudel, 2007).

Lastly, our results did not support *Hypothesis 6a* and *6b*, on the negative effect of national net exports of agricultural products and on the ambiguous effect of net exports of roundwood on forest; neither *Hypothesis 7*, on the negative impact on forest from the export of agricultural products to high-income countries (Tables 3.5 and 3.7). Although, we acknowledge the role that trade of agricultural and forest products has on forest dynamics at the national level as well as its effects on differences in forest quality across countries (Crespo Cuaresma et al., 2017; Kastner et al., 2011; Leblois et al., 2017; P. Meyfroidt et al., 2010). Furthermore, recent studies have shown the indirect effects of trade coming from countries experiencing a forest transition (Pendrill et al., 2019), the reduction in trade's potential to decrease humans' impact on land ecosystems (Roux et al., 2021) and the deforestation-related emissions driven by international trade (Worldwide Governance Indicators, 2017). However, the aggregate nature of our data and the methods we used did not allow us to test for indirect land use effects from one country to another such the deforestation displacement.

We consider the crucial role of forest in mitigating climate change and the contribution of international mechanism such as REDD+ policies (reducing emissions from deforestation and forest degradation) to it. Our results show a negative impact of agricultural expansion on forest cover dynamics. Thus, REDD+ policies could be tailored in decreasing the opportunity cost of alternative agricultural activities such as agroforestry projects which would indirectly reduce the expansion of agricultural areas (Angelsen, 2010). This goes in line with the smallholder, tree-based land use intensification pathway suggested by Meyfroidt and Lambin (2011). Moreover, the coefficients of the ECTs showed how land dynamics are faster in low-income economies. The amount of REDD+ policies should be enlarged on such countries and specifically tailored for their local features. Pro-active governance in these countries could prevent rapid forest losses. Our results support the conclusion that governance institutions can play a strong role for controlling forest cover changes.

Through this work, we have also contributed to overcome some of the current literature's limitations such the lack of a more systematic focus on causal analysis, a dis- entangling of long- and short-run dynamics, and a closer examination across different geographical and temporal scales (Oldekop, 2020). Our contribution resides on the investigation of the long- and short-run cause relationships of forest dynamics through the use of an ECM and a cross-country panel dataset. However, further analyses could extend our work by (i) an improvement of the dataset used, not only through the increase in the time frame but also improvements on each variable's measurement, especially the ones related to governance quality and trade; (ii) the use of alternative methods to explore indirect effects in a cointegration structural model. For assessing mediation and indirect effects, we could have used more sophisticated approaches building on recent advances in structural equation models for panel datasets; however, structural equation models that account for cointegration relations are less well developed (Davidson, 1998; Hsiao, 1997). In addition, we acknowledge the existence of other theories and forest transition paths that we could not test with our data. An example is the "smallholder, tree-based land use intensification" forest transition path which requires more refined measures of tree cover including outside forests such as agroforestry and sylvopastures, small woodlots and orchards. Some studies have made an attempt to measure what they consider "agroforestry" as a proxy for this path, but these measures remain imperfect (Zomer et al., 2016). Another example is the citizen and consumers' attitudes towards environmental resources that could underlie the EKC and economic development path. We also recognize the relevance of issues such the effects of urban populations on forest dynamics (DeFries et al., 2010), the ecological impacts of a forest transition (Gingrich et al., 2019) or the deforestation's displacement (Pendrill et al., 2019). These issues were out of the scope of this work, but they represent important issues which require further investigation.

3.6. Conclusions

Forest land dynamics are changing at a local and global level and the implication of those changes on societies make those dynamics an important issue. Here, we assessed some of the most prominent theories about large-scale, structural forest cover dynamics— the environmental Kuznets curve (EKC), the forest transition and the ecologically unequal exchange theories—using a methodological design that addresses key statistical and conceptual issues from previous approaches. By focusing on causal analysis and different country groups, the empirical findings provided evidence of a long-run relationship be- tween forest cover, agricultural area, economic development and rural population for all our country groups, clustered according to both income levels and forest transition phases. Thus, short-run shocks are only one aspect explaining forest cover change dynamics, the other aspect, long-run dynamic equilibrium relations, has often been ignored but it represents an essential part of the forest dynamics' understanding. Furthermore, our results confirm an inverse relationship between agricultural and forest cover area, which is larger than a unit for our low-income country group but lower in magnitude for middle- and high-income countries. This indicates that agricultural area expansion in middle- and high- income countries may to some extent occur on already cleared but not previously used land or on other, non-forested types of land, while low-income countries may lack such land resources or be unable to access them because of technical or biophysical constraints. Yet, agricultural expansion on non-forested lands, such as savannas or wetlands, also induces strong environmental impacts.

A global forest transition has been hypothesized to be more challenging than achieving a local or regional forest transitions (Pendrill et al., 2019; Pfaff & Walker, 2010). In this work, we provided empirical evidence of two forest transition mechanisms, the one pointed by the *forest scarcity* and by the *state of forest policy* transition pathways. These pathways represent a feedback response from civil society and/or policy makers to a perceived scarcity of forest and they contribute to preserve the remaining forest and revert the decreasing trend. Thus, our results suggested that there is an important role for governance to play, especially when remaining forest areas are relatively low. In line with Liu et al. (2017), our results showed that the economic development pathway was also a relevant one, but with very heterogenous results across contexts. We also found an EKC for high-income countries and post-forest transition countries. This suggests that current low- and middle-income economies are experiencing different trajectories of relationships between economic development and forest cover than the high-income countries did. These heterogenous results show the complexity and context dependency of the causal relationship between GDP and deforestation. Further studies could build on our approach to go beyond aggregate measures of economic output like GDP, to explore the relations between inequality and poverty, and long and short-term forest cover dynamics.

Lastly, we tested the role of trade through the assessment of the globalization pathway and the ecologically unequal exchange theory. Despite the importance of trade in other studies explaining forest dynamics, our results did not find evidence of those theories. In addition, we decided to include the variables' indirect effects and to analyze the changes respect to the previous results. We showed that economic development affects forest area indirectly through agricultural area in low-income countries. In those countries, we observe an EKC on the long run due to those indirect effects. On the contrary, the EKC in high-income countries disappears when we included the indirect effects of agricultural area. Therefore, it is crucial to disentangle both direct and indirect effects on studies focused on land use dynamics, especially those intended to test theories related to forest transition and EKCs.

These insights contributed to improve the current literature on forest dynamics, but they can also be a useful tool to enhance land use policies. Our work explored the complexity of forest dynamics at an aggregate level; further studies could build on this framework together with more specific national or micro-level studies that could capture more nuanced socio-economic conditions to inform policy makers.

4. REGIME SHIFTS IN LAND USE FRONTIERS: AGGLOMERATION ECONOMIES AND FARMERS' EXPECTATIONS

This Chapter is in preparation for publication as:

Rodríguez García V., Abeygunawardane D., Baumann M., Carlson K., Da Silva R., Garret R., Ingalls M., Junquera V., Kronenburg García A., le Polain Y., Meyfroidt P., (n.d.). Regime shifts in land use frontiers: agglomeration economies and farmer's expectations. In preparation for *Ecology and Society*.

4.1 Introduction

Expansion of agriculture is the main direct cause of tropical deforestation and an important driver of land changes globally (Eigenbrod et al., 2020; Foley et al., 2011; Hannah et al., 2020; Lambin & Meyfroidt, 2011b; Pacheco et al. 2021). Deforestation is increasingly driven by commodity crops (e.g., coffee, soybeans, palm oil) destined for global and domestic urban markets (DeFries et al., 2010) and large-scale commercial agriculture is a major direct cause of forest clearing in the tropics. Commodity croplands could expand not only into forest areas, but they may also replace a variety of land uses and cover including existing agricultural lands important for smallholder subsistence or non-forest lands such as fallows, abandoned agriculture or low-intensity grazing lands (Meyfroidt et al., 2014). Here we focus on the expansion of commodity crops in agricultural frontier areas, defined as places with abundant land available for agriculture and a relative lack of capital and/or labor to exploit it, and rapid land use changes (Eigenbrod et al., 2020; Meyfroidt et al., 2018).

Frontiers contain areas of natural vegetation but they are also often inhabited by people who use the land for subsistence purposes and local markets. Expansion of commodity crops can occur as these local actors switch to new land use, but oftentimes they are displaced to more marginal lands or lose access to important resources by more powerful foreign actors or local elites (Kronenburg García et al., 2021) investing in large scale and commercial activities. Most of the large-scale agricultural expansion in recent decades has taken place at a very rapid rate (Müller et al., 2014; Junquera and Gret-Regamey, 2019), a characteristic that has led those changes to be described as "crop booms" (Fox and Castella, 2013; Junquera and Gret-Regamey, 2019). However, commodity crops can contribute to the livelihoods of rural people. Thus, commodity crops are a source of environmental and societal problems, but they can also generate income for poor households and play a role in food security. The contribution of cash crops to food security varies largely. Large-scale commercial agriculture contribution to local food security and pro-poor economic development is often limited (Meyfroidt et al., 2018) but positive spillovers from their coexistence with smallholder farming have been identified in some cases (Dedehouanou et al., 2013, 2013; Deininger & Xia, 2016; Herrmann, 2017; Van den Broeck et al., 2017). At the same time, agricultural expansion in frontier areas continues to be a central strategy to increase production for rural poor households (Barbier, 2012) who, below certain farm sizes, show diminishing returns for land and capital inputs making it more difficult for small farmers to be food self-sufficient (Jayne et al., 2014, 2016). As exposed here, frontiers are places of social interaction between actors of unequal power and this may lead to socio-economic conflicts in those areas (Kronenburg García et al., 2021; Pacheco, 2012).

Balancing these tradeoffs between cropland expansion, the preservation of biodiversity and ecosystem services, and the safeguard of local inhabitants' livelihoods and worldwide demands is a major global challenge that requires an understanding of the frontier dynamics. In this paper, we take the perspective of dynamic complex systems to understand what influences land use trajectories in frontier regions. We specifically aim to identify the role of two key factors, namely i) the presence of agglomeration economies - framed in this study as a case where frontiers are driven by initial conditions - and ii) the land users' projections and expectations about the future - when added to the previous mechanism. The presence of agglomeration economies has been identified as an important mechanism of frontier development and particularly for explaining non-linear trajectories and crop booms (Garrett et al., 2013; le Polain de Waroux et al., 2018). Agglomeration economies refer to positive externalities between similar firms that are closely located. Applying this to our context, when farmers expand the production of a given commodity crop (either by adopting this crop when they were already present or by moving to the frontier) and are located together, creating a cluster or agglomeration, they enjoy positive externalities in the form of transfers of knowledge, technological spillovers, a pooled market for skilled labor, large number and diversity of inputs suppliers, better access to credit, political power and other advantages. Together, these externalities contribute to increased profits from these agricultural activities. Possibly, the same results can be obtained by internal economies of scale when large-scale farms expand their production area. The presence of agglomeration economies has been particularly documented for frontier areas of the Brazilian Amazon (Garrett et al., 2013; Igliori, n.d.; Richards, 2018b). In addition to agglomeration economies, previous studies showed that land use dynamics in frontier regions are strongly shaped by the perceptions, interpretations and evaluations that land decision makers and actors involved in land changes have of the social and environmental context in which they operate (Groeneveld et al., 2017; Junquera & Grêt-Regamey, 2019; Mastrangelo et al., 2014; Meyfroidt, 2012, 2013; Schill et al., 2019). The importance of accounting for the complexity of human behavior to study the human-environment interactions is increasingly recognized but its integration into formal models is still a challenge (Schlüter et al., 2017). The role of social learning and social dynamics to adapt to a changing environment is well acknowledged, but the individual features that explain them are less taken into account (Meyfroidt, 2012). Here we focus on one of the core features of human agency, the ability to anticipate future events and to act according to expectations of this future (Bandura, 2000; Beckert, 2013; Patrick Meyfroidt, 2013). Expectations about the future have been shown to influence land use decisions in various ways, such as in the case of expectations for high profits driving investment at the Chaco frontier area (le Polain de Waroux et al., 2018), the expectations about environmental degradation and recovery in Vietnam (Meyfroidt, 2013), the farmers' and businesspeople's perceptions of land changes in Ethiopia (Ariti et al., 2015), the imitation behaviors in Northern Laos (Junquera & Grêt-Regamey, 2019) or the impact of commercial ranchers' expectations about future rainfalls in Southern Namibia (Domptail and Nuppenau, 2010; Guido et al., 2020). However, these works remain only a small representation in the land chance science and they are still rare in frontier regions. Here we argue that the inclusion of core features of human agency, especially the ability to anticipate future events and act according to those expectations, may be needed to explain how agent learn and adapt and how nonlinear land use change occurs (Mastrangelo et al., 2014; Patrick Meyfroidt, 2013; St John et al., 2010).

All of this causes frontier areas to be characterized by their complexity, which results in large uncertainties in terms of dynamics and trajectories (Bauch et al., 2016; Biggs et al., 2018; Jinlong Liu et al., 2017). Key knowledge gaps remain regarding the underlying properties of the dynamic models (Bauch et al., 2016) and for increasing the extent of generalization from case studies (Magliocca et al., 2018). In this work, we use a theoretical model from economic geography (Krugman, 1991) translated into the context of land use dynamics. We use this model to explore theoretical properties of frontier dynamics as regime shifts, and to compare different frontier areas by formalizing the role of agglomeration economies and people's expectations about future market conditions in these frontier dynamics. The model represents the frontier area as an unstable equilibrium between two land-system regimes being: i) a pre-frontier situation dominated by semi-subsistence farming and ii) a consolidated frontier dominated by commercial agriculture. Our goal is to understand the conditions under which land users will perceive one or the other land regimes as being optimal and will shift their land use accordingly, thereby driving the frontier's dynamics to one of the two land regimes. We explore analytically the behavior of the model, and we calibrate empirically with a set of five contemporary tropical deforestation and land use expansion frontiers cases. We aimed to answer the following questions: i) How do agglomeration economies and expectations shape a frontier's trajectory? ii) What are the conditions under which these factors lead to the dominance of semi-subsistence versus commodity land systems?

The rest of the paper is organized as follows: Section 2 presents a description of the five frontier case studies selected for the calibration of the model. Section 3 explains the data collection and methods used. Section 4 presents the results obtained and is followed by Section 5 where the aforementioned results are discussed. Section 6 makes some conclusions.

4.2 Frontier case studies

Most works on frontier regions have focused on local scale studies in already established and active frontiers, mainly in the humid tropics (Buchadas et al, in prep; Schröder et al., 2021). Here, we explore a range of agricultural frontier areas in different situations[4]. We selected five cases of tropical agricultural frontier regions, described in this section (Figure 4.1).

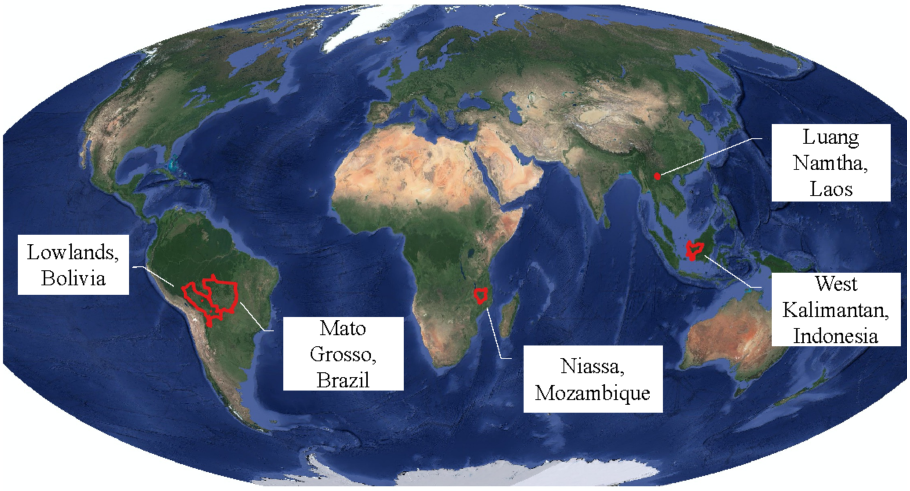


Figure 4.1: Frontier case studies

4.2.1 Niassa province in Mozambique

Niassa, in the northern part of Mozambique, is the largest province of the country with a total area of 12.9 Mha (GPN, 2017). The land use trajectory of commercial farmers in Niassa province has been described as a succession of waves and their legacies by (Kronenburg García et al., 2021). One of those waves started in 1992 with the arrival of missionary farmers and it was followed by another one of white South African commercial farmers. In 2005, large scale forestry companies invested in Niassa but did not succeed. Since 2012, a new wave of agricultural companies has been investing in the area, mostly initiated

[4] We consider three different situations here, i.e. the initial situation of "pre-frontier" or emerging frontier where the land regime is dominated by semi-subsistence agriculture, the situation in which rapid changes are happening (active frontier) and the situation in which the land regime is dominated by commercial or commodity agriculture (post-frontier or closed frontier) (Alban et al 2019).

by ex-forestry employees or service-providers. However, their profits are low and investing in Niassa remains an ongoing struggle with more failures than successes. Thus, we can consider Niassa province as a potential or potentially emerging frontier, with uncertain future trajectories (Kronenburg García et al., 2021).

Currently, only 0,29% of the province's total area is dedicated to commercial agriculture (equivalent to an area of ~37.000 ha) (DCPI, 2015). The number of actors from the post-1992 waves remains very small with only three missionary farmers, one original forestry company and a few more than seven new-wave investments (Kronenburg García et al., 2021). These farmers complain about the difficulties of making successful business. Some cotton and tobacco companies, with histories going back to the colonial period, are also active in Niassa and work through contract farming. A few agricultural companies in Niassa have made land-use decisions based on the expectation that a substantial market would develop in the neighboring province of Cabo Delgado, following large multinational investments and growing interest linked to the discovery of natural gas along the coast, but this market has not materialized yet (Kronenburg García and Meyfroidt forthcoming). Likewise, at least one forestry company had planted a tree species appropriate for an anticipated pulp industry that never developed. For this frontier case, due to the small area dedicated to commercial crops, we have not focused on one main cash crop, but instead we considered the total area dedicated to different cash crops.

4.2.2 Eastern lowlands of Santa Cruz (Bolivia)

The Department of Santa Cruz harbors 34% of the Bolivian land (37 Mha) and it is the center of the country’s major production zone contributing to 27% on average of Bolivia’s GDP (Mackey, 2011). This department contained 935000 ha dedicated to soybeans in 2014 from which, the large-scale landowners with 1000ha or more, represent only 2% of total farm units but own over 70% of the land (McKay, 2017).

Agro-industrial corporations, Santa Cruz farmers and foreign colonies (Menonites and Japanese) appeared in the lowlands area in the 1980s and they engaged in large-scale deforestation for agricultural production mainly in Santa Cruz department (Tejada et al. 2016). This process continued in the 1990s with the support of the Bolivian government and a World bank's project (Mackey, 2011). At the same time, foreign investments developed, particularly from Brazilian and Argentine producers (le Polain de Waroux, 2019; McKay & Colque, 2016). Altogether, it jumpstarted agricultural development in the region and soybean production and exports expanded dramatically. In 2006, the modernized and powerful elite of Santa Cruz agricultural producers (McKay & Colque, 2016; Redo et al., 2011) were able to use their influence and establish an alliance with the central state during the agrarian reform lead by Evo Morales (de la Vega-Leinert & Huber, 2019). Deforestation also increased, ~78% of which being happening on private land by large scale cattle ranching or mechanized agriculture, or the rest happening on communal land, by local ingenious communities and Andean peasant colonies with semi-mechanized commercial systems (de la Vega-Leinert & Huber, 2019).

Currently, there has been a continued process of expansion in the lowlands area (le Polain de Waroux et al., 2018). The large quantities of unprotected forest land at low prices, particularly in the Chiquitania, continues to be an attractive feature for large scale farmers' investments and Mennonite colonies have also been expanding significantly in the region (le Polain de Waroux, 2019).

4.2.3 Mato Grosso

Mato Grosso state is located in the southern part of the legal Amazon in Brazil. It covers an area of 90.3 Mha out of which 12.8% was dedicated to commercial crops in 2019 (MapBiomas), with 9,724,213ha (10.6%) dedicated only to soybeans (IBGE/PAM). Land use change in this state, the third largest in Brazil, is dominated by an active pioneer frontier (Arvor et al., 2018; Gil et al., 2015), turning Mato Grosso into the largest producer state of soybeans in the world since 2000 (Simoes et al., 2020).

The settlement of Mato Grosso by non-indigenous populations began in the southern regions of the state in the seventeen century. These first settlers were ranchers, they also mined gold and traded with slaves (Arvor et al 2018). Afterwards, in the 1930s, the spatial occupation of

Mato Grosso really began, and cattle ranching dominated the economy during the 1950s and 1960s. During the 1970s, the federal government intensified its occupation of the Amazon and the state and federal governments stablish agrarian reform policies (Garrett et al. 2021). As a result, the rural economic and social landscape in Mato Grosso changed between 1970 and 1990. Farmers interested in the agricultural potential of large plains came mostly from the southern regions of Brazil. During early years of settlement, rice was used in the initial stages of conversion and later as a rotation crop with artificial pasture and annual crops such soybean and maize (Garrett et al. 2013; Jepson, 2006). During the 19990s and 2000s, export agriculture and large-scale infrastructure projects happened in the Amazon and since then, vast areas of Amazon rainforest and Cerrado vegetation (Brazilian savanna) were converted, essentially to fulfill external demands for agricultural products. Between 2000 and 2005 the growth of soybean farming in Brazil was unprecedent with an increase of almost 5.0 Mha in Mato Grosso state (Amaral et al. 2021). Nevertheless, the decade 2004-2014 has been marked by a relative decoupling of deforestation and soybean production with deforestation rates decreasing 90% while crop production continued to increase (Arvor et al., 2018).

Soybean producers in Mato Grosso had an average of 1,200 ha of cultivated land in 2017, while overall, accounting for all kinds of rural farmers also including cattle ranchers, farms had an average of 1,687 ha (Bicudo Da Silva et al., 2020). At that time, up to 77% of Mato Grosso rural producers came from the South of Brazil (i.e., states of Rio Grande do Sul, Paraná, and Santa Catarina) (IMEA, 2017). Since 2017, deforestation trends have been increasing in the Amazon biome (Rajão et al., 2020), with around 18% of landowners in the State having expectations to expand their pastureland area in the next few years, and 30% planning to expand croplands – essentially (92%) over pasture and former croplands (IMEA, 2017). Among those producers planning to expand pasture or cropland in their farms, the increase of farm's profitability and future prices of agricultural products were the major causes for future expansion in the State (for 92% and 87% of the respondents respectively). Farmer's are currently facing a devaluation of the currency exchange rate against dollar which is one of the major drivers of production costs increases, which in turn, squeezes producers' profit margins (Da Silva et al. 2020). In addition, even with low interest rates, their large size and fragility of the national currency result in high investment risks.

2.4 West Kalimantan, Indonesia

West Kalimantan (14.6 Mha) is one of five provinces in Indonesian Borneo and has experienced substantial commodity crop-induced deforestation since the mid-2000s. In 2015, around 49% of Kalimantan was forested, and non-forested areas comprised small-scale agriculture including subsistence (e.g., rice, vegetables) and commercial (e.g., oil palm, rubber, black pepper) crops managed by local communities and large-scale plantations (e.g., oil palm, pulpwood) managed by corporations (Gaveau et al., 2019). The Indonesian government estimates that in 2019 West Kalimantan contained 1,86 Mha of oil palm (13% of the province) or about 13% of all Indonesian oil palm area, up from just 8% in 2011 (BPS Indonesia 2021). Although initial commercial oil palm development in West Kalimantan occurred in the 1980s (Li, 2015), oil palm expanded rapidly especially in the late 2000s (Gaveau et al., 2016), and deforestation attributed to oil palm expansion has varied across time and space (Meijaard et al., 2020).

Plantations including oil palm generated around 10-20% of total Kalimantan deforestation prior to 2000, and around 50% of total deforestation from 2005 to 2015 (Gaveau et al., 2016), yet plantation expansion has been a less important driver of forest loss in the region since circa 2012 (Gaveau et al., 2019). Zero-deforestation agreements by oil palm companies, efforts by the Indonesian government to protect forests and peatlands, fire control efforts, and lower oil palm prices in the 2000s may all contribute to this trend (Carlson et al., 2018; Heilmayr et al., 2020).

Temporal trends in oil palm expansion rates are strongly linked to palm oil market prices (Gaveau et al., 2019), and expansion has been concentrated in areas with the most favorable biophysical suitability and market infrastructure (Xin et al., 2021). Moreover, expansion typically occurs only within the vicinity of large-scale commercial mills that process fresh fruit to crude palm oil, because fruits need to be processed within 24-48 hours of harvest. Thus, market development depends on the location of mills, which are determined through land leases from the Indonesian government at the request of oil palm growing companies. These companies usually develop their own oil palm plantings near mills and often staff these plantings with laborers brought in from outside of the region, but may also source oil palm fruit from independent farmers who manage their own oil palm, or farmers that are contractually bound to sell their produce to the mill and who have a variable role in farm management. These small farmers may be long-time residents of local communities or more recent migrants. According to Indonesian law, these farmers should be paid mill-gate prices set by the government that reflect the quality of their fruit, but in practice they may receive prices that vary from this amount. Profits from oil palm vary widely with oil palm age, planting variety, time of year, antecedent weather, farm management practices, and harvesting practices, and small farmers often face higher relative costs due to economies of scale.

2.5 Luang Namtha Province, Laos

The Luang Namtha province is located in the northwest of Laos sharing a mountainous border line with China to the north and with Myanmar to the west. It has a total area of 932,500 ha and it is the province with the largest rubber area in Laos (Phasouysaingam, 2020). The mature rubber plantations in 2018 covered up to 77,120 ha (8,3% of total area)(Xiao, 2020).

In the 1990s, rubber was introduced by migrants from China into few border villages between Laos and China (Manivong & Cramb, 2007). Around 2004, the Lao government encouraged both local and foreigners (mostly Chinese) to widely establish rubber gardens and plantations (Lu & Schönweger, 2019). At the same time, the Chinese government launched the Opium Replacement Programme to support rubber investments in Laos (Lu & Schonweger, 2019) and rubber boomed in this area. Different factors triggered this, the high latex prices in the late 2000s, China becoming the world’s largest rubber consumer and the urgent need by Laos and China for a cash crop to replace opium among others (Cohen, 2009). The trend of rubber expansion slowed down significantly after 2012, in part because of

decreasing rubber prices, but rubber plantations continued to expand afterwards.
An important factor on the rubber investment decisions made by the Laos and Chinese's governments as well as the Chinese investment companies was their respective expectations about those investments. Dwyer & Vongvisouk (2019) documented this and explained how those expectations were not completely fulfilled. In the 2000s during the rubber boom, the media often showed the optimism from the government and private sector who believed that China's demand for rubber would be sustained indefinitely (Dwyer & Vongvisouk (2019). Chinese agribusiness investments boomed in Laos in the 2000s as a result of both, Laos and Chine, foreign relations and strategic economic planning of policies. On the one hand, the Lao government decided to orient its economic development strategy toward the attraction of Foreign Direct Investment in the land sector, also known as the "turning land into capital" slogan. On the other hand, and during the same time period, Chinese policies encourage companies to invest abroad, the "Go Out Policy" (Dwyer & Vongvisouk, 2019). In addition, the Chinese agribusiness investment in Laos have received special government support under the China's Opium Replacement Programme. These events acted as signals to Chinese investors for moving into Laos and they influenced the Chinese companies' expectations. Chinese investors had the expectation that the Lao government would facilitate their investments, they expected land in Laos to be easy to come by, and they also expect "greater enforceability of granted concession contracts" (Lu&Schonweger, 2019). This influenced the way companies managed the concession granting process, they assumed that the higher their political connections, the soother their projects would develop. However, certain forms of territorial control remain quite robust in Laos and the allocation process occurred at lower levels. In the end, Chinese companies received considerable areas of land but less than they expected and under specific issues that they do not expect to encounter (Lu&Schonweger, 2019).
Since mid-2011, the conditions have changed. The rubber price has crashed and there is evidence of indebtedness land sales and conversions to other crops different than rubber. Increasing land conflicts and food-security concerns have been reported. This situation caused the announce in 2004 of a ban on conversion of wet rice-

growing areas to cash-crops plantations, namely for bananas (Friis & Nielsen, 2016). All this is likely to alter investor's expectations in the area.

4.3 Methods and data

4.3.1 Model purpose and logic

We employed theory on complex systems and dynamic equations' systems to mathematically represent the frontier evolution, building on earlier studies using this approach (Bauch et al., 2016; Biggs et al., 2018; Lade et al., 2013; J. Liu et al., 2007; Milkoreit et al., 2018; D. Müller et al., 2014; Scheffer et al., 2009; Staal et al., 2020). We take each frontier area as the complex system of study. This system is represented as characterized by an unstable equilibrium between two stable states or land systems regimes. Land use dynamics evolve over time and the frontier area is expected to converge towards one or the other land regime; when a new land use regime with different economic and ecological characteristics is established, a regime shift or a critical transition[5] occurs (Biggs et al., 2018; De Alban et al., 2019; D. Müller et al., 2014; Ramankutty & Coomes, 2016).

More specifically, we adapted a model presented by Krugman (1991) to our context. We model the frontier area as an unstable equilibrium between two land systems regimes which represent: i) a pre-frontier situation dominated by semi-subsistence farming and ii) a consolidated[6] frontier dominated by commodity agriculture (also referred in the paper as the low and high equilibrium respectively). Semi-subsistence farming here comprises a range of land use activities operated mainly by smallholders and indigenous communities, using land for production of food, fuel and fiber for household and community-level consumption as well as for production destined to local markets or more distant ones, but in a limited way. The semi-subsistence land regime can also include

[5] Critical transition: it is a shift in a system's dynamical regime from its current state to a strongly contrasting state as external conditions move beyond a tipping point (Bauch et al. 2016).
[6] Also called "suspended" frontier or "closed" frontier.

more or less abundant remaining areas of natural vegetation, possibly being converted to smallholder agriculture over time. Land uses in this regime are assumed to present constant returns to scale and no agglomeration economies. In contrast, commodity agriculture here represents a more homogenized set of land uses dominated by one or a few crops almost entirely destined to markets, with an emphasis on distant (urban or international) markets. The set of actors operating commercial agriculture can be diverse (e.g., from smallholders and medium-scale farmers to large-scale operators). Production of the commodity crop is assumed to present agglomeration economies, so, profit per unit produced increases when the commodity crop covers a larger share of the frontier region. This model represents the aggregate of all decision makers in the frontier region, as a representative decision-making agent, assuming homogenous income optimizers. This assumption may not hold true for all actors, especially for smallholders and indigenous groups who often are risk minimizers. We calibrate the model for three types of representatives' agents: smallholders, medium scale farmers linked to domestic financial markets, and large-scale operators or actors funded through international capital markets. Other agents, such as public policy-makers, other private enterprises such as service providers etc, are represented indirectly in the model through the value of the key parameters.

The (representative) agents have the options of either produce a basket of crops considered essentially as semi-subsistence crops with constant returns to scale, or producing a cash crop with fixed costs and agglomeration economies. From its initial conditions, the frontier goes towards one or the other equilibrium, as reflected by the change in the variables that define the state of the system at each point. The system may end up in one of two stable equilibria with an uncertainty about the final destination of the system, depending on the conditions and perturbations. Our goal is to understand the conditions under which land users will perceive one or the other land regimes as being optimal, and will shift their land use accordingly until the final destination's trajectory. In particular, we explore the role of agglomeration economies and people's expectations about future market conditions to explain the frontier's dynamics between alternative land regimes. Although in reality both states can coexist at a given time, the model considers this as a transitory state, assuming that land uses in the frontier will converge to a stable equilibrium where the region is dominated by one or the other regimes.

4.3.2 Model formalization and assumptions

The model is represented by a system of dynamic equations, and it builds on assumption of profit maximization by farmers. At each point in time, the income of a representative farmer that could potentially grow a commodity crop (Y) is Eq 1.

Eq. 1:

$$Y = (L * 1) + S * \pi(S) - \left(\frac{1}{2\gamma}\right) * \left(\dot{S}\right)^2$$

With:

S: area cultivated under the cash crop/s of focus, relative to the total frontier area in each case.
L: the remaining area after subtracting S to the total frontier area. We assume that this area is covered by a mixture of semi-subsistence crops and natural vegetation[7] and we normalize the profit from these land uses to 1.
π: profit gain from adopting commodity crops compared to the profit from semi-subsistence crops
γ: inverse of the speed of farmer's adoption of commodity crops
$\dot{S}$ the area converted to commodity crops at each time

In addition, we made the following assumptions:

1. Two land uses: Farmers can only produce two types of goods: a basket of semi-subsistence crops, or commodity crops.

[7] We acknowledge that in reality, this area (total frontier area minus S) includes urban areas, protected areas and other land types not covered by subsistence/semi-subsistence crops.

2. Presence of agglomeration economies[8]:
   These agglomeration economies correspond to the fact that the farmers' profit gains from adopting cash crops (π) is higher when the area planted with cash crops at the frontier (S) is higher. We do not distinguish here between scale economies that are internal or external to the farm, i.e., these agglomeration or scale economies depend only on the area planted with the target crop, irrespective of whether this area is achieved by a large number of small producers or by a few large-scale farms.

$$\pi = \pi(S) \text{ and } \frac{\partial \pi}{\partial S} > 0$$

3. Presence of adjustment costs: These reflect costs to transition from the semi-subsistence crops to the target cash crops, per land unit, assumed to have a quadratic form (Eq.2) following Krugman (1991). The presence of these costs implies that the farmer's choice is an investment decision, in which the farmer is not only interested in current returns but in the future returns, which depend on the decisions of other farmers creating agglomeration economies.

Eq. 2:

$$Adjustmentcost = \left(\frac{1}{2\gamma}\right) * \left(\dot{S}\right)^2$$

3. Farmers' finance and interest rate: All farmers are able to borrow at a given interest rate (r). The profit maximization of the representative farmer implies that she will invest in cash crops until the moment when the cost of borrowing money for investing, or the opportunity cost of keeping her own money in the bank (r*q) would be equal to the benefit of such an

[8] We are applying this model an aggregate level but if could be also applied for only one farmer. In such a case, the agglomeration economies would be equivalent to economies of scale.

investment $(\pi + q^{\cdot})$[9]. Therefore, in equilibrium, we have the following arbitrage condition.

$$rq = \pi + q^{\cdot}$$

q being the discounted sum of all expected future farmers' profits and $q^{\cdot}$the change of q with time, $q^{\cdot} = \delta q/\delta t$.

1. Profit: We defined the farmer's profit[10] function including agglomeration economies by the following equation (Eq.3):

Eq. 3:

$$\pi = \frac{\mu\alpha L - F(1-\mu S)}{1 - S\mu\alpha}$$

μ being the productivity gain by adopting the cash crops; F the fixed cost of adoption; L the land that could be potentially used for growing cash crops.

4.3.2.1 Solving the model

Taking those assumptions into account, the farmer's objective is to maximize her present value profit (H)[11] (Eq. 4).

Max H

Eq. 4:

$$H = \int_0^{\infty} \left[\left(L + S * \pi(S)\right) e^{-rt} \frac{-1}{2\gamma} * \left(\dot{S}\right)^2 \right] \text{dt}$$

And the maximization results (see Annex 2) show that:

[9] Also called "arbitrage free" or "no-arbitrage" condition/assumption on the financial jargon.
[10] According to Matsuyama (1995).
[11] We assume that moving costs are paid at the time of planning. Thus, we do not discount them. See annex.

Eq. 5: $\dot{S} = \gamma q$ being $q(t) = \int_T^{\infty} \pi e^{-r(t-T)}$ dt

These results, together with the following arbitrage condition from the investment decision: $rq = \pi + q$ (*Eq. 6*); They allow us to set the dynamic version of the model, which consists of a dynamic system of two equations:

$$\begin{cases} \dot{S} = \gamma q \\ q^{\cdot} = rq - \pi \end{cases}$$

The solution to this system of equations requires the following second-degree equation being equal to zero (see Appendix).

$$\lambda^2 - r\lambda + \beta\gamma = 0 \quad \leftrightarrow \quad \lambda = \frac{r \pm \sqrt{(r)^2 - 4\beta\gamma}}{2}$$

The values of λ for which Equation xx is equal to zero could be real or complex. If $r^2 > 4\beta\gamma$ then, λ is a real number. If $r^2 < 4\beta\gamma$ then, λ is a complex number. We do not aim to calculate the precise value of λ, but mainly aim to identify if this is a real or complex number which indicates what determines frontier trajectories (see Section 4.3.4).

4.3.3 Key parameters: definition and quantification.

As pointed by our model, the solutions to the farmer's maximization problem and, in turn, the frontier dynamics, ultimately depend on three parameters: r, β and γ. As we analyze the model at the aggregate level, what matters are the aggregate value of these parameters, whether this reflects a set of homogenous farmers or an average or dominant trend. We consider three types of farmers or investors for each frontier depending on the way they finance themselves: The first type corresponds to very large investors who either do not need to finance their land investments or rely international sources of funding. The second type corresponds to the very small farmers (smallholders) which are either subsidized with special interest rates or privately financed, usually at very high interest rates. The third one corresponds to all categories in between, i.e. medium or large farmers having mainly access to national-level financial markets. In this section we define those parameters r, β and γ, and we explain how we collected the data to quantify them through: i) online interviews to experts on the selected frontier regions-for the agglomeration economies and adjustment cost parameter-, and ii) literature and online resource reviews for the interest rate parameter's quantification.

The interviews were carried out by the first author from October to December 2020. During the interviews, two questionnaires on agglomeration economies and adjustment cost respectively were used to quantify a range of values for those parameters. The interviews were focused on the frontier conditions in the contemporary period, from 2014 to 2019**.** The questionnaires contained 12 and 13 questions about the agglomeration economies and adjustment cost parameters, respectively (see Annex). We interviewed a total of nine experts, two of each per frontier case with the exception of Indonesia for which we only interviewed one expert, all of them being coauthors of this work. For the cases with two experts per frontier region, we made an average between the parameters' value obtained for each of them and we used this average number. At the end, we calculated one value of each parameter β and γ per frontier case. In addition, a few local but well-connected farmers were also contacted as well as some other researchers for enriching the understanding of the contexts and contributing to validation of the data. After we obtained the values of $\beta\gamma$ for each frontier case of study, we asked the experts to compare the relative size of those parameters for each study case to the other frontiers' cases' results.

- Strength of agglomeration economies (β)

This parameter measures the strength of agglomeration economies in the frontier region. It measures the increases in profitability of cash crop production due to an increase in the farm density, corresponding to the parameter explaining the $\pi(S)$ function. The value of this parameter corresponds to an abstract "strength" of these agglomeration economies that we quantified with a questionnaire. For the design of the questionnaire, we used as a benchmark the agglomeration economies' micro-foundations described by Duranton and Puga (2003). These are: i) The sharing of capital inputs; ii) Improvement on the matching between production requirements and types of land, labor or intermediate inputs and iii) The learning mechanisms through the interaction of economic agents. We have considered that this parameter is homogeneous across the different three types of actors already mentioned since some local public goods (i.e. roads, storage facilities, slaughterhouses, service providers etc.) are shared by different type of actors.

