Norwegian University of Life Sciences

Master's Thesis 2017 30 ECTS
Faculty of social sciences
School of Economics and Business
Supervisor: Arild Angelsen

Food vs. non-food crops:

Changes in areas and yields 1992 to 2016.

Henrik Liadal Reinertsen
Master of science in Economics

Abstract

Changing demand for different agricultural products causes shifts in land use. Until recently, food production was the main agricultural practice for most countries. Today, producing non-food crops, exclusively is on the rise. This growth, along with increasing per-capita food consumption, will require large increases in crop production. However, agricultural productive land is scarce and increase in demand are modified by the yield increase. This thesis aims to explore links between production of staple crops and demand for non-food crops and discusses if higher yield enhances or reduces total agricultural area.

Following a presentation on the literature is a discussion on theories of land use, introducing von Thünen's theory of land rent. Considering this theory, I present the debate on the impacts of higher agricultural productivity, known as the Borlaug hypothesis vs. the Jevons paradox. To dig deeper into on our research question, a dataset based on the UN Food and Agricultural Organization database and the World Bank Databank was constructed. Previous information on the distribution and performance of specific crops have only been available through remote sensing. However, a new detailed dataset, where the distribution of crops' area of usage is taken into consideration, improves the analysis environmental impacts and trends in agricultural land use.

In this thesis, descriptive statistics and regressions analysis indicates two major findings. First, expansion of agricultural area from 1992-2016 has mainly been caused by increase in feed crop area, however staple crops area has also been a large contributor, especially in low/ middle-income countries. Further, feed-, fuel- and non-food crops has experienced the largest relative growth, indicating the direction of trends in agricultural production. Second, higher staple-crop yield reduces crop area, as suggested by the Borlaug hypothesis, while higher feed -and fuel crop yield increases crop area, as suggested by the Jevons paradox.

Keywords: agriculture | yield | von Thünen | Borlaug hypothesis | Jevons paradox | panel data

Sammendrag

Endringer i etterspørselen for jordbruksprodukter fører til endringer i arealbruk. Frem til nylig var matproduksjon den viktigste landbrukspraksisen i de fleste land. I dag er det imidlertid en tendens til å produsere avlinger utelukkende for markedet. Denne veksten, sammen med økende matforbruk per person, krever stor vekst i avlinger. Egnet landbruk areal er imidlertid en mangelvare, og effekten av økt etterspørsel er modifisert av produktiviteten i landbruket. Denne oppgaven analyserer sammenhenger mellom produksjon av de viktigste avlingene og etterspørsel etter avlinger produsert utelukkende for markedet, og diskuterer om høyere jordbruksproduktivitet øker eller reduserer totalt landbruksareal.

Etter en presentasjon på litteraturen presenteres landbruksteorier som kombinerer von Thünen's teori om grunnrente. I lys av denne teorien presenterer jeg debatten om virkningen av høyere jordbruksproduktivitet, kjent som Borlaug-hypotesen vs. Jevons paradoks. For å gjennomføre detaljerte analyser ble det laget et datasett basert på data fra FNs Mat og Landbruksorganisasjon (FAO) og Verdensbanken. Tidligere studier på distribusjonen og produktiviteten av spesifikke avlinger har kun vært tilgjengelig gjennom fjernmåling. Et nytt og detaljert datasett, hvor bruksområdene til avlingene er tatt i betrakting, har derfor blitt ekstrahert for å forbedre kvaliteten og mulighetene til å analysere miljøpåvirkningene og endringer i landbruket.

Analysen gir to hovedfunn. For det første er utvidelsen av landbruksarealet fra 1992-2016 hovedsakelig forårsaket av økning i produksjon av dyrefôr, i tillegg til basis-matvarer, spesielt i lav- og mellominntektsland. Videre har dyrefôr, biodrivstoff, og ikke matvarer opplevd den største prosentvise veksten i jordbruksareal. For det andre vil høyere produktivitet for basis-matvarer redusere jordbruksarealet, som foreslått i Borlaug-hypotesen, mens høyere produktivitet for fôr- og biodrivstoff gir økt jordbruksareal, som foreslått av Jevons paradoks.

Acknowledgement

After completing a bachelor's degree in International environmental and development studies, environment and resource economics naturally caught my interest as I entered the field of economics at the Norwegian University of Life Science. In addition to general interest of the subjects, the opportunity to work with both national and international issues – and contributing to future sustainable development - motivated me. Resource economics thereby became a natural choice of subject for my master thesis.

Today, I would like to express my sincere thanks and appreciation all those who have advised and helped me in various ways in the completion of this thesis. A special thanks to my thesis supervisor Arild Angelsen who has been an inspiration throughout my whole Master's degree and who first suggested the idea for my thesis. I would also like to thank Sarah Ephrid Tione for her assistance in econometrics and guidance in STATA.

Finally, I would like to thank my family, friends, girlfriend and Brotherhood Tuck for all the motivation, love, support, and good times through my whole study career.

Ås, May 15th, 2018.

____

Henrik Liadal Reinertsen

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Changing demand for different agricultural products causes shifts in land use. Until recently, food production was the main agricultural practice in most countries. Today, however, producing crops for non-food use is on the rise. In 2016, only 55 % of food crops calories ended up directly on our tables, whereas 36 % were used as livestock feed and 14 % ended up as biofuel or other cash products (Foley, 2016). Changing demand and new market opportunities create large land use changes and an expansion of non-food crops.

Is this a matter of concern? Economic development theorists have suggested that to enhance economic growth, countries should develop strategies in favor of non-food products at the expense of food production. This is because demand for non-food products increases with more wealth, relative to demand for food products (Engel's law) (Baffes and Etienne, 2014). However, expanding agricultural land create other concerns, causing deforestation and carbon emissions. While producing to feed its own population may dominate climate in policy making, these concerns are strongest for cash crops (Wiggins et al., 2015).

The link between demand and area change is modified in two ways. First, increased demand is modified by the basic price mechanisms: higher demand partially increases production, and partially the price, with the final production increase being determined by the demand and supply elasticities. Second, area change is modified by change in land productivity (=yield), given by the identity: production = yield * area (FAO, 2017, Ray et al., 2013, Ewers et al., 2009).

Prior to the 20th century and the Green Revolution, almost all increase in food production was obtained by bringing new land into production. But agricultural land is scarce and our ability to supply food, feed and fuel, while maintaining environmental services depends on our cultivation practices. Two different outcomes of yield-enhancing policies are suggested: agricultural expansion and deforestation is reduced, or agricultural encroachment is stimulated. Although it is intuitive that intensification to increase production on existing cropland is the best way of reducing agricultural encroachment, this is not necessarily accepted scientifically (Byerlee et al., 2014). While investments in staple food crops has resulted in net-land saving of 20-30 million hectares (Evenson and Rosegrant, 2003, Stevenson et al., 2013), extensification of feed -and fuel crops, at the expense of pastures and natural vegetation, has been major drivers of agricultural expansion and environmental degradation (Nepstad and Stickler, 2008).

1.2 Problem statement and hypothesis

This thesis aims to answer the following three research questions:

1. *What are the trends of area changes for non-food crops and staple crops?*
2. *Does higher yield of staple crops and non-food crops enhance or reduce total agricultural area?*

We will compare and study developments of five crop groups to identify linkages between agricultural area and productivity. Based on FAO classifications and Ewers et al. (2009), the five crop groups are as follows:

1. *Staple crops* (barley, maize, sorghum, rice, wheat, cassava, potato, soybean, taro, yam, banana, plantain)
2. *Non-staple food crops* (coffee, sugar cane, sugar beet, tea, cocoa, sweet potato, cottonseed, tomatoes, watermelon, onions, apples, cucumbers, grapes, oranges, green bean, chickpea, lentil, cow pea, pigeon pea, brassica, millet, sunflower seed, coconut, groundnut, olives)
3. *Feed crops* (rye, oats, green maize)
4. *Fuel crops* (rapeseed, palm oil)
5. *Non-food crops* (rubber, jute, tobacco)

(FAOSTAT, 2018a, Ewers et al., 2009)

We use detailed data of agricultural production for the period 1992-2016 for ten of the largest producing countries of staple crops. The study was inspired by Ramankutty et al. (2008) and Monfreda et al. (2008), who studied geographical distribution of agricultural lands and introduced the importance of assessing the consequences of agricultural expansion and intensification. Higher yield is commonly believed to reduce expansion of agricultural area. Considering this, we are also particularly interested to see if higher yield of staple crops and non-food crops has different impacts on total agricultural area. Studying the effects of increased agricultural productivity was inspired by Angelsen (2017) and Byerlee (2013), who introduced the debate between the Jevons paradox and the Borlaug hypothesis. Angelsen and Kaimowitz (2001) finds that there are examples of both expanding agricultural areas caused by increased agricultural productivity, and vice-versa.