- Speed of adjustment[12] (γ)

This term relates changes in the area dedicated to cash crops with changes in expected profits (Eq.3 $\dot{S} = \gamma q$). The value of this parameter represents an abstract concept ("the speed of adjustment") and mathematically, it represents the inverse of the marginal cost of adjustment. The adjustment costs include all costs incurred by the farmer at the moment she decides to invest in the region to grow a cash crop or to turn a piece of land (considered here as being already used for the semi-subsistence crop) into cash crops. This term includes many different aspects of costs in terms of labour and various forms of capital – financial, but possibly also social or cultural - needed to expand cash crops area at the frontier (Singh, 2008). These different aspects include possible efforts to identify suitable land and secure tenure or use rights on it, clearing and preparing the soil, managing the creation of a new agricultural system; intermediary services employed to transform the land, psychological costs from changing social relationships and culture, adaptation to a new language, and others. They also reflect risks taken by farmers, non-market costs from imperfect information, and geographical barriers. In this case, we have also considered that this parameter is homogeneous across the three different types of actors. We performed a sensitivity analysis by increasing and decreasing 50% of the adjustment cost parameter for each actors' group in each frontier. Result indicated that only for one type of actor in one frontier case, the frontier dynamic's outcome shifted. For all the other cases, results were

[12] In Krugman (1991) this parameter is referred as the "speed of adjustment". Mathematically, it represents the inverse of the marginal cost of adjustment.

robust. Therefore, even if adjustment costs are likely to be different for the three types of actors included in this model, we decided to keep them homogeneous based on the sensitivity analysis's' results.

- Interest rate (r)

This parameter was quantified through a review of literature and online resources, as well as some contacts' consultation. In our model this parameter can be interpreted either as the discount rate of each representative farmer on her decision to invest on a cash crop, or as the interest rate a farmer would be charged for a loan used on the cash crops' investment. For each of these three types of farmers at each frontier, we chose to use data on real interest rates so, we collected data on the nominal interest rate and countries' inflation[13]. Data on inflation rates were taken from the World Bank with the exception of Bolivia from which we take the data from the Bolivian official government's website because the latest source had more accurate data (see detailed information on Annex). Data on the nominal interest rates for the three types of actors were collected from different sources for each frontier case (see Annex). This data was also revised with farmer' and experts' knowledge. For the cases in which different sources were indicating slightly different values for the same group of actors, we have computed an average of such values. For the cases of farmers financed internationally, we use the used the EURIBOR rates as a proxy for international interest rates and we set them at 1% real interest rates.

For the three parameters, we set the lower limit of the range of parameter's values to be zero. The upper limit of this interval corresponded to the higher interest rates that we found among our frontier cases. This corresponded to a maximum value of 32% reported by Mozambican medium size farmers during a two-day workshop held in Maputo in October 2018 and confirmed lately by email with one of the medium size farmers.

[13] Risk is not included on the real interest rate computation.

4.3.4 The two model's solutions

Based on the above, we formalize the solutions of the system according to the following two cases[14]:

Case 1, "Frontier driven by initial conditions" ($r^2 > 4\beta\gamma$):

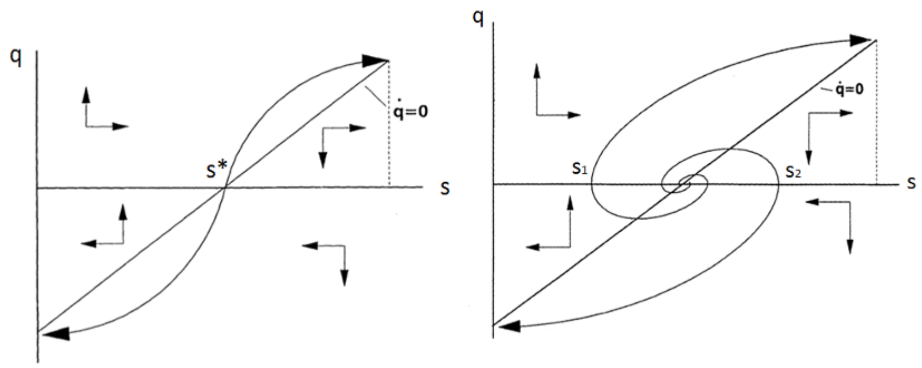
If r^2 is higher than $4\beta\gamma$, two real roots exist and the trajectory towards the equilibrium has an "s" shape (see Figure 2). For the cases when S is higher (lower) than S^*, the frontier will converge to the high equilibrium (low). In this case, the initial conditions of the system – the expected future profits (q) and the percentage of cash crops at the frontier (S), as a measure of the agglomeration economies – are sufficient to determine the equilibrium towards which the frontier dynamics will converge. The percentage of area dedicate to cash crops (S) at any point in time can be estimated via satellite image or production data. The expected profit can be determined by observable economic variables such past profits and interest rates; data about the change on expected profits ($q^{\cdot}$) is not needed in this case. However, the model in itself is not able to determine ex-ante the level of S (S*) and q at which the dynamics changes from one to the other equilibrium. This can only be determined ex-post or through different approaches.

Case 2, "Frontier driven by initial conditions and people´s expectations" ($r^2 < 4\beta\gamma$):

If r^2 is lower than $4\beta\gamma$, two complex roots exist and the trajectory towards the equilibrium will have the shape of two spirals criss-cross (see Figure 2). If initial conditions of S are between $S_1 - S_2$, the frontier could converge to either the high or low equilibriums (a pre-frontier situation dominated by semi-subsistence crops, and a consolidated frontier dominated by commercial agriculture, respectively). The one that is reached depends on the change of farmers' expectations about future profits ($q^{\cdot}$). Inside the range $S_1 - S_2$, there are more than one set of expectations (more than one set of $q^{\cdot}$ and $\dot{S}$) that

[14] We do not consider the third case (r2= 4βγ) because that would mean that the two equilibria are equal while in reality, we know that they are different.

leads to each of the two equilibriums. A small shock on the farmers profit's expectations at each point in time could drive a change in the spiral and thus the final equilibrium. Therefore, not only the initial conditions but also the people's expectations are needed to determine which equilibrium will be the final one; but for that an infinite precise information about the parameters and variables will be required. One of these variables is the change in expected future profit perceived by farmers ($q^{\cdot}$), which at each point in time depends on factors such social norms, past experiences, social networks, etc. (Schill et al., 2019). The end-point is thus likely unpredictable. In addition, the values of the interval $S_1 - S_2$ are not determined by the model. Thus, the model does not allow us to determine if a frontier region driven by the expectation dynamics lies inside or outside the $S_1 - S_2$ interval. However, if we do not observe extreme values of S at the frontier case, we will assume that the frontier lies inside the interval; if a frontier driven by the expectations would have extreme values of S (very big or very low), the dynamics would be very similar to the case in which frontiers are driven by initial conditions from a practical (non theoretical) point of view.



Source: Modified from Krugman, 1991.

Figure 4.2: system's representation of the two types of frontier dynamics: "Frontier driven by initial conditions" (left) and "Frontier driven by initial conditions and people's expectations" (right). Being

“q” the discounted future farmer’s profit and “S” the percentage of area dedicated to cash crops respect to total frontier area.

4.4 Results and comparison of the case studies

In this section we present the results obtained from applying the collected data to our theoretical model. According to it, each of the three types of actors at each frontier case will be driven by either the initial conditions or by the initial conditions and the people’s expectations depending on the parameter’s value of each frontier and actor group (Table 4.1). Yet, implementing our model to each frontier also revealed specific dynamics that are not explicitly captured in this stylized model (Tab. 4.1, last column). Based on these parameters, each frontier case can be situated in a space that indicates which of the frontier dynamics dominates there (Fig. 3).

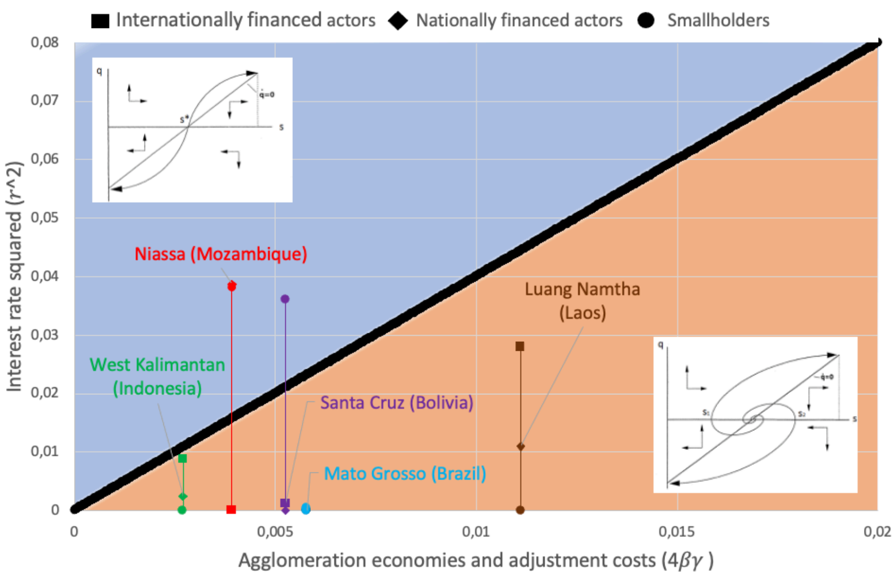


Figure 4.3. Types of frontiers dynamics according to parameters’ value. The space colored in blue color (top left of the graph) represents the parameter’s space for which the dynamics are driven by initial conditions. The space colored in orange color (bottom right of the

graph) represents the parameter's space for which the dynamics are driven by initial conditions and people's expectations. Own source.

Table 4.1 Parameters and system's solutions

CASES	Type actors (financing)	Agglomeration economics (β)	Inverse Adjustment costs (γ)	Real interest rate (r)[15]	Initial conditions vs expectations[16]	Key simplifications[17]
Mozambique, Niassa	International	0,0575	0,0681	1,00 %	Expectations	
	National			19,70%	Initial conditions	
	Smallholders			19,70%	Initial conditions	
Brazil, Mato Grosso	International	0,068	0,085	1,00 %	Expectations	Major differences between soy and cattle frontiers in agglomeration economics and
	National			0,95 %	Expectations	

[15] Data on nationally financed actors and smallholders' interest rates are an average of those data from different sources (see Annex). Data on smallholder's interest rates have been based on the smallholders eligible for receiving credit. Nevertheless, we acknowledge that, for all frontier regions, some smallholders may not have access to credit.

[16] The term "initial conditions" refers to Case 1 on previous section: "Frontiers driven by initial conditions". The term "Expectations" refers to Case 2 on previous section: "Frontiers driven by initial conditions and farmer's expectations".

[17] This column includes processes that are specific to the frontier cases and that are not well captured by the model.

	Smallholders			-2,00 %	Expectations	adjustment costs.
Indonesia, West Kalimantan	International National Smallholders	0,06	0,045	1,00 % 9,40 % 4,90 %	Expectations Expectations Expectations	Not all actors experience the same agglomeration economy (smallholder agglomeration is very different than plantations)
Laos, Luang Namtha	International National Smallholders	0,1325	0,08375	1,00 % 10,60% 17,10%	Expectations Expectations Expectations	Agglomeration economies may strongly depend on crop type and biophysical conditions (e.g., uplands vs. paddy land), given the different cooperation strategies.
	International	0,07125	0,07375	1,00 %	Expectations	Major differences

Eastern lowlands of Santa Cruz (Bolivia)	National			3,50 %	Expectations	between soy and cattle frontiers in agglomeration economics and adjustment costs.
	Smallholders			19%	Initial conditions	

4.4.1 Niassa, Mozambique

Niassa province is the only one together with the frontier in Santa Cruz driven by initial condition for the nationally financed actors and the smallholders, a situation cause by a combination of a high value interest rates and a low value agglomeration economies and adjustment costs. All together it makes the r^2 to be larger than $4\beta\gamma$ (see Table 4.1 and Figure 4.2). This frontier case indeed presents both one of the highest values of interest rates for the nationally financed actors, and the lowest value of the agglomeration economies' parameter which was corroborated by other field and modelling evidence in Niassa province (Abeygunawardane et al., 2020). Therefore, when considering smallholders and nationally financed actors, the frontier is driven predominantly by initial conditions, while when considering the internationally financed actors, both initial conditions and people's expectations drive frontier dynamics. This indicates that the emergence of a large-scale commercial crop frontier in Niassa province would be mainly driven by either an inflow of international financed actors driven by high expectations, or by exogenous interventions (e.g., international or domestic public support and programs) that either push the area of commercial crops above an (unknown) S* value or modify the economic conditions (interest rate, agglomeration economies) for smallholders and nationally financed actors.

There is evidence that internationally financed actors (South Africa, Zimbabwe, UK, Sweden/Norway) made land use decisions and investments (partly) based on the expectations of a new market in Cabo Delgado and the development of a pulp industry in Niassa (Kronenburg García et al. in preparation). However, these expectations have been challenged by the growing political instability in the region accentuated by the COVID crises, and most of those investments have been scaling down their operations. Furthermore, the conditions regarding the area under commercial crops (S in the model) remain very low, at 0,29% of the province area. Unpublished evidence gathered by coauthors during fieldwork in years 2016-2018 reveals that the expected profit (q) reported by nationally financed farmers is also very low; they claimed that farming would be profitable only over the long term. Following the model results, and considering the high interest rates and low presence of commercial farming ("S") in that landscape, the land dynamics of the nationally financed actors and the smallholders are driven by their present conditions regarding q and S; the conditions at the present moment point to the hypothesis that this frontier is going towards the low equilibrium level if all actors were driven by the initial conditions. Nevertheless, the internationally financed actors are also driven by the actors' expectations, and they could change very rapidly their land investments. If they would decide to actively invest in Niassa, or public interventions would modify the economic conditions (interest rate, agglomeration economies), then smallholders and nationally-financed actors might be driven towards the high equilibrium.

4.4.2 Eastern lowlands of Santa Cruz, Bolivia

This frontier case is driven by initial conditions and people's expectations (Table 4.1 and Figure 4.2) for the nationally and internationally financed actors while smallholders are driving by initial conditions. This case is characterized by being one of the cheapest frontiers to invest in (its cost parameter's value is one of the highest together with the frontier in Mato Grosso).
Expectations about future profits from farmers and other land use agents such as investors are relevant for the frontier outcome. There is evidence of investment made for speculative purposes in this area coming from Argentinean investors such the CRESUD company investments which occupied 45,000 ha in the lowlands in 2008 (le Polain de Waroux, 2019). However, more data about the farmer's

expectations would be needed to understand land use and cover changes in the Bolivian lowlands' frontier.

4.4.3 Mato Grosso

According to our analysis, this frontier presents lower adjustment costs than the other frontier cases (Table 4.1). This may reflect the fact that commercial agriculture activity is well developed and easy to set up. Although, adjustment costs have not been always low and they are still highly influenced by broader institutional factors (Garrett et al., 2013). And, in reality, these adjustment costs are substantially higher among some cultural groups that are not motivated by profit maximization, but instead social status and lifestyle, such as indigenous groups and cattle ranchers to move to soy farming (Garrett et al. 2017). In addition, the interest rates for the nationally financed actors are the lowest compared to the other frontier cases which contributes to the frontier being driven by those actors' expectations.
In 2017, around 18% of landowners in the State having expectations to expand their pastureland area in the next few years, and 30% planning to expand croplands – essentially (92%) over pasture and former croplands (IMEA, 2017). Cattle ranching has an intermediate status in this frontier; on the one hand, it is the historical land use that preceded the soybean frontier, and presents limited agglomeration economies compared to soy farming, on the other hand, it is still a commercial crop requiring upfront investments and behaving in some ways as our stylized "commercial crop".

This frontier is also an example of how citizens' votes and public policies influence the decision-making parameters of land use actors. The environmental success stories in the legal Amazon during the 2000s through the Brazilian government's enforcement of forest conservation laws in combination with supply chain governance have strongly mitigated farmers and investors's expectations of deriving profit from land use expansion, while encouraging land use intensification (Garrett et al., 2018b; zu Ermgassen et al., 2020). However, since 2012 clearing rates showed an increasing trend (Carvalho et al., 2019) and the recent political moves have facilitated the resurgence of deforestation rates e.g., the presidential campaign on rural development of the current president, Jair Bolsonaro, in which the dismantlement of environmental laws was proposed in return for his election (Klinger & Mack, 2020).

4.4.4 West Kalimantan, Indonesia

This case appears also driven by the expectation's dynamics. Thus, not only the initial conditions on the area dedicated to cash crops (S) and the expected profit of farmers (observed at one point in time) matter but also the changes of farmers expectations about future profits. This case is particular because the quantification of agglomeration economies is complexified by the specific interactions between smallholders and large companies. Large companies (which are most of the time financed internationally) develop their oil palm plantations and activities in places where no one else have grown oil palm yet and function essentially by internalizing scale economies. In contrast, smallholders are dependent on the presence of some large companies' processing facilities (e.g., a mill), and thus cannot stay in business without their presence in the surroundings. Here, we estimated a single parameter for the agglomeration economies across these two groups, whether these agglomeration economies are mostly internal (for large companies) or external (for smallholders). According to our data, this frontier presents the lowest agglomeration economies, which might be an underestimation due to this homogeneity assumption: Indeed, the smallholders group in this frontier have in reality one of the most extreme cases of agglomeration economies as they depend on large-scale companies for the processing of the fruit to oil. On the contrary, the results are an overestimation of the internationally financed actors who can develop in places where oil palm does not exist yet.

This frontier case is the only one close to the frontier space limit between the two types of dynamics (for the international financed actors only). However, the internationally financed actors have been facing good profit margins (Prafitri et al. 2017). This might suggest that even if those actors would be driven by initial conditions, they would likely be driven towards the high equilibrium.

4.4.5 Luang Namtha Province, Laos

This case presents the strongest agglomeration economies and the lowest adjustment cost. This means that the interest rate would have to be very high (around 45%) for this case to be driven by the initial conditions, which is the case for none of the three types of actors explored here. This can be illustrated by key land decisions in Luang Namtha Province driven by the expectations of the Laos and Chinese's governments, as well as from the Chinese companies investing in this area. Lao government decided to orient its economic development strategy toward the attraction of Foreign Direct Investment in the land sector, with the expectations of developing the rural areas of the country thanks to those foreign investments. In addition, the Chinese government implemented the China`s Opium Replacement Programme, expecting to reduce the opium production in northern Myanmar and Laos (and thus, opium consumption in China); for that, they provide direct financial and logistical support to Chinese businesses investments in commercial agriculture in those areas (Lu & Schönweger, 2019). Chinese companies' investments in Northern Laos were also driven, among other things, by their expectations about acquiring large plots of "empty land" that they could use for their own interests (Dwyer & Vongvisouk, 2019). In addition, some local farmers decided to plant rubber as a means to secure their land tenure and to protected themselves from the expected external interventions (Kramp et al., 2020). Thus, they took part on the frontier development by mimicking the rubber plantation in nearby rubber concessions; a behavior that has been also observed in other works (Junquera et al., 2020).

4.5 Discussion

This work allowed us to characterize the frontier dynamics for the five selected frontier areas. The results show that only in Niassa for the smallholders and nationally-financed actors, and in the eastern lowlands of Santa Cruz for the smallholders, the frontier dynamics are driven mainly by initial conditions. The outcomes in the other frontier cases, are driven by both initial conditions and people's expectations. Thus, in addition to the initial conditions of the area regarding the percentage of commercial crop/s and the actors' future profits, how actors' expectations about their future profits change over time is central to determining frontier trajectories. Those two types of frontiers' dynamics separated by a precise limit in the parameter space are stylized representations that highlight the key processes that drive frontiers; in reality, frontiers dynamics are more complex and the distinction between frontiers driven mainly by initial conditions or also by people's expectations is less clear-cut.

Assessing and measuring people's expectations in a robust way in frontiers contexts is difficult and would require much more specific approaches and data than used in this paper. However, it is not beyond reach, and several works have already been exploring the role of people's expectations in driving land use dynamics in frontiers. Speculative investments from Argentinean investors started the frontier in the Bolivian lowlands (le Polain de Waroux, 2019). Similarly, rubber investments made under the Chinese agrobusiness companies' expectations of an "empty land" triggered the rubber boom in northern Laos (Dwyer & Vongvisouk, 2019). Our model directly represents only the expectations of the farmers and land users themselves. Yet, in reality, these expectations are themselves influenced by the expectations, decisions and actions of a wide range of other actors, including policymakers, actors in financial markets, and others. The Brazilian Amazon (of which the frontier in Mato Grosso is part of) has experienced visible and quite rapid land changes that coincided with cyclical changes to Brazilian environmental policies since the 1980's. The most recent chapter is the political campaign on rural development of Brazil's president Jair Bolsonaro, who created expectations by promising to dismantle previous environmental laws should he be elected (Klinger & Mack, 2020), and the materialization of those expectations after his arrival to the presidency. Conversely, there is a broad consensus that environmental governance in the period 2004-2012 has driven reductions in the Amazonian deforestation (Thaler et al., 2019). Thus, voters' expectations through their electoral choice are another form of expectations that impact land use changes, if the policies they support effectively translate into changing the decision-making conditions on the ground. These voters' expectations and subsequent political choices in turn influence the farmers' expectations and parameters of the system's dynamics described by our model.

The frontier region represented in this model may represent a range of situations that differ from the two stable states represented here (the semi-subsistence farming and commercial agriculture), as long as one land regime is assumed to have no adjustment costs (i.e., it is present in the landscape at the beginning), and no agglomeration economies (corresponding to the L below). The other land regime has to have adjustment costs and agglomeration economies instead (corresponding to the S below). These two land regimes do not explicitly include natural reserves or inaccessible forest areas, which are represented as a part of the mosaics of land uses originally present in the landscape (the "L" share). Similarly, the different nature of the agriculture in each area may influence farmers decisions (e.g., in Kalimantan individuals or households may choose not to convert all their land to a single commodity as a risk reduction strategy). Our analysis only explores the different nature of agriculture regarding adjustment cost and agglomeration economies, but further work could be done on this. In addition, the model captures different ways in which agglomeration economies can arise for different actors, such as smallholders' cooperatives (Verhofstadt & Maertens, 2014), social business initiatives (Maertens & Velde, 2017), medium scale farmers' entrepreneur partnerships (Jepson et al., 2006) or large scale farms which are big enough to internalize the agglomeration economies (equivalent to internal economies of scale) in its own (Dharmawan et al., 2020). The model can be adjusted to different time periods by updating the parameters. This reflects changes in structural conditions (economic, technological, political...) (Coppola et al., 2018) but also the learning and adaptation capacity of humans over time (Groeneveld et al., 2017) and the evolving nature of human behavior (Schill et al., 2019). Nevertheless, this model remains highly stylized and thus faces limitations. People's expectations are only one aspect of the cognitive and behavioral factors that result in land use decisions (Schill, 2019), and imperfectly capture the co-evolution of behaviors and social-ecological contexts in frontier areas. Our work here assumes homogeneity of the agglomeration economies and adjustment cost across actors, yet, beyond facing different material and institutional conditions regarding adjustment costs and agglomeration economies, different actors also have distinct values, beliefs and agentic capacity to change the structural conditions of their context or to respond to undesirable conditions (Schill, 2019). Some of those agents being indigenous groups (Russo et al. 2021). The representative agents in this model make decisions based on an income optimization logic. This might work well for a number of land use actors, but others such as indigenous groups or smallholders practicing semi-subsistence agriculture often use other decision-making logics such as risk minimization of "satisficing" logics – i.e., ensuring a certain level of material output (Meyfroidt et al. 2018; Malek et al., 2019). Perspectives from psychology, sociology, political science, anthropology and economics can help to improve the understanding of human behavior in the context of social-ecological systems (Schill, 2019) and the representation of behavioral theories in land use and social-ecological systems models (Groeneveld et al., 2017).

In each frontier, some specific processes or factors were imperfectly captured in our stylized model (Table 4.1). Further, the process of agglomeration economies represented implies that profit per unit farmed with the commercial crop increases when the total area of commercial crops (S) increases. Yet, in reality, other dynamics could occur. First, the logic of Ricardian land rent theory would suggest that, in a landscape, the most suitable lands are used first, so that when the total area of commercial crops increases, more marginal lands have to be used and thus the marginal profit might decrease (Meyfroidt et al. 2018; Lambin, 2012). Second, frontiers development is often driven by speculative and anticipative dynamics of rent capture, where initial movers in a landscape can secure land at a cost that can be considerably lower than its potential rent. In such a view, those land use actors that operate the initial shares of commercially-cropped areas would obtain a higher profit than those latecomers who develop the subsequent areas (le Polain de Waroux et al. 2018).

This work also serves to design better policies in frontier regions, based on the identification of the two types of dynamics. If the trajectory appears mainly driven by initial conditions, policies to steer land use dynamics in one direction or another should focus on modifying these structural conditions including interest rates, agglomeration economies and adjustment costs. Instead, when changes in actors' expectations play a key role, which is the case for most of the frontiers and actors analyzed here, this provides an additional policy lever. Policymakers or other actors (NGOs, international agencies…) may seek to steer land use trajectories by fostering certain visions about the future of a region. Policies could create contexts and influence their change over time and space (Schill, 2019). Therefore, our work could help policy makers to design policies that create or support contexts that could serve as "attractors", guiding human actions and relations towards a desired direction. Policy making in complex land systems can be facilitated by the identification of "early warning signals", i.e. indicators that the system is about to cross certain thresholds triggering rapid changes or regime shifts (Bauch et al., 2016; Müller et al., 2014). Yet, identifying such early warning signals is uneasy (Boettiger & Hastings, 2013). Our model contributes to address this by setting the basis for building up an early warning system. It points to agglomeration economies and changes in farmers expectations as ways in which signals will be transmitted. Therefore, good measurement of both will contribute to the detection of warning signals used to anticipate potential land regime shifts. The level of agglomeration economies in an area can be observed by fieldwork and statistical approaches, and might thus be easiest to monitor. The changes on farmer's expectations are instead more difficult to quantify and they require a continuous measurement.

4.6 Conclusions

Agricultural frontiers can be analyzed as spaces and times of transition between multiple stable land system dynamics or, in other words, regime shifts. These areas are characterized by their complexity with multiple processes interacting at different scales, their uncertainty and the diversity of actors present in them, which complicates the understanding of a frontier's trajectories. Here, we shine some light on this complexity by analyzing when frontier dynamics can be understood essentially based on their initial conditions, in particular of interest rates, agglomeration economies and adjustment costs, or as also driven by changes in the expectations of the land users. The results show that the expectation type of dynamics is the more common in frontier regions, with the exception of Niassa province in Mozambique, for smallholders and nationally-financed actors, where very unfavorable structural economic and institutional conditions prevail. in all the other frontiers studied here – namely, Mato Grosso (Brazil), the Bolivian lowlands, West Kalimantan (Indonesia) and Luang Province (Laos) –, land users' expectations would play a key role. This stylized model strongly simplifies the complex realities in frontiers but allows to extract key insights on the role of human agency and the cognitive process on land decisions, seen especially through the human's ability to anticipate future events and project outcome expectations. This work also identifies two warning signals of shifts from one land regime to another, the level of agglomeration economies and the people's expectations.

This work highlights that policy-makers and other actors intervening to steer land use dynamics in frontiers may focus not only on modifying the structural conditions including interest rates, agglomeration economies and adjustment costs, but also by fostering certain visions about the future of a region.

This approach could be extended to different places and periods of time. The model can be modified to represent different decision-making logics and heterogeneity of actors, and can rely on more detailed data on the key variables and parameters obtained through different approaches such as surveys and remote sensing. Progress in understanding frontier areas or other human-environmental complex systems can be strengthened by more in-depth collaborations across disciplines, including those dedicated to the study of human behavior from contrasting perspectives.

5. CONCLUSION

A key challenge of the 21st century consists of satisfying the humans' increasing demands for land-based products without also undermining the ecosystems upon which we depend for life (Hannah et al., 2020). Global demand for agricultural crops is increasing, if it continues to do so, some expansion of agricultural land may be inevitable over the coming decades. Balancing the increase in agricultural production to satisfy current and future demands, along with the conservation of the biodiversity and ecosystem structures and the safeguard of poor smallholders' livelihoods is a major challenge for humanity (Hannan et al., 2020; Foley et al., 2011). This challenge involves important and complex environmental, social and economic tradeoffs which can be associated with different land systems' trajectories. This dissertation explains the agricultural and forest dynamics which are the base for these unavoidable tradeoffs. First, Chapter 2 investigates the causal relationships between changes in cropland area expansion and agricultural intensification measured through yields and total factor productivity. It does so through the assessment of three major theories explaining changes to agricultural land: land sparing, the rebound effect and induced intensification. Secondly, Chapter 3 expanded upon the previous work by focusing on forest dynamics. In this Chapter, I assessed some of the major hypotheses explaining forest dynamics: the environmental Kuznets curve, the forest transition and the ecologically unequal exchange. Finally, I explore agricultural frontier dynamics in Chapter 4 through the development of a framework that identifies the key parameters explaining frontier dynamics and it allows for several comparisons among different agricultural frontier areas.

This conclusion first summarizes the rationale, objectives and major outcomes of this dissertation. In section 5.3, I discuss the main contributions – both methodological and theoretical – of the present work to the existing body of literature on different fields of research, as well as the implications for policymakers and land planners. Finally, in section 5.4, I address some limitations of the approach adopted in the previous chapters and propose future research perspectives in the continuity of this dissertation.

5.1 Long and short run agricultural and forest dynamics

Under the growing societal demands for land-based products and the limited nature of land availability, a decrease of this demand and an increase in the volume of production will be required. The models used in this dissertation capture implicitly the demand for land as land changes are the result of both, demand and supply for land-based products. In addition, those models include the national GDP per capita which also captures much of the drivers of growth in demand and consumption of land-based products, in addition to a measure of population density. However, this dissertation specifically focuses on the production or supply side, more specifically on agricultural expansion or intensification and forest expansion or contractions. The existing literature has identified agricultural expansion as the main driver of deforestation and forest fragmentation (FAO 2019; Hosonuma et al., 2012; Curtis et al. 2018). The link between agricultural and forest dynamics is thus evident, although as pointed out in Chapter 3, agricultural land and forest dynamics are not perfectly symmetric and the causal mechanisms behind the relationship between forest and agriculture are still not fully clarified. This dissertation examined those mechanisms through the operationalization and testing of the main theories and hypotheses about the relations between agricultural and forest changes. This section presents the main conclusions and findings extracted from them.

Firstly, Chapters 2 and 3 confirmed the existence of a long-run dynamic equilibrium relationship between forest cover area, agricultural area, agricultural intensity (measured through both technological progress and increase of yields per land unit), economic development and rural population density. Long-run dynamic equilibrium relations have often been overlooked; however, as shown in these chapters, they represent an essential part of the understanding of forest dynamics, with short-run shocks being only one of several aspects that help explain them.

Secondly, this dissertation provides a broad scale quantification of the inverse relationship between agricultural and forest cover area, which appears in all the country groups of our sample over the long run. This relationship was larger than a unit for the low-income country group and lower in magnitude for middle- and high-income countries. This indicates that agricultural area expansion in middle and high-income countries may, to some extent, occur on already cleared but not previously used land or on other, non-forested types of land; while low-income countries may lack such land resources – possibly because land already cleared is directly used for agriculture – or be unable to access them because of technical or biophysical constraints. In these results, we include pastures inside our definition of agricultural area. Instead, results about the rebound effect, land sparing and induce intensification theories from Chapter 2 do not consider pasture areas because the inclusion of it would have probably blurred the results.

In addition to those general conclusions, this dissertation provides findings on specific the following theories:

- Land sparing and rebound effect

This dissertation provided evidence of the land sparing hypothesis over the long run for some staple cereals such as wheat, rice and maize. Yet, even for these crops, low absolute values in the short-term effects of yield increase indicate a partial rebound effect that reduces the land sparing. Evidence of the rebound effect hypothesis was also found, mostly in low- and middle-income countries. For all crops aggregated, yields had a rebound effect in middle-income countries over the short run. Many countries in this group, such as in Latin America, have export-oriented agriculture strongly engaged in global commodity markets, and land available for cropland expansion. Crop-specific results showed a rebound effect for several commodities (including bananas, coffee, sugarcane, rubber, flex crops, tropical fruits, palm oil in Africa, soybean in South American frontier countries) with high price-elasticity of demand in which middle-income countries are dominant players.

Over the long-run, the evidence for a rebound effect is less strong, still existing for some crops such as rubber and flex crops but disappearing for most other crops. This may be due to the fact that over the long run, rebound effects may decline as demand for many products saturates, or with stronger constraints on cropland expansion.

In addition, results show that intensification driven by technological progress was generally more likely to result in a rebound effect than intensification due to yield increase driven by increased input use. These results confirm the importance of continuing agricultural research as a strategy to increase yields (Coomes et al. 2019). An example of this is the rebound effect by TFP improvements show for low-income countries over both the long and the short run. In contrast, our results show a lack of effect of TFP on cropland area in high-income countries. In those countries, technological progress mostly affects the efficiency of other inputs, in particular labor, rather than land. In the short run, neither the land sparing or rebound effect hypotheses were validated for high-income countries, indicating that other forces such as land use zoning and agricultural policies including subsidies may have stronger effects on cropland area changes.

- Induced intensification

The induced intensification hypothesis, by which cropland intensity increase if local pressure on land increases too, was validated for low-, middle- and high-income groups over the long run for our sample of countries and years, which represent a crucial period of intensification such as through the Green Revolution (Pellegrini & Fernández, 2018; James R. Stevenson et al., 2013). When including short run results, we found that an increase in population pressure caused cropland expansion in the short run in low and middle-income countries (while the opposite was true for low-income countries over long run). This suggests that although the short-term response of increasing population pressure is generally to expand cropland, over the long-run induced intensification may reduce cropland expansion in low-income countries. This shows the importance of testing for stationary and to distinguish between the long and short run relationships when using long time series data. We also found support for the induced intensification hypothesis for staple crops such as cereals and oil palm in Africa, as well as for labor-intensive crops such as cocoa and coffee in the short run. This may be due to the fact that staple crops respond more directly to local demand than non-staples such as high value crops, and thus are more likely to see their production increasing through intensification if expansion is constrained.

In contrast, cropland expansion induced an increase in yields for several key commodity crops over the long-run (cocoa, sugarcane, oil palm, rubber, the group of high-elasticity crops and flex crops) or the short-run (staple crops in Sub-Saharan Africa, soybean, banana). This suggests the presence of agglomeration economies for these crops which is especially relevant for this dissertation as Chapter 4 investigates the presence of agglomeration in frontier regions and its effects on the frontier development process. Induced intensification and agglomeration economies are not mutually exclusive intensification pathways, but our results suggest that in some circumstances one dominates.

- The environmental Kuznets curve for deforestation

This dissertation explored the EKC hypothesis applied to forest cover change as a U-shaped relationship between forest cover and GDP per capita over time. Operationally, we thus used the quadratic form of GDP per capita as an aggregate measure of all structural changes coming from economic development. Yet, multiple mechanisms are bundled in GDP measures and that is the reason we included additional variables that allow us to test some of these main mechanisms.

Our analysis shows that old-fashioned ways of testing the EKC with only a measure of GDP or though the addition of control variables but without checking for their stationarity nature results on bias estimates and erroneous conclusions on the EKC. This dissertation made advances on this, and the results were a rejection of the EKC for most of the countries groups, with the exception of the EKC for high-income countries in the long run and post-forest transition countries in the short run. These results show the low explanatory power of GDP in the present of other control variables. In addition, the rejection of the EKC for low and middle-income economies suggest that those countries are experiencing different trajectories of relationships between economic development and forest cover. Finally, this dissertation also showed no evidence for a global EKC (for all countries included). The methods used and the inclusion of the control variables makes the results more robust than previous assessments of the EKC.

In relation to the inclusion of control variables, it is important to note that the world is highly integrated via international trade and forest loss and gains cannot be simply explained by national dynamics (Pendrill et al 2019b). In this thesis, the EKC for deforestation has been tested as only accounting for forest gains and lost at the country/ies of study. Yet, we have considered the trade of forest products as a mechanism that influence forest changes in the short run. It is widely accepted on the literature that local forest transition may be facilitated by forest loss shifting to other regions through increased imports of forest and/or agricultural products (Paff and Walker, 2010). For the results obtained in this dissertation that show an EKC for deforestation, the effect of forest products' imports has been taken into account through the variable net exports of forest products on the short run, yet the amount of forest products imports is only an imperfect measure of the corresponding impact on forest. The average of this variable is indeed negative for high income countries while positive for middle- and low-income countries. However, we did not find a significant effect of this variable on forest cover.

- Forest transition

The literature has described the processes of forest transition through interacting pathways which are: the economic development path, the globalization path, the forest scarcity path, the state forest policy path and the smallholder's pathway (Jinlong Liu et al., 2017; Patrick Meyfroidt & Lambin, 2011; Thomas K. Rudel et al., 2005; Meyfroidt et al., 2018). In this dissertation we operationalized and tested all those pathways except for the smallholder's one. The results found evidence only for some of those pathways, mainly the forest scarcity and the state of forest policy pathways. The economic development pathway was also a relevant one, but with very heterogenous results across contexts.

In addition, and even though the literature has identified mechanisms affecting forest dynamics such the agricultural intensification or international trade in land-based products (Meyfroidt et al., 2018), the variables representing these two mechanisms were, surprisingly, not statistically significant if indirect effects were not included. This may be due to the fact that the mechanisms behind these theories on forest changes are more interlinked than those proposed by theories focusing on the agricultural land. To address this issue, we identified the indirect impacts between the variables proposed by each forest-centered theory; also, we excluded the variable on agricultural area during a part of our analysis because changes in agricultural areas are expected to be a key mediator through which other underlying affect forest cover. We compare both results, but we still did not found evidence for the globalization pathway.

As for short-run dynamics, we show that for low- and high-income countries, and pre-transition countries, the main direct factor influencing forest cover was changes in agricultural area. In contrast, short-run forest dynamics in middle-income countries seem more complex, with a wider range of variables (agricultural producer price, government quality and annual average temperature) having small effects on forest cover area. The middle- income countries' group gather a heterogeneous sample of countries; some studies with a smaller sample of middle-income countries have also identified a diversity of social and political variables that are important factors for land dynamics in those countries (Ceddia et al., 2014c; Imai et al., 2018).

5.2 Agricultural frontiers dynamics

A variety of modelling approaches have been applied to study agricultural frontier areas and regime shifts. There have also been some attempts to find commonalities across specific frontier cases but, they all remain to be well-known socio-economic factors such population growth or natural disasters and unpredictable institutional and political events (Rindfuss et al., 2007). A major barrier to understanding complex systems such as frontier regions has been the insufficient attention given to long temporal perspectives on complex system behavior that can provide insights through both analog and evolutionary approaches (Dearing et al., 2010). Through the use of one of the most common approaches to the study of regime shifts, Chapter 4 tackled both aspects: the identification of parameters that would allow us to compare any frontier region, and the role of previous events on providing insight into site-specific system properties.

Our findings point to three key parameters that ultimately determine the agricultural frontier areas' trajectories are: i) the strength of agglomeration economies in those areas, ii) the costs from adopting commercial crops in those areas and iii) the interest rates charged for agricultural investment in those areas. In addition, Chapter 4's framework allows us to compare different frontier regions according to these key parameters. The framework classified a frontier's dynamics based on its initial conditions, or as depending also on the expectations of the land users. Our results showed that the expectation type of dynamics is the more common type in frontier regions, as only the Niassa province (Mozambique) and Santa Cruz in Bolivia and only for some actors showed to be dominated by initial conditions. All the other frontier cases considered in this Chapter– namely Mato Grosso (Brazil), the Bolivian lowlands, West Kalimantan (Indonesia) and Luang Province (Laos)- were dominated by the expectation type of dynamics. Frontier areas dominated by the expectation type of dynamics are influenced by their current level of agglomeration economies, and also by the land decision-makers' expectations about future agricultural profits at the frontier area of study (in frontier areas' dynamics dominated initial conditions, the land decision-makers' expectations are not relevant for the frontier land trajectory). Therefore, these results provide evidence of the important role of human agency and the cognitive process on land decisions; especially through the human ability to anticipate future events and project outcome expectations. Furthermore, this work proposes two warning signals to explore the regime shift on agricultural frontier regions: the level of agglomeration economies and the decision makers' expectations about future profits. Even if those results are based on a particular type of potential regime shift - a pre-frontier situation dominated by semi-subsistence farming shifting to a consolidated frontier dominated by commercial agriculture - they can easily be applied to other types of land regime shifts.

5.3 Main contributions of dissertation

5.3.1 Methodological contributions

This dissertation benefited from a multidisciplinary approach that rigorously used quantitative empirical econometric analysis, theoretical dynamic modelling as well as qualitative interviews with experts on frontier case studies. More specifically, this dissertation emphasizes the need for addressing the problem of spurious correlations when data are stationary, and it advocates for the use of co-integration or an equivalent method to solve this problem on long time series. This is an important methodological contribution for the field of land use and cover changes where the problem of spurious regressions and its detection through stationarity tests are often ignored. In addition, this dissertation shows that the use of cointegration is especially useful for its distinction between the long and short run dynamics and the corresponding progress in causal analysis thought the distinction between the long run systemic causality and the short run Granger causality. The progress made in causal analysis is especially relevant as the field of land system sciences is lacking a structured method for establishing causal explanations; studies often rely on terms like “drivers”, “spatial determinants” or “factors” (Meyfroidt, 2016) and particularly on cases where the same initial cause (e.g., agricultural intensification) triggers several mechanisms with counteracting effects (e.g., the land sparing and rebound effects). As mentioned above, I adopted the cointegration methodology to address the problem of spurious regression by which we obtain unilateral causal short-run effects (also called Granger causality) and long-run effects representing the net effect of a system of causal relationships. A way to explain the difference between the long and short run relationships is through the effect of population density on the agricultural land. In the short run, an increase of population density may lead to an expansion of cropland, as increasing population requires more food that. However, over the long run the increase in population may induce intensification and reduce the cropland expansion, as agricultural techniques adjust to this new population pressure and leads to improved productivity. This shows the importance of testing for stationary and to distinguish between the long and short run relationships when using long time series data. This approach has been proven to be adequate for Chapters 2 and 3 of this dissertation. However, depending on the research question and the scale of analysis this approach may not be the most adequate. Other methods to explore causal relationships are agent-based modelling, equilibrium models or structural equations systems (Verbur et al., 2016, Filatova et al., 2016). However, those methods are not adequate for non-stationary time series data. Furthermore, this dissertation went one step further on the investigation of these long-run effects. I added to our work a test that allowed us to identify the direction (sign) of the long run causal effects, noting that mutual causation could occur (Canning and Pedroni, 2008). Furthermore, other effects may complicate causal analysis such as interaction between variables (e.g., mediation, moderation) or temporal effects such as the path dependency effect or self-reinforcing sequences (Meyfroidt, 2016). Those effects are also considered in this dissertation; mediator variables are explored in Chapter 3, and Chapter 4 includes an example of self-reinforcing sequency (the agglomeration economies) and its analysis together with the influence of the initial conditions on a frontier region. In addition to this, the other main methodological contribution of this dissertation resides on the synthesis effort to operationalize and assess the major theories and hypotheses regarding these changes.

5.3.2 Theoretical contributions

This dissertation has brought important theoretical contributions to the study of socio-ecological systems. These are twofold. First of all, I provided rigorous empirical testing of the major theories and hypotheses existent in the literature to explain forest and agricultural changes. Secondly, this dissertation has made progress in the theoretical generalization of processes explaining land changes in agricultural frontier regions. Both contributions are especially relevant in the field of land system sciences where progress in theoretical generalizations has largely lagged behind (Meyfroidt, 2016) despite its success in producing empirical observations, descriptions and contextual explanations (Dearing et al., 2010). I did so though an exhaustive and laborious work of operationalizing theories with sometimes little theoretical support. Furthermore, the analysis of frontier regions in Chapter 4 has provided a theoretical framework that could be used on any land frontier region. This framework also allows for comparison of frontier areas based on well-defined criteria as well as a systematic way of comparing those areas. The proposed framework and model on Chapter 4 may bring important insights and inspiration for future research on the theoretical understanding of frontier regions.

5.3.3 Contributions beyond the land system science

Some continued expansion of agricultural land seems inevitable over the coming decades. How to best guide or reduce this expansion so that it takes place in areas where the environmental and social cost will be lowest is still a challenge for humanity (Stevenson et al., 2012). This challenge has been the focus not only of land system science but also of other fields such agricultural and development economies, rural studies, ecology, or food systems science among others; and it is also tackled by this thesis. As such, this thesis also contributes to those fields in different manners. A common topic among the agricultural economists is the production functions and this dissertation could well represent a global production function of agricultural and forest products as well as it can represent global scale production behaviors. Also related to the agricultural economics' field are the global trends on agricultural and forest areas and products that this dissertation analyzes. In addition, this dissertation provides a global understanding of the trade-offs and environmental consequences of development through the analysis of land changes due to economic development, which is also of interest for the field of ecology and development studies.

As a way of achieving the human challenge of satisfying land demand while keeping minimizing environmental and social impacts, different farming systems, strategies, and business model have been proposed (agroecology, conservation agriculture, organic farming, urban agriculture, agricultural intensification, crop diversification etc) (Ramankutty et al., 2019; Tamburini et al., 2019). This dissertation does not explore these alternatives explicitly. However, the analysis and results can contribute to a better understanding of this problematic and a global picture of it.

5.3.4 Policy relevance

The social and environmental transformation that characterized land changes may have severe consequences for ecosystems and people. This is the reason why policy makers and land planners may be interested in designing interventions to promote sustainable land uses. The understanding of causal mechanisms is crucial to attain this goal (Meyfroidt, 2016). The present dissertation has made progress in causal analysis of changes in agricultural and forest areas. Thus, the outcome of this dissertation has the following potential normative implications.

This dissertation can be used to produce transformative changes towards a sustainable land system through policies and land decision based on this dissertation's results. Firstly, Chapters 2 and 3 showed a long-run dynamic equilibrium relationship between forest cover area, economic development, agricultural area, agricultural intensification and rural population density. Thus, any intervention in one of those variables may alter the others in complex ways over the long run. Secondly, this dissertation shows that intensification is unlikely to result in actual land sparing for nature, with the exception of staple crops over the long run, unless accompanied by appropriate policies and other interventions to control and orient cropland expansion. Also, agricultural intensification driven by technological progress was generally more likely to result in a rebound effect than intensification due to yield increase driven by increased input use. All this indicates that policies supporting agricultural intensification (especially if driven by technological progress) should be accompanied by policies protecting important environmental areas if a safeguard of the latter was to be achieved. Furthermore, our results confirmed that low- and middle-income economies are experiencing different trajectories of relationships between economic development and forest cover than those of high-income countries. Development strategies in those regions should thus not rely on previous experiences of high-income countries regarding the effect of economic development. Our findings also emphasize the importance of governance as a positive feedback mechanism for forest, especially when a region has experienced previous periods of deforestation or where remaining forest areas are relatively low. Policies could build on the framework developed by Chapter 2 and 3; this framework, together with more specific national or micro-level studies could help policy makers to make decisions on specific contexts. The framework could also be applied to improve land-use policies targeted to specific agricultural crops.

Thirdly, this dissertation has major implications for the decision-making process in frontier regions. Those areas have been identified as regions where regulations may have lower impacts and where monitoring and enforcement may be more challenging (Börner et al., 2015). Although, it has been suggested that governance in frontier regions has an important role in its development (Thaler et al 2019). An improved theoretical framework and a clearer understanding of system dynamics in those areas could help identify land system vulnerabilities to experience land transitions (Ramankutty & Coomes, 2016). In addition, as these transitions continue to happen, policy-makers need to find effective ways of managing the circumstances in which regime shifts occur (mitigation), or of reducing any negative consequences of regime shifts that cannot be avoided (adaptation) (Filatova et al., 2016). The theoretical framework proposed in Chapter 4 contributes to both: i) to identify land systems vulnerabilities to experience land transitions through the identification of early warning signals, and ii) to help policy makers manage frontier regions through better design of policies. The two types of frontier dynamics identified in this dissertation contribute to a better design of policies in the following way: If the trajectory appears mainly driven by initial conditions, policies to steer land use dynamics in one direction or another should focus on modifying the structural conditions including interest rates, agglomeration economies and adjustment costs. Instead, when frontiers are driven by the actors' expectations, this provides an additional policy lever. Policy-makers or other actors (NGOs, international agencies…) may seek to steer land use trajectories by fostering certain visions about the future of a region. Chapter 4's analysis sets the basis for building up an early warning system. It points to agglomeration economies and changes in farmers expectations as ways in which signals will be transmitted, which could thus be a focus of efforts for researchers to collect data and provide better understanding of these two relevant variables. Therefore, good measurement of both will contribute to the detection of warning signals used to anticipate potential land regime shifts. The level of agglomeration economies in an area can be observed by fieldwork and statistical approaches, and might thus be easiest to monitor. The changes on farmer's expectations are instead more difficult to quantify and they require a continuous measurement.

Lastly**,** the role of climate change may have some implications for this dissertation. The effect of climate change has been taken into account since climate changes have impacted the data used in this dissertation. Thus, the variables used have been moving and changing together with climate change and they will continue to do so as reflected in the long run equilibrium. However, there may be a possibility that the variables will not change gradually with climate change but abruptly. In such case, the long run dynamic equilibrium relationship might change, and the models used in this dissertation would have to be re-estimated after such a structural break. This may happen if, for example, human behavior starts to be more reactive to climate change and this ends up on a different long run equilibrium relationship. However, this is not captured in our data and it will take time to have significant changes of human behavior reflected in data.

In relation to this, some works have proposed that policies should consider the long-term impact that human behavior has on them and that interventions should recognize that human motivation is socially constructed (Schill et al., 2019). Understanding this dynamic interplay between agency and structure means that studies should move beyond the purely bio-physical or economic considerations of human-environment challenges in order to encompass broader and often previously unacknowledged socio-institutional, political and knowledge-based dimensions of interventions. Following this reasoning, policies could instead create contexts and influence their change over time and space (Schill et al., 2019). To design policies according to level of agglomeration economies and farmers' expectations on a frontier area – as part of the frontier areas' context-goes in line with the approached suggested by this stream of literature.

5.4 Limitations and future research perspectives

Specific limitations are discussed in more depth in each chapter, but in general this dissertation had several limitations. First of all, the use of econometric techniques in Chapters 2 and 3 implies that new econometric methods will constantly appear as econometrics is permanently under development. Furthermore, the analysis used on Chapters 2 and 3 could have been improved by a different estimation of the Error Correction Model, such as the one proposed by Johansen (1988) which allows us to separate the long-run relationships among variables into several equations. I expect results based on the Johansen method to be consistent with the results exposed in this dissertation. Nevertheless, future works could replicate those results with the use of improved econometric techniques and also via the use of alternative methods to explore indirect effects in a cointegration approach e.g., sophisticated structural equation models for panel datasets.

Another limitation of this dissertation is related to the scale of analysis. The macro scale analyses used in Chapters 2 and 3 hindered the exploration of processes that occur at lower scales levels -subnational, regional or local - which may diverge from the national dynamics but that are relevant to understand certain land use processes. In relation to the scale of analysis, one limitation of this dissertation resides on the data used on each chapter. Data used in Chapters 2 and 3 was based on well-known global databases such as FAOSTAT or the World Bank Indicators that sometimes have inconsistencies and data gaps. The data used on the macro level analysis on this dissertation could be improved by an increase in the time frame and also improvements on each variable's measurement, especially the ones related to governance quality and trade. Future research based on the approaches seen in Chapters 2 and 3 or using similar databases will benefit from the above data improvements. Furthermore, the data collected for the parametrization of Chapter 4's model could have been improved in two ways. First at all, the data collection could have been performed not only to experts on each frontier area, but also to some of the actors operating on them. Also, and following Chapter's 4 analysis, interviews to the three types of actors identified- internationally financed, nationally financed and smallholders- would have improved the analysis by having a more differentiated data of the agglomeration economies and the actors' expectations parameters. In the future, I would like to see a continuation of Chapter 4’s work with improvements for a more detailed questionnaire and the use of other data sources rather than direct interviews e.g., consultation of media records to estimate current actors' expectations. In the case of this dissertation, both improvements of our data would have required much more time and resources than what is currently available. As well as macro level analysis, modelling approaches bring some limitations regarding the hypotheses and assumptions formulated. I acknowledge that hypotheses in Chapter 4 may look not too realistic, but I show in this dissertation the explanatory power of the model. Another limitation of theoretical exercises such as the one I have done in Chapter 4 is on the interpretation of the results. Therefore, a lot of caution is required when interpreting those results. This caution has been used for the interpretation I made of this dissertation’s results.

Finally, and apart from the specific limitations explained on the respective Chapters, there were some aspects of the land system dynamics that this dissertation did not fully cover. The telecoupling perspective is not explicit in Chapters 2 and 3 although, it is implicit as land changings occurring in one country are supposed to happen at the detriment of other countries (also, I used countries’ aggregates for data on agricultural land and forest which are an aggregate of national land changes). In addition, changes in demand from distant consumers, the citizens' and consumers’ attitudes towards environmental resources or the effect of urban population on land changes were not accounted for directly. Those factors, which are part of the land’s demand, could have been included if time series national level data on them would be easily accessible. However, I argue that to some extent they have been implicitly accounted for, as the land use changes studied here are the result of both, land’s demand and supply. Another aspect that could be explored in greater depth in future research, is the influence of socio-cultural and environmental contexts on the actors’ decisions in frontier regions as well as the co-evolution of this context with the actor’s behavior over the long run. The inclusion of the actor’s context may be relevant if that could be translated into a differentiation of the actor’s land decision according to a set of contexts. This could be done with the help of experts on disciplines such as psychology, anthropology or political sciences. However, the inclusion of the socio-cultural and environmental actors' contexts may bring a bigger complexity than the benefits (in terms of different/better results) obtained from its inclusion.

APPENDIX Chapter 2

This Appendix contains the following elements:

Supplementary text
Supplementary Tables S1 to S14
Supplementary Figures S1 to S12
Supplementary References

SUPPLEMENTARY INFORMATION TEXT:

Countries' income groups:

The dataset covered 161 countries, some existing only over a subset of the period. We classified the countries included in our database into three income groups. Prior that we excluded very small countries, and those with very little cropland or potential for cropland expansion. Specifically, we excluded: (i) countries below 250.000 ha of land area and (ii) countries in which the total cropland and forest area together represented less than 10% of their total area. After those countries were excluded, we ran the analysis by separating countries in three income country groups. These groups follow the 2016 World Bank classification. The "Low-income countries" group corresponds to the "Low-income economies" World Bank's category, which includes countries with a Gross National Income (GNI) per capita calculated using the World Bank Atlas method of 1,005$ or less. The "Middle-income countries" group includes the "Lower-middle-income economies" and "Upper-middle-income economies" categories which include countries with a GNI per capita between 1,006$-4,035$ and 4,036-12,235$. The "High-income countries" correspond to the "High-income economies" category which includes countries with GNI per capita of 12,056$ or more.

Low-income countries:

Afghanistan, Benin, Burkina Faso, Burundi, Central African Republic, Democratic Republic of Congo, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Haiti, Liberia, Madagascar, Malawi, Mozambique, Nepal, Niger, North Korea, Rwanda, Senegal, Sierra Leone, Somalia, South Sudan, Tanzania, Togo, Uganda, Zimbabwe.

Middle-income countries:

Albania, Angola, Argentina, Armenia, Azerbaijan, Bangladesh, Belarus, Belize, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Cambodia, Cameroon, Cape Verde, China, Colombia, Congo, Costa Rica, Cote d'Ivoire, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Fiji, Gabon, Georgia, Ghana, Guatemala, Guyana, Honduras, India, Indonesia, Iran, Iraq, Jamaica, Kazakhstan, Kenya, Kyrgyz Republic, Laos, Lebanon, Lesotho, Macedonia, Malaysia, Mexico, Moldova, Montenegro, Morocco, Myanmar, Namibia, Nicaragua, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Romania, Russia, Samoa, Solomon Islands, South Africa, Sri Lanka, Sudan, Suriname, Swaziland, Syria, Tajikistan, Thailand, Timor, Tunisia, Turkey, Turkmenistan, Ukraine, Uzbekistan, Vanuatu, Venezuela, Vietnam, Yugoslavia, Zambia.

High-income countries:

Australia, Austria, Bahamas, Belgium, Brunei, Canada, Chile, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, French Polynesia, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Netherlands, New Caledonia, New Zealand, Norway, Poland, Portugal, Puerto Rico, Slovak Republic, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan, Trinidad and Tobago, United Kingdom, United States, Uruguay.

FAOSTAT data:

Country-level variables cover the 1961-2016 period (Table S1). This period was the longest available for the required data, thus enabling the analysis to more adequately capture the potential long-term equilibrium effects. We have not tested in depth the sensitivity of the results to different time subperiods, in the same way as we have not explored the results of all possible subsets of crops and regions. This would deserve further efforts. Yet, preliminary explorations suggest that our key results for the key subsets of crops and regions are robust to the time periods observed. Yet, the shorter the time period, the more difficult it is to identify potential long-term equilibrium, hence the choice to use the longest time period possible.

Control variables: Countries' weather conditions and prices

Both cropland area and intensity can be influenced by weather conditions. Several climate datasets exist including from the Climate Research unit at the University of East Anglia, Postdam Institute for Climate Impact Research and the Global Agroecological Zoning (GAEZ) from IIASA/FAO. As all these datasets are largely consistent for the purposes here, we used the World Bank precipitation and temperature datasets (https://climateknowledgeportal.worldbank.org/) which are already processed to produce yearly average values at country level.

Country and time fixed effects do not fully control for variations on domestic prices. Thus, we explored the effects of including a producers' price index, from FAOSTAT, for the analysis merging all crops for classes of country income. However, this variable masked in a number of cases the effects of the cropland area and intensity variables. This result was expected because prices act to some extent as a mediating variable, or a mechanism, between cropland area and yield. Indeed, prices capture partly the effect of cropland area and yields changes (supply). There are of course other reasons for prices changes, but supply, in other words area and yields changes, are a key component of prices (note that here, we capture only domestic supply here, therefore the prices also include information from foreign supply, accounting for market integration, and from demand, domestic and abroad). The variable of GDP per capita is included precisely to control for domestic demand. Thus, we decided not to include the producer price index variable on the model.

Testing the time series properties of the data

We assessed the time series properties of our data by first testing the stationarity of our variables using the Fisher unit root test, which allows for gaps in the panel series (StataCorp, 2017). This test performs the Augmented Dickey-Fuller unit root tests on each panel to verify whether a variable has a unit root or, equivalently, that the variable follows a random walk. It performs a unit-root test on each panel (made of each variable) separately and combines the p-values to obtain an overall test (StataCorp, 2017). We performed this test by income groups. The null hypothesis was that all panels contained a unit roots. The alternative was that at least one panel was stationary for each variable. We included time trend and one lag. We explored specifications with different lag structures up to four lags, but this introduced collinearity between the lag variables and the other variables. When the variables were non-stationary, we continued by regressing them and testing for the stationarity of the error term of that regression, which indicates cointegration.

We tested for panel cointegration using the Kao test. This test assumes a cointegrating vector that is homogenous across all panels, and estimates panel specific means not allowing for a time trend (StataCorp, 2017). The null hypothesis is no cointegration between the dependent and independent variable, the alternative hypothesis that the variables are cointegrated in all panels. This test reports different t statistics that differ in how they formulate the hypothesis and how they control for serial correlation in the error. We chose the Modified Dickey-Fuller t statistics for our results.

Long-run causality test

The presence of cointegration is sufficient to prove the existence of at least one non-spurious long-run causal relationship between the variables. Yet, the estimated coefficients represent the net effect of a system of relationships and cannot be directly interpreted as unilateral causal effects (Granger causality) but rather as a systemic causality. Those coefficients are unbiased but not robust and it is therefore impossible to use them directly for statistical inference on the long-run direction of the causality (Canning & Pedroni, 2008; Stock, 1987). We thus used the approach developed by Canning and Pedroni (2008) to test for the direction of long-run causality in cointegrated panel data, extending their two-variables design to a five-variables model. Their test identifies the sign of the ratio of error correction coefficients under a transformed specification. Here, we exemplify the approach with the long- and short-run models for yield and cropland described above (Eq. S1-S6).

We first transformed the long- and short-run equations (Eq. S1-S4), to isolate γ_7 and γ'_7 (Eq. S5). These parameters always have a negative sign as they are the ECT, bounded between 0 and -1. Thus, we find that the sign of the ratio $\left(\frac{\gamma_7}{\gamma'_7}\right)$ is the opposite than that of the long-run coefficient β_1 (Eq. S6). We replicated the same transformation to estimate the sign for $\beta_2, \beta_3, \beta_4, \beta_5$ *and* β_6. We tested the sign of the ratio using a Wald test with a chi-squared statistics at a significant level of 1%, 5% or 10% (StataCorp, 2017).

Eq.S1

$$\begin{aligned} LogYield_{i,t} = {} & \propto_{1i} + \beta_1 LogCropland_{i,t} + \beta_2 TFP_{i,t} \\ & + \beta_3 LogGDPcapita_{i,t} + \beta_4 LogRPDensity_{i,t} \\ & + \beta_5 AvgRain_{i,t} + \beta_6 AvgTemperature_{i,t} \\ & + \beta_7 \mu_{yieldi,t} \end{aligned}$$

Eq. S2

$$\begin{aligned} & LogCropland_{i,t} \\ & = \frac{\alpha_{1i}}{-\beta_1} + \frac{1}{\beta_1} LogYield_{i,t} + \frac{\beta_2}{-\beta_1} TFP_{i,t} \\ & + \frac{\beta_3}{-\beta_1} LogGDPcapita_{i,t} + \frac{\beta_4}{-\beta_1} LogRPDensity_{i,t} \\ & + \frac{\beta_5}{-\beta_1} AvgRain_{i,t} + \frac{\beta_6}{-\beta_1} AvgTemperature_{i,t} \\ & + \frac{\beta_7}{-\beta_1} \mu_{yieldi,t} \end{aligned}$$

Transformed short-run equations (Eq.S3-S4):

Eq. S3

$$\begin{aligned} \Delta LogYield_{i,t} = {} & \gamma_1 \Delta LogCropland_{i,t-1} + \gamma_2 \Delta TFP_{i,t-1} \\ & + \gamma_3 \Delta LogGDPcapita_{i,t-1} + \gamma_4 \Delta LogRPDensity_{i,t-1} \\ & + \gamma_5 \Delta AvgRain_{i,t-1} + \gamma_6 \Delta AvgTemperature_{i,t-1} \\ & + \gamma_7 \hat{\mu}_{yieldi,t-1} + \varepsilon_{i,t} \end{aligned}$$

Eq. S4

$$\begin{aligned}
&\Delta LogCropland_{i,t} \\
&= \frac{1}{-\beta_1}\Delta LogYield_{i,t-1} + \frac{\gamma'_2}{-\beta_1}\Delta TFP_{i,t-1} \\
&+ \frac{\gamma'_3}{-\beta_1}\Delta LogGDPcapita_{i,t-1} \\
&+ \frac{\gamma'_4}{-\beta_1}\Delta LogRPDensity_{i,t-1} + \frac{\gamma'_5}{-\beta_1}AvgRain_{i,t-1} \\
&+ \frac{\gamma'_6}{-\beta_1}AvgTemperature_{i,t-1} + \frac{\gamma'_7}{-\beta_1}\hat{\mu}_{yieldi,t-1} + \varepsilon_{i,t}
\end{aligned}$$

Eq. S5

$$\frac{\gamma_7}{\frac{\gamma'_7}{-\beta_1}} = \left(\frac{\gamma_7}{\gamma'_7}\right)\cdot(-\beta_1)$$

Eq. S6

$$sign\left(\frac{\gamma_7}{\gamma'_7}\right) = -sign(\beta_1)$$

TABLES:

Table S1. Definition and source of variables

Variable's name	Definition	Unit	Source
Yield	Aggregate of all crops' harvested production / Harvested area for all crops	Tonnes/ha	FAOSTAT, http://www.fao.org/faostat/en/#data
$Yield_c$	Specific crop's harvested production/	Tonnes/ha	FAOSTAT, http://www.fao.org/faostat/en/#data

	Specific crop's harvested area		
Cropland area	(Sum of Arable land and Permanent crops / Countries' Land Area)*100	%	FAOSTAT, http://www.fao.org/faostat/en/#data
Cropland area$_c$	Specific crop's harvested Area / Countries' Land Area	%	FAOSTAT, http://www.fao.org/faostat/en/#data
TFP growth[18]	The ratio of an output index (total amount of crop and livestock output) to an index of land and nonland inputs (all land, labor, capital and material resources employed in farm production).	Unitless (ratio of outputs and inputs expressed in monetary terms). The change rate uses 1961 as the baseline year.	United States Department of Agriculture, http://www.ers.usda.gov/data-products/international-agricultural-productivity.aspx
GDP per capita	Gross domestic product divided by	Constant U.S. dollars (converted from domestic	The World Bank, https://databan

[18]To reduce potential index number bias in TFP growth estimates, cost shares are varied by decade whenever such information is available. For outputs, base year prices are fixed (the base period for output prices is 2004-06). Source: https://www.ers.usda.gov/data-products/international-agricultural-productivity/documentation-and-methods/

	midyear population. (named NY.GDP.PCAP.KD at the World Bank)	currencies using 2010 official exchange rates).	k.worldbank.org/data /source/world-development-indicators
Rural population density	Rural Population / countries' Land Area	Persons / ha	FAOSTAT, http://www.fao.org/faostat/en/#data
AvgRain	Average rainfall per year (computed from monthly data)	mm	World Bank, https://climateknowledgeportal.worldbank.org/
AvgTemperature	Average temperature per year (computed from monthly data)	Degreess Celsius	World Bank, https://climateknowledgeportal.worldbank.org/

Table S2. Description of the data subsets

We selected subsets of main crops that are representative of distinct conditions regarding our hypotheses. We did not aim to be exhaustive and present the results for all possible crops and regions. We focused particularly on developing and tropical regions where the largest overall agricultural expansion occurred over the recent decades. The selected crops were wheat, rice, maize (each individually, as well as grouping these three crops), banana, cocoa, coffee, sugarcane, oil palm, soybean, rubber, and a group of flex crops including sugarcane, oil palm and soybean. Maize is a flex crop in some regions, but in other regions or subregions within a country, this is predominantly a key staple food crop, with a notable share of production directly consumed by producers. Hence, we expect that its behavior would appear somewhat mixed, in between that of a staple crop with inelastic demand, and that of a flex crop with elastic demand. In order to characterize more cleanly these effects, we preferred thus to exclude it from the flex crops group. We also assessed specific groups of these crops in subsets covering specific regions of interest, as follows: a group of major staple crops in Sub-Saharan Africa, tropical fruits in tropical countries, oil palm in Sub-Saharan Africa – considered as being mostly a staple crop in this region over the study period –, oil palm and rubber in Southeast Asia, a group of high price-elasticity crops in tropical countries, and soybean in South America.

Group crops' names	**Crops included and their FAOSTAT's code in brackets**	**Groups of countries included in each crop group (Table S3)**
Wheat	Wheat (15)	Wheat's main producer countries
Rice	Rice (27)	Rice's main producer countries
Maize	Maize (56)	Maize's main producer countries
Wheat, rice and maize	Wheat (15), Rice (27) & Maize (56)	Sum of wheat, rice and maize's main producer countries
Banana	Banana (486)	Banana's main producer countries
Cocoa	Cocoa (661)	Cocoa's main producer countries
Coffee	Coffee (656)	Coffee's main producer countries
Sugarcane	Sugarcane (156)	Sugarcane's main producer countries
Oil palm	Oil palm (254)	Oil palm's main producer countries; African countries
Soybean	Soybean (236)	Soybean' main producer countries;

		South American producers
Rubber	Rubber (836)	Rubber's main producer countries
FlexCrops	Sugarcane (156), palm oil (254) and soybeans (236)	Flex crops' main producer countries
Staple crops	Barley (44), buckwheat (89), canary seed (101), cereals (108), fonio (94), grain mixed (103), maize (56), millet (79), oats (75), quinoa (92), rice paddy (27), rye (71), sorghum (83), triticale (97), wheat (15), roots and tubers (149), cassava (125)	Sub-Saharan African countries
Tropical Fruits crops	Lemons and limes (497), avocados (572), bananas (486), mangoes, mangosteens, guavas (571), papayas (600) and pineapples (574).	Tropical producer countries
Oil palm and Rubber	Oil palm (254) and Rubber (836)	Southeast Asia producer countries
High-elasticity crops	Cocoa (661), coffee (656), banana (486), sugar (156), oil palm (254), rubber (836), soybean (236)	Tropical countries

Table S3. Groups of countries by crop production

We identified the main crop producer countries as those countries that together, by decreasing order of importance, produced at least 80% of the total production of each crop in 1961 or 2014.

Groups	Countries included
Main wheat producers	China, India, Russia, United States, France, Canada, Germany, Pakistan, Australia, Ukraine, Turkey, United Kingdom, Kazakhstan, Poland, Iran, Italy, Argentina, Romania
Main rice producers	China, India, Indonesia, Bangladesh, Vietnam, Thailand, Myanmar, Philippines, Brazil, Japan, United States, Cambodia, Pakistan, Nigeria, South Korea, Nepal, Laos, Madagascar, Sri Lanka
Main Asian rice producers	Bangladesh, Cambodia, China, India, Indonesia, Japan, Laos, Myanmar, Nepal, Pakistan, Philippines, South Korea, Sri Lanka, Thailand, Vietnam.
Main maize producers	United States, China, Brazil, Argentina, Ukraine, India, Mexico, Indonesia, France, South Africa, Romania, Russia, Yugoslavia
Main soybeans producers	United States, Brazil, Argentina, China
South American soybean producers	Brazil, Argentina, Uruguay, Bolivia, Paraguay
Main cocoa producers	Cote d'Ivoire, Ghana, Indonesia, Brazil, Cameroon, Ecuador, Nigeria

Main coffee producers	Brazil, Vietnam, Colombia, Indonesia, Ethiopia, India, Honduras, Guatemala, Peru, Cote d'Ivoire, Angola, Mexico, El Salvador
Main oil palm producers	Nigeria, Democratic Republic of Congo, Indonesia, Guinea, Ghana, Malaysia
Main rubber producers	Thailand, Indonesia, Vietnam, India, China, Malaysia, Sri Lanka
Main sugarcane producers	Brazil, India, China, Thailand, Pakistan, Mexico, Colombia, Australia, Cuba, United States, Philippines, Indonesia, Puerto Rico, Argentina
Main flexcrops producers	Argentina, Australia, Brazil, China, Colombia, Cuba, Democratic Republic of Congo, Ghana, Guinea, India, Indonesia, Malaysia, Mexico, Nigeria, Pakistan, Philippines, Puerto Rico, Thailand, United States
Main banana producers	India, China, Philippines, Brazil, Indonesia, Ecuador, Guatemala, Angola, Tanzania, Costa Rica, Mexico, Vietnam, Rwanda, Colombia, Burundi, Venezuela, Honduras, Panama, Bangladesh, Kenya, Dominican Republic, Malaysia
Main cotton producers	India, China, United States, Pakistan, Brazil, Uzbekistan, Russia, Mexico
Tropical countries	Angola, Bahamas, Belize, Benin, Bolivia, Brazil, Brunei, Burkina Faso, Myanmar, Burundi, Cambodia, Cameroon, Central African Republic, Colombia, Congo, Costa Rica, Cote d'Ivoire, Cuba, Democratic Republic of Congo, Dominican Republic, Timor, Ecuador, El Salvador, Equatorial Guinea, Eritrea, Ethiopia, French Guiana, Gabon, Ghana, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, India, Indonesia,

	Jamaica, Kenya, Laos, Liberia Madagascar, Malawi, Malaysia, Mexico, Mozambique, Nicaragua, Niger, Nigeria, Panama, Paraguay Peru, Philippines, Puerto Rico, Rwanda, Senegal, Sierra Leone, Somalia, Sudan, Suriname, , Suriname, Tanzania, Thailand, Gambia, Togo, Trinidad and Tobago, Uganda, Venezuela, Vietnam, Zambia
Sub-Saharan African countries	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Congo, Cote d'Ivoire, Democratic Republic of Congo, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
African countries	Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cape Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, Cote d'Ivoire, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Togo, Tunisia, Uganda, Tanzania
Southeast Asia oil palm and rubber producers countries	Malaysia, Indonesia, Thailand, Sri Lanka, Vietnam, India and China

Table S4. Panel Unit Root Test.

	Low-income		Middle-income		High-income	
Variables	N Countries	Test statistic	N Countries	Test statistic	N Countries	Test statistic
Log Yield	28	-1.3230*	84	-3.2989***	43	-5.2890***
Log Cropland	27	0.5752	84	3.5827	43	0.6951
TFP	26	1.1016	80	3.6367	34	0.0096
Log RPDensity	27	-7.2270***	83	-1.0570	43	0.9140
Log GDP capita	24	0.0516	74	-0.4301	37	3.2894
AvgRain	28	-17.3102***	81	-31.6696***	40	-25.6767***
AvgTemperature	28	-12.8455***	81	-27.1232***	40	-21.5714***

Note: The test statistics are distributed as an inverse normal under the null hypothesis of “all panels contain unit roots” and the alternative “At least one panel is stationary”. *** denotes statistical significance at the 1% level and ** at 5%, * at 10%.

Table S5. Kao Panel Cointegration Test.

	Low-income		Middle-income		High-income	
Variables	N Countries	Test statistic	N Countries	Test statistic	N Countries	Test statistic
Log Cropland, LogYield, TFP, LogGDPcapita, LogRPDensity, AvgRain and	22	0.1826	69	0.8904	31	2.3830***

AvgTemperature						

Note: The statistics are constructed using the Modified Dickey-Fuller t statistics. Null hypothesis is “No cointegration” and the alternative hypothesis “All panels are cointegrated”. *** denotes statistical significance at the 1% level and ** at 5%, * at 10%. Note that for high-income countries, the Kao test was not significant, but the ECT term was significant and negative, confirming the presence of cointegration despite the results of the Kao test.

Table S6. Additional explanatory variables explored

Variable name	Definition	Time Period	Countries covered	Source
General Governance	An average of the following 5 variables: Control of corruption, Government Effectiveness, Regulatory quality, Rule of Law, Voice and Accountability.	1996-2015	213	www.govindicators.org.
Imports Agri Products	Imports of agricultural products (in tons or area of land embodied)	1986-2011	198	Data from Kastner et al., (2014).
Exports Agri products	Exports of agricultural products (in tons or area of land embodied)	1986-2011	198	Data from Kastner et al., (2014).