Hence, the following hypotheses are put forward:

- H1: The increase in total agricultural area is driven by fuel and non-food crop, not staples.
- H2: The yield area relationship differs between crop types. (H2.1) For staple food, higher yield results in lower crop area; (H2.2) for non-food and fuel crops, higher yield results in larger crop area.

1.3 Structure

The thesis has six chapters. Following this general introduction, Chapter 2 present the justification of the research topic. Chapter 3 reviews the literature and theoretical framework to explain why the production of certain types of crops are becoming more popular than others, and the different outcomes of higher agricultural productivity. Chapter 4 introduces the data and variables used, before the methods are presented. Chapter 5 presents and discusses the results of the analysis. Limitations of the analysis are briefly discussed at the end of this chapter, before we end with a conclusion and a summary.

2 Background to topic

This chapter will present the necessary background knowledge for this thesis. First, I give a general background on the role of agriculture in economic development and the historical perspectives on agricultural production. Second, I discuss briefly the distinction between staple crops and crops grown for sale to return profit - hereafter referred to as “cash crops”, as well as why farmers may prefer one over the other.

2.1 The role of agriculture in economic development

The majority of people living in rural areas in poor countries depend on agriculture for their livelihood, directly or indirectly, yet information on distribution of specific crops are limited (Anderson et al., 2014, World Bank, 2008). To study resource management and land degradation scientists has been using remote sensing analysis, i.e., obtaining information on crops typically aircrafts or satellites (NOAA, 2017). Supporters of remote sensing analysis argue that, while subnational statistics provide limited information on cropland trends, remote sensing has proved capable of providing reliable data on a timely basis to a fraction of the cost of traditional methods of collecting data (Nellis et al., 2009, Anderson et al., 2014). However, critics say that remote sensing products are ill suited for many applications due to insufficient resolution and lack of dependable and consistent remote sensing systems (Nellis et al., 2009). Nevertheless, to examine agricultural trends, recent studies have incorporated data from remote sensing with available information from statistical surveys[1] (Fischer et al., 2012, Leff et al., 2004, Monfreda et al., 2008, Portmann et al., 2010, You et al., 2014, Ramankutty et al., 2008)

Besides providing food, agriculture also offer business opportunities through high-value products in both domestic and international markets. Thanks to market liberalization and technological change, agriculture has become a major cause of economic growth and poverty reduction for many countries. For example in China, rapid growth in agriculture has caused rural poverty to drop 45% in a few years (World Bank, 2008). Additionally, traditional crops have recently obtained greater attention through expanded areas of usage, such as feed-maize exports and sugar cane for biofuels (De la Torre Ugarte, 2006).

[1] Presentation of empirical studies using remote sensing in Appendix D

However, agriculture has not always been considered an engine of growth (Tiffin and Irz, 2006). In 1857, Ernst Engel observed that poor families spent a greater proportion of their total expenditure on food, rather than manufactured goods. He concluded therefore that the wealthier the nation, the smaller the proportion of food to total expenditure. This became known as Engel's Law, caused by less than unitary income elasticity of food commodities (Figure 2.1) (Laitner, 2000, Baffes and Etienne, 2014).

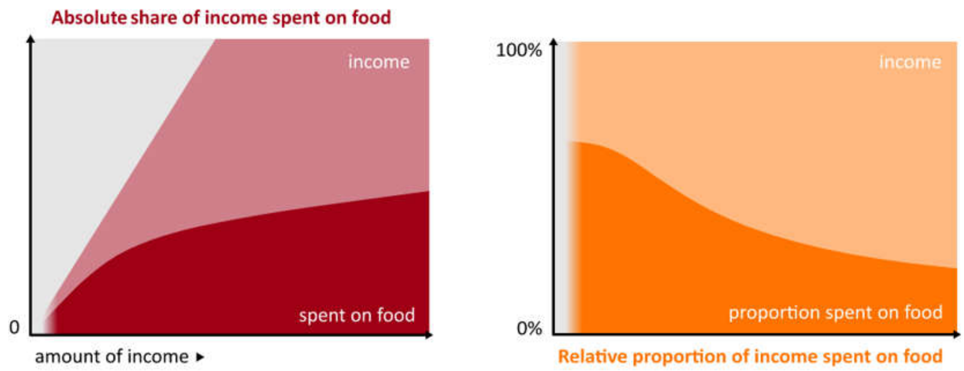


Figure 2.1: Engels law (Kraft, 2018).

Engel's observations formed several competing views attempting to explain and forecast the long-term behavior of the terms-of-trade (ToT) faced by developing countries. The most important view argued, "ToT will follow a downward path because income growth leads to smaller demand increases in primary commodities than manufacture products" (Baffes and Etienne, 2014, p. 2), an outcome which is consistent with Engel's law. Several decades later, the American economist Charles Kindleberger argued, "the ToT move against agricultural and raw material countries as the world's standard of living increases and as Engels's law of consumption operates. The elasticity of demand for wheat, cotton, sugar, coffee, and bananas is low with respect for income" (Kindleberger, 1943, p. 349). Countries should therefore switch from production of food crops to products with higher income elasticity, such as biofuels, to promote economic growth. Kindleberger statements were later empirically tested by Raul Prebisch and Hans Singer and is expressed as the "Prebisch-Singer thesis". The thesis states that over time, the ToT would turn against countries who exported primary goods and imported manufactures (Eicher and Staatz, 1998). Countries should therefore base their development strategies on import substitution of manufactured goods rather than promotion of agricultural exports, as economic growth provided by agriculture and other primary exports is very limited (Cuddington et al., 2002, Eicher and Staatz, 1998).

An alternative to this approach is the urban-industrial impact model, better known as the location model. The location model was first formulated in Germany by J.H. von Thünen to explain geographic variations in the location and production of agricultural commodities in an industrializing economy[2]. It suggests that industrial development simulate agricultural development by increasing demand for farm products, higher product prices, land values, and rates of land use (Eicher and Staatz, 1998). Later it was used to explain the performance of factor and product markets linking agriculture and non-agriculture sectors in regions of rapid urban-industrial development (Ruttan and Hayami, 1972).

In the 1960s, led by the "Father of the Green Revolution", Norman Borlaug, the perception of agriculture as a tool for development changed. Through new technologies, based on fertilizer-responsive grain cultivars and high-yielding varieties of crops, the agricultural sector was able to provide employment for the growing rural labor force, while simultaneously, provide wages to expand the industrial labor force. Hence, it became possible to achieve both employment and economical profits from the agricultural sector (Eicher and Staatz, 1998). However, with rising resource scarcity and worries of environmental loss, concerns about the present agricultural practices advanced. Land is scarce and staple crop yields are not growing fast enough to provide food for a growing population. Hence, our ability to supply the growing demand for food, feed and fuel, while maintaining the current landscape of environmental services lies on our ability to produce more on less land. Only through more efficient production of staple crops, cash crops and non-food crops, are countries going to be able to feed its population while making economic progress (World Bank, 2008).

2.2 Staple crops and cash crops

Production of staple crops is not up to speed to satisfy the rapidly growing population (FAO, 2017, Harvey, 2013, Ray et al., 2013). Staple crops are food that is eaten regularly, even daily, and in such quantities that it constitutes the dominant part of a population's diet. Most people in the world live on a diet based on at least one or more of the following crops: rice, wheat, maize, barley, sorghum, cassava, potato, taro, yam, banana, plantain and soybeans. Together, these crops provide close to 90 % of world's food energy intake, whereas more than 50 % of these comes from only three "mega-crops": rice, wheat and maize (IDRC, 2010, O'Connor, 2014). To satisfy the rapidly growing population, and changes in per capita consumption and diets, global staple crop production needs to double by 2050. However, a study shows that the three "mega-crops" only increases by 0.9 % to 1.6 % a year, far below what is needed to meet

[2] Discussed further in Chapter 3

projected demands for 2050. Yield improvements are insufficient to keep up with the (project) demand increase (Ray et al., 2013).

The food sector is less competitive on the international market than non-food crops. Even with globalization, much of the staple crop sector remains largely non-tradable, producing mainly for the domestic market (Figure 2.2) (World Bank, 2008). A World Bank (2008) report, "Agriculture for development", argues that staple crops such as cassava, yams and sorghum are rarely traded on the international market due to low international prices and trade barriers such as formal trade barriers, poor infrastructure, high transportation - and marketing costs and trade policies favoring cash crops. This is especially the case for rural areas and land-locked countries, where the is more isolated from the global markets, and exports often unprofitable. Because food prices are inelastic, production is price sensitive and therefore less tempting for entering the international market (World Bank, 2008).