Tradehigh	Exports of agricultural products send to high-income countries/total exports of agricultural products	1986-2011	198	Own calculations using data from Kastner et al., (2014).
RatioValueag Products	Value of agricultural products' exports (FAO)/ agricultural area embodied in agricultural Products 'exports	1986-2011	198	Own calculations using data from FAO and data presented in Kastner et al., (2014).
EmploAg	Employment in agriculture (% of total employment) (modeled ILO estimate) (SL.AGR.EMPL.ZS)	1991-2016	217	The World Bank, http://databank.worldbank.org/data/
Net Trade	Imports agricultural products per agricultural area - Exports agricultural products per agricultural area	1986-2011	198	Data from Kastner et al., (2014).
GDPcapSq	The square of the variable GDP (constant 2010 US$) (NY.GDP.PCAP.KD)	1960-2016	217	The World Bank, http://databank.worldbank.org/data/

Table S7. Results for yield as dependent variable, per income levels and crops and regional groups

Yield	Aggregated			
Subsets	All countries together	Low-income	Middle-income	High-Income
Long-run				
logCropland	-0.115	-0.243	-0.264	-0.0568
TFP	**0.845**	0.869	**0.854**	**0.999**
logGDPcap	**(-)0.0419**	**-0.0212**	0.0695	-0.180
logRPopD	0.379	0.00665	0.232	0.333
Rain_avg	-0.000313	0.000365	-0.000377	-0.000888
Temp_avg	-0.0169	-0.0973	-0.00184	-0.0227
Constant	-82.97	-84.73	-85.27	-95.71
N	5382	957	3010	1415
Short-run				
LD.logCropland	-0.00673	0.109	0.0855*	-0.157**
	(0.0374)	(0.1009)	(0.0518)	(0.0649)
LD.TFP	0.0960**	0.170*	0.0583	0.183*
	(0.0430)	(0.1030)	(0.0540)	(0.1017)
LD.logGDPcap	0.0357	-0.0832	0.0541	0.110
	(0.0266)	(0.0514)	(0.0336)	(0.0909)
LD.logRPopD	-0.255**	0.223	-0.404*	-0.133
	(0.1226)	(0.2655)	(0.2115)	(0.1822)
LD.rain_avg	-0.0000177	-0.000344*	0.00000617	0.000204
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
LD.temp_avg	0.00720**	0.0171*	0.00662	0.00464
	(0.0030)	(0.0095)	(0.0047)	(0.0041)
L.residuals	-0.172***	-0.227***	-0.186***	-0.153***
	(0.0081)	(0.0236)	(0.0112)	(0.0157)
Constant	0.0114***	0.00225	0.0150***	0.00163
	(0.0019)	(0.0063)	(0.0032)	(0.0040)
N	5260	935	2941	1384

Table S7. Cont'. Results for yield as dependent variable, per income levels and crops and regional groups

Yield	Specific crops and regions						
Subsets	Stable crops Sub-Sahara Africa	Tropical fruits	Oil palm in Africa	Oil palm and rubber in South Asia	High-elasticity crops	Soybean in South America	Flex crops
Long-run							
logCropland	-0.206	-0.276	0.0214	-0.336	**(+)-0.0778**	0.934	**0.0910**
TFP	**0.984**	0.662	0.0227	0.198	0.692	-0.0306	0.271
logGDPcap	-0.0253	0.0446	0.0276	-0.397	-0.0927	-0.568	0.190
logRPopD	0.0920	0.350	0.0236	1.401	-0.531	-1.817	0.199
Rain_avg	0.00164	-0.000264	-0.000307	-0.000656	-0.000212	0.00362	-0.000595
Temp_avg	-0.122	0.0544	-0.00773	-0.103	-0.0743	-0.101	-0.0509
Constant	-92.66	-63.07	-0.165	-8.666	-64.51	-4.376	-23.93
N	1545	2571	838	188	455	259	909
Short-run							
LD.logCropland	0.178***	0.0215	-0.0715***	0.0381	-0.0357	0.268***	-0.00922
	(0.0377)	(0.0151)	(0.0242)	(0.1155)	(0.0674)	(0.0660)	(0.0401)
LD.TFP	0.482***	0.102	-0.0892	0.102	-0.0367	-0.714	0.140

	(0.1470)	(0.0713)	(0.0737)	(0.3290)	(0.1596)	(0.5805)	(0.1750)
LD.logGDPcap	-0.0155	0.0194	-0.0391	0.174	-0.121*	-0.408	0.00191
	(0.0674)	(0.0388)	(0.0305)	(0.1758)	(0.0680)	(0.4916)	(0.0955)
LD.logRPopD	0.205	0.184	0.476**	0.618	-0.378	3.988	-0.109
	(-0.3942)	(0.1704)	(-0.2106)	(0.4612)	(0.2976)	(2.8691)	(0.2868)
LD.rain_avg	-0.000737**	-0.0000553	-0.0000588	0.000145	-0.000132	0.00207	-0.000281
	(0.0003)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0014)	(0.0002)
LD.temp_avg	0.0321**	0.00609	-0.00513	0.0242	0.000790	0.0472	-0.00838
	(0.0133)	(0.0066)	(0.0074)	(0.0177)	(0.0132)	(0.0472)	(0.0116)
L.residuals	-0.300***	-0.0795***	-0.181***	-0.0906***	-0.125***	-0.324***	-0.224***
	(0.0216)	(0.0093)	(0.0217)	(0.0283)	(0.0237)	(0.0888)	(0.0221)
Constant	0.00212	0.000437	-0.00496	0.0263**	0.0212***	0.126***	0.00262
	(0.0088)	(0.0034)	(0.0044)	(0.0132)	(0.0056)	(0.0269)	(0.0058)
N	1512	2514	820	184	442	254	890

Table S8. Results for yield as dependent variable for specific crops

Yield	Wheat	Rice	Maize	Wheat. Rice and Maize	Banana
Long-run					
logCropland	**(-)0.101**	0.173	0.0517	-0.150	-0.217
TFP	-0.119	**0.0608**	0.0357	0.237	0.354
logGDPcap	0.358	0.150	0.0771	0.106	0.428
logRPopD	0.230	-0.296	-0.131	0.301	0.728
Rain_avg	0.00139	0.000323	0.00252	0.000436	-0.000300
Temp_avg	-0.0163	-0.000447	-0.113	-0.0244	0.0360
Constant	9.178	98.18	-2.204	-20.15	-34.35
N	719	846	523	1459	1010
Short-run					
LD.logCropland	-0.157***	0.0315	-0.118**	-0.0851***	0.0573*
	(0.0502)	(-0.0300)	(0.0590)	(0.0279)	(-0.0297)
LD.TFP2	0.496**	0.219*	-0.306	0.0562	0.0781
	(2.36)	(1.76)	(-1.00)	(0.60)	(0.43)
LD.logGDPcap	0.337**	0.0155	0.465**	0.223***	-0.0300

	(2.30)	(0.24)	(2.53)	(3.78)	(-0.29)
LD.logRPopD	0.382	0.196	-0.853	0.122	0.159
	(0.7074)	(0.2178)	(0.9074)	(0.2259)	(0.3878)
LD.rain_avg	-0.000244	-0.0000618	-0.000321	-0.000108	-0.000256
	(0.0006)	(0.0001)	(0.0005)	(0.0001)	(0.0002)
LD.temp_avg	0.00879	-0.000205	0.0470***	0.00426	-0.00891
	(0.0085)	(0.0068)	(0.0172)	(0.0044)	(0.0137)
L.residuals	-0.659***	-0.173***	-0.606***	-0.209***	-0.0948***
	(0.0415)	(0.0204)	(0.0469)	(0.0175)	(0.0161)
Constant	-0.000423	0.0110***	0.0204*	0.0133***	0.0114
	(-0.06)	(2.59)	(1.94)	(3.92)	(1.52)
N	703	829	512	1428	989

Table S8. Cont'. Results for yield as dependent variable for specific crops

Yield	Cocoa	Coffee	Sugarcane	Oil Palm	Soybean	Rubber
Long-run						
logCropland	**0.109**	0.0221	**0.00580**	**(+)-0.0395**	1.169	**0.234**
TFP	1.197	1.439	0.811	0.240	0.635	-0.364
logGDPcap	**0.436**	0.0446	0.0177	0.247	-0.194	0.689
logRPopD	1.421	0.395	0.258	0.136	**1.071**	0.884
Rain_avg	-0.00109	0.0000704	-0.000597	0.00115	0.00133	-0.000398
Temp_avg	-0.0276	-0.0137	-0.0497	-0.0151	-0.0743	-0.113
Constant	-122.2	-144.5	-75.97	-23.25	-69.16	31.93
N	371	606	722	240	212	324
Short-run						
LD.logCropland	-0.310	-0.0182	-0.00258	0.0466	0.199***	0.00226
	(0.3163)	(0.0607)	(0.0369)	(-0.0605)	(-0.0726)	(-0.0753)
LD.TFP2	0.103	-0.261	0.357*	-0.0182	1.145*	0.584**
	(1.14)	(-0.80)	(1.93)	(-0.12)	(1.81)	(2.29)
LD.logGDPcap	0.354*	0.500**	0.0371	-0.0338	0.259	-0.193
	(1.85)	(2.44)	(0.33)	(-0.38)	(0.74)	(-1.37)

LD.logRPopD	3.784***	2.375**	-0.154	-0.135	2.729*	0.566
	(1.0367)	(1.0463)	(0.2567)	(0.4821)	(1.6053)	(0.4131)
LD.rain_avg	0.00000182	-0.000280	-0.0000450	0.000226	0.0000748	-0.000157
	0	(0.0003)	(0.0002)	(0.0002)	(0.0025)	(0.0002)
LD.temp_avg	0.0196	-0.0317	-0.00325	0.0137	-0.0251	-0.00960
	(0.0272)	(0.0240)	(0.0108)	(0.0144)	(0.0375)	(0.0130)
L.residuals	-0.223***	-0.410***	-0.189***	-0.400***	-0.462***	-0.0906***
	(0.0433)	(0.0358)	(0.0231)	(0.0656)	(0.1138)	(0.0234)
Constant	-0.0430**	-0.0229	-0.00179	0.00311	0.0306	0.0101
	(-2.48)	(-1.49)	(-0.31)	(0.36)	(1.39)	(1.05)
N	364	593	707	235	208	317

Table S9. Results for cropland as dependent variable per income levels and crops and regional groups

Cropland	Aggregated			
Subsets	All countries together	Low-income	Middle-income	High-Income
Long-run				
logYield	-0.0860	-0.154	-0.153	-0.0408
TFP	**0.113**	**0.627**	0.186	0.204
logGDPcap	-0.0675	0.0802	-0.0742	-0.0651
logRPopD	0.358	**-0.148**	0.0667	0.104
Rain_avg	0.000129	0.000181	-0.0000337	0.000567
Temp_avg	**-0.0378**	-0.0149	**-0.0823**	**0.0284**
Constant	-13.69	-67.14	-20.44	-23.99
N	5382	957	3010	1415
Short-run				
LD.logYield	0.00887*	-0.00183	0.0219***	-0.0112
	(0.0051)	(0.0114)	(0.0067)	(0.0110)
LD.TFP	0.0962***	0.151***	0.0899***	0.0113
	(0.0160)	(0.0341)	(0.0197)	(0.0419)

LD.logGDPcap	0.0177*	0.0317*	0.00809	-0.0328
	(0.0101)	(0.0175)	(0.0126)	(0.0381)
LD.logRPopD	0.118**	0.505***	0.197**	-0.0666
	(0.0465)	(0.0899)	(0.0779)	(0.0757)
LD.rain_avg	-0.0000279	-0.000107	0.00000546	-0.000125
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
LD.temp_avg	0.00159	0.00180	0.00400**	-0.000692
	(0.0011)	(0.0033)	(0.0017)	(0.0017)
L.residuals	-0.0343***	-0.0372***	-0.0422***	-0.0378***
	(0.0036)	(0.0102)	(0.0055)	(0.0076)
Constant	0.00203***	0.00148	0.00301**	-0.00515***
	(0.0007)	(0.0021)	(-0.0012)	(0.0017)
N	5260	935	2941	1384

Table S9. Cont'. Results for cropland as dependent variable per income levels and crops and regional groups

Cropland	Specific crops and regions						
Subsets	Stable crops Sub-Sahara Africa	Tropical fruits	Oil palm in Africa	Oil palm and rubber in South Asia	High-elasticity crops	Soybean in South America	Flex crops
Long-run							
logYield	-0.240	-0.450	0.181	-0.119	-0.0421	0.903	**0.681**
TFP	-0.0228	0.567	-0.133	-0.651	0.0103	-0.437	1.582
logGDPcap	0.0426	0.0171	-0.0431	-0.384	0.325	0.180	-0.313
logRPopD	0.366	0.440	0.0860	-0.512	**(-)0.0396**	1.583	0.447
Rain_avg	0.00104	0.000435	-0.000655	-0.000508	-0.000134	-0.00518	-0.0000921
Temp_avg	-0.130	-0.0509	-0.341	-0.0308	0.0291	0.00339	-0.226
Constant	16.19	-46.88	29.07	79.10	6.116	56.54	-143.8
N	1545	2571	838	188	455	259	909
Short-run							
LD.logYield	0.0615***	0.159***	0.0525	0.0151	0.0814**	0.246***	0.116***
	(0.0201)	(-0.0196)	(0.0482)	(0.0444)	(0.0355)	(0.0454)	(0.0314)
LD.TFP	0.714***	0.130	0.275***	-0.333*	0.0876	-0.378	0.356**

	(0.0971)	(0.0813)	(0.1054)	(0.1970)	(0.1200)	(0.4846)	(0.1618)
LD.logGDPcap	0.132**	-0.0383	-0.0705	0.0185	0.250***	0.169	0.144
	(0.0543)	(0.0445)	(0.0438)	(0.1028)	(0.0484)	(0.4122)	(0.0960)
LD.logRPopD	1.115***	0.182	0.316	-0.532*	0.821***	2.654	0.114
	(0.3159)	(0.1936)	(0.3010)	(0.2756)	(0.2155)	(2.3910)	(0.2651)
LD.rain_avg	-0.000163	-0.000182*	-0.0000547	0.000134	0.0000937	0.000978	0.000515***
	(0.0003)	(0.0001)	(0.0002)	(-0.0001)	(0.0001)	(0.0012)	(0.0002)
LD.temp_avg	0.0571***	0.00131	0.00128	0.000777	0.00796	0.0101	0.0115
	(0.0108)	(0.0077)	(0.0107)	(0.0111)	(0.0100)	(0.0388)	(0.0106)
L.residuals	-0.164***	-0.0902***	-0.00990	0.0377	-0.0853***	-0.0657	-0.0330***
	(0.0164)	(0.0084)	(0.0106)	(0.0283)	(0.0234)	(0.0773)	(0.0075)
Constant	-0.0125*	0.0217***	0.00679	0.0407***	0.00607	0.107***	0.0155***
	(0.0070)	(0.0037)	(0.0062)	(-0.0076)	(0.0041)	(0.0222)	(0.0053)
N	1512	2514	820	184	442	255	890

Table S10. Results for Cropland as dependent variable for specific crops

Subsets	Wheat	Rice	Maize	Wheat. Rice and Maize	Banana	Rubber
Long-run						
logYield	0.122	**-0.643**	0.0533	**-0.301**	-0.384	**0.618**
TFP	**-0.544**	**0.510**	**(+)-0.307**	-0.0154	0.977	0.994
logGDPcap	-0.0323	-0.105	0.338	0.0700	1.030	-0.105
logRPopD	0.436	0.844	-0.299	0.524	1.193	1.592
Rain_avg	0.00465	0.000386	0.00178	0.00102	0.0000230	-0.000217
Temp_avg	0.00577	-0.00498	0.0472	-0.0181	-0.00802	-0.190
Constant	66.02	-38.47	36.59	13.71	-95.99	-85.36
N	719	846	523	1459	1010	324
Short-run						
LD.logYield	-0.0826***	-0.0519	-0.0379	0.000168	0.0965***	0.0905**
	(0.0236)	(0.0387)	(0.0298)	(0.0168)	(0.0345)	(0.0417)
LD.TFP	-0.318**	0.596***	0.380	0.0744	0.588***	-0.348*
	(0.1551)	(0.1422)	(0.2317)	(0.0896)	(0.1967)	(0.1871)

LD.logGDPcap	0.173	-0.00695	0.0774	0.0454	-0.0199	0.0329
	(1.59)	(-0.09)	(0.57)	(0.80)	(-0.18)	(0.31)
LD.logRPopD	0.539	0.363	0.635	0.216	0.461	0.00932
	(0.5183)	(0.2521)	(0.6978)	(0.2182)	(0.4191)	(0.3107)
LD.rain_avg	-0.00165***	0.000216	-0.00246***	-0.000185	-0.0000946	0.0000475
	(0.0005)	(0.0001)	(0.0004)	(0.0001)	(0.0002)	(0.0001)
LD.temp_avg	-0.00738	0.0290***	-0.00676	0.00125	0.00350	0.00239
	(0.0064)	(0.0076)	(0.0133)	(0.0042)	(0.0146)	(0.0096)
L.residuals	-0.197***	-0.0591***	-0.257***	-0.0873***	-0.0732***	-0.0225**
	(0.0268)	(0.0132)	(0.0354)	(0.0117)	(0.0132)	(0.0108)
Constant	0.00359	-0.00152	0.00284	0.00350	0.000706	0.0254***
	(0.0057)	(0.0049)	(0.0081)	(0.0032)	(0.0078)	(0.0070)
N	703	829	512	1428	989	317

Notes for tables S7-S10:
For the short run: *** denotes statistical significance at the 1% level. ** at 5% and * at 10%. Standard errors are reported in parentheses for the short-run regressions.
For the long run: the estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated. Instead, the coefficients in bold letters indicate that the test based on the approach from Canning and Pedroni (2008) is significant. The sign of the coefficient can thus be interpreted as statistically significant. Brackets indicate an inconsistency between the sign obtained from the regression and the one pointed out by the test (which is indicated within brackets). In these cases, the result was ambiguous.

Table S10. Cont'. Results for Cropland as dependent variable for specific crops

Croland	Cocoa	Coffee	Sugarcane	Oil Palm	Soybean	Rubber
Long-run						
logYield	0.548	0.0432	0.0189	-0.855	0.803***	**0.618**
TFP	-1.260	-1.163	1.619	0.252	-0.726***	0.994
logGDPcap	2.094	1.195	0.355	1.562	0.178***	-0.105
logRPopD	0.810	0.489	1.014	-2.019	-1.415***	1.592
Rain_avg	0.000501	-0.000802	-0.00118	0.000176	-0.000412	-0.000217
Temp_avg	0.260	-0.00699	-0.0409	-0.0746	0.0429	-0.190
Constant	114.3	117.0	-154.5	-25.17	76.35***	-85.36
N	371	606	722	240	212	324
Short-run						
LD.logYield	0.0558*	-0.0301	0.269***	0.0944	0.214***	0.0905**
	(0.0305)	(0.0241)	(0.0415)	(0.0689)	(0.0338)	(0.0417)
LD.TFP	0.0595	0.477**	0.171	0.167	0.305	-0.348*
	(0.1859)	(0.2129)	(0.2085)	(0.1670)	(0.3813)	(0.1871)

LD.logGDPcap	-0.0935	0.0144	0.194	0.0963	0.420**	0.0329
	(-0.83)	(0.11)	(1.53)	(1.00)	(1.99)	(0.31)
LD.logRPopD	-1.100*	0.359	-0.0511	0.718	1.821*	0.00932
	(0.6044)	(0.6774)	(0.2839)	(0.5279)	(0.9686)	(0.3107)
LD.rain_avg	0.000347	-0.000118	0.000655***	0.000126	-0.00126	0.0000475
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0014)	(0.0001)
LD.temp_avg	-0.00604	-0.0397**	0.0129	0.00564	-0.00463	0.00239
	(0.0159)	(0.0157)	(0.0122)	(0.0161)	(0.0232)	(0.0096)
L.residuals	-0.0425***	-0.132***	-0.0544***	0.0115	0.0865	-0.0225**
	(0.0116)	(0.0172)	(0.0145)	(-0.0153)	(-0.0874)	(0.0108)
Constant	0.0415***	0.00701	0.00478	0.0357***	0.0316**	0.0254***
	(0.0100)	(0.0102)	(0.0065)	(0.0091)	(0.0133)	(0.0070)
N	364	593	707	235	208	317

Notes for tables S7-S10:
For the short run: *** denotes statistical significance at the 1% level. ** at 5% and * at 10%. Standard errors are reported in parentheses for the short-run regressions.
For the long run: the estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated. Instead, the coefficients in bold letters indicate that the test based on the approach from Canning and Pedroni (2008) is significant. The sign of the coefficient can thus be interpreted as statistically significant. Brackets indicate an inconsistency between the sign obtained from the regression and the one pointed out by the test (which is indicated within brackets). In these cases, the result was ambiguous.

Table S11. Results without TFP. For yield as dependent variable, per income levels and crops and regional groups

Yield	Aggregated			
Subsets	All countries together	Low-income	Middle-income	High-Income
Long-run				
logCropland	-0.0805	-0.0430	-0.179	0.0371
logGDPcap	0.178	0.134	0.174	-0.180
logRPopD	0.289	**0.0113**	0.220	0.328
Rain_avg	-0.000320	0.000112	-0.000357	-0.00160
Temp_avg	-0.00624	-0.128	0.0155	-0.0124
Constant	0.382	2.761	-0.464	4.862
N	5678	998	3142	1538
Short-run				
LD.logCropland	0.00820	0.0413	0.0714**	-0.190***
	(0.0283)	(0.0960)	(0.0332)	(0.0644)
LD.logGDPcap	0.0199	-0.123**	0.0543	0.0741
	(0.0265)	(0.0532)	(0.0333)	(0.0893)

LD.logRPopD	-0.275**	0.363	-0.578***	-0.0956
	(0.1244)	(0.2792)	(0.2109)	(0.1875)
LD.rain_avg	0.00000720	-0.000352*	0.0000501	0.000191
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
LD.temp_avg	0.00458	0.0192*	0.00608	0.0000242
	(0.0030)	(0.0100)	(0.0047)	(0.0024)
L.residualsYieldL	-0.143***	-0.225***	-0.146***	-0.118***
	(0.0072)	(0.0223)	(0.0096)	(0.0134)
Constant	0.0126***	0.000656	0.0174***	0.00480
	(0.0019)	(0.0066)	(0.0031)	(0.0037)
N	5417	953	2998	1466

Table S11. Con't. Results without TFP. For yield as dependent variable, per income levels and crops and regional groups

Yield	Specific crops and regions						
Subsets	Stable crops Sub-Sahara Africa	Tropical fruits	Oil palm in Africa	Oil palm and rubber in South Asia	High-elasticity crops	Soybean in South America	Flex crops
Long-run							
logCropland	-0.272	-0.239	0.0175	-0.349	-0.0901	0.925	**0.108**
logGDPcap	0.0394	0.162	0.0298	-0.406	0.0798	-0.614	0.244
logRPopD	0.250	0.353	0.0420	1.112	-0.591	-1.747	0.185
Rain_avg	0.00124	-0.000170	-0.000357	-0.000573	-0.0000366	0.00297	-0.000374
Temp_avg	-0.162	0.0519	-0.00592	-0.0927	-0.0486	-0.100	-0.0458
Constant	7.265	2.126	2.107	11.06	2.956	-6.677	2.470
N	1597	2628	856	192	466	264	926
Short-run							
LD.logCropland	0.102***	0.0213	-0.0696***	0.0343	-0.0313	0.248***	-0.00818
	(0.0301)	(0.0152)	(0.0243)	(0.0283)	(0.0960)	(0.0332)	(0.0644)
LD.logGDPcap	-0.0293	0.0167	-0.0391	0.164	-0.133*	-0.348	-0.00632

	(0.0681)	(0.0388)	(0.0305)	(0.1726)	(0.0679)	(0.7102)	(0.1053)
LD.logRPopD	0.164	0.112	0.467**	0.595	-0.345	7.647*	-0.142
	(0.4000)	(0.1697)	(0.2104)	(0.4508)	(0.2974)	(4.1113)	(0.2898)
LD.rain_avg	-0.000670**	-0.0000607	-0.0000597	0.000138	-0.000142	-0.00242	-0.000317
	(0.0003)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0020)	(0.0002)
LD.temp_avg	0.0358***	0.00648	-0.00536	0.0237	-0.0000530	-0.0598	-0.00831
	(0.0137)	(0.0066)	(0.0074)	(0.0176)	0	(0.0672)	(0.0117)
L.residualsYieldL	-0.262***	-0.0739***	-0.181***	-0.0930***	-0.122***	-0.279**	-0.219***
	(0.0198)	(0.0088)	(0.0216)	(0.0282)	(0.0228)	(0.1257)	(0.0219)
Constant	0.00638	0.00227	-0.00513	0.0290***	0.0204***	0.0826**	0.00467
	(0.0090)	(0.0032)	(0.0044)	(0.0103)	(0.0052)	(0.0369)	(0.0053)
N	1530	2514	820	184	442	259	890

Table S12. Results without TFP. For yield as dependent variable, for specific crops

Yield	Wheat	Rice	Maize	Wheat.Rice and Maize	Banana
Long-run					
logCropland	0.111	-0.195	0.0483	-0.155	-0.174
logGDPcap	0.332	0.0730	0.0719	0.137	0.435
logRPopD	0.224	0.225	-0.128	0.257	0.582
Rain_avg	0.00148	0.000254	0.00241	0.000511	-0.000178
Temp_avg	-0.0148	0.00778	-0.109	-0.0185	0.0454
Constant	-2.691	2.593	1.403	3.254	0.439
N	735	863	534	1490	1031
Short-run					
logCropland	-0.155***	0.0177	-0.111*	-0.0898***	0.0577*
	(0.0503)	(0.0300)	(0.0584)	(0.0279)	(0.0296)
LD.logGDPcap	0.327**	0.0234	0.485***	0.218***	-0.0337
	(0.1460)	(0.0632)	(0.1830)	(0.0591)	(0.1021)
LD.logRPopD	0.187	0.158	-0.825	0.0853	0.0997

	(0.6926)	(0.2164)	(0.8967)	(0.2245)	(0.3835)
LD.rain_avg	-0.000247	-0.0000744	-0.000300	-0.000114	-0.000258
	(0.0007)	(0.0001)	(0.0005)	(0.0001)	(0.0002)
LD.temp_avg	0.00897	-0.000525	0.0450***	0.00370	-0.00953
	(0.0085)	(0.0066)	(0.0171)	(0.0044)	(0.0136)
L.residuals	-0.680***	-0.158***	-0.609***	-0.203***	-0.0968***
	(0.0410)	(0.0196)	(0.0469)	(0.0173)	(0.0159)
Constant	0.00713	0.0141***	0.0153*	0.0144***	0.0132**
	(0.0410)	(0.0196)	(0.0469)	(0.0173)	(0.0159)
N	703	829	512	1428	989

Table S12. Con't. Results without TFP. For yield as dependent variable, for specific crops

Subsets	Cocoa	Coffee	Sugarcane	Oil Palm	Soybean	Rubber
Long-run						
logCropland	**0.0888**	-0.0854	**0.0977**	0.0175	1.110	**0.229**
logGDPcap	**0.505**	0.375	0.0948	0.0298	-0.0921	0.624
logRPopD	0.904	0.293	0.184	0.0420	0.424	1.139
Rain_avg	-0.000924	-0.0000649	-0.000356	-0.000357	0.00257	-0.000813
Temp_avg	-0.0372	-0.0362	-0.0415	-0.00592	-0.0509	-0.131
Constant	-3.130	-1.407	3.731	2.107	-7.757	-3.613
N	378	619	735	856	216	331
Short-run						
logCropland	0.133	0.00711	-0.00335	-0.0696***	0.273**	0.00829
	(0.0905)	(0.0646)	(0.0372)	(0.0243)	(0.1119)	(0.0754)
LD.logGDPcap	0.379**	0.376*	0.0104	-0.0391	0.444	-0.193
	(0.1905)	(0.2112)	(0.1156)	(0.0305)	(0.5163)	(0.1419)
LD.logRPopD	3.937***	2.465**	-0.210	0.467**	0.938	0.404

	(1.0388)	(1.0811)	(0.2561)	(0.2104)	(2.3450)	(0.4081)
LD.rain_avg	0.00000746	-0.000245	-0.000104	-0.0000597	-0.00491	-0.000169
	(0.0004)	(0.0004)	(0.0002)	(0.0001)	(0.0035)	(0.0002)
LD.temp_avg	0.0190	-0.0290	-0.00362	-0.00536	-0.0653	-0.0105
	(0.0275)	(0.0248)	(0.0110)	(0.0074)	(0.0578)	(0.0130)
L.residuals	-0.198***	-0.302***	-0.165***	-0.181***	-0.319**	-0.0776***
	(0.0406)	(0.0329)	(0.0217)	(0.0216)	(0.1603)	(0.0231)
Constant	-0.0492***	-0.0259*	0.00370	-0.00513	0.0143	0.0212**
	(0.0406)	(0.0329)	(0.0217)	(0.0216)	(0.1603)	(0.0231)
N	364	593	707	820	212	317

Table S13. Results without TFP. For cropland as dependent variable, per income levels and crops and regional groups

Cropland	Aggregated			
Subsets	All countries together	Low-income	Middle-income	High-Income
Long-run				
logYield	**(+)-0.0501**	-0.0245	-0.0952	0.0190
logGDPcap	-0.0545	0.196	-0.0585	-0.0562
logRPopD	0.335	**-0.155**	0.0584	0.0750
Rain_avg	-0.00000423	0.000210	-0.000211	0.000673
Temp_avg	**-0.0378**	-0.0137	**-0.0866**	**0.0327**
Constant	-2.585	-5.322	-1.890	-3.853
N	5678	998	3142	1538
Short-run				
LD.logYield	0.0163**	0.00151	0.0360***	-0.0115
	(0.0064)	(0.0094)	(0.0099)	(0.0106)

LD.logGDPcap	0.0229*	0.0236	0.0156	-0.0207
	(0.0131)	(0.0182)	(0.0186)	(0.0370)
LD.logRPopD	0.107*	0.537***	0.149	-0.0544
	(0.0605)	(0.0936)	(0.1164)	(0.0756)
LD.rain_avg	-0.0000682**	-0.000114	-0.0000460	-0.000123
	(3E-05)	(7E-05)	(4E-05)	(8E-05)
LD.temp_avg	0.00142	0.00326	0.00328	-0.000510
	(0.0014)	(0.0034)	(0.0026)	(0.0017)
L.residuals	-0.0418***	-0.0351***	-0.0551***	-0.0375***
	(0.0045)	(0.0097)	(0.0075)	(0.0075)
Constant	0.00215**	0.00144	0.00282*	-0.00510***
	(0.0009)	(0.0023)	(0.0017)	(0.0015)
N	5417	953	2998	1466

Table S13. Con't. Results without TFP. For cropland as dependent variable, per income levels and crops and regional groups

Cropland	Specific crops and regions						
Subsets	Stable crops Sub-Sahara Africa	Tropical fruits	Oil palm in Africa	Oil palm and rubber in South Asia	High-elasticity crops	Soybean in South America	Flex crops
Long-run							
logYield	**(+)-0.247**	-0.363	0.155	-0.145	-0.0461	0.940	0.864
logGDPcap	0.0419	0.108	-0.0477	-0.375	0.329	0.249	-0.0371
logRPopD	0.366	0.429	0.0810	0.698	**(-)0.00364**	1.661	0.410
Rain_avg	0.000983	0.000600	-0.000537	-0.00107	-0.000140	-0.00426	0.000881
Temp_avg	-0.140	-0.0614	-0.331	-0.0679	0.0332	-0.0224	-0.225
Constant	14.16	9.282	15.56	15.22	6.971	12.74	11.68
N	1597	2628	856	192	466	264	926
Short-run							
LD.logYield	0.0668***	0.158***	0.0478	0.0130	0.0856**	0.247***	0.118***
	(0.0190)	(0.0195)	(0.0483)	(0.0433)	(0.0349)	(0.0454)	(0.0316)
LD.logGDPcap	0.117**	-0.0409	-0.0717	0.0450	0.253***	0.205	0.129

	(0.0549)	(0.0445)	(0.0437)	(0.1047)	(0.0485)	(0.4020)	(0.0963)
LD.logRPopD	1.085***	0.105	0.339	-0.480*	0.831***	2.636	0.0527
	(0.3191)	(0.1909)	(0.3000)	(0.2727)	(0.2153)	(2.3748)	(0.2635)
LD.rain_avg	-0.000175	-0.000190*	-0.0000395	0.000149	0.0000933	0.00102	0.000468**
	(3E-04)	(1E-04)	(2E-04)	(1E-04)	(1E-04)	(1E-03)	(2E-04)
LD.temp_avg	0.0667***	0.00181	0.00210	0.000453	0.00811	0.00836	0.0121
	(0.0108)	(0.0075)	(0.0105)	(0.0113)	(0.0099)	(0.0398)	(0.0107)
L.residuals	-0.185***	-0.0880***	-0.0114	-0.00785	-0.0871***	-0.0650	-0.0308***
	(0.0165)	(0.0082)	(0.0108)	(0.0262)	(0.0234)	(0.0765)	(0.0072)
Constant	-0.00860	0.0239***	0.00732	0.0326***	0.00698*	0.102***	0.0205***
	(0.0071)	(0.0036)	(0.0063)	(0.0059)	(0.0039)	(0.0211)	(0.0048)
N	1530	2514	820	184	442	255	890

Table S14. Results without TFP. For cropland as dependent variable, for specific crops

Cropland	Wheat	Rice	Maize	Wheat. Rice and Maize	Banana
Long-run					
logYield	0.145	**-0.587**	0.0533	**-0.297**	**(+)-0.328**
logGDPcap	-0.116	-0.0238	0.298	0.0736	1.163
logRPopD	0.456	0.706	-0.285	0.517	0.832
Rain_avg	0.00444	0.000573	0.00157	0.000931	0.000260
Temp_avg	-0.00300	-0.00307	0.0396	-0.0158	0.00183
Constant	12.29	11.92	6.248	12.09	0.350
N	735	863	534	1490	1031
Short-run					
LD.logYield	-0.0778***	-0.0475	-0.0405	-0.000442	0.0944***
	(0.0234)	(0.0393)	(0.0300)	(0.0221)	(0.0347)

LD.logGDPcap	0.218**	0.00731	0.0634	0.0437	-0.0295
	(0.1090)	(0.0731)	(0.1349)	(0.0568)	(0.1093)
LD.logRPopD	0.678	0.181	0.516	0.196	0.199
	(0.5215)	(0.2479)	(0.6880)	(0.2154)	(0.4146)
LD.rain_avg	-0.00168***	0.000196	-0.00245***	-0.000182	-0.0000865
	(0.0005)	(0.0001)	(0.0004)	(0.0001)	(0.0002)
LD.temp_avg	-0.00685	0.0285***	-0.00442	0.00122	0.00101
	(0.0063)	(0.0077)	(0.0134)	(0.0042)	(0.0144)
L.residuals	-0.196***	-0.0585***	-0.244***	-0.0868***	-0.0604***
	(0.0262)	(0.0130)	(0.0351)	(0.0116)	(0.0126)
Constant	-0.00238	0.00737*	0.00914	0.00465	0.0121*
	(0.0051)	(0.0044)	(0.0068)	(0.0029)	(0.0073)
N	703	829	512	1428	989

Notes for tables S11-S14:
For the short run: *** denotes statistical significance at the 1% level. ** at 5% and * at 10%. Standard errors are reported in parentheses for the short-run regressions.
For the long run: the estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated. Instead, the coefficients in bold letters indicate that the test based on the approach from Canning and Pedroni (2008) is significant. The sign of the coefficient can thus be interpreted as statistically significant. Brackets indicate an inconsistency between the sign obtained from the regression and the one pointed out by the test (which is indicated within brackets). In these cases, the result was ambiguous

Table S14. Con't. Results without TFP. For cropland as dependent variable, for specific crops

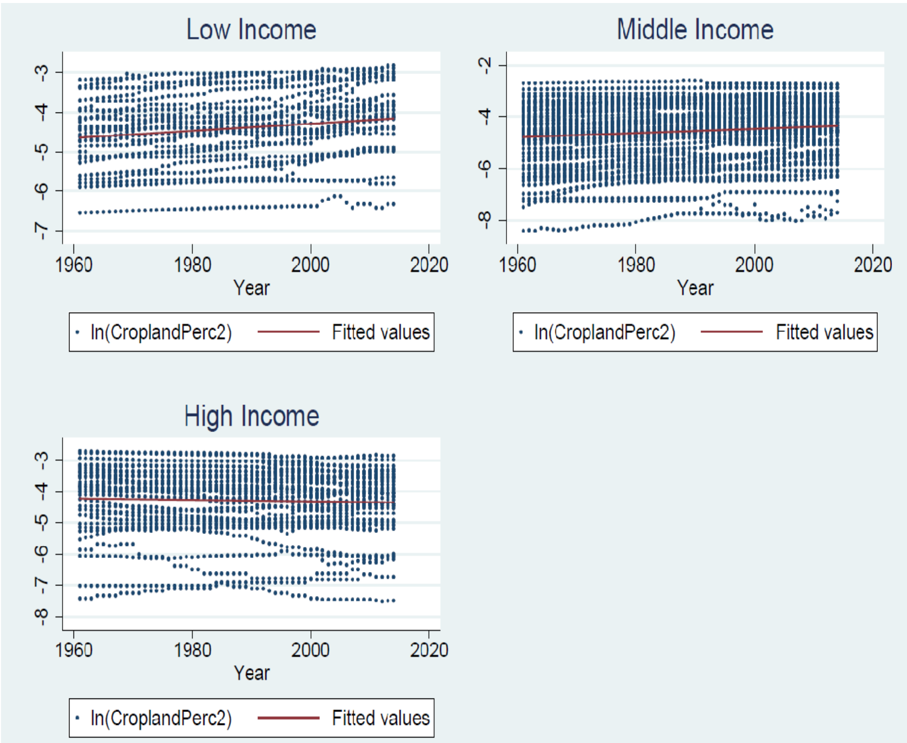
Subsets	Cocoa	Coffee	Sugarcane	Oil Palm	Soybean	Rubber
Long-run						
logYield	0.402	**-0.135**	0.334	-0.768	0.853	**0.642**
logGDPcap	2.129	1.074	0.560	1.609	0.0576	0.0156
logRPopD	1.455	0.675	0.962	-2.107	-0.778	0.623
Rain_avg	0.000249	-0.000776	-0.000821	0.000177	-0.00203	0.000223
Temp_avg	0.237	-0.00232	-0.0114	-0.108	0.0127	-0.161
Constant	-10.92	1.339	4.180	0.441	6.656	12.84
N	378	619	735	245	216	331
Short-run						
LD.logYield	0.0578*	-0.0214	0.265***	0.0982	0.215***	0.0848**
	(0.0303)	(0.0240)	(0.0415)	(0.0687)	(0.0343)	(0.0418)
LD.logGDPcap	-0.103	0.0274	0.178	0.0859	0.363*	0.0412
	(0.1120)	(0.1305)	(0.1271)	(0.0954)	(0.2039)	(0.1056)

LD.logRPopD	-1.179*	0.120	-0.0884	0.679	1.540*	0.0980
	(0.6015)	(0.6667)	(0.2852)	(0.5264)	(0.9222)	(0.2970)
LD.rain_avg	0.000350	-0.000111	0.000626***	-0.000143	-0.00130	0.0000584
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0014)	(0.0001)
LD.temp_avg	-0.00555	-0.0409***	0.0126	0.00683	-0.00406	-0.00205
	(0.0159)	(0.0157)	(0.0122)	(0.0159)	(0.0226)	(0.0098)
L.residuals	-0.0439***	-0.130***	-0.0500***	-0.0134	0.0598	-0.0264**
	(0.0113)	(0.0166)	(0.0136)	(0.0152)	(0.0777)	(0.0104)
Constant	0.0432***	0.0161*	0.00755	0.0381***	0.0377***	0.0184***
	(0.0096)	(0.0094)	(0.0058)	(0.0088)	(0.0114)	(0.0059)
N	364	593	707	235	208	317

Notes for tables S11-S14:
For the short run: *** denotes statistical significance at the 1% level. ** at 5% and * at 10%. Standard errors are reported in parentheses for the short-run regressions.
For the long run: the estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated. Instead, the coefficients in bold letters indicate that the test based on the approach from Canning and Pedroni (2008) is significant. The sign of the coefficient can thus be interpreted as statistically significant. Brackets indicate an inconsistency between the sign obtained from the regression and the one pointed out by the test (which is indicated within brackets). In these cases, the result was ambiguous

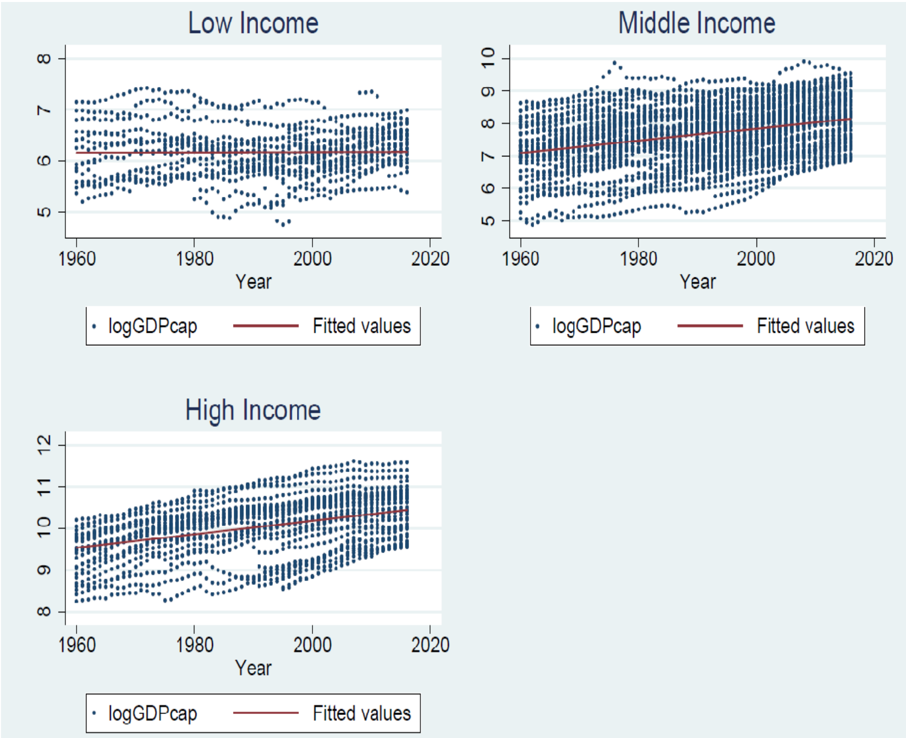
FIGURES:

Fig. S1[19] Temporal evolution of cropland area in low-, middle- and high-income countries.



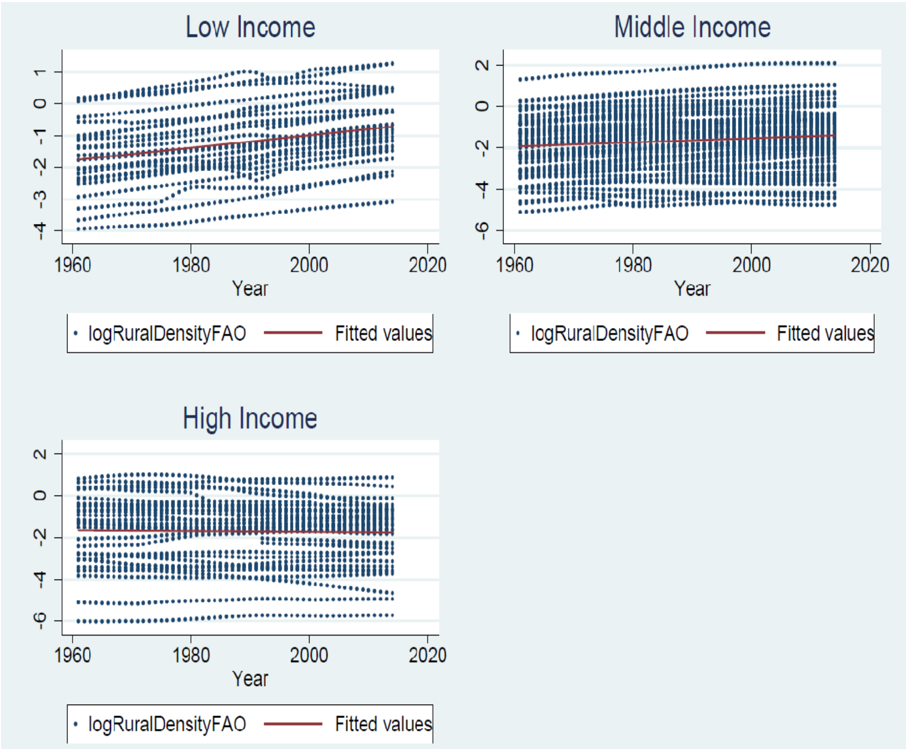
19 The variable on the vertical access corresponds to the logarithm of "Cropland area" (in Table S1).

Fig. S2[20] Temporal evolution of GDP per capita (in logarithm form) for low-, middle- and high-income countries.



20 The variable on the vertical access corresponds to the logarithm of "GDP per capital" (in Table S1).

Fig. S3[21] Temporal evolution of rural population density (in logarithm form) for low-, middle- and high-income countries.



[21] The variable on the vertical access corresponds to the logarithm of "Rural population density" (in Table S1).

Fig. S4[22] Temporal evolution of yield (in logarithm form) for low-, middle- and high-income countries.

22 The variable on the vertical access corresponds to the logarithm of "Yield" (in Table S1).

Fig. S5 Temporal evolution of average rainfall for low-, middle- and high-income countries.

Fig. S6 Temporal evolution of average temperature for low-, middle- and high-income countries.

Fig. S7 Temporal evolution of yield (in logarithm form) for wheat, rice, Maize, Soybeans and cocoa.

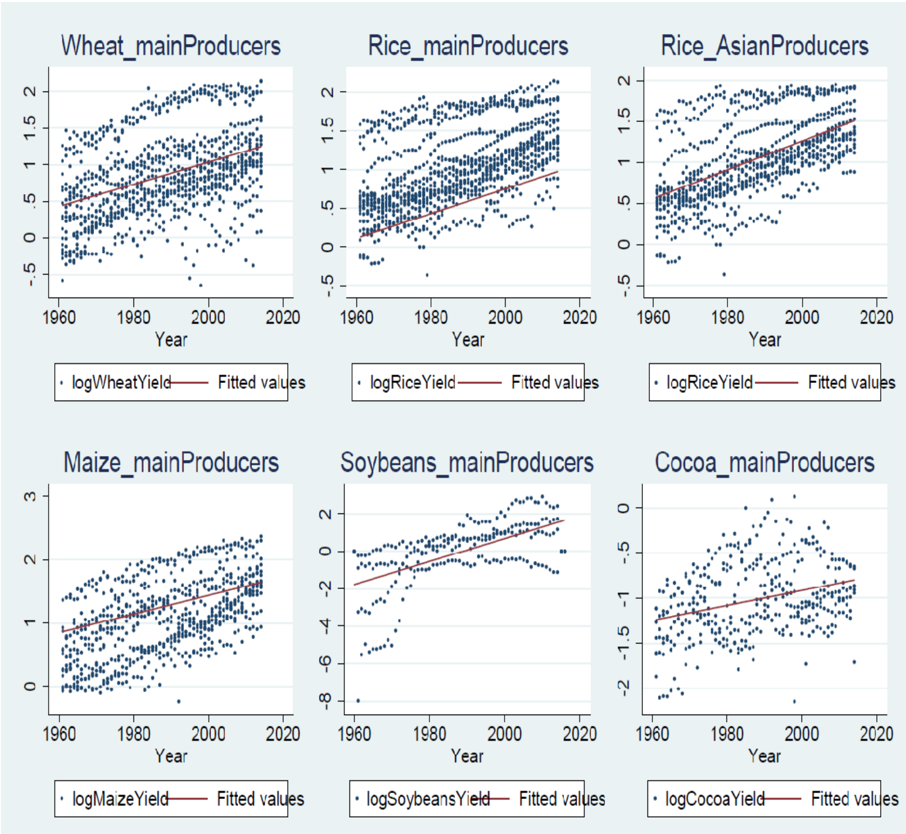


Fig. S8 Temporal evolution of yield (in logarithm form) for coffee, oil palm, rubber, cassava, banana and sugarcane.

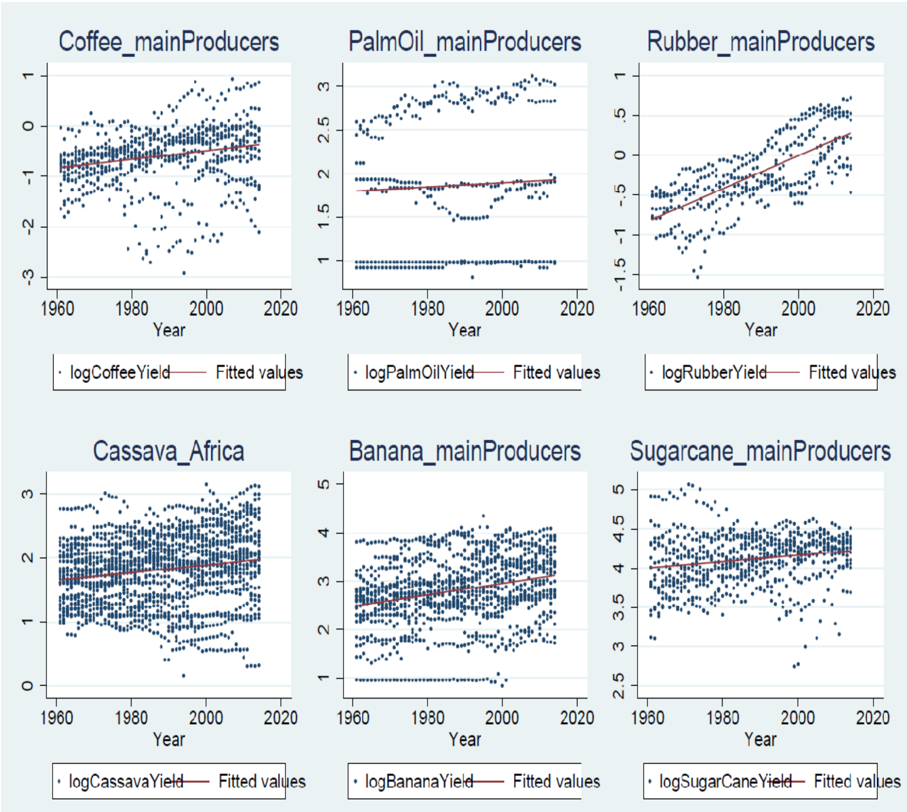


Fig. S9 Temporal evolution of yield (in logarithm form) for cotton, tropical fruits and staple crops.

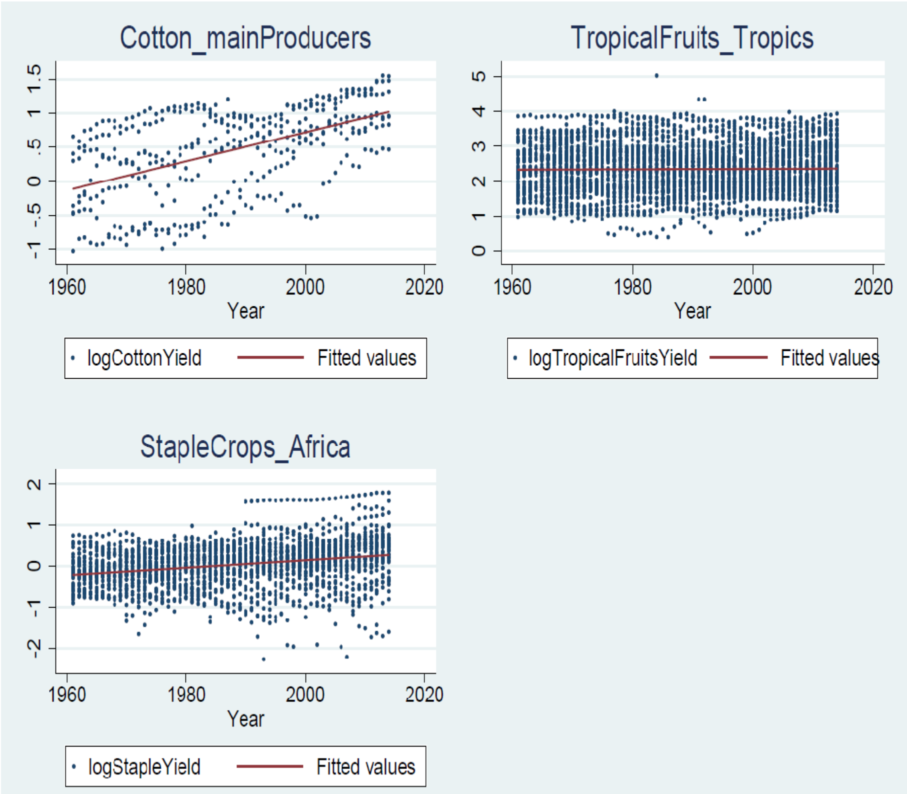
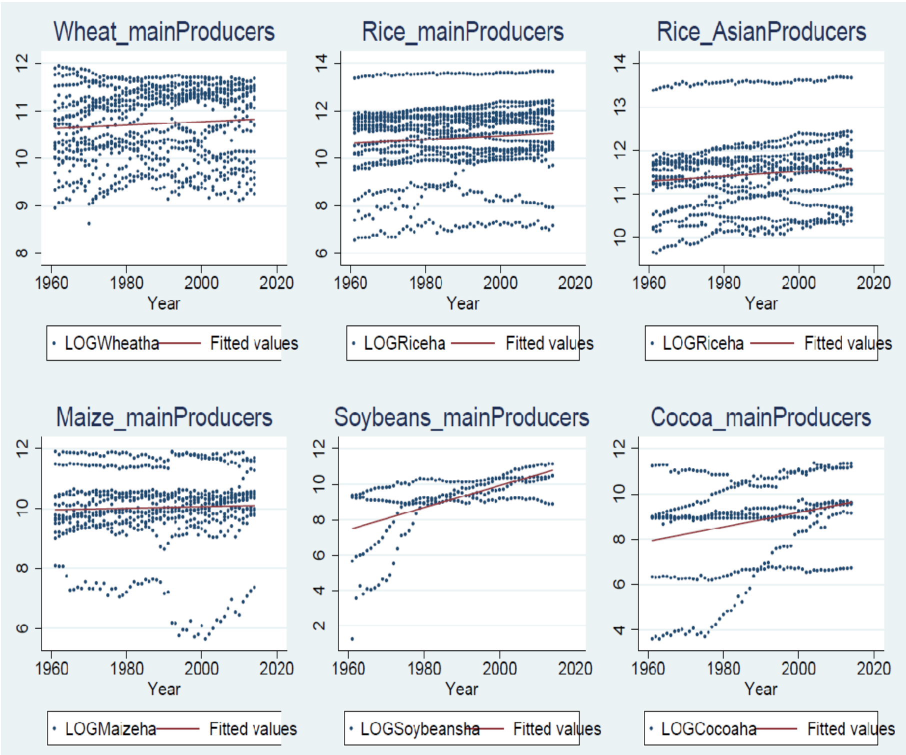


Fig. S10[23] Temporal evolution of cropland (in logarithm form) for wheat, rice, Maize, Soybeans and cocoa.



23 The variables on the vertical access for S8-S10 correspond to the logarithm of "Cropland areac" (in Table S1).

Fig. S11 Temporal evolution of cropland (in logarithm form) for coffee, oil palm, rubber, cassava, banana and sugarcane.

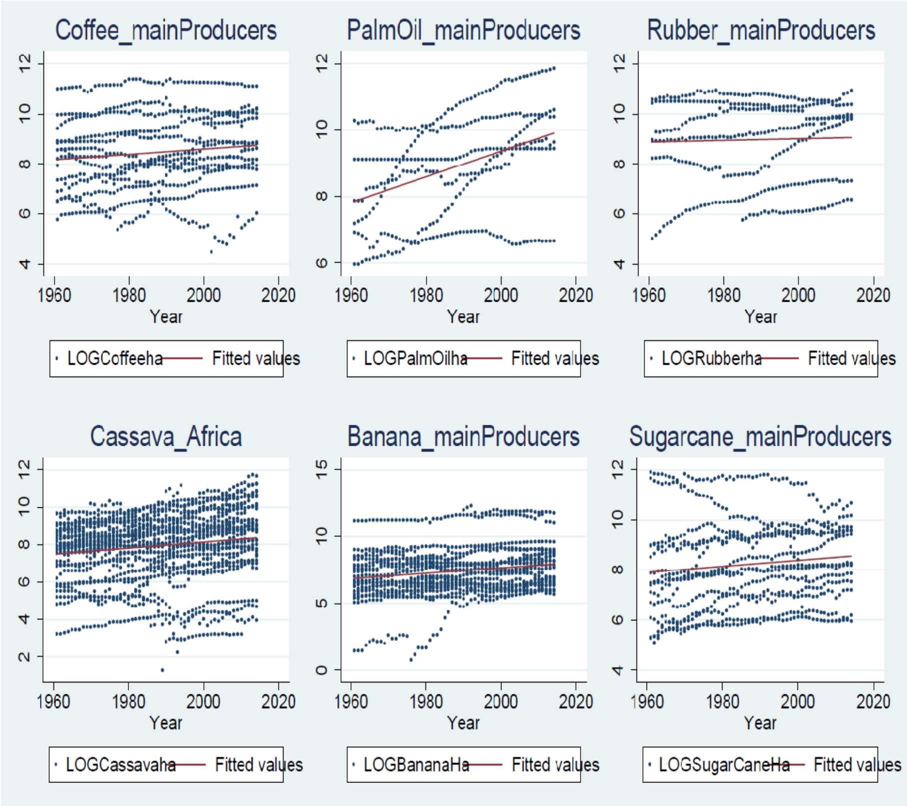
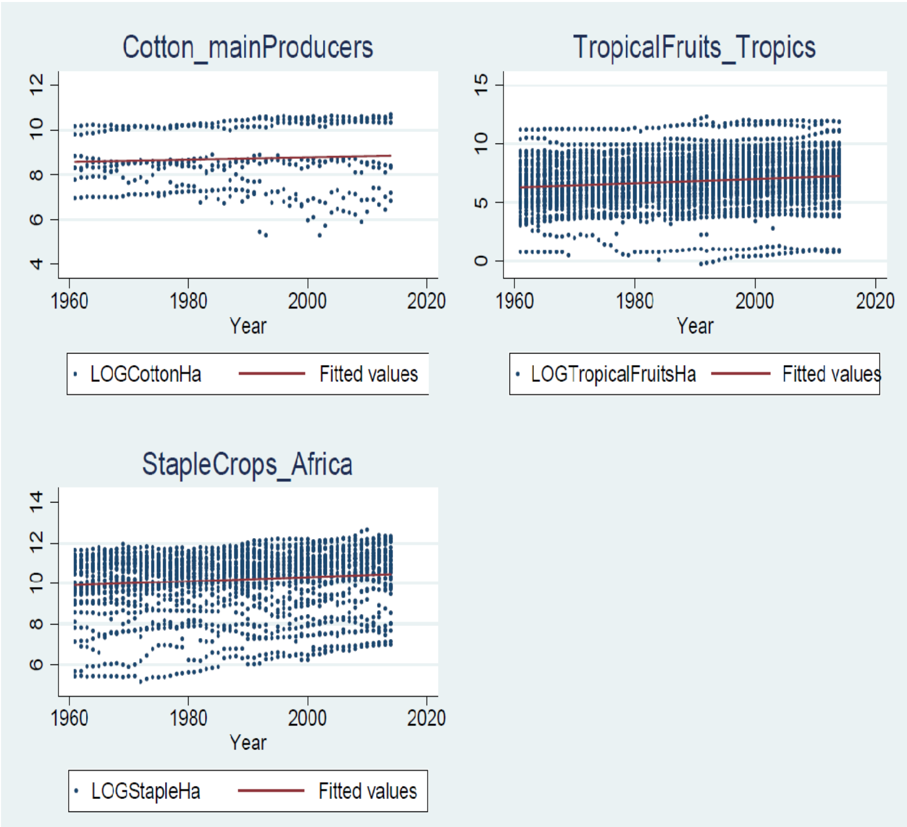


Fig. S12 Temporal evolution of cropland (in logarithm form) for cotton, tropical fruits and staple crops.



APPENDIX Chapter 3

Appendix A Variables' description

Table A1. Definition and source of variables.

Variable's name	Definition	Units	Source
Forest (%)	Coefficient between: i) Forest cover (ha) (the sum of tree-covered areas[24] and mangroves[25] categories) and ii) land area (ha) (country area excluding area under inland waters and coastal waters).	%	Land Cover CCI Product User Guide Version 2.0 (2017). Download it from http://www.fao.org/faostat/en/#data/LC. Reference: [14]
Agr land (%)	Coefficient between: i) Agricultural cover area defined as the sum of three CCI categories: herbaceous crops,[26] wood	%	Land Cover CCI Product User Guide Version 2.0 (2017). Download it from http://www.fao.org/faostat/en/#data/LC. Reference: [14]

[24] This category includes any geographical area dominated by natural tree plants with a cover of 10 per cent or more. Other types of plants (shrubs and/or herbs) can be present, even with a density higher than that of trees. Areas planted with trees for afforestation purposes and forest plantations are included in this class. This class includes areas seasonally or permanently flooded with freshwater but excludes coastal mangroves.

[25] This category includes any geographical area dominated by woody vegetation (trees and/or shrubs) with a cover of 10 per cent or more that is permanently or regularly flooded by salt and/or brackish water located in the coastal areas or in the deltas of rivers.

[26] This category is composed of a main layer of cultivated herbaceous plants (graminoids or forbs). It includes herbaceous crops used for hay. All the non-perennial crops that do not last for more than two growing seasons and crops like sugar cane, where the upper part of the plant is regularly harvested while the root system can remain for more than one year in the field, are included in this class.

	crops,[27] and grassland[28] and ii) land area.		
Rural pop density	Population living in rural areas over the land area of the country	Persons/ ha	The World Bank, https://databank.worldbank.org/data/source/world-development-indicators Reference: [35]
GDP cap and GDP2 cap	Gross domestic product divided by midyear population.	Constant U.S. dollars (converted from domestic currencies using 2010 official exchange rates).	The World Bank, https://databank.worldbank.org/data/source/world-development-indicators Reference: [35]
TFP[29]	The ratio of an output index (total amount of crop and livestock output) to an index of land and nonland inputs (all land, labor, capital and material resources employed in farm production).	Unitless (ratio of outputs and inputs expressed in monetary terms). The change rate uses 1961 as the baseline year.	United States Department of Agriculture, http://www.ers.usda.gov/data-products/international-agricultural-productivity.aspx Reference: [69]
Yield	Aggregate of all crops' harvested	Tonnes/ha.	FAOSTAT (FAO, 2018)

27 This category is composed of a main layer of permanent crops (trees or shrub crops) and includes all types of orchards and plantations (fruit trees, coffee and tea plantation, oil palms, rubber plantation, Christmas trees, etc.).

28 This category includes any geographical area dominated by natural herbaceous plants (grasslands, prairies, steppes and savannahs) with a cover of 10 per cent or more, irrespective of different human and/or animal activities, such as grazing or selective fire management. Woody plants (trees and/or shrubs) can be present, assuming their cover is less than 10 per cent.

29 To reduce potential index number bias in TFP growth estimates, cost shares are varied by decade whenever such information is available. For outputs, base year prices are fixed (the base period for output prices is 2004-06). Source: https://www.ers.usda.gov/data-products/international-agricultural-productivity/documentation-and-methods/

	production / harvested area for all crops.		http://www.fao.org/faostat/en/#data Reference: [11]
Agr (PIN)	Producer Price Index (2004-2016 = 100). It measures the average annual change over time in the selling prices received by farmers (prices at the farm gate or at the first point of sale).		FAOSTAT (FAO, 2018). http://www.fao.org/faostat/en/#data/PI Reference: [11]
Agri empl (%)	Agricultural employment: Employment in agriculture (% of total employment) (modeled ILO estimate).	%	The World Bank, http://databank.worldbank.org/data/ Reference: [35]
Agr prod net ex	Area of land embodied on exports of agricultural products (ha) minus imports agricultural products (ha).	ha	Own calculations using the data from [27].
For prod net ex	Exports of roundwood (m3) minus imports of roundwood (m3)	m3	FAOSTAT (FAO, 2018) Reference: [11]
Government quality	An average of the following four index: Control of corruption, Government Effectiveness, Regulatory quality, Rule of Law, Voice and Accountability.	Unitless	Worldwide Governance Indicators, www.govindicators.org Reference: [70]

Trade high	Exports of agricultural products send to high income countries divided by the total exports of agricultural products.	%	Own calculations using data from [27].
Avg Temperature	Average temperature per year (computed from monthly data)	mm	World Bank, https://climateknowledgeportal.worldbank.org/ Reference: [35]
Avg Rain	Average temperature per year (computed from monthly data)	Degrees Celsius	World Bank, https://climateknowledgeportal.worldbank.org/ Reference: [35]

Appendix B Countries' classification

We classified the countries included in our database into three income groups following the WB's classification (2018): low-income economies, with a maximum gross national income (GNI) per capita of US$ 1,005; middle income economies, with a GNI between US$ 1,006 and 12,235; high income economies, with a GNI equivalent or greater than US$ 12,236. Moreover, only countries with at least 1 million hectares of tree coverage in 2000 have been selected, in order to reduce heterogeneity in our sample. According to this classification, we have included the following countries in each income group:

- Low-income economies: Afghanistan, Benin, Burkina Faso, Central African Republic, Congo, Dem.Rep., Ethiopia, Guinea, Guinea Bissau, Kyrgyzstan, Liberia, Madagascar, Malawi, Mozambique, Nepal, Niger, Sierra Leone, Tanzania, Togo, Uganda, Zimbabwe

- Middle-income economies: Angola, Argentina, Azerbaijan, Bangladesh, Belarus, Belize, Bhutan, Bolivia, Bosnia and Herzegovina, Brazil, Bulgaria, Cambodia, Cameroon, China, Colombia, Congo, Rep., Costa Rica, Cote d'Ivoire, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Fiji, Gabon, Georgia, Ghana, Guatemala, Guyana, Honduras, India, Indonesia, Iran, Islamic Rep., Kazakhstan, Kenya, Lao PDR, Malaysia, Mexico, Morocco, Myanmar, Namibia, Nicaragua, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Romania, Russian Federation, Senegal, South Africa, Sri Lanka, Sudan, Suriname, Thailand, Ukraine, Venezuela, RB, Vietnam, Zambia.
- High-income economies: Australia, Austria, Canada, Chile, Croatia, Czech Republic, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Japan, Korea, Rep., Latvia, Lithuania, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, Uruguay

Countries' classification according to the forest transition phases

We adopted Pendrill's classification [26] to cluster countries according to their forest transition pattern (see Figure A1). Countries are divided into four categories based on the following variables: percentage of total forest coverage, net gross forest loss, and net forest cover area. More specifically, early transition countries are those with a percentage of forest greater than 50% and with net forest cover greater than -0.2%; pre-transition countries are those with a net forest cover lower than -0.2% and a positive net gross forest loss; late transition countries still have a net forest cover lower than -0.2%, but also a negative gross forest loss or a positive gross forest loss juxtapose to a forest coverage lower than 15%; eventually, post transition countries are those with a net forest cover greater than -0.2% and a forest coverage lower than 50% or a positive net forest cover. Some few countries that are missing in the Pendrill et al. (2019)'s classification, have been classified based on either the Hosonuma et al. (2012) [71] classification and/or their historical trend in forest cover. Due to sample consistency we decided to merge the two categories of early and pre-forest transition into a cathegory called "pre-transtion". According to this classification, we have included the following countries in each income group:

- Pre-transition: Angola, Belize, Bolivia, Cambodia, Cameroon, Congo, Dem. Rep., Congo, Rep., Equatorial Guinea, Guinea, Guinea-Bissau, Guyana, Honduras, Indonesia, Liberia, Madagascar, Malawi, Myanmar, Papua New Guinea, Paraguay, Peru, Senegal, Sri Lanka, Suriname, Tanzania, Zambia.
- Late transition countries: Afghanistan, Argentina, Australia, Benin, Brazil, Burkina Faso, Colombia, Ecuador, El Salvador, Ethiopia, Guatemala, Mozambique, Namibia, Nicaragua, Niger, Nigeria, Pakistan, Panama, Uganda, Venezuela, RB, Zimbabwe.
- Post transition countries: Austria, Azerbaijan, Bangladesh, Belarus, Bhutan, Bosnia and Herzegovina, Bulgaria, Canada, Central African Republic, Chile, China, Costa Rica, Cote d'Ivoire, Croatia, Cuba, Czech Republic, Dominican Republic, Estonia, Fiji, Finland, France, Gabon, Georgia, Germany, Ghana, Greece, Hungary, India, Iran, Islamic Rep., Italy, Japan, Kazakhstan, Kenya, Korea, Rep., Kyrgyzstan, Lao PDR, Latvia, Lithuania, Malaysia, Mexico,

Morocco, Nepal, New Zealand, Norway, Philippines, Poland, Portugal, Romania, Russian Federation, Sierra Leone, Slovak Republic, Slovenia, South Africa, Spain, Sudan, Sweden, Switzerland, Thailand, Togo, Turkey, Ukraine, United Kingdom, United States, Uruguay, Vietnam.

Eventually, in both classifications (income and forest transition clusters) the following countries have been excluded due to a general lack of data: Democratic Republic of Korea, Montenegro, Serbia, Serbia and Montenegro, and Somalia.

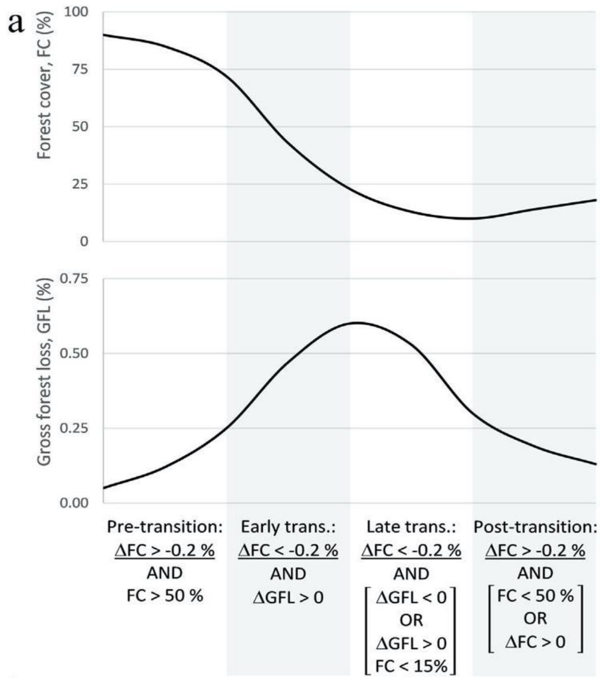


Figure 1. Forest transition countries' classification. Source: Pendrill et al. [26] (p.6).

Appendix C. Model with control variables in the long run

Table A2. Long run, income clusters, net agricultural exports as additional control variable.

Income clusters: Long run equations

	(1) Low	(2) Low	(3) Middle	(4) Middle	(5) High	(6) High
GDP cap (log)	.199	-.383	-.061	-.018	-.56	.023
GDP2 cap (log)	-.014	.03	.004	.001	.033	-.002
Rural pop (log)	.009	-.138	.019	-.035	-.085	-.066
Agr Land (log)	-1.54		-.243		-.611	
Agr prod net ex (ha)	.000	.000	.000	.000	.000	.000
Cons	-3.765	-.282	-1.093	-.985	.484	-1.161
Observations	390	390	1216	1216	583	583
R-squared	.562	.053	.18	.028	.411	.107
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Note: The long run coefficient estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated on the tables.

Table A3. Short run, income clusters, net agricultural exports as additional control variable.

Income clusters: Short run equations

	(1) Low	(2) Low	(3) Middle	(4) Middle	(5) High	(6) High
LD.GDP cap (log)	.127	.222	-.014	.008	.013	.048
	(.16)	(.136)	(.041)	(.04)	(.113)	(.105)
LD.GDP2 cap (log)	-.011	-.018	.001	0	0	-.001
	(.014)	(.011)	(.003)	(.002)	(.006)	(.005)
LD.Rural pop (log)	.081	.124	-.038	-.044	.018	.005
	(.199)	(.167)	(.031)	(.03)	(.03)	(.027)
LD.Agr Land (log)	-.527***		-.01		-.113**	
	(.113)		(.013)		(.049)	
LD.Agr empl	0	-.001	0	0	0	0
	(.001)	(.001)	(0)	(0)	(0)	(0)
D.Agr prod ex. (ln)	-.006	-.001	0	0	.006	.007
	(.014)	(.012)	(.001)	(.001)	(.006)	(.005)
LD.Agr prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.For prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.TFP	-.719	3.189	-.23	-.603	.488	.31
	(2.975)	(2.52)	(.505)	(.498)	(.741)	(.694)
LD.Agr (PIN) (log)	-.018	-.004	.007**	.007**	.002	.001

	(.018)	(.015)	(.003)	(.003)	(.006)	(.006)
LD.Yield (log)	-.006	-.017	0	0	.006	.006*
	(.013)	(.011)	(.002)	(.002)	(.003)	(.003)
LD.Trade high	0	-.001	0**	0**	0	0**
	(.001)	(.001)	(0)	(0)	(0)	(0)
LD.Government	-.016	-.01	-.002	-.005	-.006	-.007
	(.021)	(.017)	(.003)	(.003)	(.008)	(.008)
DL(Gov/L.For)	-.006	-.003	.001*	0	-.002	-.003
	(.01)	(.008)	(0)	(0)	(.005)	(.005)
LD.Temperature	-.003	-.003	-.001*	-.001*	0	0
	(.003)	(.002)	(0)	(0)	(0)	(0)
LD.Rainfall	0**	0*	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
L	-.324***	-.373***	-.128***	-.125***	-.087***	-.123***
	(.04)	(.025)	(.011)	(.01)	(.018)	(.015)
Cons	.035***	.018***	-.003***	-.003***	-.007***	-.007***
	(.006)	(.004)	(0)	(0)	(.001)	(.001)
Observations	226	226	771	771	364	364
R-squared	.389	.563	.191	.211	.131	.216

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table A4: Long run, forest transition clusters, net agricultural exports as addition to control variables.

Forest transition clusters: Long run equations

	(1) Early_& _Pre	(2) Early_& _Pre	(3) Late	(4) Late	(5) Post	(6) Post
GDP cap (log)	-.123	-.102	.069	.082	.048	.000
GDP2 cap (log)	.008	.007	-.003	-.005	-.004	-.007
Rural pop (log)	-.045	-.04	.123	.052	.017	-.008
Agr Land (log)	-.243		-.481		-.502	
Agr prod net ex (ha)	.000	.000	.000	.000	.000	.000
Cons	-.592	-.183	-2.14	-1.658	-1.769	-1.393
Observations	499	499	410	410	1280	1280
R-squared	.277	.177	.424	.096	.178	.027
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Note: The long run coefficient estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated on the tables.

Table A5. Short run, forest transition clusters, net agricultural exports as additional control variable.

Forest transition clusters: Short run equations

	(1) Early_& _Pre	(2) Early_& _Pre	(3) Late	(4) Late	(5) Post	(6) Post
LD.GDP cap (log)	.064	.078	-.084	-.046	-.132***	-.144***
	(.056)	(.056)	(.058)	(.053)	(.036)	(.036)
LD.GDP 2 cap (log)	-.004	-.004	.005	.002	.008***	.009***
	(.004)	(.004)	(.004)	(.003)	(.002)	(.002)
LD.Rural pop (log)	-.158	-.14	.027	.033	.012	-.007
	(.166)	(.165)	(.047)	(.042)	(.026)	(.026)
LD.Agr Land (log)	-.06*		-.04		-.086**	
	(.032)		(.037)		(.043)	
LD.Agr empl	0	0	-.001**	-.001	0	0
	(.001)	(.001)	(0)	(0)	(0)	(0)
D.Agr prod ex. (ln)	.005	.007	-.001	0	0	0
	(.009)	(.009)	(.006)	(.005)	(.001)	(.001)
LD.Agr prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.For prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.TFP	1.667	.068	-.202	-.059	.034	-.11
	(1.664)	(1.641)	(1.144)	(1.024)	(.549)	(.54)
LD.Agr (PIN) (log)	.013	.012	.001	.003	.005	.006
	(.011)	(.011)	(.007)	(.006)	(.004)	(.004)

LD.Yield (log)	-.009	-.011	-.004	-.004	.001	0
	(.01)	(.01)	(.005)	(.004)	(.002)	(.002)
LD.Trade high	0	0	0	0	0***	0***
	(0)	(0)	(0)	(0)	(0)	(0)
LD.Government	-.018	-.026**	.004	.003	-.001	-.003
	(.011)	(.011)	(.009)	(.008)	(.004)	(.004)
DL(Gov/ L.For)	.002*	-.001	.001	.001	0	-.001
	(.001)	(.001)	(.006)	(.005)	(.002)	(.002)
LD.Temperature	-.002	-.001	-.002*	-.002*	0	0
	(.002)	(.002)	(.001)	(.001)	(0)	(0)
LD.Rainfall	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
L	-.314***	-.302***	-.201***	-.206***	-.171***	-.165***
	(.029)	(.027)	(.024)	(.017)	(.01)	(.009)
Cons	-.006**	-.011***	.021***	.023***	-.006***	-.007***
	(.003)	(.003)	(.003)	(.002)	(0)	(0)
Observations	307	307	254	254	800	800
R-squared	.336	.338	.304	.433	.3	.32

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table A6. Long run, income clusters, net agricultural exports and net exports of forests products as additional control variables.

	(1) Low	(2) Low	(3) Middle	(4) Middle	(5) High	(6) High
GDP cap (log)	.171	-.422	-.063	-.02	-.495	.061
GDP2 cap (log)	-.012	.034	.005	.001	.028	-.006
Rural pop (log)	-.009	-.136	.019	-.035	-.068	-.048
Agr Land (log)	-1.51		-.244		-.579	
Agr prod ex. (ln)	.003	.009	0	0	.079	.09
_cons	-3.698	-.293	-1.084	-.973	-.813	-2.528
Observations	390	390	1216	1216	583	583
R-squared	.562	.06	.179	.026	.524	.253
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Table A7. Short run, income clusters, net agricultural exports and net exports of forests products as additional control variables.

	(1) Low	(2) Low	(3) Middle	(4) Middle	(5) High	(6) High
LD.GDP cap (log)	.122	.237*	-.014	.007	.023	.05
	(.159)	(.137)	(.041)	(.041)	(.113)	(.105)
LD.GDP 2 cap (log)	-.01	-.02*	.001	0	-.001	-.002
	(.013)	(.011)	(.003)	(.002)	(.006)	(.005)
LD.Rural pop (log)	.058	.178	-.04	-.046	.021	.006
	(.199)	(.168)	(.031)	(.03)	(.03)	(.027)
LD.Agr Land (log)	-.523***		-.01		-.109**	
	(.113)		(.013)		(.049)	
LD.Agr empl	0	-.001	0	0	0	0
	(.001)	(.001)	(0)	(0)	(0)	(0)
D.Agr prod ex. (ln)	-.005	-.004	0	0	.008	.01*
	(.014)	(.012)	(.001)	(.001)	(.006)	(.005)
LD.Agr prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.For prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.TFP	-.807	3.61	-.233	-.606	.469	.198
	(2.962)	(2.54)	(.505)	(.499)	(.737)	(.689)
LD.Agr (PIN) (log)	-.018	-.005	.007**	.007**	0	-.002
	(.018)	(.015)	(.003)	(.003)	(.006)	(.006)
LD.Yiel	-.006	-.015	0	-.001	.006*	.006**

d (log)						
	(.013)	(.011)	(.002)	(.002)	(.003)	(.003)
	0	-.001	0**	0**	0	0*
LD.Trade high						
	(.001)	(.001)	(0)	(0)	(0)	(0)
	-.015	-.012	-.002	-.005	-.007	-.01
LD.Government						
	(.021)	(.018)	(.003)	(.003)	(.008)	(.008)
	-.005	-.004	.001*	0	-.003	-.005
DL(Gov/ L.For)						
	(.01)	(.009)	(0)	(0)	(.005)	(.005)
	-.003	-.003	-.001*	-.001*	0	0
LD.Temperature						
	(.003)	(.002)	(0)	(0)	(0)	(0)
	0**	0*	0	0	0	0
LD.Rainfall						
	(0)	(0)	(0)	(0)	(0)	(0)
L	-.326***	-.372***	-.127***	-.125***	-.104***	-.142***
	(.039)	(.025)	(.011)	(.01)	(.021)	(.016)
_cons	.036***	.015***	-.003***	-.003***	-.009***	-.009***
	(.006)	(.004)	(0)	(0)	(.002)	(.001)
	226	226	771	771	364	364
Observations						
R-squared	.395	.558	.19	.209	.137	.228

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table A8. Long run, forest transition clusters, net agricultural exports and net exports of forests products as additional control variables.