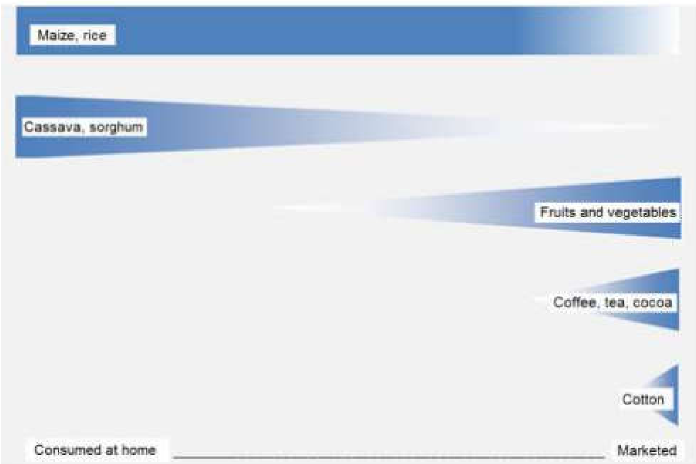


Figure 2.2: Distribution of domestically and internationally consumed staple food crops and cash crops. On the x-axis, "consumed at home" represents staple crops and "marketed" represents cash crops. Source: (Achterbosch et al., 2014, p. 19).

Thus, the poverty-reducing effect of producing staple crops depend on the net marketing position of the poor and the price elasticity of food demand. In countries where staple crops constitute the majority of crop production and is non-tradable, gains in staple crop production increases aggregate food supply and reduces food price. Consequently, wages of unskilled workers as well as prices of inputs are kept low, making the non-food tradable sector, such as biofuel, more competitive. However, increasing staple crop productivity usually reduces overall poverty as more than half of poor rural households are net food buyers (World Bank, 2008, CEPR, 2010).

The term “cash crops” is defined as crops grown for sale in an agricultural market, and can be linked to a process of agricultural commercialization (Achterbosch et al., 2014, p. 7). It implies strengthened market orientation in farming where food-crops and non-food crops are primarily produced for selling at the domestic or international market, including major export crops such as, cocoa, rubber, palm oil, tea and tobacco (Barbier, 1989). In earlier times, cash crops constituted a small part of farm’s total yield, while today, producing the majority of crops for sale is trending in developed countries (World Bank, 2008). Cash crops are favored for their potential contribution to growth, employment and trade balance, and can provide the basis of industrial development through intersectoral linkages. On the other hand, cash crops have also been criticized for additional drawbacks in terms of crop dependency and food security, especially in relation to the “food-first” movement. In contrast to staple crop producers, however, cash crops producers are benefiting from globalization because of larger markets and increased competition (World Bank, 2008). Recently, high energy prices and desire to mitigate climate change has generated new markets for agriculture in terms of production of biofuel. However, in the long run, cash crops have been associated with secular incline in the terms of trade, i.e., improving comparative advantage. This argument is used to suggest that resources should be transferred out of agriculture and into manufacturing (Prebisch-Singer thesis) (Maxwell and Fernando, 1989).

3 Literature Review and Theoretical Framework

The most relevant theory and theoretical framework on agricultural area and yield trends will be presented in this chapter, based on Angelsen (2007), Angelsen and Kaimowitz (2001), Alcott (2005), Borlaug (2007), Byerlee et al. (2014), and McNally et al. (2014). First, the von Thünen theory of land rent is presented. This part also discusses whether increased agricultural productivity enhances forest preservation or encourages agricultural encroachment. This is known as the Borlaug hypothesis vs. the Jevons paradox. Second, I present theories on the political economy of land use, looking at how political factors, rather than market factors, influences changes in land use.

3.1 Theories of land use

The challenge of meeting the growing food demand in a world of limited suitable land and degrading environment have raised questions of land use changes and awareness of a controversy between increased yield and increased area of production (Ray et al., 2013, Edgerton, 2009). According to the von Thünen model, land is allocated to the use which yields the highest rent, where rent is determined by location. Hence, shifts in land use become a question of changes in rent of forests vs. rent of agriculture.

The economics of land use argues that land is allocated to the use with the highest land rent (surplus or profit). Several factors such as crops prices, input costs, technologies, etc., determines the rent of alternative land. Many of these depending directly or independently on location of land (Angelsen, 2007). In his 1826 seminal work, *The Isolated State*, Johann von Thünen examined how land rent, as determined by distance to commercial center (the city), shapes land uses. He asked, "Under these conditions what kind of agriculture will develop and how will the distance to the city affect the use of land if this is chosen with the utmost rationality?". Where the "utmost rationality" assumption implies that land with highest yield is chosen first (Angelsen, 2007, Von Thünen, 1966). To model how land rents can stimulate returns to alternative land uses, in a context where agriculture is well integrated in markets, we assume two kinds of land-use: agriculture and forest (Stevenson et al., 2011).

Agricultural rent

Land rent from agricultural activities (r) is modeled as a function of distance (d) from the market as follows:

(1) $r^a = p^a y^a \quad wl^a \quad qk^a \quad v^a d,$

where (p^a) is the price at which the products are sold in a central market, (y^a) is the agricultural production per hectare (yield), the labor (l^a) and capital (k^a) required per hectare is fixed, with inputs including wages (w) and annual costs of capital (q). The fixed input, wage, implies that labor can move freely in and out of the agricultural sector. Transportation costs is the sum of the costs per kilometer (v^a) and the distance from the central market (d) (Angelsen, 2010). As distance to market increases, agricultural rent decreases, and the agricultural frontier is where $r^a = 0$, that is, where agricultural expansion is no longer profitable. The frontier is thus defined by $d = (p^a y^a \quad wl^a \quad qk^a)/v^a)$, and is presented below (Figure 3.1).

Ignoring forest rent, deforestation takes place up to distance A, depending on agricultural prices, road quality and off-farm employment opportunities (Angelsen and Kaimowitz, 1999). Higher output prices and better technology that increases yield or reduces costs, as well as better road quality and shortage of off-farm employments, are drivers of agricultural expansion and deforestation. Hence, the agricultural rent curve moves to the right. Higher wages, reflecting the costs of hiring labor or the use of family labor work in the opposite direction. (Angelsen, 2010).

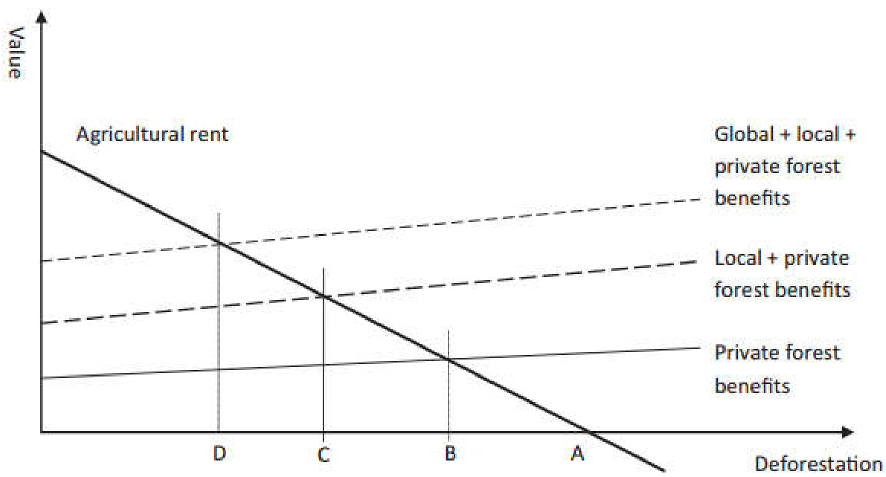


Figure 3.1: Agricultural- and forest rents and forest rent capture. Source: Angelsen (2010)

As displayed in Figure 3.1, higher forest rent and lower agricultural rent lead to less forest being put under agricultural use. To reduce agricultural rent, three agricultural policies are possible (Angelsen, 2010). First, referred to as the “improved Gabonese recipe” for forest conservation, heavy taxation on agricultural exports, neglect of rural roads and limited support for smallholders were central to reduce agricultural rent and preserve forest. Second, an important extension of the simple von Thünen model is to differentiate between extensive and intensive agriculture, where “intensive” is understood as improved productive inputs. Hence, encouraging higher supply from intensive agricultural can been be considered a forest conservation policy because it will pull labor from the extensive sector and thereby reduce extensive agriculture rent. However, the outcome is not guaranteed in cases where the dominant crop in the intensive sector is internationally traded and if improved technology in the intensive sector halter the labor pull effect. Third, where tenure is endogenous and property rights are weak, deforestation becomes a strategy to declare ownership, and forest is cleared prematurely to establish property rights. Hence, establishing clear property rights that reduces agricultural rents are proven to be important for forest conservation (Angelsen, 2010).

Forest rent

Forest rent aims to reflect the value of products and services generated by standing forests. There are three main types of forest rent: first, private forest products, such as timer and a large number of nontimber forest products (NTFP) (*extractive* forest rent), second, local public goods, such as water catchment and pollination services (*protective* forest rent); and third, carbon sequestration and storage and biodiversity maintenance (protective forest rent) (Angelsen, 2010). Forest rent can be written as

(2) $$r^f = (p^t y^t \quad w l^t \quad q k^t \quad v^t d) + p^l y^l + p^g y^g$$

The extractive rent increases when prices of timber and NTFP (p^t) is high, technological progress (y^t, l^t, k^t); and lower costs of labor (w), capital (q), and transportation (v^t). In contrast, protective rent increases when prices of local (p^l) and global (p^g) public goods is high, and will therefore lead to less forest being turned over to agricultural use (Angelsen, 2010). Historically, forest scarcity path has been linked to higher *extractive* forest rent, however, in the future it could be driven by *protective* forest rent (Angelsen, 2010). Because forest is a public good in nature, increase in protective forest rent does not impact deforestation unless land users can capture some share of it. This can be done by internalizing the externalities by moving land use decisions to a higher scale and creating a market for public goods. One key factor is the difference between managed forests and open access forests. For managed forests, there are certain costs in terms of defining and enforcing property rights. Beyond a certain distance, the rent becomes lower than the costs and open access takes over (Angelsen, 2007). Hence, it would be easier for agriculture to expand into open access forests, as there are no managed forests to act as buffers. However, such outcome depends on how that rent is being captured.

Protected areas, institutional mechanisms and payment mechanisms enable land users to capture higher share of the local and global benefits provided by forests (Angelsen, 2010). Large tropical forests are characterized by weak, unclear property rights, making them open access in reality. Hence, land users have no incentives to include forest rent in their decisions. In Figure 3.1, if private properties are established, we move from point A to point B. At this point, some forest rent is captured by users and more forest is conserved. A popular argument for reducing deforestation is by establishing individual property rights. However, this will not solve the problem of local and global goods. Establishing clear and secure individual- or community level property (point C, Figure 3.1) is therefore necessary for the establishment of systems for payments for environmental services (PES) and global forest rent capturing system (point D. Figure 3.1). This will encourage more sustainable management than open access with positive effects on degradation. Lastly, various types of protected areas (PAs) have significantly reduced deforestation. Successful PAs are expected to have similar effects on agricultural yield as policies of property rights, however because there is less assurance that the least productive land is saved for agriculture, PAs lead to high land use rents, but can also lead to higher loss of agricultural production per hectare forest saved (Angelsen, 2010).

Basic model with two sectors

Now, consider a market where both intensive and extensive agriculture sector (land use) produces the same product. The technological change in the intensive sector increases market supply. How will that affect the land rent of the extensive sector, and thereby the agricultural frontier and deforestation? (Angelsen, 2007). The answer can be found by considering the inverse price elasticity of demand. Let x_1 be the output from the intensive sector, and x_2 be the output from the extensive one (and $x = x_1 + x_2$). Hence,

(3) $$\frac{1}{e_1} = \frac{\partial p}{x_1}\frac{x_1}{p} = \frac{\partial p}{x_1}\frac{x_1}{p}\frac{x}{x} = \frac{\partial p}{x_1}\frac{x_1}{p}\frac{x_1}{x} = \frac{1}{e}\frac{x_1}{x}$$

Except for the extreme cases of demand being perfectly elastic, an increase in supply in the intensive sector (x_1), due to change in technology, will result in a drop in the price, and thereby a contraction of the extensive sector. The magnitude of price reduction, however, depends on the market's demand elasticity (e) and share of intensive factor in overall production $\left(\frac{X1}{X}\right)$. If $\left(\frac{X1}{X}\right)$ is high and (e) is low, this becomes in line with Norman Borlaug, the "father of the Green Revolution", and his arguments of increasing yield as a tool for land sparing, sometimes referred to as the Borlaug hypothesis.

Borlaug Hypothesis

Historically, cropland expansion has been the major source of growth in agricultural productions. Due to technological improvements, yield-increase, rather than area expansion, has allowed for the increase of world food demand to be met without increasing existing cropland (Byerlee et al., 2014). Norman Borlaug argued that increasing productivity of agriculture is the best farmland can do to control deforestation and reduce demand for new farmland. When debating critics of the Green revolution, he stated "If the global cereal yields of 1950 still prevailed in 2000, we would have needed nearly 1.2 billion more hectares of the same quality, instead of the 660 million hectares used, to achieve 2000's global harvest. Moreover, had environmentally fragile land been brought into agricultural production, the soil erosion, loss of forests and grasslands, reduction in biodiversity, and extinction of wildlife species would have been disastrous" (Borlaug, 2007, p. 359). Hence, increases in yield are saving new agricultural land and ecosystems from exploitation by maintaining natural areas (Stevenson et al., 2011).

To assess Borlaug's arguments, Angelsen (2010) uses a simple identity, known as the global food equation (GFE) which links population and food consumption per capita with agricultural yield and land area:

(4) $$Pop \ \left(\frac{Food}{Pop}\right) \equiv \left(\frac{Food}{Ag\ land}\right) \ Ag\ land$$

The equation states that, without an increase in yield, agricultural area must expand to feed a growing population and meet growing food consumption. Hence, GFE has been used to argue for the Borlaug hypothesis (Angelsen and Kaimowitz, 2001). Using this equation, Balford and co-authors estimated that agricultural land in developing countries will increase by 2-49% between 2000 and 2050, depending on assumptions of population growth. (Balmford et al., 2005). However, it is important to clarify that the estimates of land saving do not consider effects of prices. In cases where yield does not increase, food prices would have increased and altering the agricultural expansion. Additionally, the GFE does not create any direct link between agricultural areas and forest areas, nor does it account for trade between countries and the fact that much of agricultural production is non-food crops. By further decomposing GFE, Angelsen (2010) develops a national deforestation equation (NDE):

$$(5) \qquad Pop \ \left(\frac{Food\ cons}{Pop}\right) \equiv \left(\frac{Food\ cons}{Food\ prod}\right) \ \left(\frac{Food\ prod}{Ag\ prod}\right) \ \left(\frac{Ag\ prod}{Ag\ land}\right) \ \left(\frac{Ag\ land}{Forest}\right) \ Forest$$

Or

$$(6) \qquad deforestation \approx pop\ growt \ + \ food\ cons\ per\ capita$$
$$sufficency\ ratio\ (inverse) \qquad food\ s\ \ are$$
$$ag\ yield \qquad \left(\frac{ag}{forest\ ratio}\right)$$

Agricultural yield is one of many factors affecting deforestation, and changes in yield have an indirect effect on these factors. First, international trade of agricultural products has increasingly become a larger part of a country's economy. Higher yields in developing countries have boosted their competitiveness and raised self-sufficiency (Francois et al., 2005). Second, with increasing popularity of crops being produced for non-food purposes, such as biofuel, a lower share of agricultural output is being consumed as food, and deforestation enhanced for farmers to produce for profits. Third, land is not only being exploited for forest, cropland and pasture. Large areas of fallow, savannah, bush and other land categories are available for agricultural expansion. For example, Waggoner and Ausubel (2001) finds that changes in cropland and forest area remained uncorrelated between 1900 and 1995. Hence,   (ag. / forest) ratio is not stable. Other potential impacts of increased agricultural yield include the price effect on food consumption per capita and increased population because of increased food consumption (Malthusian effect). (Angelsen, 2010). Angelsen (2010) concludes that the GFE and NDE are useful in producing a consistent accounting network but are potentially dangerous as policy analysis as they ignore how yield change influences other factors through behavioral and market changes.