	(1) Early_& _Pre	(2) Early_& _Pre	(3) Late	(4) Late	(5) Post	(6) Post
GDP cap (log)	.005	.004	.066	.079	.049	.084

GDP2 cap (log)	0	0	-.003	-.005	-.004	-.007
Rural pop (log)	-.061	-.061	.126	.059	.017	-.008
Agr Land (log)	-.279		-.489		-.502	
Agr prod ex. (ln)	-.078	-.049	.003	.005	.001	.001
_cons	-.066	.069	-2.176	-1.693	-1.778	-1.401
Observations	499	499	410	410	1280	1280
R-squared	.319	.193	.423	.073	.178	.027
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Table A9. Short run, forest transition clusters, net agricultural exports and net exports of forests products as additional control variables.

	(1) Early_&_Pre	(2) Early_&_Pre	(3) Late	(4) Late	(5) Post	(6) Post
LD.GDP cap (log)	.026	.056	-.081	-.044	-.132***	-.144***
	(.056)	(.056)	(.058)	(.053)	(.036)	(.036)
LD.GDP 2 cap (log)	-.001	-.003	.004	.002	.008***	.009***
	(.004)	(.004)	(.004)	(.003)	(.002)	(.002)
LD.Rural pop (log)	-.013	-.044	.016	.013	.012	-.007
	(.166)	(.165)	(.047)	(.043)	(.026)	(.025)
LD.Agr Land (log)	-.061*		-.041		-.086**	
	(.032)		(.037)		(.043)	
LD.Agr empl	0	0	-.001**	0	0	0

	(.001)	(.001)	(0)	(0)	(0)	(0)
D.Agr prod ex. (ln)	-.005	.001	0	.002	0	0
	(.009)	(.009)	(.006)	(.005)	(.001)	(.001)
LD.Agr prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.For prod net ex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
LD.TFP	2.084	.232	-.179	.016	.037	-.106
	(1.68)	(1.646)	(1.143)	(1.033)	(.549)	(.539)
LD.Agr (PIN) (log)	.019*	.015	.001	.004	.005	.006
	(.012)	(.011)	(.007)	(.006)	(.004)	(.004)
LD.Yield (log)	-.014	-.013	-.004	-.005	.001	0
	(.01)	(.01)	(.005)	(.004)	(.002)	(.002)
LD.Trade high	0	0	0	0	0***	0***
	(0)	(0)	(0)	(0)	(0)	(0)
LD.Government	-.012	-.023**	.004	.004	-.001	-.003
	(.011)	(.011)	(.009)	(.008)	(.004)	(.004)
DL(Gov/ L.For)	.002*	-.001	.001	.002	0	-.001
	(.001)	(.001)	(.006)	(.005)	(.002)	(.002)
LD.Temperature	-.002	-.001	-.002**	-.002*	0	0
	(.002)	(.002)	(.001)	(.001)	(0)	(0)
LD.Rainfall	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
L	-.318***	-.304***	-.2***	-.195***	-.171***	-.165***
	(.03)	(.027)	(.023)	(.017)	(.01)	(.009)
_cons	-.006**	-.011***	.021***	.022***	-.006***	-.007***
	(.003)	(.003)	(.003)	(.002)	(0)	(0)

Observations	307	307	254	254	800	800
R-squared	.328	.335	.305	.422	.301	.32

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Appendix D Long run causality test

The presence of cointegration is sufficient to prove the existence of at least one non-spurious long-run causal relationship between the variables. Yet, the estimated coefficients represent the net effect of a system of relationships and cannot be directly interpreted as unilateral causal effects (Granger causality) but rather as a systemic causality. These long run coefficients are superconsistent – i.e., the distribution of the estimated coefficients converges to the real value as the sample size tends to infinity [36] - but their standard errors are large and unreliable, and thus not robust, so that statistical inference is not straightforward [48,49] . We thus used the approach developed by Canning and Pedroni [48] to test for the direction of long-run causality in cointegrated panel data, extending their two-variables design to a five-variables model. Their test identifies the sign of the ratio of error correction coefficients under a transformed specification. Here, we exemplify the approach with the long and short run models previously described on the methodology section.
We first transformed the long and short-run equations (Equations A1-A4), to isolate λ_{14} and λ'_{14} (Equation A5). These parameters always have a negative sign as they are the ECT, bounded between 0 and -1. Thus, we find that the sign of the ratio $\frac{(\lambda_{14}}{\lambda'_{14})}$ is the opposite than that of the long-run coefficient β_1 (Equation A6). We replicated the same transformation to estimate the sign for β_2, β_3 and β_4. We tested the sign of the ratio using a Wald test with a chi-squared statistics at a significant level of 1%, 5% or 10% [44].

Long run equation:

$$LogForest_{i,t} = \propto_{1i} + \beta_1 LogGDPcap_{i,t} + \beta_2 LogGDPcap2_{i,t} + \beta_3 LogAgri_{i,t} + \beta_4 LogRuralPopDensity_{i,t} + \mu_{forest\ i,t}$$

(Eq. A1)

If we substitute $LogGDPcap_{i,t}$:

$$LogGDPcap_{i,t} = \frac{\propto'_{1i}}{-\beta_1} + \frac{1}{-\beta_1} LogForest_{i,t} + \frac{\beta_2}{-\beta_1} LogGDPcap2_{i,t} + \frac{\beta_3}{-\beta_1} LogAgri_{i,t} + \frac{\beta_4}{-\beta_1} LogRuralPopDensity_{i,t} + \frac{1}{-\beta_1} \mu_{forest\ i,t}$$

(Eq. A2)

Transformed short run equations:

$$\Delta LogForest_{i,t} = \lambda_1 \Delta LogGDP_{i,t-1} + \lambda_2 \Delta LogGDP2_{i,t-1} + \beta\lambda_3 \Delta LogAgri_{i,t-1} + \beta\lambda_4 \Delta LogRuralPopDensity_{i,t-1} + \lambda_5 \Delta log(AgriEmp)_{i,t-1} + \lambda_6 \Delta TFP_{i,t-1} + \lambda_7 \Delta LogYield_{i,t-1} + \lambda_8 \Delta LogCummulativeForest_{i,t-1} + \lambda_9 \Delta NetExportsAgri_{i,t-1} + \lambda_{10} \Delta TermsTrade_{i,t-1} + \lambda_{11} \Delta TradeGDP_{i,t-1} + \lambda_{12} \Delta Goverment_{i,t-1} + \lambda_{13} \Delta TradeHigh_{i,t-1} + \lambda_{14} \mu^{\wedge}{}_{forest\ i,t-1} + \varepsilon_{i,t}$$

(Eq. A3)

$$
\begin{aligned}
\Delta LogGDP_{i,t} = {} & \frac{1}{-\beta_1}\,\Delta LogForest_{i,t} + \frac{\lambda'_2}{-\beta_1}\Delta LogGDP2_{i,t-1} \\
& + \frac{\lambda'_3}{-\beta_1}\Delta LogAgri_{i,t-1} \\
& + \frac{\lambda'_4}{-\beta_1}\Delta LogRuralPopDensity_{i,t-1} \\
& + \frac{\lambda'_5}{-\beta_1}\Delta log(AgriEmp)_{i,t-1} + \frac{\lambda'_6}{-\beta_1}\Delta TFP_{i,t-1} \\
& + \frac{\lambda'_7}{-\beta_1}\Delta LogYield_{i,t-1} \\
& + \frac{\lambda'_8}{-\beta_1}\Delta LogCummulativeForest_{i,t-1} \\
& + \frac{\lambda'_9}{-\beta_1}\Delta NetExportsAgri_{i,t-1} \\
& + \frac{\lambda'_{10}}{-\beta_1}\Delta TermsTrade_{i,t-1} + \frac{\lambda'_{11}}{-\beta_1}\Delta TradeGDP_{i,t-1} \\
& + \frac{\lambda'_{12}}{-\beta_1}\Delta Goverment_{i,t-1} + \frac{\lambda'_{13}}{-\beta_1}\Delta TradeHigh_{i,t-1} \\
& + \frac{\lambda'_{14}}{-\beta_1}\mu\hat{}_{forest\ i,t-1} + \varepsilon_{i,t}
\end{aligned}
$$

(Eq. A4)

Ratio of the coefficients must be non-zero if a long run relationship is to hold:

$$\frac{\lambda_{14}}{\frac{\lambda'_{14}}{-\beta_1}} = \left(\frac{\lambda_{14}}{\lambda'_{14}}\right)\cdot(-\beta_1)$$

(Eq. A5)

λ_{14} and λ'_{14} always have negative sign (they are the EC terms). Thus, the sign of the ratio of the short-term coefficients ($\frac{\lambda_{14}}{\lambda'_{14}}$) is opposite of each of the long run coefficient (β_1).

$$sign\left(\frac{\lambda_{14}}{\lambda'_{14}}\right) = -sign(\beta_1)$$

Long run results using causality test:

Their approach has proven to be useful in other studies [13]. Nevertheless, we perform this test for our long run coefficients (see Tables S2 and S3) and we obtained inconclusive results and very few significant results. Thus, we decided not to include those results on our analysis and using the estimators only for our sample. The long run coefficients and their signs are unbiased for our sample, which include most forested countries, and can thus be used to draw conclusions for that specific sample, without making inference beyond.

Table A10. Long run results, income clusters.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Low	Low	Middle	Middle	High	High
GDP cap (log)	.168	-.357	**(+)-0.071**	**(+)-0.030**	-.406	.072
GDP2 cap (log)	-.013	.029	.005	**0.002**	.024	-.004
Rural pop (log)	-.025	-.108	.02	-.031	**(+)-.073**	-.072
Agr Land (log)	-1.357		-.258		**(+)-0.515**	
Constant	-3.461 (.417)	-.352 (.525)	-1.081 (.105)	-.937 (.114)	-.014 (.407)	-1.474 (.462)
Observations	470	470	1455	1455	703	703
R-squared	.489	.042	.187	.027	.36	.126
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Table A11. Long run results, forest transition clusters.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Pre	Pre	Late	Late	Post	Post
GDP cap (log)	-.136	-.087	.073	.095	.047	.089
GDP2 cap (log)	.009	.006	-.003	-.006	-.003	-.006
Rural pop (log)	-.019	**(+)-.024**	.108	.051	.014	-.015
Agr Land (log)	-.244		-.475		**(+)-.472**	
_cons	-.492 (.176)	-.207 (.184)	-2.182 (.155)	-1.728 (.188)	-1.76 (.096)	-1.459 (.102)

Observations	599	599	493	493	1536	1536
R-squared	.283	.181	.382	.062	.168	.019
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: The long run coefficient estimators are unbiased for the sample but not robust and thus cannot be directly used for statistical inference. For that reason, the significance is not indicated on the tables. Instead, the coefficients in bold letters indicate that the test based on the approach from Canning and Pedroni [48] is significant . The sign of the coefficient can thus be interpreted as statistically significant. If brackets are present, those indicate an inconsistency between the sign obtained from the regression and the one pointed out by the test (which is indicated within brackets). In these cases, the result was ambiguous.

Appendix E Assessment of indirect effects in the short run.

The objective of this procedure is to assess the indirect impacts of some of the independent variables on the dependent one, forest cover, through other the others independent variables ("mediator"). Thus, we want to identify which part of the explanatory power of each variable came from the indirect impacts of other variables. For that, we perform the following steps for for each variable of fuccus that we would like to asses. We show it thanks the following example with only two independent variables. This five steps process is as follows:

1) We specify our basic model (direct effects):

$$Y = \lambda_1 X_1 + \lambda_2 X_2 + \mu$$

2) We identify the indirect effects of the independent variables according to our causal diagram. In this example, we assume that X_2 has an indirect effect on Y through the variable X_1.

3)

$$X_2 = \quad {}_1 X_1 + residual_{x2}$$

4) We substitute X_2 in Y:

$$Y = \lambda_1 X_1 + \lambda_2 (\quad {}_1 X_1 + residual_{x2}) + \mu$$

$$Y = (\lambda_1 + \lambda_2 \quad {}_1) X_1 + \lambda_2 residual_{x2} + \mu$$

5) We compare λ_1 with $\lambda_1 + \lambda_2 \lambda_1$.

Note 1: $residual_{x2}$ and X_1 are orthogonal as it is assumed by the OLS method. Note 2: the error terms (μ) in both, equation on step 1 and equation in step 2, are equal as both equations are linear combinations of the same matrix.

Here we show the auxiliary equations and auxiliary figures[30] for each independent variable of focuss:

Estimating total effect of "GDP" variable (λ'_1):

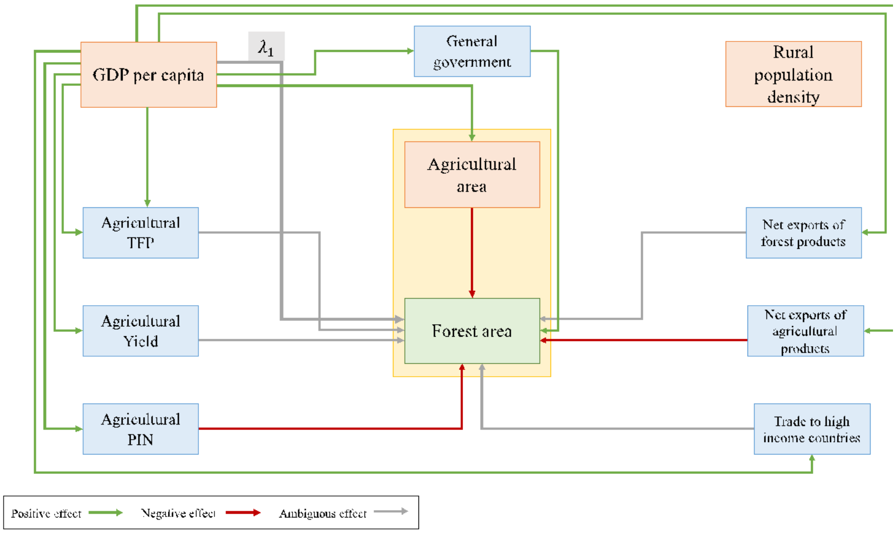
$\Delta LogAgri_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{Agri}$
$\Delta LogYield_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{Yield}$
$\Delta TFP_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{TFP}$
$\Delta Goverment_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{Goverment}$

$\Delta NetExportForest_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{ExportFo}$
$\Delta NetExportsAgri_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{ExportAgri}$
$\Delta TradeHigh_{i,t-1} = \Delta LogGDP_{i,t-1} + residual_{TradeHigh}$
$\Delta LogAgri(PIN) = \Delta LogGDP_{i,t-1} + residual_{AgriPIN}$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta LogGDP_{i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 residual_{Agri_{i,t-1}} + \lambda'_4 residual_{ExportAgri_{i,t-1}} \\ & + \lambda'_5 residual_{ExportFo_{i,t-1}} + \lambda'_6 residual_{TFP\,i,t-1} \\ & + \lambda'_7 residual_{Yield\,i,t-1} + \lambda'_8 residual_{AgriPIN} \\ & + \lambda'_9 residual_{TradeHigh_{i,t-1}} + \lambda'_{10} residual_{Goverment\,i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu^\wedge{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A2.

[30] The graphs only show indirect effects of the variable of study on forest through the ones to which it has a direct effect

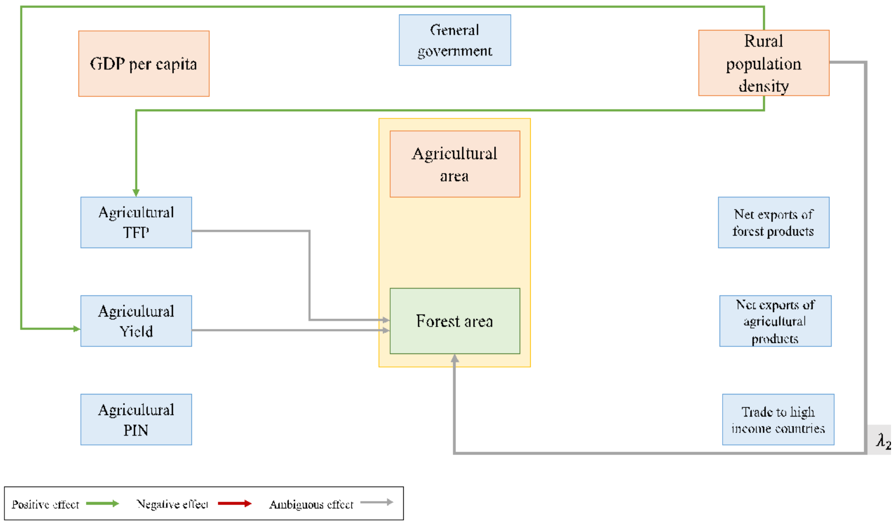


<u>Estimating total effect of "RurPopDensity" variable (λ'_2)</u>

$$\Delta TFP_{i,t-1} = \Delta LogRuralPopDensity_{i,t-1} + residual_{TFP}$$
$$\Delta LogYield_{i,t-1} = \Delta LogRuralPopDensity_{i,t-1} + residual_{Yield}$$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta LogGDP_{i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 \Delta LogAgri_{i,t-1} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 residual_{TFP\,i,t-1} \\ & + \lambda'_7 residual_{Yield\,i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu\hat{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A3.

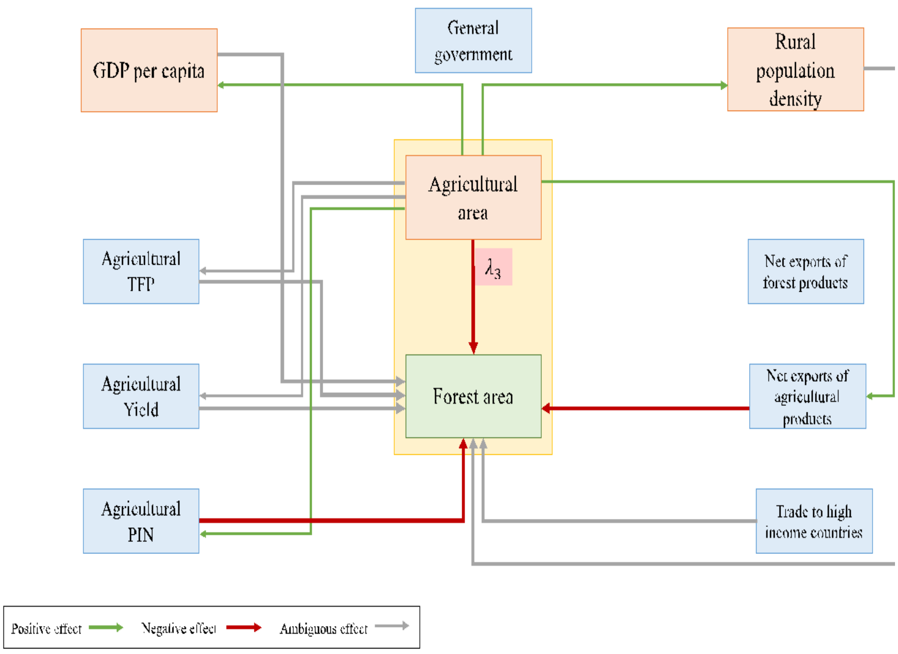


Estimating total effect "Agri" variable (λ'_3):

$\Delta LogGDP_{i,t-1} = \Delta LogAgri_{i,t-1} + residual_{GDP}$
$\Delta TFP_{i,t-1} = \Delta LogAgri_{i,t-1} + residual_{TFP}$
$\Delta LogYield_{i,t-1} = \Delta LogAgri_{i,t-1} + residual_{Yield}$
$\Delta LogRuralPopDensity_{i,t-1} = \Delta LogAgri_{i,t-1} + residual_{ruralpopde}$
$\Delta NetExportsAgri_{i,t-1} = \Delta LogAgri_{i,t-1} + residual_{ExAgri}$
$\Delta LogAgri(PIN)_{i,t-1} = \Delta LogAgri_{i,t-1} + residual_{AgriPIN}$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 residual_{GDP\,i,t-1} + \lambda'_2 \Delta residual_{ruralpopde\,i,t-1} \\ & + \lambda'_3 \Delta LogAgri_{i,t-1} + \lambda'_4 residual_{ExAgri\,i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 residual_{TFP\,i,t-1} \\ & + \lambda'_7 residual_{Yield\,i,t-1} + \lambda'_8 residual_{AgriPIN\,i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu^{\wedge}{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

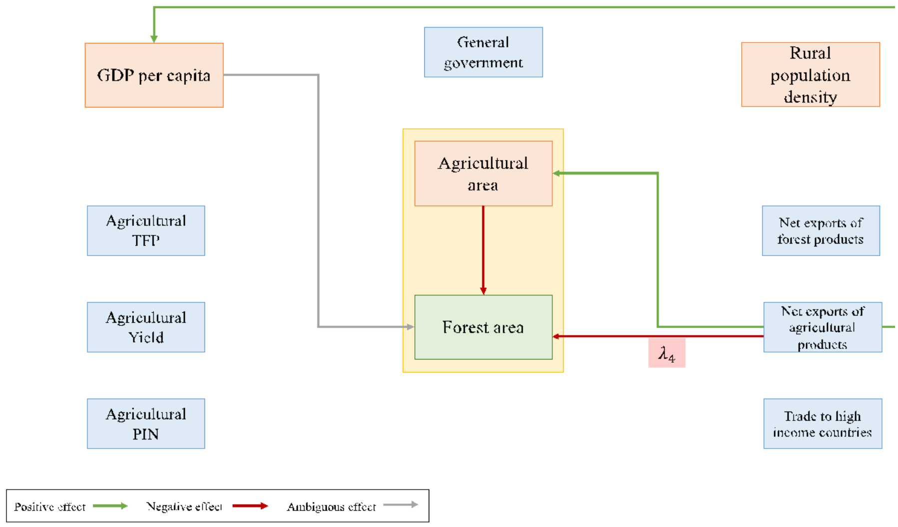
Figure A4.



<u>Estimating total effect "Net exports Agri" variable (λ'_4):</u>

$$\Delta LogAgri_{i,t-1} = \Delta NetExportsAgri_{i,t-1} + residual_{Agri}$$
$$\Delta LogGDP_{i,t-1} = \Delta NetExportsAgri_{i,t-1} + residual_{GDP}$$
$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 \Delta residual_{Agri_{i,t-1}} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 \Delta TFP_{i,t-1} \\ & + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu^{\wedge}{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A5.

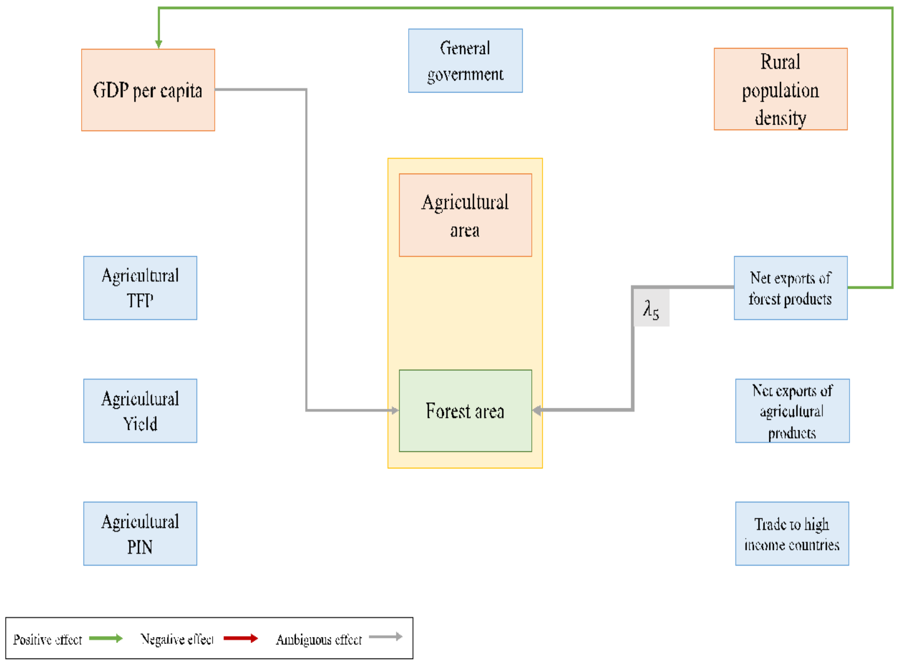


Total effect "Net exports forest" variable (λ'_5):

$$\Delta LogGDP_{i,t-1} = \Delta NetExportForest_{i,t-1} + residual_{GDP}$$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 \Delta LogAgri_{i,t-1} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 \Delta TFP_{i,t-1} \\ & + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu\hat{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A6

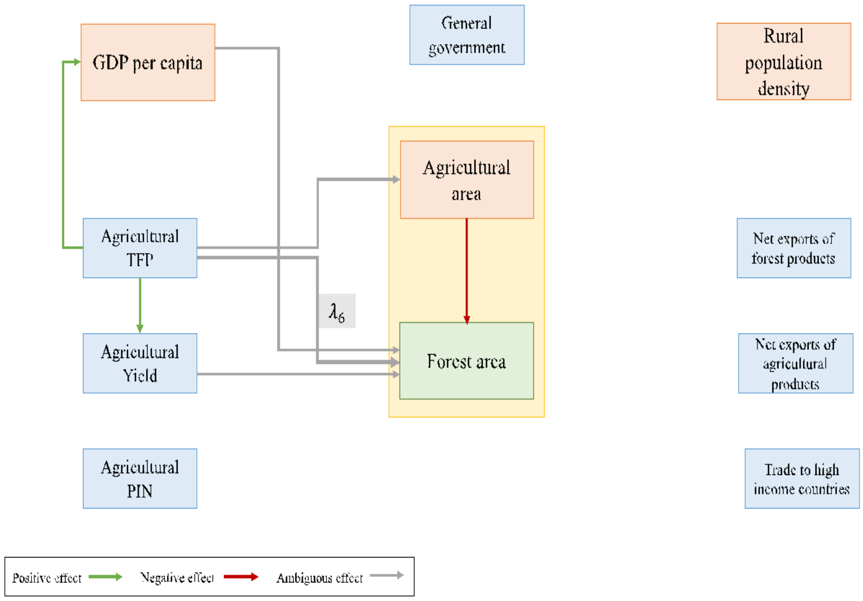


<u>Estimating total effect of “TFP” variable (λ'_6):</u>

$\Delta LogGDP_{i,t-1} = \Delta TFP_{i,t-1} + residual_{GDP}$
$\Delta LogYield_{i,t-1} = \Delta TFP_{i,t-1} residual_{Yield}$
$\Delta LogAgri_{i,t-1} = \Delta TFP_{i,t-1} + residual_{Agri}$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 \Delta residual_{Agri_{i,t-1}} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 \Delta TFP_{i,t-1} \\ & + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \\ & + \lambda'_7 residual_{Yield\,i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu\hat{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A7:



<u>Estimating total effect of "yield" variable (λ'_7):</u>

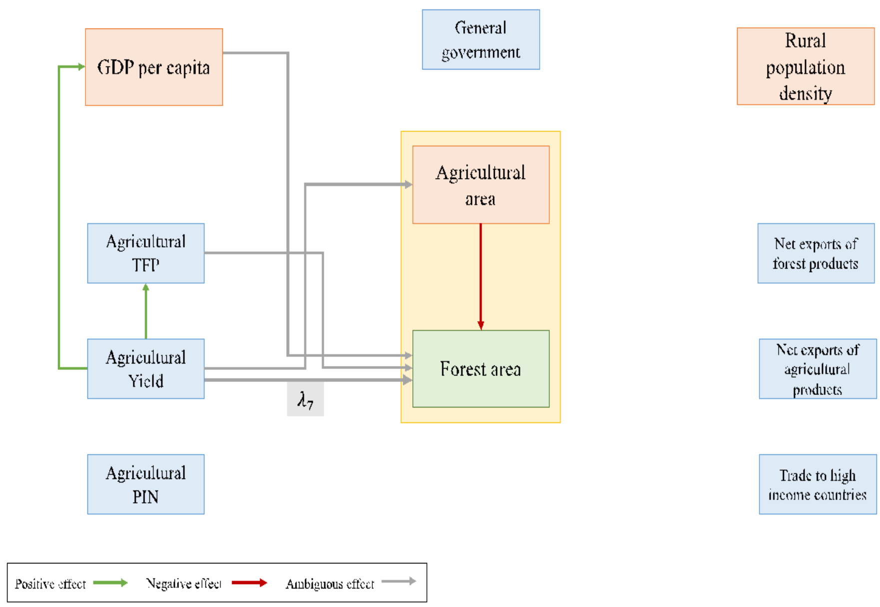
$$\Delta LogGDP_{i,t-1} = \Delta LogYield_{i,t-1} + residual_{GDP}$$

$$\Delta TFP_{i,t-1} = \Delta LogYield_{i,t-1} residual_{TFP}$$

$$\Delta LogAgri_{i,t-1} = \Delta LogYield_{i,t-1} + residual_{agri}$$

$$
\begin{aligned}
\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\
& + \lambda'_3 \Delta residual_{Agri\,i,t-1} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\
& + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 residual_{TFP\,i,t-1} \\
& + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\
& + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\
& + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\
& + \lambda'_{13} \mu^{\wedge}{}_{i,t-1} + \varepsilon_{i,t}
\end{aligned}
$$

Figure A8

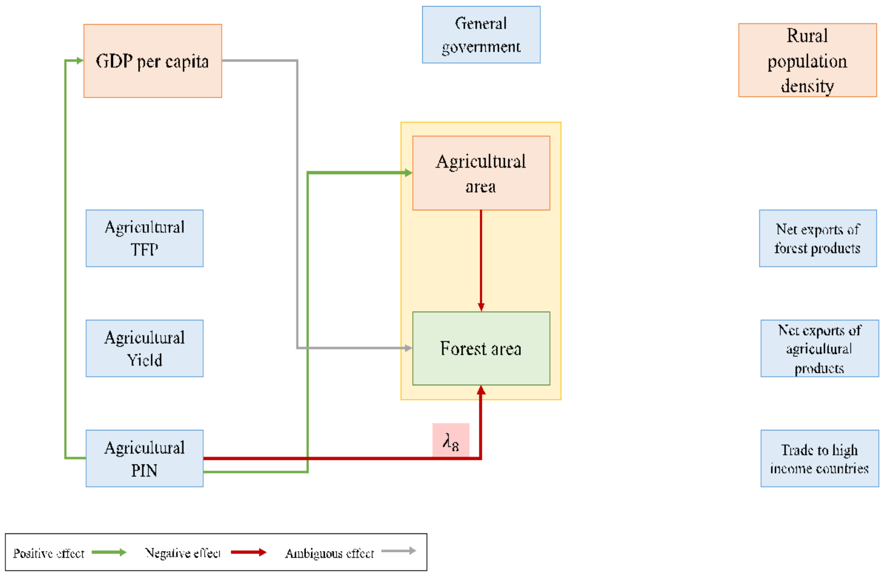


<u>Estimating total effect of "AgriPIN" variable (λ'_8)</u>

$$\Delta LogGDP_{i,t-1} = \Delta LogAgri(PIN)_{i,t-1} + residual_{GDP}$$
$$\Delta LogAgri_{i,t-1} = \Delta LogAgri(PIN)_{i,t-1} + residual_{Agri}$$

$$
\begin{aligned}
\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\
& + \lambda'_3 \Delta residual_{Agri_{i,t-1}} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\
& + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 \Delta TFP_{i,t-1} \\
& + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \\
& + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\
& + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\
& + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\
& + \lambda'_{13} \mu\hat{}_{i,t-1} + \varepsilon_{i,t}
\end{aligned}
$$

Figure A9:



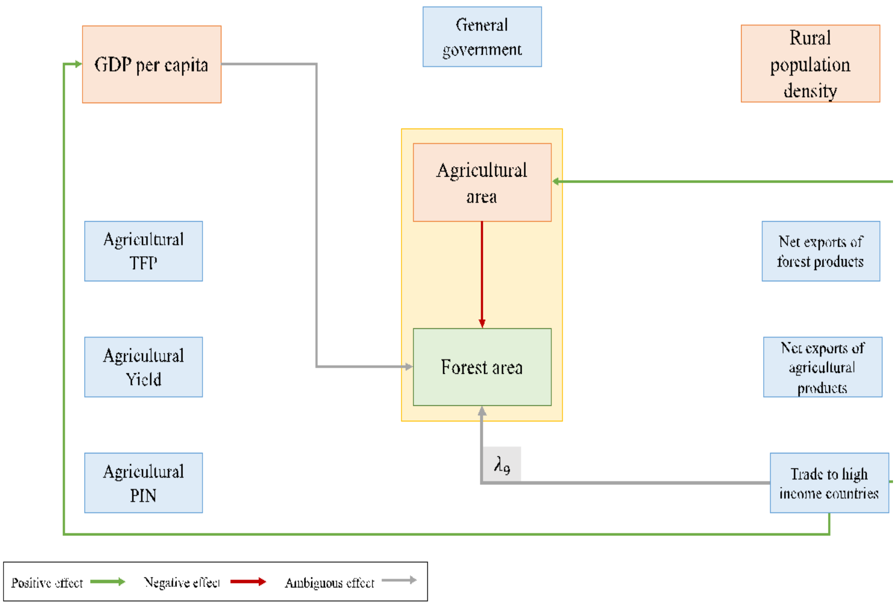
<u>Estimating total effect of "trade high" variable (λ'_9)</u>

$$\Delta LogGDP_{i,t-1} = \Delta TradeHigh_{i,t-1} + residual_{GDP}$$

$$\Delta LogAgri_{i,t-1} = \Delta TradeHigh_{i,t-1} + residual_{Agri}$$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 \Delta residual_{Agri_{i,t-1}} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 \Delta TFP_{i,t-1} \\ & + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \\ & + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu\hat{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A10:

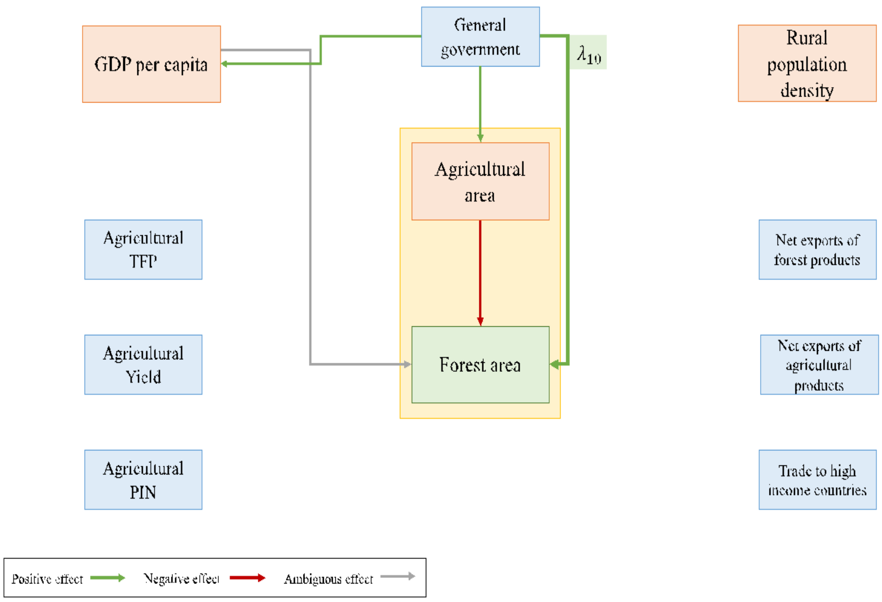


Estimating total effect "Government" variable (λ'_{10})

$\Delta LogGDP_{i,t-1} = \Delta Goverment_{i,t-1} + residual_{GDP}$
$\Delta LogAgri_{i,t-1} = \Delta Goverment_{i,t-1} + residual_{Agri}$

$$\begin{aligned}\Delta LogForest_{i,t} = {} & \lambda'_1 \Delta residual_{GDP\,i,t-1} + \lambda'_2 \Delta LogRuralPopDensity_{i,t-1} \\ & + \lambda'_3 \Delta LogAgri_{i,t-1} + \lambda'_4 \Delta NetExportsAgri_{i,t-1} \\ & + \lambda'_5 \Delta NetExportForest_{i,t-1} + \lambda'_6 \Delta TFP_{i,t-1} \\ & + \lambda'_7 \Delta LogYield_{i,t-1} + \lambda'_8 \Delta LogAgri(PIN)_{i,t-1} \\ & + \lambda'_9 \Delta TradeHigh_{i,t-1} + \lambda'_{10} \Delta Goverment_{i,t-1} \\ & + \lambda'_{11} \Delta Temperature(avg) + \lambda'_{12} \Delta Rainfall(avg) \\ & + \lambda'_{13} \mu\hat{}_{i,t-1} + \varepsilon_{i,t}\end{aligned}$$

Figure A11

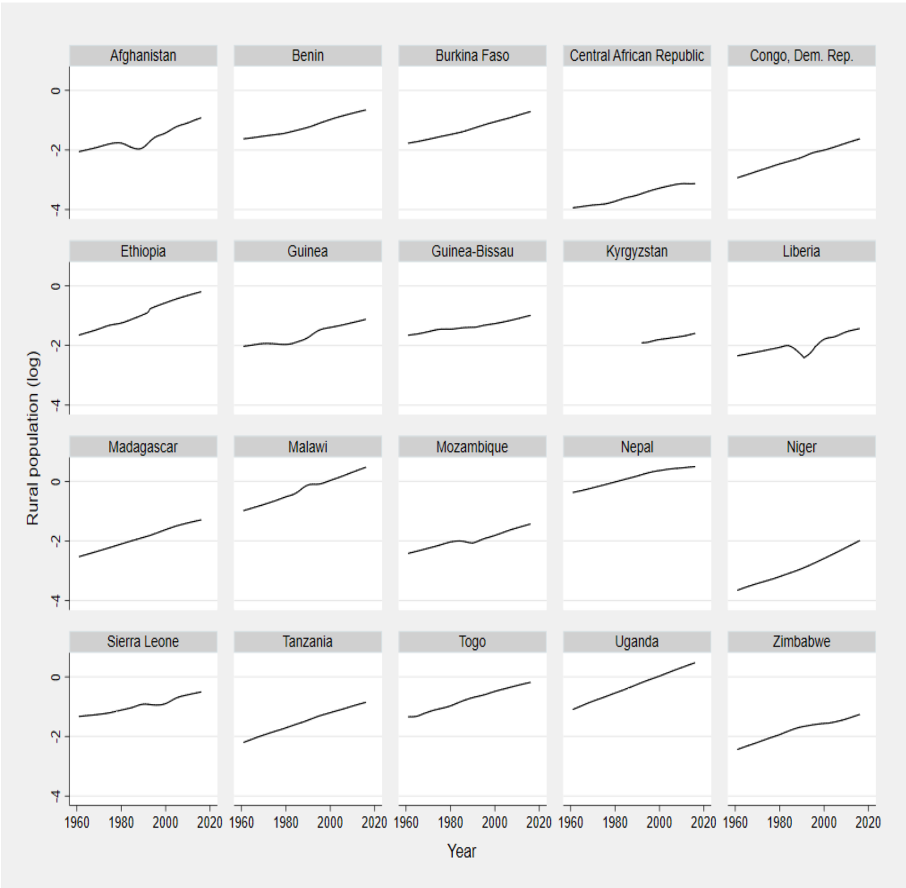


Appendix F. Temporal evolution

Figure A12. Temporal evolution of rural population density, high-income group.