Jevons Paradox

When new innovation improve agricultural productivity or reduce costs for producers compared to other land-use practices, agricultural expansion is encouraged (Alcott, 2005, Stevenson et al., 2011). Therefore, it exists an apparent paradox that the adoption of technology to prevent agricultural expansion, could, under some circumstances, lead to the opposite of what is previously suggested. The general principle of the paradox was introduced by William Jevons in the context of coal and its relationship to new technologies: "Economy multiplies the value and efficiency of our chief material...[and] renders the employment of coal more profitable, and thus the present demand for coal is increased. . .. [If] the quantity of coal used in a blast-furnace, for instance, be diminished in comparison with the yield, the profits of the trade will increase, new capital will be attracted, the price of pig iron will fall, but the demand for it increases and eventually the greater number of furnaces will more than make up for the diminished consumption of each" (Alcott, 2005, p. 13).

In a von Thünen framework, higher yield can lead to agricultural expansion when agricultural rent is larger than forest rent. New agricultural technologies that free up labor and reduces costs can cause agriculture to become more profitable than forest, and could therefore, under some circumstances, lead to area expansion. This paradox is supported by the work by Angelsen and Kaimowitz (2001), which incorporates a number of local and national case-studies supporting these results.

3.1.1 Determining land expansion outcomes:

To sum up, Angelsen and Kaimowitz (2008) suggested three characteristics critical for determining land expansion outcomes from increased yield:

I. *Types of technologies:*

Farmers are capital and labor constrained, hence if labor -or capital-intensive technologies are implemented, land expansion tend to be constrained. However, farmers only tend to adapt such land-saving practices when land has become scarce and most of the forest is gone (Kaimowitz and Angelsen, 2008). Thus, even though new technologies will increase yield and profitability, it could either increase production at the existing land area, or provide incentives to expand crop and pasture areas (Angelsen, 2017).

II. *Output markets:*

Demand elasticity in the market and market share of the sector determine the magnitude of price effects. When yield increases, food prices decrease, hence, farmers income is reduced, but poorer consumers benefit. The price-dampening effect can be limited because total market demand is inelastic, or because its market share is low, or both (Angelsen, 2017). Further, the land expansion effects are most likely to be greatest in regions of small scale farmers with relatively low yields and high land supply elasticities (Hertel, 2012).

III. *Scale and sector adoption:*

The scale and sector adoption of new technologies is critical for the supply increase and price-dampening effects. The Green Revolution is a form of large-scale technological progress, which have had both local win-lose situations in forms of local negative impacts from decreasing prices of rice, maize and other staple crops, and global win-win effects in forms of saving major forestlands. However, looking at the labor market effects, the Green Revolution may also be linked to deforestation as some sectors are becoming less labor-intensive, freeing up labor and increasing profitability of investments in forest-clearing initiatives (Ruf, 2001). Land expansion outcomes of technological change can therefore be mixed, but to achieve win-win outcome, new government policies that can compete with agricultural land expansion must be introduced. Though, land expansion must also be considered an option to feed the world, though, with high environmental cost to biodiversity and carbon emissions (Ray et al., 2013)

Hence, the Borlaug hypothesis – that we must increase agricultural yields to meet growing food demand and thereby avoid agricultural encroachment – still holds. However, due to global product, capital and labor markets, specific agricultural technologies do not guarantee that famers will help conserving the forest. Improved agriculture and better technologies that increases agriculture on the forest frontier are highly risky and can impose great concern of encroachment. Agricultural policies that are highly labor-intensive and target low-forest areas, or crops and production systems at the forest frontier show more promise. Protected areas and payment mechanisms that enable users to capture larger shares for land rents can therefore be beneficial for forest conservation (Angelsen and Kaimowitz, 2001, Angelsen, 2010).

3.2 Political economy of land use rights

Political economy is concerned with the distribution of power and resources, and focuses on the competing interests of actors, networks and institutions. Political economy attempts to go beyond understanding economics as means of determining land use at forest-agriculture margins, and instead highlight roles of politics and how relations between power and resources shape interests and incentives (McNally et al., 2014). When examining the political economy with respect to forest and agriculture, the political structures influence the interest of actors in the agricultural sector and how forest land is used. It also recognizes that external actors influence political decisions and priorities. Policies and regulatory frameworks are important factors when determining if forest encroachment may occur. Incentives for people to migrate to rural areas through land grants, better infrastructure and subsidies on agricultural inputs are some of the political incentives that can lead to forest degradation (McNally et al., 2014).

Businesses and political interest in land-based natural resource sectors can be highly intertwined as providing access to land for agriculture is a relatively low-cost option to gain political support from powerful actors. Gaining political power often depends on delivering exclusive benefits that directly support certain actors. Powerful actors within the forestry sectors that support deforestation and forest degradation might therefore hamper protective policy changes. Hence, forest-dependent people who live in or near forests tend to be weakened by outsiders, such as national governments, commercial farmers and minders as forests-rights shifts hands (McNally et al., 2014).

Another example is governments selling agricultural land rights to foreign investors (Rulli et al., 2013). The World Bank claims such policies will create opportunities for poor countries through expanding agricultural sector, improve infrastructure and provide access to better technologies (World Bank, 2010). Verma (2015) and (Weingärtner, 2010) argue that through these investments, people's customary *rights* are converted into *marketable tiles,* leaving them to the forces of capitalism, as common and collective rights become private property. In fact, out of the 42.42 million hectares purchased between 2013 and 2018, only 3.47 million hectares were for food production, while 17.45 million hectares are devoted for non-food crops, and 21.49 million hectares for animal feed and fuel. Hence, in addition to reducing agricultural rent through improving infrastructure and access to technologies, food security -and small-scale farmers competitiveness is potentially reduced, both in relation to resources availability, and on product markets (Weingärtner, 2010).

4 Data and Methods

This chapter will first present methods of data collection and variables used in this analysis. Secondly, I will discuss how different explanatory variables are expected to impact agricultural area, before presenting the method used in the analyses. Our main research question is if there is a link between higher production of non-food crops and production of staple crops. Additionally, we want to explore what effects increased agricultural productivity has on area of staple crops, fuel crops and non-food crops.

4.1 Data used

The dataset developed contain area and yield data for 45 crops, categorized into five main crop groups (staple crops, non-staple food crops, feed crops, fuel crops, non-food crops), over a 25-year period from 1992-2016, collected from ten countries. We initially wanted to look at changes in crop areas from 1961-2017, but data from two of the selected countries were inaccessible. The dataset on agricultural production are extracted from a statistical database, FAOSTAT, created by the United Nations Food and Agricultural Organization. In addition to crops, the dataset includes, forest cover, GDP, unemployment and population. These variables are important when considering crop trends and the relationship between economic growth and agricultural production. Data on unemployment rate, GDP and population are extracted from the World Bank DataBank.

Crop selection

Out of 160 crops available in FAOSTAT, 45 crops were selected for this study based on their importance for world food consumption and shifts in land use practices. The crops constitute 87 % of total agricultural production in 2016 (measured in tonnes), and are main contributors to shifts in recent land use trends (FAOSTAT, 2018b). Trends in yields and agricultural areas for cereals (maize, rice, wheat, barley, sorghum, millet, oats and rye) affect global food security, whereas trends for tropical crops (coffee, bananas, palm oil and soy) influence rates of tropical deforestation. Particularly interesting is it to explore the effects the different crops has on agricultural area and why some of them, recently has become more “widespread” than others.

Crop classification

To identify trends in agricultural area and productivity, crops were categorized into five main categories : staple crops, non-staple food crops, feed crops, fuel crops and non-food crops, based on FAOSTAT (2018a) and Ewers et al. (2009)

One of the major challenges of the dataset is the multiple usage areas of crops. For example, palm oil is both a staple crop, animal feed and biofuel. To solve this issue, a VIF test for multicollinearity was performed in each regression and variables that correlated above a given limit were omitted.

Table 4.1 shows the distribution of each crop (estimated in tones for a single year) divided into ten crop classifications and the five crop groups, based on FAO classifications (FAOSTAT, 2018a). Notice that, even though a crop is characterized as a staple crop, it does not necessarily mean the majority of the crop is used for food, but rather that it constitutes the dominant part of a population's diet. Hence, the conventional use might diverge from the actual use of the crop. For example, the population in many countries of sub-Saharan Africa, Southeast Asia and Latin America subsist on maize as staple food, yet, more than half of maize supplies are used as animal feed (Nuss and Tanumihardjo, 2010).

Ideally, I would have liked to have annual and country-specific data for each of the crops, showing the distribution across the five crop categories. Those data are not readily available, and this is a major limitation of the data set I use. The implication might be an underestimation of some trends. For example, higher demand for biofuel from palm oil may cause a higher share of palm oil to be allocated to fuel. This is not captured by the data set.

Table 4.1: Distribution of crops between variables (measured in percentage of quantity).

		Food					
Crop category	**Crop**	**Staple food**	**Non-staple food**	**Feed**	**Fuel**	**Non-food**	**Source**
Cereals	Barley	15 %	-	85 %	-	-	(Akar et al., 2004, OECD, 2008)
	Maize	10 %	-	67 %	23 %	-	(Tenenbaum, 2008, Grujcic et al., 2018)
	Sorghum	42 %	-	48 %	-	10 %	(Amido, 2015, Pavelescu, 2011)
	Rice	90 %	-	2 %	8 %	-	(Nguyen, 2002)
	Wheat	70 %	-	18 %	5 %	7 %	(IGC, 2014, Vogel, 2017)
	Millets	-	80%	7 %	-	13 %	(FAO, 1996)
	Oats	-	30%	70 %	-	-	(FAOSTAT, 2013)
	Rye	-	44%	56 %	-	-	(Rye and Health, 2013)
Oilbearing crops	Coconuts	-	100 %	-	-	-	
	Soybean	19 %	-	75 %	2 %	4 %	(Islas-Rubio and Higuera-Ciapara, 2002, WWF, 2018)
	Groundnut	-	75 %	25 %	-	-	(Jimoh et al., 2012)
	Rapeseed	-	10 %	10 %	80 %	-	(Ufop, 2013)
	Sunflower seed	-	87 %	-	13 %	-	
	Palm oil fruit	-	34 %	4 %	45 %	17 %	(Tan et al., 2009, Nelsen, 2016, Buttler, 2013)
	Seed cotton	-	100 %	-	-	-	
	Olives	-	100 %	-	-	-	
Vegetables	Green Bean	-	100 %	-	-	-	
	Watermelon	-	100 %	-	-	-	
	Onion	-	100 %	-	-	-	
	Cucumber	-	100 %	-	-	-	
	Tomato	-	100 %	-	-	-	
	Brassica	-	100 %	-	-	-	
	Green maize	-	-	100 %	-	-	
Roots and tubers	Cassava	58 %	-	25 %	-	17 %	(FAO, 2000)
	Potato	100 %	-	-	-	-	
	Sweet potato	-	100 %	-	-	-	
	Taro	100 %	-	-	-	-	
	Yams	100 %	-	-	-	-	
Fruits	Apples	-	100 %	-	-	-	
	Grape	-	100 %	-	-	-	
	Orange	-	100 %	-	-	-	
	Banana	100%	-	-	-	-	
	Plantain	100 %		-	-	-	
Pulses	Chick-pea	-	100 %	-	-	-	
	Pigeon pea	-	90 %	10 %	-	-	(Mula and Saxena, 2010)
	Lentils	-	100 %	-	-	-	
	Cow pea	-	52 %	13 %	35 %	-	(CGIAR, 2016)
Stimulant crops	Tea	-	100 %	-	-	-	
	Coffee	-	100 %	-	-	-	
	Cocoa	-	100 %	-	-	-	
Sugar crops	Sugar cane	-	60 %	-	40 %	-	(Lam et al., 2009)
	Sugar beet	-	100 %	-	-	-	(FAO, 2017)
Fibers	Jute	-	-	-	-	100 %	
Other	Tobacco	-	-	-	-	100 %	
	Rubber	-	-	-	-	100 %	

Country selection

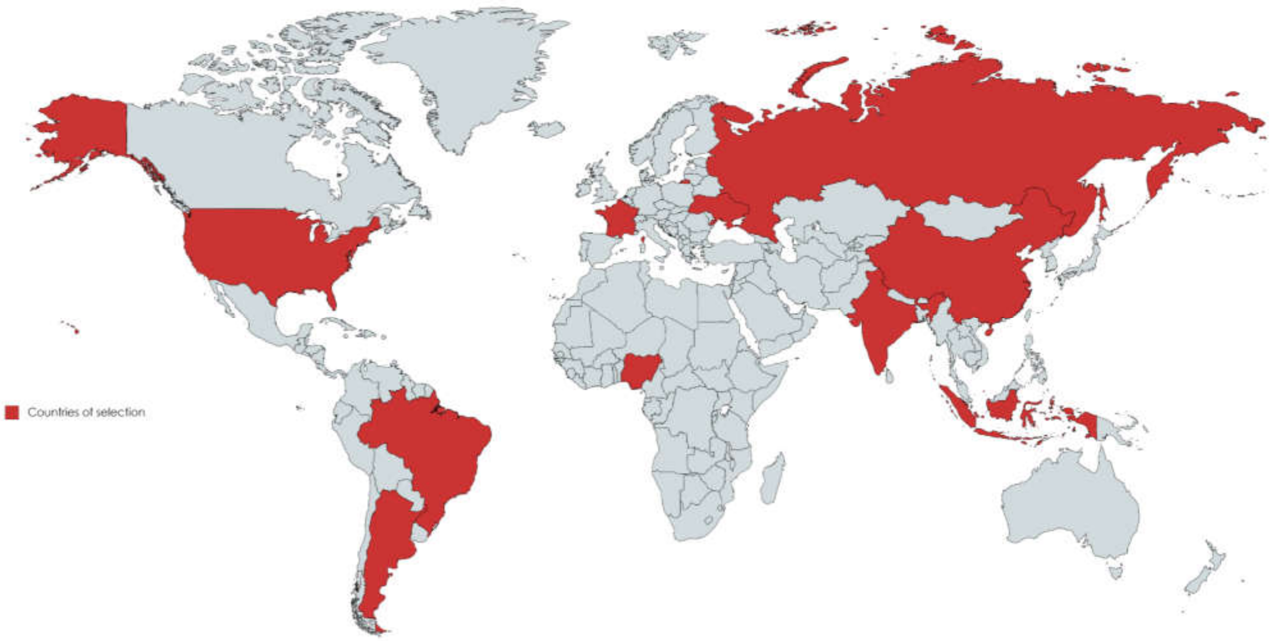


Figure 4.1: Map of study area

The study is based upon the top ten crop producing countries in the world (measured in area). The countries are as follows: (1) China, (2) United States of America, (3) India, (4) Brazil, (5) Russian Federation (6) Nigeria, (7) Indonesia, (8) Ukraine, (9) Argentina, (10) France (FAOSTAT, 2018b). Although these countries only represent 53,5 % of world total agricultural area, they produce the majority of staple crops. However, in recent years, these countries have also become major feed and fuel crop producers. Hence, recent agricultural trends are largely influence by these countries, which makes them important to analyze when studying links between staple crops and non-food crops (FAOSTAT, 2018b). Because we want to discover impacts on staple crops at a global scale, not domestically, we aggregated the data sampled on these countries to create the closest estimate. Ideally, more countries should be included, but due to time and data constraints, the study was limited to the countries listed above.

To examine geographical trends in agricultural production, the countries can be divided into three classifications, based on stages of development: developed countries (United States, Russia, Ukraine, France), middle-income countries (China, Brazil, Argentina) and poor/ low middle-income countries (India, Nigeria, Indonesia). Part of the motivation is to examine if the

Prebisch-Singer hypothesis, including whether rich countries mainly produces non-food crops for profits, whereas poorer countries mainly produce food crops.

4.2 Data collected and choice of variables

Summary statistics of variables used in this thesis are provided in Table 4.2. The table also contains information units of measurement of variables as well as how the different variables are expected to affect agricultural area. Below the table is a presentation of the relevance of the variables and a justification of their expected signs.

Table 4.2: Summary statistics for the dataset including expected signs with agricultural area as dependent variables.