Figure A13. Temporal evolution of rural population density, middle-income group.

Figure A14. Temporal evolution of rural population density, low-income group.



Appendix G. Unit root tests' results

Table A12. Unit root tests, income clusters.

Low Income Economies								
	For (log)		**Δ For (log)**		**GDP cap (log)**		**Δ GDP cap (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	42.3409	0.3703	103.0908	0	28.4854	0.9131	213.1503	0
Constant and trend	22.7278	0.9872	100.1469	0	92.284	0	174.3859	0
	Rural pop (log)		**Δ Rural pop (log)**		**Agr (log)**		**Δ Agr (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	141.6243	0	370.5433	0	54.2312	0.066	136.8267	0
Constant and trend	347.5515	0	489.7693	0	67.2637	0.0045	106.2696	0
Middle Income Economies								
	For (log)		**Δ For (log)**		**GDP cap (log)**		**Δ GDP cap (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	127.8523	0.3404	309.0733	0	75.2122	0.9997	539.318	0
Constant and trend	59.723	1	363.1026	0	190.8576	0.0001	405.9193	0
	Rural pop (log)		**Δ Rural pop (log)**		**Agr (log)**		**Δ Agr (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	174.823	0.0012	329.7181	0	172.5064	0.0018	518.1225	0
Constant and trend	415.1723	0	390.1991	0	138.4452	0.1466	422.5851	0

High Income Economies								
	For (log)		**Δ For (log)**		**GDP cap (log)**		**Δ GDP cap (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	56.0176	0.622	175.6909	0	89.148	0.0086	222.2363	0
Constant and trend	38.2678	0.9871	189.5342	0	59.0724	0.5096	192.3877	0
	Rural pop (log)		**Δ Rural pop (log)**		**Agr (log)**		**Δ Agr (log)**	
Inverse chi-squared	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	38.4231	0.9864	96.044	0.0022	46.4387	0.9004	252.7817	0
Constant and trend	89.2417	0.0085	98.3846	0.0013	96.1258	0.0021	200.7362	0

Note: the statistics are constructed using a Fisher-type test based on an inverse chi-squared test where the null hypothesis H0 is "All panels contain unit roots" and the alternative H1 is "At least one panel is stationary" [53].

Table A13. Unit rot tests, forest clusters**.**

Pre Transition Economies								
	For (log)		**Δ For (log)**		**GDP cap (log)**		**Δ GDP cap (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	25.4884	0.9939	151.3153	0	10.0256	1	160.2524	0
Constant and trend	25.8388	0.9929	121.4375	0	28.8431	0.9775	121.9925	0
	Rural pop (log)		**Δ Rural pop (log)**		**Agr (log)**		**Δ Agr (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	145.1717	0	93.3585	0	91.9331	0.0001	174.0693	0
Constant and trend	179.279	0	102.9771	0	65.6635	0.0299	139.9722	0

Late Transition Economies								
	For (log)		**Δ For (log)**		**GDP cap (log)**		**Δ GDP cap (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value

Constant	89.4357	0.0418	161.6554	0	51.541	0.9314	318.8514	0
Constant and trend	33.5529	0.9999	221.0974	0	118.776	0.0001	238.9642	0
	Rural pop (log)		**Δ Rural pop (log)**		**Agr (log)**		**Δ Agr (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	89.7655	0.0398	399.3701	0	86.1799	0.0675	292.8106	0
Constant and trend	345.0809	0	439.3871	0	111.2495	0.0007	228.1819	0

Post Transition Economies

	For (log)		**Δ For (log)**		**GDP cap (log)**		**Δ GDP cap (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	111.2867	0.3949	274.8844	0	131.2791	0.0634	495.6008	0
Constant and trend	61.3268	0.9999	310.2488	0	194.5949	0	411.7363	0
	Rural pop (log)		**Δ Rural pop (log)**		**Agr (log)**		**Δ Agr (log)**	
Inverse chi-squared	P Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Constant	119.9333	0.2036	303.577	0	95.0633	0.8085	440.8511	0
Constant and trend	327.6055	0	435.9887	0	124.9217	0.1269	361.4367	0

Table A14. Cointegration tests, income countries groups.

	Low income		***Middle income***		***High income***	
Constant	**Statistic**	**p-value**	Statistic	p-value	Statistic	p-value
Modified Phillips-Perron t	3.0233	0.0013	4.9632	0	5.1023	0
Constant and trend						
Modified Phillips-Perron t	3.4613	0.0003	5.7663	0	5.4084	0
Without agricultural area						
	Low income		***Middle income***		***High income***	

Constant	Statistic	p-value	Statistic	p-value	Statistic	p-value
Modified Phillips-Perron t	3.578	0.0002	5.3606	0	4.5923	0
Constant and trend						
Modified Phillips-Perron t	4.2702	0	5.7043	0	5.2077	0

Note: the statistics of the Pedroni test are constructed using a Modified Phillips-Perron t statistic where the null hypothesis H0 is "No cointegration" and the alternative H1 is "All panels are cointegrated" [44].

Table A15. Cointegration test, forest transition clusters.

	Pre transition		***Late transition***		***Post transition***	
Constant	**Statistic**	**p-value**	Statistic	p-value	Statistic	p-value
Modified Phillips-Perron t	2.8929	0.0019	4.6014	0	5.8305	0
Constant and trend						
Modified Phillips-Perron t	3.3326	0.0004	5.565	0	5.5252	0
Without agricultural area						
	Pre transition		***Late transition***		***Post transition***	
Constant	Statistic	p-value	Statistic	p-value	Statistic	p-value
Modified Phillips-Perron t	2.4393	0.0074	4.5979	0	5.3066	0
Constant and trend						
Modified Phillips-Perron t	3.6286	0.0001	5.099	0	6.8802	0

Note: the statistics of the Pedroni test are constructed using a Modified Phillips-Perron t statistiscs where the null hypothesis H0 is "No cointegration" and the alternative H1 is "All panels are cointegrated" [44].

APPENDIX Chapter 4

Annex 1: Solving continuous maximization problem

The farmer's objective is to maximize their present value[31] output (Y) at any t in time (H):

Max H

$$H = \int_0^{\infty} Ye^{-rt}$$

$$Y = (L) + S * \pi(S) - \left(\frac{1}{2\gamma}\right) * \left(\dot{S}\right)^2$$

Assuming that moving cost are paid at the time of planning (they are not discounted)[32] thus, we do not discount them. Then, we have:

$$H = \int_0^{\infty} \left[\left((L) + S * \pi(S)\right) * e^{-rt} - \left(\frac{1}{2\gamma}\right) * \left(\dot{S}\right)^2 \right] dt$$

We use the Euler equiation to solve this optimal control problem. For that, we use the assumption that s(t) is the optimal trajectory around which other small deviation's trajectories (z(t)) exist:

$$z(t) = s(t) + \varepsilon\eta(t)$$

[31] Discount rate in continuous time: PVt=FV1+rn^nt ; 1+rnnt =ert

[32] This assumption is made because it simplifies the solution of the maximization problem. Nevertheless, it is a realistic assumption; many time farmers bear the cost of moving to the frontier at the present moment, not in the future. E.g. searching information cost.

Where $\eta(0) = \eta(\infty) = 0$ and $z(t) = s(t)$

If we substitute $\dot{S}$ by $\dot{Z}$ and we call the function $J(\varepsilon)$:

$$J(\varepsilon) = \int_0^\infty \left[\left((L) + Z * \pi\right) * e^{-rt} - \left(\frac{1}{2\gamma}\right) * \left(\dot{Z}\right)^2\right] dt$$

We want to maximize $J(\varepsilon)$, that happens when $\varepsilon = 0$ or what it is the same, when $\frac{\partial J(\varepsilon)}{\partial \varepsilon} = 0$. We applied a special case of Leibniz's rule: $\frac{d}{dx}(\int_a^b f(x,t)dt) = \int_a^b \frac{\partial}{\partial x} f(x,t)dt$ and we the integral become:

$$\left(\frac{\partial J(\varepsilon)}{\partial \varepsilon}\right) = 0 = \int_0^\infty \left[(\eta(t) * \pi) * e^{-rt} - \frac{1}{2\gamma} * \dot{Z} * 2 * \dot{\eta}\right] dt$$

Now, we applied the opposite change than previously, we substitute $\dot{Z}$ by $\dot{S}$:

$$\left(\frac{\partial J(\varepsilon)}{\partial \varepsilon}\right) = 0 = \int_0^\infty \left[(\eta(t) * \pi) * e^{-rt} - \frac{1}{\gamma} * \dot{S} * \dot{\eta}\right] dt$$

If we solve this integral by parts (using the integration rule: $\int u * dv = u * v - \int v * du$ for the second part):

$$\left(\frac{\partial J(\varepsilon)}{\partial \varepsilon}\right) = 0 = \int_0^\infty [(\eta(t) * \pi) * e^{-rt}]dt - \int_0^\infty \frac{1}{\gamma} * \dot{S} * \dot{\eta}\, dt$$

$$\left(\frac{\partial J(\varepsilon)}{\partial \varepsilon}\right) = 0 = \int_0^\infty [(\eta(t) * \pi) * e^{-rt}]dt - \left[\eta(t)\frac{\dot{S}}{\gamma}\right] + \int_0^\infty \eta(t)\frac{\ddot{S}}{\gamma}\, dt$$

Taking into account that $\eta(0) = \eta(\infty) = 0$ and taking $\eta(t)$ as common factor, we have:

$$\left(\frac{\partial J(\varepsilon)}{\partial \varepsilon}\right) = 0 = \int_0^\infty \eta(t)\left[\pi e^{-rt} + \frac{\ddot{S}}{\gamma}\right] dt$$

The previous integral will be equal to zero if, and only, if one of the terms are equal to zero. We know that $\eta(t)$is not always zero. Thus, the expression between brackets should be zero for any $\eta(t)$:

$$\int_T^\infty -\frac{\ddot{S}}{\gamma} dt = \int_T^\infty \pi e^{-r(t-T)} dt$$

As final result:

$$\dot{S} = \gamma \int_T^\infty \pi e^{-r(t-T)} dt$$

Annex 2: **Solving the dynamic equation system**

$$\begin{cases} \dot{q} = rq - \dfrac{\mu\alpha L - F(1-\mu s)}{1 - s\mu\alpha} \\ \\ \dot{s} = \gamma q \end{cases}$$

To be able to solve this system of dynamic equations by the method of the characteristic values and vectors, we transform the nonlinear equation into a linear one using the Taylor[33] approximation round s^*.

Thus, $\dot{q}= rq - \frac{\mu F}{1-\alpha F+\mu\alpha^2 L}(s - s^*)$ being $\beta = \frac{\mu F}{1-\alpha F+\mu\alpha^2 L}$ and $M = \dot{S} - S^*$.

$$\begin{cases} \dot{q}= rq - \beta M \\ \dot{M} = \gamma q + 0M \end{cases}$$

We transform this system of equations into a matrix form:

$$\begin{matrix} \dot{q} \\ M \end{matrix} = \begin{pmatrix} r & -\beta \\ \gamma & 0 \end{pmatrix} \begin{matrix} q \\ M \end{matrix}$$

The Fundamental theorem of algebra[34] states that every polynomial equation having complex coefficients and degree ≥ 1 has at least one complex root. We could apply this theorem to our system of equations if we transform it into a polyonomy.

The characteristic polyonomy of a (small) matrix called A, it is the function p(λ) = det (A-λI) where λ is the characteristic value and I is the identity matrix. All the characteristics values of a matrix A could be calculated by solving the equation P_A(λ)= 0.

Being A= $\begin{pmatrix} r & -\beta \\ \gamma & 0 \end{pmatrix}$

[33] Taylor: f(s*)+ f'(s*)/1! (s-s*)
[34] Theorem explanation: http://mathworld.wolfram.com/FundamentalTheoremofAlgebra.html

P (λ)= det $[(A-\lambda I)]$

det $\begin{pmatrix} r & -\beta \\ \gamma & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$;

P (λ)= 0; det $\begin{vmatrix} r-\lambda & -\beta \\ \gamma & -\lambda \end{vmatrix} = 0 \Leftrightarrow$ (r$-\lambda$) $(-\lambda) - \gamma(-\beta)$= 0 $\Leftrightarrow$
$\lambda^2 - r\lambda + \beta\gamma = 0$

$$\lambda = \frac{r \pm \sqrt{(-r)^2 - 4\beta}}{2}$$

$r^2 > 4\beta$ $\rightarrow$ λ is a real root

$r^2 < 4\beta$ $\rightarrow$ λ is a complex root

Once we know λ, we are able to solve the dynamic system of two equations. However, we stop the problem solving here because it is enough for the model to reach the conclusion that λ could have a real or complex root.

Annex 3: Parameters' interpretation

- Extreme values of the parameters*

	Min	Max
r (r^2)	0	0'32 (0,1)
β	0	0,16
γ	0	0,16

*Those extreme values have been chosen by first setting a realistic maximum value of r in our frontier cases (our max value 0,32 was reported by Mozambican medium size farmers during my fieldwork there, and it was confirmed lately by email). A minimum of zero makes sense from the mathematics point of view (in case we have a negative interest rate (r), as the term is squared, a negative value squared is a positive number so, higher than zero). The min and max values for β and γ were chosen following the previous data for the r parameter (interest rate). As r^2 is related to $4\beta\gamma$ by an inequality, their min and max have to be the same for both sedes (the max value of r^2 (0,1) is equal to the max value of $4\beta\gamma$ ((4*0,16*0,16)= 0,1).

- How to interpret the values of r, β, γ.

	Interpretation
r^2	The value of this parameter is the real interest rate. We interpreted as such.
β	The value of this parameter represents an abstract concept: the strength agglomeration economics; It corresponds to the parameter explaining the $\pi(S)$ function. Taking our definition of agglomeration economics (see introduction), we will interpret the value of this parameter relatively to the maximum value (0,16) and also, relatively to the values of other frontier cases. In the hypothetical case that β =0 , it means that there is no agglomeration economics at the frontier area. So, the farmer's income (π) does not increase with S (the area other than his, dedicated to cash crops), her income only depends on herself (her decision). In the hypothetical case that β =0,16 (max value), it means that farmer's profits increase by its max according to the β value. β is composed by other parameters (see Annex 3 below). Thus, how much the profit increase when β =0,16 is not a trivial calculation.
γ	The value of this parameter represents an abstract concept: inverse of the marginal adjustment costs or "the speed of adjustment" (a parameter of the cost equation); see equation:$Adjustment\ cost = \left(\frac{1}{2\gamma}\right) * (S^{\cdot})^2$. Taking our definition of adjustment costs (see model description), we will interpret the value of this parameter relatively to the maximum value (0,16) and also, relatively to the values at others frontier cases. Note: γ is the value of the INVERSE marginal costs of adjustment experienced by the farmer. So, the higher this value, the LOWER the speed of adjustment. In the hypothetical case that γ =0 (the highest possible costs), It means there is no change of the frontier area (no increase of the cash crop area at the frontier). This can be seen from the equation: $s^{\cdot} = \gamma q\ when\ \gamma = 0;\ s^{\cdot} = 0$. The closer γ is to zero, the smallest the adjustment are at any period until nothing changes. In the hypothetical case that γ =0,16 .This means that in this frontier area, the farmers/investors incur very small costs so, the conversion to cash crops ($s^{\cdot}$) occurs easier/faster. In reality that would mean that the

	land at the frontier becomes as flexible as a financial assets (very easy to trade)

E.g. Beta is equal to 0.07: This value indicates that the strength of the agglomeration economics is around half of its maximum value (0,08 would be 50% of the total area cover by cash crops). It has to be noted that the maximum value in each frontier could mean/represent different things in reality.

E.g. if gamma is equal to 0.073: This value indicates that the adjustment costs at this frontier are happening at an speed which almost half of the maximum speed. The adjustment costs in this frontier are neither too slow nor too fast according to the min and max but it is also interesting to compare it with the other frontier cases.

Annex 3: Questionnaires

The way we quantify the value of the parameters for each frontier region was thorugh interviews to experts on those regions. In those interviews, I asked questions to the experts following two questionnaires with 12 and 13 questions about the agglomeration economies and adjustment cost parameters, respectively.

The experts were asked to answer those questions thinking about each specific frontier area of expertise. The questions were close, with four possible answers on a scale from the highest to the lowest value; in addition, each question has to be rated on its relative importance respects to the others. The questionnaire was repeated for twice for two time periods: a "recent" time frame after the 2000s and a "past" one after the 2000s. The reason we selected the year 2000 is because it was a relevant date for most of the frontiers in terms of important changes.

<u>Data collection for the agglomeration economics' and adjustment costs parameters:</u>

Regarding the three types of farmers that we identified (international financed, nationally financed and smallholders), we assumed homogeneity across those types for the agglomeration economies and adjustment cost's parameters but not for the interest rates. For the design of the questionnaire on agglomeration economies we used as a benchmark, the agglomeration economies' micro-foundations described by Duranton and Puga (2003). These are: i) The sharing of capital inputs; ii) Improvement on the matching between production requirements and types of land, labor or intermediate inputs and iii) The learning mechanisms through the interaction of economic agents. For the case of adjustment costs, we based our questions on costs pointed by the literature as well as the author's knowledge from their fieldwork's experience in one of the frontier areas of study.

Those answered were pondered (see table below); The agglomeration economics and the adjustment costs' parameters were set to have a minimum of cero score and a maximum of 0,16 score (with a 0,013 and 0,01 per question for each parameter respectively).

Cathegories				**Importance of this on the farmer's profit**			
low (1)	**2**	**3**	**High (4)**	**1**	**2**	**3**	**4**
0,0025	0,005	0,005	0,01	Cat+0,0025	*1	*1	Cat-0,0025

Questions about agglomeration economics:

1. How common is to have isolated farms? (>1h travel distant from closest neibourgh)
2. Frequency of face-to-face meetings with other farmers
3. Exchange of useful information from interaction with other actors different than farmers at the frontier
4. Presence of agricultural inputs suppliers at the frontier
5. Degree of specialization of inputs suppliers
6. Support/AUXILIARY services provided at the frontier (e.g. Agricultural research, agricultural extension, Agricultural credit, Agricultural Marketing, etc.)

7. Do farmers have access to information from previous farmers experiences?
8. Own machinery shared with (or rented to) neighbor farmers
9. Amount of local infrastructure built by the frontier adopters
10. Local facilities owned commonly by group of neighbors' farmers (warehouse, dryer machine etc.)
11. Easiness to find workers with the right skills for the activities required?
12. How common is to find a business partners between 2 o more farmers at the frontier

Questions about adjustment costs:

1. How is the quality of life at the frontier respect to the non-frontier area of origin? (health, education, security, amenities etc)
2. Difficulty to find information about starting commercial agriculture at the frontier area?
3. Legal difficulties to acquire land at the frontier (institutions, bureaucracy etc)
4. How different is the price of the land plot at the frontier?
1. Distance of land plot at the frontier from nearest city (low: 1h; high: >5h)
2. Distance from the land plot at the frontier to the closer inputs' suppliers' market? (low: 1h; high: >5h vehicle travel distance)
3. Problems to use the land due to local community surrounded
4. Costs paid for the labor and machinery needed to clear land
5. Costs for inputs needed from the start of operations till the sale of products (labor, machinery, buildings, biological or chemical inputs etc)
6. Are the labor surveillance and land maintenance more costly compare to the non-frontier area?
7. How would you rate the following risks?
    Breaking the land deal's conditions
    Unexpected events harmful for the farm (animal attacks, robberies etc)
    Uncertainty about the supply of inputs (e.g., contracts)

8. ¿Could you think about other costs of moving to the frontier not asked on the previous questions?

Data on interest rates

Frontier cases	Type	Interest rate NOMINAL	Average Inflaction (2014-2019) (Source world Bank except for Bolivia)	Real Interest rate
North Mozambique	International	between +1 and -1	6,3% annual	
	Local Very small	Max 32% min 20% (avg 26%)		19,70%
ST_Bolivia	Grande, mediana. 6%	between +1 and -1	3%	
	Pequeña. 7%	6,50%		3,50%
	Micro. 22%	22,00%		19,00%
Mato Grosso Brasil	Otros. 8%	between +1 and -1	5,80%	
	Medianos 6%	6,75		0,95%
	Pequeños. 3,8%	3,80%		-2%
		between +1 and -1	3,60%	

Indonesia, oilpalm, Kalimantan	Big 12% Medium 14% Small 7%-10%= 8,5%	13% 8,50%		9,40% 4,90%
North Laos, rubber	APBank. 12% Commertial banks. 15% Money lenders. 20%	13,50% 20%	2,90%	10,60% 17.1%

Sources of interest rates' data:

Frontier cases	Data Sources
North Mozambique	Kobus interview For the max: https://omrmz.org/omrweb/wp-content/uploads/Observador-Rural-11.pdf For the min: Mosca, J. & Dada, Y. (2013). Contributo para o Estudo dos Determinantes da Produção Agrícola. Observador Rural No 5. Observatório do Meio Rural (OMR), Maputo. https://omrmz.org/omrweb/wp-content/uploads/Observador-Rural-05.pdf
ST_Bolovia	Articulo 5, Decreto Supremo 2055, que reglamenta la Ley de Servicios Financieros (New banking law, Nueva ley bancaria)

	Banco mercantil santa cruz: https://www.bmsc.com.bo/Documents/Educación%20Financiera/Banco%20Ayuda%20-%20Sector%20Productivo.pdf Banco nacional de bolivia: https://boliviaemprende.com/noticias/rigen-tasas-maximas-del-6-al-115-en-creditos-productivos
Mato Grosso Brasil	Gov_Plano Safra 2019/2020: https://www.gov.br/agricultura/pt-br/assuntos/noticias/plano-safra-2019-2020-entra-em-vigor-nesta-segunda-feira
Indonesia, oilpalm, Kalimantan	Pramudya, E. P., Hospes, O., & Termeer, C. J. A. M. (2017). Governing the palm-oil sector through finance: the changing roles of the Indonesian State. Bulletin of Indonesian Economic Studies, 53(1), 57-82.
North Laos, rubber	Manivong, V., & Cramb, R. A. (2007). Economics of smallholder rubber production in Northern Laos (No. 418-2016-26497).

REFERENCES

Abeygunawardane, D., García, A. K., Sun, Z., Müller, D., Sitoe, A., & Meyfroidt, P. (n.d.). Resource frontiers and agglomeration economies: The varied logics of land-based investing in Southern and Eastern Africa. 34.

Al Mamun, Md., Sohag, K., Hannan Mia, Md. A., Salah Uddin, G., & Ozturk, I. (2014). Regional differences in the dynamic linkage between CO2 emissions, sectoral output and economic growth. *Renewable and Sustainable Energy Reviews*, *38*, 1–11. https://doi.org/10.1016/j.rser.2014.05.091

Amaral, D. F., de Souza Ferreira Filho, J. B., Chagas, A. L. S., & Adami, M. (2021). Expansion of soybean farming into deforested areas in the amazon biome: the role and impact of the soy moratorium. *Sustainability Science*, 1-18.

Alcott, B. (2005). Jevons' paradox. *Ecological Economics*, *54*(1), 9–21. https://doi.org/10.1016/j.ecolecon.2005.03.020

Angelsen, A. (2010). Policies for reduced deforestation and their impact on agricultural production. *Proceedings of the National*

Academy of Sciences, *107*(46), 19639–19644. https://doi.org/10.1073/pnas.0912014107

Angelsen, A., & Kaimowitz, D. (1999a). Rethinking the Causes of Deforestation: Lessons from Economic Models. *The World Bank Research Observer*, *14*(1), 73–98. https://doi.org/10.1093/wbro/14.1.73

Ariti, A. T., van Vliet, J., & Verburg, P. H. (2015). Land-use and land-cover changes in the Central Rift Valley of Ethiopia: Assessment of perception and adaptation of stakeholders. *Applied Geography*, *65*, 28–37. https://doi.org/10.1016/j.apgeog.2015.10.002

Arvor, D., Daugeard, M., Tritsch, I., De Mello-Thery, N. A., Thery, H., & Dubreuil, V. (2018). Combining socioeconomic development with environmental governance in the Brazilian Amazon: The Mato Grosso agricultural frontier at a tipping point. *Environment, Development and Sustainability*, *20*(1), 1–22. https://doi.org/10.1007/s10668-016-9889-1

Baltagi, B. H. (2005). *Econometric analysis of panel data* (3rd ed). J. Wiley & Sons.

Bandura, A. (2000). *SOCIAL COGNITIVE THEORY: An Agentic Perspective*. 28.

Barbier, E. B. (2012). Scarcity, frontiers and development: Scarcity, frontiers and development. *The Geographical Journal*, *178*(2), 110–122. https://doi.org/10.1111/j.1475-4959.2012.00462.x

Barbier, E. B., Burgess, J. C., & Grainger, A. (2010). The forest transition: Towards a more comprehensive theoretical framework. *Land Use Policy*, *27*(2), 98–107. https://doi.org/10.1016/j.landusepol.2009.02.001

Barbier, E. B., Delacote, P., & Wolfersberger, J. (2017). The economic analysis of the forest transition: A review. *Journal of Forest Economics*, *27*, 10–17. https://doi.org/10.1016/j.jfe.2017.02.003

Barbier, E. B., & Tesfaw, A. (2015). Explaining forest transitions: The role of governance. *Ecological Economics*, *119*, 252–261. https://doi.org/10.1016/j.ecolecon.2015.09.010

Baron, R. M., & Kenny, D. A. (n.d.). *The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations*. 10.

Bauch, C. T., Sigdel, R., Pharaon, J., & Anand, M. (2016). Early warning signals of regime shifts in coupled human–environment systems. *Proceedings of the National Academy of Sciences*, *113*(51), 14560–14567. https://doi.org/10.1073/pnas.1604978113

Beckert, J. (2013). Imagined futures: Fictional expectations in the economy. *Theory and Society*, *42*(3), 219–240. https://doi.org/10.1007/s11186-013-9191-2

Bicudo Da Silva, R. F., Batistella, M., Moran, E., Celidonio, O. L. D. M., & Millington, J. D. A. (2020). The Soybean Trap: Challenges and Risks for Brazilian Producers. *Frontiers in Sustainable Food Systems*, *4*, 12. https://doi.org/10.3389/fsufs.2020.00012

Biggs, R., Peterson, G. D., & Rocha, J. C. (2018). The Regime Shifts Database: A framework for analyzing regime shifts in social-ecological systems. *Ecology and Society*, *23*(3), art9. https://doi.org/10.5751/ES-10264-230309

Boettiger, C., & Hastings, A. (2013). From patterns to predictions. *Nature*, *493*(7431), 157–158. https://doi.org/10.1038/493157a

Boserup, E. (1965) *The Conditions of Agricultural Growth* (Aldine, Chicago).

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (Vol. 424). New York: Wiley. ISBN 0471810339.

Börner, J., Kis-Katos, K., Hargrave, J., & König, K. (2015). Post-crackdown effectiveness of field-based forest law enforcement in the Brazilian Amazon. *PLoS One*, *10*(4), e0121544.

Bruckner, M., Giljum, S., Lutz, C., & Wiebe, K. S. (2012). Materials embodied in international trade – Global material extraction and consumption between 1995 and 2005. *Global Environmental Change*, *22*(3), 568–576. https://doi.org/10.1016/j.gloenvcha.2012.03.011

Buchadas, A., Baumann, M., Meyfroidt, P., Kuemmerle, T., Uncovering major types of deforestation frontiers across the world's tropical dry forests, in prep.

Byerlee, D., Stevenson, J., & Villoria, N. (2014a). Does intensification slow crop land expansion or encourage deforestation? *Global*

Food Security, *3*(2), 92–98. https://doi.org/10.1016/j.gfs.2014.04.001

Canning, D., & Pedroni, P. (2008). Infrastructure, long-run economic growth and causality tests for cointegrated panels. *Manchester School*, *76*(5), 504–527. https://doi.org/10.1111/j.1467-9957.2008.01073.x

Caravaggio, N. (2020). A global empirical re-assessment of the Environmental Kuznets curve for deforestation. *Forest Policy and Economics*, *119*, 102282. https://doi.org/10.1016/j.forpol.2020.102282

Carlson, K. M., Heilmayr, R., Gibbs, H. K., Noojipady, P., Burns, D. N., Morton, D. C., Walker, N. F., Paoli, G. D., & Kremen, C. (2018). Effect of oil palm sustainability certification on deforestation and fire in Indonesia. *Proceedings of the National Academy of Sciences*, *115*(1), 121–126. https://doi.org/10.1073/pnas.1704728114

Carvalho, W. D., Mustin, K., Hilário, R. R., Vasconcelos, I. M., Eilers, V., & Fearnside, P. M. (2019). Deforestation control in the Brazilian Amazon: A conservation struggle being lost as agreements and regulations are subverted and bypassed.

Perspectives in Ecology and Conservation, *17*(3), 122–130. https://doi.org/10.1016/j.pecon.2019.06.002

Castella, J.C., Boissau, S., Trung, T.N., Quang, D.D., 2005. Agrarian

transition and lowland-upland interactions in mountain areas in northern Vietnam: applica- tion of a multi-agent simulation model. Agric. Syst. 86 (3), 312e332.

Ceddia, M. G., Bardsley, N. O., Gomez-y-Paloma, S., & Sedlacek, S. (2014a). Governance, agricultural intensification, and land sparing in tropical South America. *Proceedings of the National Academy of Sciences*, *111*(20), 7242–7247. https://doi.org/10.1073/pnas.1317967111

Ceddia, M. G., Sedlacek, S., Bardsley, N. O., & Gomez-y-Paloma, S. (2013). Sustainable agricultural intensification or Jevons paradox? The role of public governance in tropical South America. *Global Environmental Change*, *23*(5), 1052–1063. https://doi.org/10.1016/j.gloenvcha.2013.07.005

Coelli, T. J., & Rao, D. S. P. (2005). Total factor productivity growth in agriculture: A Malmquist index analysis of 93 countries, 1980–2000. *Agricultural Economics*, *32*(s1), 115–134. https://doi.org/10.1111/j.0169-5150.2004.00018.x

Cohen, P. T. (2009). The post-opium scenario and rubber in northern Laos: Alternative Western and Chinese models of development. *International Journal of Drug Policy*, *20*(5), 424–430. https://doi.org/10.1016/j.drugpo.2008.12.005

Coomes et al. - 2019—Leveraging total factor productivity growth for su.pdf. (n.d.).

Coomes, O. T., Barham, B. L., MacDonald, G. K., Ramankutty, N., & Chavas, J.-P. (2019). Leveraging total factor productivity growth for sustainable and resilient farming. *Nature Sustainability*, *2*(1), 22–28. https://doi.org/10.1038/s41893-018-0200-3

Coppola, A., Ianuario, S., Chinnici, G., Di Vita, G., Pappalardo, G., & D'Amico, M. (2018). Endogenous and Exogenous Determinants of Agricultural Productivity: What Is the Most Relevant for the Competitiveness of the Italian Agricultural Systems? *Agris On-Line Papers in Economics and Informatics*, *10*(2), 33–47. https://doi.org/10.7160/aol.2018.100204

Crespo Cuaresma, J., Danylo, O., Fritz, S., McCallum, I., Obersteiner, M., See, L., & Walsh, B. (2017). Economic Development and Forest Cover: Evidence from Satellite Data. *Scientific Reports*, *7*(1), 40678. https://doi.org/10.1038/srep40678

Cropper, M., & Griffiths, C. (2021). *The Interaction of Population Growth and Environmental Quality*. 6.

Culas, R. J. (2012). REDD and forest transition: Tunneling through the environmental Kuznets curve. *Ecological Economics*, *79*, 44–51. https://doi.org/10.1016/j.ecolecon.2012.04.015

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science*, *361*(6407), 1108–1111. https://doi.org/10.1126/science.aau3445

Davidson, J. (1998). Structural relations, cointegration and identification: Some simple results and their application. *Journal of Econometrics*, *87*(1), 87–113. https://doi.org/10.1016/S0304-4076(98)00007-4

Daskalov, G.M., Grishin, A.N., Rodionov, S., Mihneva, V., 2007.

Trophic cascades triggered by overfishing reveal possible mechanisms of ecosystem regime shifts. Proc. Natl. Acad. Sci. 104 (25), 10518e10523.

Da Silva, R. F., Batistella, M., Moran, E., Celidonio, O. L. D. M., & Millington, J. D.(2020). The soybean trap: challenges and risks for Brazilian producers. *Frontiers in sustainable food systems*, *4*, 12.

De Alban, J., Prescott, G., Woods, K., Jamaludin, J., Latt, K., Lim, C., Maung, A., & Webb, E. (2019). Integrating Analytical Frameworks to Investigate Land-Cover Regime Shifts in Dynamic Landscapes. *Sustainability*, *11*(4), 1139. https://doi.org/10.3390/su11041139

De la Vega-Leinert, A. C., & Huber, C. (2019). The Down Side of Cross-Border Integration: The Case of Deforestation in the Brazilian Mato Grosso and Bolivian Santa Cruz Lowlands. *Environment: Science and Policy for Sustainable Development*, *61*(2), 31–44. https://doi.org/10.1080/00139157.2019.1564214

Dearing, J. A., Braimoh, A. K., Reenberg, A., Turner, B. L., & van der Leeuw, S. (2010). Complex Land Systems: The Need for Long

Time Perspectives to Assess their Future. *Ecology and Society*, *15*(4), art21. https://doi.org/10.5751/ES-03645-150421

Dedehouanou, S. F. A., Swinnen, J., & Maertens, M. (2013). Does Contracting Make Farmers Happy? Evidence from Senegal. *Review of Income and Wealth*, *59*, S138–S160. https://doi.org/10.1111/roiw.12041

DCPI (2015). Anuário de Estatísticas Agrárias 2015. Ministério da Agricultura e Segurança Alimentar. Direcção de Planificação e Cooperação Internacional (DPCI). Maputo, Mozambique. https://1drv.ms/x/s!AkiYdW2a_CqSg-U6GXEfipsuZ-nF7w?e=l8h541

DeFries, R. S., Rudel, T., Uriarte, M., & Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*, *3*(3), 178–181. https://doi.org/10.1038/ngeo756

Deininger, K., & Xia, F. (2016). Quantifying Spillover Effects from Large Land-based Investment: The Case of Mozambique. *World Development*, *87*, 227–241. https://doi.org/10.1016/j.worlddev.2016.06.016

Dharmawan, A. H., Mardiyaningsih, D. I., Komarudin, H., Ghazoul, J., Pacheco, P., & Rahmadian, F. (2020). Dynamics of Rural Economy: A Socio-Economic Understanding of Oil Palm Expansion and Landscape Changes in East Kalimantan, Indonesia. *Land*, *9*(7), 213. https://doi.org/10.3390/land9070213

Dorninger, C., Hornborg, A., Abson, D. J., von Wehrden, H., Schaffartzik, A., Giljum, S., Engler, J.-O., Feller, R. L., Hubacek, K., & Wieland, H. (2021). Global patterns of ecologically unequal exchange: Implications for sustainability in the 21st century. *Ecological Economics*, *179*, 106824. https://doi.org/10.1016/j.ecolecon.2020.106824

Duan, Q., & Tan, M. (2019). Spatial and Temporal Variations of Forest Cover in Developing Countries. *Sustainability*, *11*(6), 1517. https://doi.org/10.3390/su11061517

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063-2117). Elsevier.

Dwyer, M., & Vongvisouk, T. (2019). The long land grab: Market-assisted enclosure on the China-Lao rubber frontier. *Territory,* *Politics, Governance*, *7*(1), 96–114. https://doi.org/10.1080/21622671.2017.1371635

Eboli, F., Parrado, R., Roson, R., 2010. Climate-change feedback on economic growth: explorations with a dynamic general equilibrium model. Environ. Dev. Econ. 15 (5), 515e533.

Egli, L., Meyer, C., Scherber, C., Kreft, H., & Tscharntke, T. (2018). Winners and losers of national and global efforts to reconcile agricultural intensification and biodiversity conservation. *Global Change Biology*, *24*(5), 2212–2228. https://doi.org/10.1111/gcb.14076

Eigenbrod, F., Beckmann, M., Dunnett, S., Graham, L., Holland, R. A., Meyfroidt, P., Seppelt, R., Song, X.-P., Spake, R., Václavík, T., & Verburg, P. H. (2020). Identifying Agricultural Frontiers for Modeling Global Cropland Expansion. *One Earth*, *3*(4), 504–514. https://doi.org/10.1016/j.oneear.2020.09.006

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: Anthropogenic biomes of the world. *Frontiers in Ecology and* *the Environment*, *6*(8), 439–447. https://doi.org/10.1890/070062

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, *55*(2), 251. https://doi.org/10.2307/1913236

ESA. (2017). Land Cover CCI: Product user guide, version 2.0. UCL Geomatic (Université Catholique de Louvain), Friederich-Schille-Universität Jena, Wageningen University, Max-Planck-Institut für Meteorologie, JRC European Commission, Met Office. *Retrieved from Http://Maps.Elie.Ucl.Ac.Be/CCI/Viewer/Download/ESACCI-LC-Ph2-PUGv2_2.0.Pdf*. Retrieved from http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Ewers, R. M. (2006). Interaction effects between economic development and forest cover determine deforestation rates. *Global Environmental Change*, 9.

Ewers, R. M., Scharlemann, J. P. W., Balmford, A., & Green, R. E. (2009). Do increases in agricultural yield spare land for nature?

Global Change Biology, *15*(7), 1716–1726. https://doi.org/10.1111/j.1365-2486.2009.01849.x

FAO. (2019). *FAOSTAT. Food and Agriculture Organization of the United Nations, Rome, Italy*. http://www.fao.org/faostat/en/#data

Filatova, T., Polhill, J. G., & van Ewijk, S. (2016). Regime shifts in coupled socio-environmental systems: Review of modelling challenges and approaches. *Environmental Modelling & Software*, *75*, 333–347. https://doi.org/10.1016/j.envsoft.2015.04.003

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

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

Friis, C., & Nielsen, J. Ø. (2016). Small-scale land acquisitions, large-scale implications: Exploring the case of Chinese banana investments in Northern Laos. *Land Use Policy*, *57*, 117–129. https://doi.org/10.1016/j.landusepol.2016.05.028

Fuglie, K. (2015). Accounting for growth in global agriculture. *Bio-Based and Applied Economics*, *Vol 4 No 3*, 201-234 Pages. https://doi.org/10.13128/BAE-17151

García, V. R., Gaspart, F., Kastner, T., & Meyfroidt, P. (2020). Agricultural intensification and land use change: Assessing country-level induced intensification, land sparing and rebound effect. *Environmental Research Letters*, *15*(8), 085007. https://doi.org/10.1088/1748-9326/ab8b14

Garrett, Rachael D., Lambin, E. F., & Naylor, R. L. (2013). The new economic geography of land use change: Supply chain configurations and land use in the Brazilian Amazon. *Land Use Policy*, *34*, 265–275. https://doi.org/10.1016/j.landusepol.2013.03.011

Garrett, R.D., Koh, I., Lambin, E. F., le Polain de Waroux, Y., Kastens, J. H., & Brown, J. C. (2018). Intensification in agriculture-forest frontiers: Land use responses to development and conservation policies in Brazil. *Global Environmental Change*, *53*, 233–243. https://doi.org/10.1016/j.gloenvcha.2018.09.011

Garrett, R.D., T. Gardner, T. Fonseca, S. Marchand, J. Barlow, D. E. de Blas, J. Ferreira, A. C. Lees, and L. Parry. (2017). Explaining the Persistence of Low Income and Environmentally Degrading Land Uses in the Brazilian Amazon. Ecology and Society 22, no. 3 https://doi.org/10.5751/ES-09364-220327.