Variable name	**Description**	**Unit**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Expected effect on tot. agricultural area**
$TotalSCa_{it}$	Total staple crop area	Hectare (ha)	2.67e+07	2.30e+07	3797564	8.38e+07	Dependent variable (H2.1)
$TotalFeeda_{it}$	Total feed crop area	Hectare (ha)	1.91e+07	1.46e+07	3404262	5.47e+07	Dependent variable (H2.2)
$TotalFuela_{it}$	Total fuel crop area	Hectare (ha)	7399862	8009310	623416.6	3.06e+07	Dependent variable (H2.2)
$TotalNFa_{it}$	Total non-food crop area	Hectare (ha)	967941	1090542	0	3845429	Dependent variable (H2.2)
$TotalSCy_{it}$	Total staple crop yield	Tones/ha	4.554227	1.916633	.7548507	9.022795	-
$TotalNSCy_{it}$	Total non-staple crop yield	Tones/ha	10.17396	9.841919	1.156265	44.11067	-
$TotalFeedy_{it}$	Total feed crop yield	Tones/ha	3.272358	1.898057	.8175378	7.692615	+
$TotalFuely_{it}$	Total fuel crop yield	Tones/ha	7.590296	6.993058	.6877608	32.72535	+
$TotalNFy_{it}$	Total non-food crop yield	Tones/ha	1.460954	.7241201	0	5	+
$forest_{it}$	Total forest area	Hectare (ha)	205866.2	254473.3	6993	815135.6	±
$Unem_{it}$	Unemployment rate	Percent (%)	7.6672	3.196215	2.8	19.6	+
gdp_{it}	Gross domestic product per capita	U.S. dollar ($)	10733.62	14379.59	153.6467	57638.16	±
pop_{it}	Total population		3.52e+08	4.38e+08	3.37e+07	1.38e+09	+

Variables

Agricultural areas

Agricultural areas consists of four crop categories: $TotalSCa_{it}$, $TotalFeeda_{it}$, $TotalFuela_{it}$ and $TotalNFa_{it}$. For Hypothesis 2, we ask if higher yield of staple crops and non-food crops enhance or reduce total agricultural area. Here, staple-, feed-. Fuel-, and non-food crop area act as dependent variable as we want to verify if increased yield encourages land saving or agricultural encroachment, depending on crop group.

Forest area

Total forest area, called $forest_{it}$, is an independent variable. It is defined as "land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 %, or trees able to reach these thresholds in situ. It does not, however, include land that is predominantly under agriculture or urban land use" (FAOSTAT, 2018b). Total forest variable area has a missing observation value equal to zero because multiple countries has unreported number for the year 2016. To deal with this issue, I used the average of the two last years as an estimate for 2016.

The expected sign for total forest area is unknown because large forest area is suggestive of having a large area for agricultural expansion, suggesting a positive link. On the other hand, because an increase in forest area reduces the total agricultural area (Benhin, 2006). Keep in mind that $forest_{it}$ does not include land that is under agricultural use, including agroforestry.

Yield

Yields ($TotalSCy_{it}$,$TotalNSCy_{it}$,$TotalFeedy_{it}$,$TotalFuely_{it}$ and $TotalNFy_{it}$) are the main independent variables in our study. Crop yield was calculated for all countries from 1992 – 2016 as the total crop production of the 42 crops divided by the sum of area (ha) under those crops in time t and country i. We refer to this throughout the paper as yield. To correct for market fluctuations, a factor of crop prices has been included in the calculation of yield.

To aggregate across crops (e.g., to create yield for staple crops), one must use weights for the different crops. On option is to just use tones, but this obscures the fact that some crops are more valuable per kg than others. And, a yield increase can occur by shifting to more valuable crops. Using value is therefore a better alternative, but this can also be misleading if one uses annual prices: a price increase can then be confused with a productivity increase. Better methods, which we used, is therefore to calculate average prices for the period, and use these as weights for all years in the aggregation. The prices were calculated from prices collected at (Index Mundi, 2018) in time period 1992 – 2017.

Staple crops are compared to price of rice (which has the weight of 1), non-staple food with the price of cocoa, fuel crops with the price of palm oil, and non-food crops with the price of cotton. There was no available data on prices for crops defined as feed crops, hence, the price of palm oil was used here as well.

The expected sign of yields is both negative and positive because it depends whether or not the changes in agricultural proactive, which drive the yield change, free up labor or capital for other kinds of agricultural practices. Hence, we expect that an increase in yield can both increase and decrease agricultural area, depending on crop group (Alcott, 2005, Borlaug, 2007).

Unemployment rate

Unemployment, called $Unem_{it}$, refers to the share of the labor force in country i that is without work, but available for, and seeking employment. Unemployment rate data is obtained from the World Bank DataBank and is measured as a percentage of total labor force (modeled ILO estimate) (World Bank, 2018b).

Because agriculture is a labor-intensive industry in many of the countries studied, we expect $Unem_{it}$ to have a positive impact on agricultural area. This is because we expect an increase in unemployment rate to increase agricultural area as more cheap labor is available for agricultural production (Roser, 2018).

GDP

GDP per capita, called gdp_{it}, is the gross domestic product divided by midyear population of each country. It is measured in current US$ and is the sum of gross value added by all resident produced in the economy plus any product taxes and minus any subsidies not included in the value of the product (World Bank, 2018a). The economic role of agricultural seems more important in developing countries than in developed countries and can therefore have major impact on development of agricultural practices and trends in agricultural areas. GDP is an independent variable, however the expected effects are uncertain. Because higher GDP advocate new technology that improves yield, it can either reduce or enhance agricultural expansion, depending on crop group. Further, GDP growth may also increase demand for agricultural products, and thereby lead to more land expansion (Sali, 2012). Hence, the overall expected effect of GDP growth is uncertain.

Population

Population, called pop_{it}, includes all residence in a country regardless of legal status or citizenship. It is used as an independent variable because increase in human population influences exploitation of natural resources and social infrastructure greatly.

We expect population to have a positive sign because an increase in population is expected to enhance agricultural encroachment: population growth increases the demand for agricultural products, and more labor is available, and more people compete for the same resources (land) (Angelsen, 1999).

4.3 Adopted model for the study

To answer our research questions, we first created four separate regression models, appropriate to each of the different assumptions in Hypothesis 2. This is called the original model. Secondly, based on the results from the original model, we ended up with our main model. The process of how the model was chosen is discussed in this subsection.

Original models

To test which functional form were appropriate for our model, both linear functional form (lin-lin model) and log transformed functional form (log-log model) was tested. Based on the different diagnostic tests addressing issues of serial correlation, heteroskedasticity, misspecification and normality, descriptive statistics is suitable for Hypothesis 1, while regression analysis with log-log specification is preferable for Hypothesis 2. The estimated coefficients for Hypothesis 2 is interpreted as elasticities. (Wooldridge, 2015, Benoit, 2011).

The models are given as follows:

Model 1: Panel data for Hypothesis 2.1

$$\begin{aligned}\log(TotalSCa_{it}) \\ = \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ + \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ + \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + \beta_{10} \ln(year_t) + \alpha_i + u_{it}\end{aligned}$$

Model 1: Panel data for Hypothesis 2.2 with fuel crop area as dependent variable

$$\begin{aligned}\log(TotalFuela_{it}) \\ &= \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ &+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ &+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + \beta_{10} \ln(year_t) + \alpha_i + u_{it}\end{aligned}$$

Model 1: Panel data for Hypothesis 2.2 with feed crop area as dependent variable

$$\begin{aligned}\log(TotalFeeda_{it}) \\ &= \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ &+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ &+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + \beta_{10} \ln(year_t) + \alpha_i + u_{it}\end{aligned}$$

Model 1: Panel data for Hypothesis 2.2 with non-staple food crop area as dependent variable

$$\begin{aligned}\log(TotalNFa_{it}) \\ &= \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ &+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ &+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + \beta_{10} \ln(year_t) + \alpha_i + u_{it}\end{aligned}$$

Where α_i (i = 1…n) is a variable capturing all unobserved, time-constant factors that affect y_{it} for each entity (*n* entity-specific intercepts). u_{it} is the error term, often called idiosyncratic error term, or time-varying error as representing unobserved factors that varies across subjects and time (Wooldridge, 2015). The unobserved effect also captures geographical features, such as the location of agricultural areas and historical factors where countries may have different agricultural areas and yield for historical reasons.

There are differences of variables across countries and between countries, which is subscripted by i where i=1,…,N. The subscript t represents time periods where t = 1,…,T. Variables that subscript i are time-invariant and depend only on the country specific effects, while variables that subscript t are time-variant and varies across countries in specific period of time. Variables who denote both i and t, varies across countries and across time, hence they are indexed with the subscript it. In regard to our model, all the variables vary across time and across countries due to different stages of development and geographical locations (Wooldridge, 2015).