Garrett, R. D., Cammelli, F., Ferreira, J., Levy, S. A., Valentim, J., & Vieira, I. (2021). Forests and Sustainable Development in the Brazilian Amazon: History, Trends, and Future Prospects. *Annual Review of Environment and Resources*, *46*.

Gaveau, D. L. A., Locatelli, B., Salim, M. A., Yaen, H., Pacheco, P., & Sheil, D. (2019). Rise and fall of forest loss and industrial plantations in Borneo (2000–2017). *Conservation Letters*, *12*(3). https://doi.org/10.1111/conl.12622

Gaveau, D. L. A., Sheil, D., Husnayaen, Salim, M. A., Arjasakusuma, S., Ancrenaz, M., Pacheco, P., & Meijaard, E. (2016). Rapid conversions and avoided deforestation: Examining four decades of industrial plantation expansion in Borneo. *Scientific Reports*, *6*(1), 32017. https://doi.org/10.1038/srep32017

Geist, H., McConnell, W., Lambin, E. F., Moran, E., Alves, D., &

Rudel, T. (2006). Causes and trajectories of land-use/coverchange. In E. F. Lambin & H. J. Geist (Eds.), Land-use and land-cover change: Local processes and global impacts. Berlin: Springer.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences*, *107*(38), 16732–16737. https://doi.org/10.1073/pnas.0910275107

Gil, J., Siebold, M., & Berger, T. (2015). Adoption and development of integrated crop–livestock–forestry systems in Mato Grosso, Brazil. *Agriculture, Ecosystems & Environment*, *199*, 394–406. https://doi.org/10.1016/j.agee.2014.10.008

Gingrich, S., Lauk, C., Niedertscheider, M., Pichler, M., Schaffartzik, A., Schmid, M., Magerl, A., Le Noë, J., Bhan, M., & Erb, K. (2019). Hidden emissions of forest transitions: A socio-ecological reading of forest change. *Current Opinion in Environmental Sustainability*, *38*, 14–21. https://doi.org/10.1016/j.cosust.2019.04.005

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F.,... & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *science*, *327*(5967), 812-818.

Granger, C. W. (1988). Some recent development in a concept of causality. *Journal of Econometrics*, *39*(1–2), 199–211.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*(1), 121–130. https://doi.org/10.1016/0304-4076(81)90079-8

Granger, C. W. J., & Newbold, P. (2001). Spurious regression in econometrics. *ECONOMETRIC SOCIETY MONOGRAPHS*, *33*, 109–118.

Grassini, P., Eskridge, K. M., & Cassman, K. G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nature Communications*, *4*(1), 2918. https://doi.org/10.1038/ncomms3918

Green, R.E., Cornell, S.J., Scharlemann, J.P., Balmford, A., 2005.

Farming and the fate of wild nature. Science 307 (5709), 550-555.

Groeneveld, J., Müller, B., Buchmann, C. M., Dressler, G., Guo, C., Hase, N., Hoffmann, F., John, F., Klassert, C., Lauf, T., Liebelt, V., Nolzen, H., Pannicke, N., Schulze, J., Weise, H., & Schwarz, N. (2017). Theoretical foundations of human decision-making in agent-based land use models – A review. *Environmental Modelling & Software*, *87*, 39–48. https://doi.org/10.1016/j.envsoft.2016.10.008

Guido, Z., Zimmer, A., Lopus, S., Hannah, C., Gower, D., Waldman, K., Krell, N., Sheffield, J., Caylor, K., & Evans, T. (2020). Farmer forecasts: Impacts of seasonal rainfall expectations on agricultural decision-making in Sub-Saharan Africa. *Climate*

Risk Management, *30*, 100247. https://doi.org/10.1016/j.crm.2020.100247

Hannah, L., Roehrdanz, P. R., K. C., K. B., Fraser, E. D. G., Donatti, C. I., Saenz, L., Wright, T. M., Hijmans, R. J., Mulligan, M., Berg, A., & van Soesbergen, A. (2020). The environmental consequences of climate-driven agricultural frontiers. *PLOS ONE*, *15*(2), e0228305. https://doi.org/10.1371/journal.pone.0228305

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hansen, J. W., Marx, S. M., & Weber, E. U. (2004). The role of climate perceptions, expectations, and forecasts in farmer decision making: the Argentine Pampas and South Florida: Final Report of an IRI Seed Grant Project.

Heck, V., Hoff, H., Wirsenius, S., Meyer, C., & Kreft, H. (2018). Land use options for staying within the Planetary Boundaries –

Synergies and trade-offs between global and local sustainability goals. *Global Environmental Change*, *49*, 73–84. https://doi.org/10.1016/j.gloenvcha.2018.02.004

Heilmayr, R., Carlson, K. M., & Benedict, J. J. (2020). Deforestation spillovers from oil palm sustainability certification. *Environmental Research Letters*, *15*(7), 075002. https://doi.org/10.1088/1748-9326/ab7f0c

Herrmann, R. T. (2017). Large-Scale Agricultural Investments and Smallholder Welfare: A Comparison of Wage Labor and Outgrower Channels in Tanzania. *World Development*, *90*, 294–310. https://doi.org/10.1016/j.worlddev.2016.10.007

Hertel, T. W., Ramankutty, N., & Baldos, U. L. C. (2014a). Global market integration increases likelihood that a future African Green Revolution could increase crop land use and CO2 emissions. *Proceedings of the National Academy of Sciences*, *111*(38), 13799–13804. https://doi.org/10.1073/pnas.1403543111

Hertwich, E. G. (2005). Consumption and the rebound effect: An industrial ecology perspective. *Journal of Industrial Ecology*, *9*(1–2), 85–98.

Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., Angelsen, A., & Romijn, E. (2012). An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters*, *7*(4), 044009. https://doi.org/10.1088/1748-9326/7/4/044009

Hsiao, C. (1997). Cointegration and Dynamic Simultaneous Equations Model. *Econometrica*, *65*(3), 647. https://doi.org/10.2307/2171757

Hyde, W.F. (2012). *The global economics of forestry*.

Igliori, D. C. (n.d.). *Deforestation, Growth and Agglomeration Effects: Evidence from Agriculture in the Brazilian Amazon*. 36.

Imai, N., Furukawa, T., Tsujino, R., Kitamura, S., & Yumoto, T. (2018). Factors affecting forest area change in Southeast Asia during 1980-2010. *PLOS ONE*, *13*(5), e0197391. https://doi.org/10.1371/journal.pone.0197391

IMEA, (2017) http://www.agroicone.com.br/wp-content/uploads/2019/11/relatorio-imea-input.pdf

Iqbal, M. Z., & Babcock, B. A. (2018). Global growing-area elasticities of key agricultural crops estimated using dynamic heterogeneous panel methods. *Agricultural Economics*, *49*(6), 681–690. https://doi.org/10.1111/agec.12452

Jadin, I., Meyfroidt, P., & Lambin, E. F. (2016). International trade, and land use intensification and spatial reorganization explain Costa Rica's forest transition. *Environmental Research Letters*, *11*(3), 035005.

Jayne, T. S., Chamberlin, J., & Headey, D. D. (2014). Land pressures, the evolution of farming systems, and development strategies in Africa: A synthesis. *Food Policy*, *48*, 1–17. https://doi.org/10.1016/j.foodpol.2014.05.014

Jayne, T. S., Chamberlin, J., Traub, L., Sitko, N., Muyanga, M., Yeboah, F. K., Anseeuw, W., Chapoto, A., Wineman, A., Nkonde, C., & Kachule, R. (2016). Africa's changing farm size distribution patterns: The rise of medium-scale farms. *Agricultural Economics*, *47*(S1), 197–214. https://doi.org/10.1111/agec.12308

Jepson, W. (2006). Private agricultural colonization on a Brazilian frontier, 1970–1980. *Journal of Historical Geography*, *32*(4), 839–863. https://doi.org/10.1016/j.jhg.2004.12.019

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3

Jorgenson, A. K. (2006). Unequal Ecological Exchange and Environmental Degradation: A Theoretical Proposition and Cross-National Study of Deforestation, 1990-2000*. *Rural Sociology*, *71*(4), 685–712. https://doi.org/10.1526/003601106781262016

Joshi, P., & Beck, K. (2017). Environmental Kuznets curve for deforestation: Evidence using GMM estimation for OECD and non-OECD regions. *IForest - Biogeosciences and Forestry*, *10*(1), 196–203. https://doi.org/10.3832/ifor2066-009

Junquera, V., & Grêt-Regamey, A. (2019). Crop booms at the forest frontier: Triggers, reinforcing dynamics, and the diffusion of knowledge and norms. *Global Environmental Change*, *57*, 101929. https://doi.org/10.1016/j.gloenvcha.2019.101929

Junquera, V., Meyfroidt, P., Sun, Z., Latthachack, P., & Grêt-Regamey, A. (2020). From global drivers to local land-use change: Understanding the northern Laos rubber boom. *Environmental*

Science & Policy, *109*, 103–115. https://doi.org/10.1016/j.envsci.2020.04.013

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, *90*(1), 1–44. https://doi.org/10.1016/S0304-4076(98)00023-2

Kastner, T., Erb, K.-H., & Nonhebel, S. (2011). International wood trade and forest change: A global analysis. *Global Environmental Change*, *21*(3), 947–956. https://doi.org/10.1016/j.gloenvcha.2011.05.003

Klingler, M., & Mack, P. (2020). Post-frontier governance up in smoke? Free-for-all frontier imaginations encourage illegal deforestation and appropriation of public lands in the Brazilian Amazon. *Journal of Land Use Science*, *15*(2–3), 424–438. https://doi.org/10.1080/1747423X.2020.1739765

Koch, N., zu Ermgassen, E. K. H. J., Wehkamp, J., Oliveira, F., & Schwerhoff, G. (2017). Agricultural Productivity and Forest

Conservation: Evidence from the Brazilian Amazon. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3031416

Kramp, J., Suhardiman, D., & Keovilignavong, O. (2020). (Un)making the upland: Resettlement, rubber and land use planning in Namai village, Laos. *The Journal of Peasant Studies*, 1–23. https://doi.org/10.1080/03066150.2020.1762179

Kronenburg García, A., Meyfroidt, P., Abeygunawardane, D., & Sitoe, A. (2021). *Waves and legacies: The making of an investment frontier in Niassa, Mozambique* [Preprint]. AfricArXiv. https://doi.org/10.31730/osf.io/cvs3b

Krugman, P. (1991). History Versus Expectations. *The Quarterly Journal of Economics*, *106*(2), 651. https://doi.org/10.2307/2937950

Kuznets, S. (1955). *Economic Growth and Income Inequality*. 31.

Lade, S. J., Tavoni, A., Levin, S. A., & Schlüter, M. (2013). Regime shifts in a social-ecological system. *Theoretical Ecology*, *6*(3), 359–372. https://doi.org/10.1007/s12080-013-0187-3

Lambin, E. F., & Meyfroidt, P. (2010). Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use*

Policy, *27*(2), 108–118. https://doi.org/10.1016/j.landusepol.2009.09.003

Lambin, E. F., & Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences*, *108*(9), 3465–3472. https://doi.org/10.1073/pnas.1100480108

Lambin, E. F., Meyfroidt, P., Rueda, X., Blackman, A., Börner, J., Cerutti, P. O., Dietsch, T., Jungmann, L., Lamarque, P., & Lister, J. (2014). Effectiveness and synergies of policy instruments for land use governance in tropical regions. *Global Environmental Change*, *28*, 129–140.

Lambin, E. F. (2012). Global land availability: Malthus versus Ricardo. *Global Food Security*, *1*(2), 83-87.

Laurance, W. F., Sayer, J., & Cassman, K. G. (2014). Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution*, *29*(2), 107–116. https://doi.org/10.1016/j.tree.2013.12.001

Le Polain de Waroux, Y. (2019). Capital has no homeland: The formation of transnational producer cohorts in South America's commodity frontiers. *Geoforum*, *105*, 131–144. https://doi.org/10.1016/j.geoforum.2019.05.016

Le Polain de Waroux, Y., Baumann, M., Gasparri, N. I., Gavier-Pizarro, G., Godar, J., Kuemmerle, T., Müller, R., Vázquez, F., Volante, J. N., & Meyfroidt, P. (2018). Rents, Actors, and the Expansion of Commodity Frontiers in the Gran Chaco. *Annals of the American Association of Geographers*, *108*(1), 204–225. https://doi.org/10.1080/24694452.2017.1360761

Leblois, A., Damette, O., & Wolfersberger, J. (2017). What has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data. *World Development*, *92*, 82–102. https://doi.org/10.1016/j.worlddev.2016.11.012

Li TM, L. (2015). *Social impacts of oil palm in Indonesia: A gendered perspective from West Kalimantan*. Center for International Forestry Research (CIFOR). https://doi.org/10.17528/cifor/005579

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H.,

& Taylor, W. W. (2007). Complexity of Coupled Human and Natural Systems. *Science*, *317*(5844), 1513–1516. https://doi.org/10.1126/science.1144004

Liu, Jinlong, Liang, M., Li, L., Long, H., & De Jong, W. (2017). Comparative study of the forest transition pathways of nine Asia-Pacific countries. *Forest Policy and Economics*, *76*, 25–34. https://doi.org/10.1016/j.forpol.2016.03.007

Lu, J., & Schönweger, O. (2019). Great expectations: Chinese investment in Laos and the myth of empty land. *Territory, Politics, Governance*, *7*(1), 61–78. https://doi.org/10.1080/21622671.2017.1360195

Magliocca, N. R., Ellis, E. C., Allington, G. R. H., de Bremond, A., Dell'Angelo, J., Mertz, O., Messerli, P., Meyfroidt, P., Seppelt, R., & Verburg, P. H. (2018). Closing global knowledge gaps: Producing generalized knowledge from case studies of social-ecological systems. *Global Environmental Change*, *50*, 1–14. https://doi.org/10.1016/j.gloenvcha.2018.03.003

Manivong, V., & Cramb, R. A. (n.d.). *Economics of Smallholder Rubber Production in Northern Laos*. 15.

Malek, Ž., Douw, B., Van Vliet, J., Van Der Zanden, E. H., & Verburg,

P. H. (2019). Local land-use decision-making in a global context. *Environmental Research Letters*, *14*(8), 083006.

Mastrangelo, M. E., Gavin, M. C., Laterra, P., Linklater, W. L., & Milfont, T. L. (2014). Psycho-Social Factors Influencing Forest Conservation Intentions on the Agricultural Frontier: Factors influencing conservation intentions. *Conservation Letters*, *7*(2), 103–110. https://doi.org/10.1111/conl.12033

Mather, A. S. (1992). The forest transition. *Area, 367-379.*

Mather, A.S., Needle, C.L., 1998. The forest transition: a theoretical basis. Area 30 (2), 117–124.

Maertens, M., Minten, B., & Swinnen, J. (2012). Modern food supply chains and development: Evidence from horticulture export sectors in sub-Saharan Africa. *Development Policy Review*, *30*(4), 473-497.

Martens, P., Rotmans, J. (Eds.), (2002). Transitions in a Globalising World. Swets & Zeitlinger, Lisse, p. 135.

Maertens, M., & Velde, K. V. (2017). Contract-farming in staple food chains: The case of rice in Benin. *World Development*, *95*, 73-87.

MapBiomas. https://plataforma.brasil.mapbiomas.org/ Last accessed: 09/03/2021.

McKay, B., & Colque, G. (2016). Bolivia's soy complex: The development of 'productive exclusion.' *The Journal of Peasant Studies*, *43*(2), 583–610. https://doi.org/10.1080/03066150.2015.1053875

McKay, B. M. (2017). Agrarian Extractivism in Bolivia. *World Development*, *97*, 199–211. https://doi.org/10.1016/j.worlddev.2017.04.007

Meijaard, E., Brooks, T. M., Carlson, K. M., Slade, E. M., Garcia-Ulloa, J., Gaveau, D. L. A., Lee, J. S. H., Santika, T., Juffe-Bignoli, D., Struebig, M. J., Wich, S. A., Ancrenaz, M., Koh, L. P., Zamira, N., Abrams, J. F., Prins, H. H. T., Sendashonga, C. N., Murdiyarso, D., Furumo, P. R., … Sheil, D. (2020). The environmental impacts of palm oil in context. *Nature Plants*, *6*(12), 1418–1426. https://doi.org/10.1038/s41477-020-00813-w

Meyfroidt, P., Roy Chowdhury, R., de Bremond, A., Ellis, E. C., Erb, K.-H., Filatova, T., Garrett, R. D., Grove, J. M., Heinimann, A., Kuemmerle, T., Kull, C. A., Lambin, E. F., Landon, Y., le Polain de Waroux, Y., Messerli, P., Müller, D., Nielsen, J. ø., Peterson, G. D., Rodriguez García, V., … Verburg, P. H. (2018). Middle-range theories of land system change. *Global Environmental Change*, *53*, 52–67. https://doi.org/10.1016/j.gloenvcha.2018.08.006

Meyfroidt, P., Rudel, T. K., & Lambin, E. F. (2010). Forest transitions, trade, and the global displacement of land use. *Proceedings of the National Academy of Sciences*, *107*(49), 20917–20922. https://doi.org/10.1073/pnas.1014773107

Meyfroidt, Patrick. (2013). Environmental Cognitions, Land Change and Social-Ecological Feedbacks: Local Case Studies of Forest Transition in Vietnam. *Human Ecology*, *41*(3), 367–392. https://doi.org/10.1007/s10745-012-9560-x

Meyfroidt, Patrick. (2016). Approaches and terminology for causal analysis in land systems science. *Journal of Land Use Science*, *11*(5), 501–522. https://doi.org/10.1080/1747423X.2015.1117530

Meyfroidt, Patrick, Carlson, K. M., Fagan, M. E., Gutiérrez-Vélez, V. H., Macedo, M. N., Curran, L. M., DeFries, R. S., Dyer, G. A., Gibbs, H. K., Lambin, E. F., Morton, D. C., & Robiglio, V. (2014). Multiple pathways of commodity crop expansion in tropical forest landscapes. *Environmental Research Letters*, *9*(7), 074012. https://doi.org/10.1088/1748-9326/9/7/074012

Meyfroidt, Patrick, & Lambin, E. F. (2011). Global Forest Transition: Prospects for an End to Deforestation. *Annual Review of Environment and Resources*, *36*(1), 343–371. https://doi.org/10.1146/annurev-environ-090710-143732

Milkoreit, M., Hodbod, J., Baggio, J., Benessaiah, K., Calderón-Contreras, R., Donges, J. F., Mathias, J.-D., Rocha, J. C., Schoon, M., & Werners, S. E. (2018). Defining tipping points for social-ecological systems scholarship—An interdisciplinary literature review. *Environmental Research Letters*, *13*(3), 033005. https://doi.org/10.1088/1748-9326/aaaa75

Müller, D., Sun, Z., Vongvisouk, T., Pflugmacher, D., Xu, J., & Mertz, O. (2014). Regime shifts limit the predictability of land-system change. *Global Environmental Change*, *28*, 75–83. https://doi.org/10.1016/j.gloenvcha.2014.06.003

Müller, R., Müller, D., Schierhorn, F., Gerold, G., & Pacheco, P. (2012). Proximate causes of deforestation in the Bolivian lowlands: An analysis of spatial dynamics. *Regional Environmental Change*, *12*(3), 445–459. https://doi.org/10.1007/s10113-011-0259-0

Narayan, P. K., & Smyth, R. (2005). Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests. *Energy Policy*, *33*(9), 1109–1116. https://doi.org/10.1016/j.enpol.2003.11.010

Nanni, A.S., Grau, H.R., 2017. Land-use redistribution compensated for ecosystem service losses derived from agriculture expansion, with mixed effects on biodiversity in a NW Argentina watershed. Forests 8 (8), 303.

Oberlack, C., Sietz, D., Bürgi Bonanomi, E., de Bremond, A., Dell'Angelo, J., Eisenack, K., Ellis, E. C., Epstein, G., Giger, M., Heinimann, A., Kimmich, C., Kok, M. T., Manuel-Navarrete, D., Messerli, P., Meyfroidt, P., Václavík, T., & Villamayor-Tomas, S. (2019). Archetype analysis in sustainability research: Meanings, motivations, and evidence-based policy making. *Ecology and Society*, *24*(2), art26. https://doi.org/10.5751/ES-10747-240226

Ogundari, K., Ademuwagun, A. A., & Ajao, O. A. (2017). Revisiting Environmental Kuznets Curve in Sub-Sahara Africa: Evidence from deforestation and all GHG emissions from agriculture. *International Journal of Social Economics*, *44*(2), 222–231. https://doi.org/10.1108/IJSE-02-2015-0034

Oldekop, J. A. (n.d.). Forest-linked livelihoods in a globalized world. *Nature Plants*, 8.

Oldekop, J. A., Rasmussen, L. V., Agrawal, A., Bebbington, A. J., Meyfroidt, P., Bengston, D. N., Blackman, A., Brooks, S., Davidson-Hunt, I., Davies, P., Dinsi, S. C., Fontana, L. B., Gumucio, T., Kumar, C., Kumar, K., Moran, D., Mwampamba, T. H., Nasi, R., Nilsson, M., … Wilson, S. J. (2020). Forest-linked livelihoods in a globalized world. *Nature Plants*, *6*(12), 1400–1407. https://doi.org/10.1038/s41477-020-00814-9

Pacheco, P. (2012). Actor and frontier types in the Brazilian Amazon: Assessing interactions and outcomes associated with frontier expansion. *Geoforum*, *43*(4), 864–874. https://doi.org/10.1016/j.geoforum.2012.02.003

Pacheco, P., Mo, K., Dudley, N., Shapiro, A., Aguilar-Amuchastegui, N., Ling,P.Y.,Anderson, C. and Marx, A. (2021). Deforestation fronts: Drivers and responses in a changing world. WWF, Gland, Switzerland.

Panayotou, T. (n.d.). *Chapter 2 economic growth and the environment*. 28.

Panayotou, T. (1993). *Empirical tests and policy analysis of environmental degradation at different stages of economic development*.

Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes* (4th ed). McGraw-Hill.

Pedroni, P. (1999). *Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors*. 18.

Pellegrini, P., & Fernández, R. J. (2018). Crop intensification, land use, and on-farm energy-use efficiency during the worldwide spread of the green revolution. *Proceedings of the National Academy of Sciences*, *115*(10), 2335–2340.

Pendrill, F., Persson, U. M., Godar, J., & Kastner, T. (2019). Deforestation displaced: Trade in forest-risk commodities and the prospects for a global forest transition. *Environmental Research Letters*, *14*(5), 055003. https://doi.org/10.1088/1748-9326/ab0d41

Perman, R., & Stern, D. I. (2003). Evidence from panel unit root and cointegration tests that the Environmental Kuznets Curve does not exist: Existence of Environmental Kuznets Curve. *Australian Journal of Agricultural and Resource Economics*, *47*(3), 325–347. https://doi.org/10.1111/1467-8489.00216

Pfaff, A., & Walker, R. (2010). Regional interdependence and forest "transitions": Substitute deforestation limits the relevance of local reversals. *Land Use Policy*, *27*(2), 119–129. https://doi.org/10.1016/j.landusepol.2009.07.010

Phalan, B., Green, R., & Balmford, A. (2014). Closing yield gaps: Perils and possibilities for biodiversity conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1639), 20120285–20120285. https://doi.org/10.1098/rstb.2012.0285

Phalan, Benjamin. (2018). What Have We Learned from the Land Sparing-sharing Model? *Sustainability*, *10*(6), 1760. https://doi.org/10.3390/su10061760

Phasouysaingam, A. (2020). *Rubber Plantation and Risk Management in Northern of Laos PDR*. *28*, 12.

Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling: MONTHLY IRRIGATED AND RAINFED CROP AREAS. *Global Biogeochemical Cycles*, *24*(1), n/a-n/a. https://doi.org/10.1029/2008GB003435

Potapov PV et al. Eastern Europe's forest cover dynamics from 1985 to 2012 quantifiedfrom the full Landsat archive. Remote Sens. Environ 159, 28–43 (2015).

Prafitri, T., Rachmina, D., & Maulana, T. N. A. (2017). The effect of working capital on the profitability of palm oil plantation companies. *Indonesian Journal of Business and Entrepreneurship (IJBE)*, *3*(2), 111-111.

Pretty, J. (2018). Intensification for redesigned and sustainable agricultural systems. *Science*, *362*(6417), eaav0294. https://doi.org/10.1126/science.aav0294

Ramankutty, N., & Coomes, O. T. (2016). Land-use regime shifts: An analytical framework and agenda for future land-use research. *Ecology and Society*, *21*(2), art1. https://doi.org/10.5751/ES-08370-210201

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000: GLOBAL AGRICULTURAL LANDS IN 2000. *Global Biogeochemical Cycles*, *22*(1), n/a-n/a. https://doi.org/10.1029/2007GB002952

Ramankutty, N., Ricciardi, V., Mehrabi, Z., & Seufert, V. (2019). Trade-offs in the performance of alternative farming systems. *Agricultural Economics*, *50*, 97-105.

Rasmussen, L. V., Coolsaet, B., Martin, A., Mertz, O., Pascual, U., Corbera, E., Dawson, N., Fisher, J. A., Franks, P., & Ryan, C.

M. (2018). Social-ecological outcomes of agricultural intensification. *Nature Sustainability*, *1*(6), 275–282. https://doi.org/10.1038/s41893-018-0070-8

Rajão, R., Soares-Filho, B., Nunes, F., Börner, J., Machado, L., Assis, D., ... & Figueira,D. (2020). The rotten apples of Brazil's agribusiness. *Science*, *369*(6501), 246-248.

Redo, D., Millington, A. C., & Hindery, D. (2011). Deforestation dynamics and policy changes in Bolivia’s post-neoliberal era. *Land Use Policy*, *28*(1), 227–241. https://doi.org/10.1016/j.landusepol.2010.06.004

Rice, J. (2007). Ecological Unequal Exchange: International Trade and Uneven Utilization of Environmental Space in the World System. *Social Forces*, *85*(3), 1369–1392.

Richards, P. (2018). It’s not just where you farm; it’s whether your neighbor does too. How agglomeration economies are shaping new agricultural landscapes. *Journal of Economic Geography*, *18*(1), 87–110. https://doi.org/10.1093/jeg/lbx009

Rindfuss, R. R., Entwisle, B., Walsh, S. J., Mena, C. F., Erlien, C. M., & Gray, C. L. (2007). Frontier Land Use Change: Synthesis, Challenges, and Next Steps. *Annals of the Association of*

American Geographers, *97*(4), 739–754. https://doi.org/10.1111/j.1467-8306.2007.00580.x

Roux, N., Kastner, T., Erb, K.-H., & Haberl, H. (2021). Does agricultural trade reduce pressure on land ecosystems? Decomposing drivers of the embodied human appropriation of net primary production. *Ecological Economics*, *181*, 106915. https://doi.org/10.1016/j.ecolecon.2020.106915

Rudel, Thomas K., Coomes, O. T., Moran, E., Achard, F., Angelsen, A., Xu, J., & Lambin, E. (2005). Forest transitions: Towards a global understanding of land use change. *Global Environmental Change*, *15*(1), 23–31. https://doi.org/10.1016/j.gloenvcha.2004.11.001

Rudel, Thomas K., Schneider, L., Uriarte, M., Turner, B. L., DeFries, R., Lawrence, D., Geoghegan, J., Hecht, S., Ickowitz, A., & Lambin, E. F. (2009). Agricultural intensification and changes in cultivated areas, 1970–2005. *Proceedings of the National Academy of Sciences*, *106*(49), 20675–20680.

Russo Lopes, Gabriela, Mairon G. Bastos Lima, and Tiago N. P. dos Reis. “Maldevelopment Revisited: Inclusiveness and Social Impacts of Soy Expansion over Brazil’s Cerrado in Matopiba.”

World Development 139 (March 1, 2021): 105316. https://doi.org/10.1016/j.worlddev.2020.105316.

Ruttan, V. W. (1997). Induced Innovation, Evolutionary Theory and Path Dependence: Sources of Technical Change. *The Economic Journal*, *107*(444), 1520–1529. https://doi.org/10.1111/j.1468-0297.1997.tb00063.x

Ruttan, V. W., & Hayami, Y. (1984). Toward a theory of induced institutional innovation. *The Journal of Development Studies*, *20*(4), 203–223. https://doi.org/10.1080/00220388408421914

Satake, A., & Rudel, T. K. (2007). Modeling the forest transition: forest scarcity and ecosystem service hypotheses. *Ecological Applications*, *17*(7), 2024–2036. https://doi.org/10.1890/07-0283.1

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53–59. https://doi.org/10.1038/nature08227

Schill, C., Anderies, J. M., Lindahl, T., Folke, C., Polasky, S., Cárdenas, J. C., Crépin, A.-S., Janssen, M. A., Norberg, J., & Schlüter, M.

(2019). A more dynamic understanding of human behaviour for the Anthropocene. *Nature Sustainability*, *2*(12), 1075–1082. https://doi.org/10.1038/s41893-019-0419-7

Schlüter, M., Baeza, A., Dressler, G., Frank, K., Groeneveld, J., Jager, W., Janssen, M. A., McAllister, R. R. J., Müller, B., Orach, K., Schwarz, N., & Wijermans, N. (2017). A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecological Economics*, *131*, 21–35. https://doi.org/10.1016/j.ecolecon.2016.08.008

Schröder, J. M., Ávila Rodríguez, L. P., & Günter, S. (2021). Research trends: Tropical dry forests: The neglected research agenda? *Forest Policy and Economics*, *122*, 102333. https://doi.org/10.1016/j.forpol.2020.102333

Sencer Atasoy, B. (2017). Testing the environmental Kuznets curve hypothesis across the U.S.: Evidence from panel mean group estimators. *Renewable and Sustainable Energy Reviews*, *77*, 731–747. https://doi.org/10.1016/j.rser.2017.04.050

Shafik, N., & Bandyopadhyay, S. (1992). Economic growth and environmental quality: Time-series and cross-country evidence. *World Bank Publications*, *Vol. 904*.

Shaver, I., Chain-Guadarrama, A., Cleary, K. A., Sanfiorenzo, A., Santiago-García, R. J., Finegan, B., Hormel, L., Sibelet, N., Vierling, L. A., & Bosque-Pérez, N. A. (2015). Coupled social and ecological outcomes of agricultural intensification in Costa Rica and the future of biodiversity conservation in tropical agricultural regions. *Global Environmental Change*, *32*, 74–86.

Sherrick, B. J., Sonka, S. T., Lamb, P. J., Mazzocco, M. A. (2000).

Decision-makerexpectations and the value of climate prediction information: conceptual considerations and preliminary evidence. Meteorological Applications(7), 377-386.

Simoes, R., Picoli, M. C. A., Camara, G., Maciel, A., Santos, L., Andrade, P. R., Sánchez, A., Ferreira, K., & Carvalho, A. (2020). Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017. *Scientific Data*, *7*(1), 34. https://doi.org/10.1038/s41597-020-0371-4

Singh, N. (2008). Transaction costs, information technology and development. *Indian Growth and Development Review*.

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*(7720), 639–643. https://doi.org/10.1038/s41586-018-0411-9

Sorrell, S. (2007). *The Rebound Effect: An assessment of the evidence for economy-wide energy savings from improved energy efficiency*. UK Energy Research Centre London.

Spera, S. A., Cohn, A. S., VanWey, L. K., Mustard, J. F., Rudorff, B. F., Risso, J., & Adami, M. (2014). Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. *Environmental Research Letters*, *9*(6), 064010. https://doi.org/10.1088/1748-9326/9/6/064010

St John, F. A. V., Edwards-Jones, G., & Jones, J. P. G. (2010). Conservation and human behaviour: Lessons from social psychology. *Wildlife Research*, *37*(8), 658. https://doi.org/10.1071/WR10032

Staal, A., Fetzer, I., Wang-Erlandsson, L., Bosmans, J. H. C., Dekker, S. C., van Nes, E. H., Rockström, J., & Tuinenburg, O. A. (2020). Hysteresis of tropical forests in the 21st century. *Nature*

Communications, *11*(1), 4978. https://doi.org/10.1038/s41467-020-18728-7

StataCorp. (2017). Stata 15 Base Reference Manual. *College Station, TX: Stata Press)*.

Stevenson, J. R., Villoria, N., Byerlee, D., Kelley, T., & Maredia, M. (2013). Green Revolution research saved an estimated 18 to 27 million hectares from being brought into agricultural production. *Proceedings of the National Academy of Sciences*, *110*(21), 8363–8368. https://doi.org/10.1073/pnas.1208065110

Stock, J. H. (1987). Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors. *Econometrica*, *55*(5), 1035. https://doi.org/10.2307/1911260

Tamburini, G., Bommarco, R., Wanger, T. C., Kremen, C., van der Heijden, M. G.,

Liebman, M., & Hallin, S. (2020). Agricultural diversification promotes multiple ecosystem services without compromising yield. *Science advances*, *6*(45), eaba1715.

Thaler, G. M., Viana, C., & Toni, F. (2019). From frontier governance to governance frontier: The political geography of Brazil's

Amazon transition. *World Development*, *114*, 59–72. https://doi.org/10.1016/j.worlddev.2018.09.022

Thomson, A. M., Ellis, E. C., Grau, Hé. R., Kuemmerle, T., Meyfroidt, P., Ramankutty, N., & Zeleke, G. (2019). Sustainable intensification in land systems: Trade-offs, scales, and contexts. *Current Opinion in Environmental Sustainability*, *38*, 37–43. https://doi.org/10.1016/j.cosust.2019.04.011

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Tilman, David, Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, *418*(6898), 671–677. https://doi.org/10.1038/nature01014

Tramberend, S., Fischer, G., Bruckner, M., & van Velthuizen, H. (2019). Our Common Cropland: Quantifying Global Agricultural Land Use from a Consumption Perspective. *Ecological Economics*, *157*, 332–341. https://doi.org/10.1016/j.ecolecon.2018.12.005

Turner, B. L., & Ali, A. S. (1996). Induced intensification: Agricultural change in Bangladesh with implications for Malthus and Boserup. *Proceedings of the National Academy of Sciences*, *93*(25), 14984–14991.

Van den Broeck, G., Swinnen, J., & Maertens, M. (2017). Global value chains, large-scale farming, and poverty: Long-term effects in Senegal. *Food Policy*, *66*, 97–107. https://doi.org/10.1016/j.foodpol.2016.12.003

Verhofstadt, E., & Maertens, M. (2014). Smallholder cooperatives and agricultural performance in Rwanda: Do organizational differences matter? *Agricultural Economics*, *45*(S1), 39–52. https://doi.org/10.1111/agec.12128

Verburg, P. H., Dearing, J. A., Dyke, J. G., Van Der Leeuw, S., Seitzinger, S., Steffen, W., & Syvitski, J. (2016). Methods and approaches to modelling the Anthropocene. *Global Environmental Change*, *39*, 328-340.

Villoria, N. (2019a). Consequences of agricultural total factor productivity growth for the sustainability of global farming: accounting for direct and indirect land use effects. *Environmental Research Letters*, *14*(12), 125002.

Villoria, N. B. (2019b). Technology spillovers and land use change: empirical evidence from global agriculture. *American Journal of Agricultural Economics*, *101*(3), 870-893.

Villoria, N. B., Byerlee, D., & Stevenson, J. (2014). The Effects of Agricultural Technological Progress on Deforestation: What Do We Really Know? *Applied Economic Perspectives and Policy*, *36*(2), 211–237. https://doi.org/10.1093/aepp/ppu005

WB. (2018). *World Development Indicators. World Bank.*

Waggoner PE, Ausubel JH. 2001. How much will feeding more and wealthier people encroach on forests? *Popul. Dev. Rev.* 27:239–57

Weber, E. U. (1997). Perception and expectation of climate change: Precondition foreconomic and technological adaptation. In M. Bazerman, D. Messick, A. Tenbrunsel , Wade-Benzoni, K.

(Eds.), Psychological and Ethical Perspectives to Environmental and Ethical Issues in Managemen. San Francisco: Jossey-Bass, 314-341.

Westerlund, J. (2005). New Simple Tests for Panel Cointegration. *Econometric Reviews*, *24*(3), 297–316. https://doi.org/10.1080/07474930500243019

Wolfersberger, J., Delacote, P., & Garcia, S. (2015). An empirical analysis of forest transition and land-use change in developing countries. *Ecological Economics*, *119*, 241–251. https://doi.org/10.1016/j.ecolecon.2015.08.018

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th ed). South Western, Cengage Learning.

Worldwide Governance Indicators. (n.d.). www.govindicators.org.

Xiao, C. (2020). *Sentinel-2 red-edge spectral indices (RESI) suitability for mapping rubber boom in Luang Namtha Province, northern Lao PDR*. 11.

Xin, Y., Sun, L., & Hansen, M. C. (2021). Biophysical and socioeconomic drivers of oil palm expansion in Indonesia.

Environmental Research Letters, *16*(3), 034048. https://doi.org/10.1088/1748-9326/abce83

Yao, G., Hertel, T. W., & Taheripour, F. (2018a). Economic drivers of telecoupling and terrestrial carbon fluxes in the global soybean complex. *Global Environmental Change*, *50*, 190–200. https://doi.org/10.1016/j.gloenvcha.2018.04.005

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., … Zeng, N. (2016). Greening of the Earth and its drivers. *Nature Climate Change*, *6*(8), 791–795. https://doi.org/10.1038/nclimate3004

Zomer, R. J., Neufeldt, H., Xu, J., Ahrends, A., Bossio, D., Trabucco, A., van Noordwijk, M., & Wang, M. (2016). Global Tree Cover and Biomass Carbon on Agricultural Land: The contribution of agroforestry to global and national carbon budgets. *Scientific Reports*, *6*(1), 29987. https://doi.org/10.1038/srep29987

Zoundi, Z. (2017). CO2 emissions, renewable energy and the Environmental Kuznets Curve, a panel cointegration approach.

Renewable and Sustainable Energy Reviews, *72*, 1067–1075. https://doi.org/10.1016/j.rser.2016.10.018

zu Ermgassen, E. K. H. J., Godar, J., Lathuillière, M. J., Löfgren, P., Gardner, T., Vasconcelos, A., & Meyfroidt, P. (2020). The origin, supply chain, and deforestation risk of Brazil's beef exports. *Proceedings of the National Academy of Sciences*, *117*(50), 31770–31779. https://doi.org/10.1073/pnas.2003270117