Main model

The main model was obtained by differencing the original pooled OLS model. It is a simple model, which is similar to single cross-sectional equation, but where each variable is differenced with respect to time. This is explained further in the next sub-chapter (4.5). We tested the model with different compositions of explanatory variables, and by studying the differences in the functional form of the estimated coefficients, p-values and R-squared, we ended up at our main model. The analysis in this model follows the same log-log transformation as the original model, however because of issues with our functional form, we need for differentiate across time. Additionally, first differences estimator also eliminates α_i from the model. Hence, the main models are given as follows, where "δ_0." indicates the first-difference model (Wooldridge, 2014, p. 373):

Model 1: Panel data for Hypothesis 2.1

$$\begin{aligned}\log(TotalSCa_{it}) &= \delta_0 + \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ &+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ &+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + u_{it}\end{aligned}$$

Model 1: Panel data for Hypothesis 2.2

$$\begin{aligned}\log(TotalFuela_{it}) &= \delta_0 + \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ &+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ &+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + u_{it}\end{aligned}$$

Model 1: Panel data for Hypothesis 2.2

$$\begin{aligned}\log(TotalFeeda_{it}) &= \delta_0 + \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\ &+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\ &+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + u_{it}\end{aligned}$$

Model 1: Panel data for Hypothesis 2.3

$$
\begin{aligned}
\log(TotalNFa_{it})
&= \delta_0 + \beta_0 + \beta_1 \ln(TotalSCy_{it}) + \beta_2 \ln(TotalNSCy_{it}) + \beta_3 \ln(TotalFeedy_{it}) \\
&+ \beta_4 \ln(TotalFuely_{it}) + \beta_5 \ln(TotalNFy) + \beta_6 \ln(forest_{it}) + \beta_7 \ln(pop_{it}) \\
&+ \beta_8 \ln(Unem_{it}) + \beta_9 (lngdp_{it}) + u_{it}
\end{aligned}
$$

Keep in mind that statistical results are an inherit property of the data used in the study resulting from reporting errors and inconsistencies in reporting systems[3]. The errors can be positive or negative with a mean zero, explaining some of the unexpected results in the correlation matrix.

The time trend ($year_t$) is not included in the main model. This indictor variable was not significant and could not capture any of the effects on agricultural area that the other variables could. Thus, to avoid multicollinearity, the time trend variable was excluded from main model.

4.4 Data and estimation issues

Four central data and estimation issues are identified: firstly, the data was obtained by random sampling because we are analyzing the top ten crop producing countries within these time periods. The probability of being included in this sample within these periods were not equal for all countries as the selection was based on area of agricultural production. However, because most of world's staple crops are produced in these countries, we believe that our sample is representative for the population.

Secondly, our model might suffer from endogeneity as some variables were not included in the study due to lack of information. Thus, variables considered important for the study might be omitted, and we might have a problem with omitted variable bias (Wooldridge, 2015, p. 799) However, because we do not always have the knowledge on variables influencing the dependent variables when data is not available or accessible, in practice there will always be one or more relevant variables omitted. To include all relevant variables may therefore be impossible, but as many variables as possible should be included.

Thirdly, large problems with imperfect data exists. According to Rudel et al. (2009), researchers has yet to examine actual historical episodes where agricultural intensification has restrained agricultural expansion on large areas because precise data on the global extent of abandoned

[3] Yield trend with area noise displayed in Appendix C.1

cropland do not exists. As mentioned in Chapter 3.1, satellite images can detect the presence or absence of forests but has difficulties of distinguishing idled land from active crop land on global scale. Agricultural statistics collected from FAO provide a way to investigate at a global scale, however countries use different methods in compiling data they report, hence large measurement errors, or noise, may exist when addressing impacts of land use changes. Nonetheless, it is the only global data available (Rudel et al., 2009).

Fourthly, our dataset might suffer from multicollinearity. Because crop variables have multiple areas of usage, food, feed and fuel, variation in one independent variable can cause variation in another. Hence, there is a correlation between our independent variables.

4.5 Methodology

The econometric model specification is based on models for panel data and involves regressing crop area against yield for the years 1992-2016. The data is pooled over 25 years (T) with ten countries (N) as the panel of the data. The panel data is strongly balanced, meaning that information on all the variables for all time periods exists. However, there are more time periods than countries, i.e. the number of subjects is less than the number of time periods, which restricts regression possibilities.

For panel data, four regression models are tested: the pooled model, the fixed effects model, the random effects model and the first-difference model. It is informative to compare them with each other to help us determine the nature of the biases caused by leaving unobserved effect, α_i, entirely in the error term (as the pooled model), partly in the error term (as the random effects model), or independently of the error term (as the fixed effects model and first differencing model). However, it is important to remember that, even if α_i is entirely uncorrelated with the explanatory variables in all periods of time, the standard errors and test statistics of the pooled model is usually invalid as it tend to ignore serial correlation in the composite error, $v_{it} = \alpha_i + u_{it}$ (Wooldridge, 2015).

Because ordinary least squares (OLS) disregards much of the useful information in the time-periods, we can use the same information in a pooled OLS (POLS) procedure. POLS is an "OLS estimation with independently pooled cross sections, panel data, or cluster samples, where the observations are pooled across time (or groups) as well as across the cross-sectional units" (Wooldridge, 2015, p. 800). Except for being linear in parameters and data to be obtained by random sampling, the POLS assume all coefficients area equal. Additionally, the estimator disregards any systematic observed heterogeneity between observations from different populations. This observation seem strict as it is unlikely that all the observations from one population is generated the same way (Wooldridge, 2015). For example, changes in agricultural area in Brazil is mostly likely not behaving similar as changes in agricultural area in Ukraine.

For the POLS estimator to be consistent, we assume population orthogonality, i.e. $E(x'u) = 0$, which requires weak exogeneity $cov(\varepsilon_{it}, x_{it}) = 0, \forall_i, t$, where $\varepsilon_{it} = \alpha_i + u_{it}$ and x_{it} is our explanatory variables. If the true model is pooled and the regressors are uncorrelated with the error terms, the pooled OLS regressor is consistent. However, POLS specify constant coefficients and is therefore the most restrictive panel data model. If there is unobserved heterogeneity, i.e. some unobserved factors affect the dependent variable, and correlated with some observed regressor, the POLS become inconsistent with all regressors, while other panel data models are still consistent. Thus, POLS is not very common in literature. Fixed Effects (FE), on the other hand, can be estimated if we believe that u_{it} is correlated with the x's (the time-varying explanatory variables) (Wooldridge, 2015)

Because this study considers periods of 25 years and ten different countries, the analysis of this study must include both regional and temporal scale variation. Econometrically, these variations are tested running the model as a two-way FE estimator. The model is estimated as a panel considering time and place in a fixed effects model:

(9) $Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$ i = 1,….,n, t = 1,….,T

Where α_i (i = 1…n) is the unknown intercept for each entity (*n* entity-specific intercepts). To begin analyzing our panel data it must satisfy four assumptions. The first three assumption are equal to POLS assumptions, however, assumption (FE.4) implies strict exogeneity, i.e.

(10) $E(u_{it}|x_i, c_i) = 0, t = 1, \dots, T.$

(Wooldridge, 2015, p. 459).

When using FE, we assume that something within the individual may influence or bias the predictor or outcome variables, which we need to control for. FE remove the effect of those time-invariant characteristics, α_i, so we can assess the net effect of the predictors on the outcome variable. (Torres-Reyna, 2007). In other words, FE can absorb unobserved time invariant determinants of the dependent variable. Instead, time dummies can be included in the regressor x_i. The disadvantage of the estimator is that it removes all variables that are time-invariant. Thus, it cannot yield any estimate of time-invariant variables on the dependent variable. The Random Effect (RE) estimator, on the other hand, allows us to estimate time-invariant variables on the dependent variable. It assumes that ε_{it} is uncorrelated with explanatory variable and if the assumption of $cov(\varepsilon_{it}, x_{it}) = 0$ is true, the estimated standard errors are smaller than under FE. However, if we cannot consider the observations to be random draws from a large population, for example if we have data on countries – it makes sense to assume that α_i as parameters to estimate, in which case we should use fixed effects.

For small datasets, an alternative to the FE is the First Differencing (FD) estimator. The FD achieves the same goal: eliminating α_i from the model, and is similar to single cross-sectional equation, but where each variable is differenced over time.

The first differencing model is

(11) $\quad Y_{it} = \delta_0 + \beta_1 \; x_{it} + \; u_{it}$

(Wooldridge, 2015)

In addition to the POLS assumptions and strict exogeneity, FD assumes that "each explanatory variable change over time for at least some units, and there is no exact linear relationship among regressors in the population, i.e."

(12) $\quad rankE(\; x' \; x) = k$

(Wooldridge, 2015)

The most important of these is that u_{it} must be uncorrelated with x_{it}. This assumption holds if the idiosyncratic error at each time t, u_{it} is uncorrelated with the explanatory variable in both time periods. Hence, it rules out cases where future explanatory variables react to current changes in the idiosyncratic error, as must be the case if x_{itj} is a lagged dependent variable (Wooldridge, 2015).

Compared to the FE estimators, the FD estimators is numerically equivalent for T = 2. However, when T ≥ 3, the FE and FD are not the same. For large N and small T, the choice depends on the relative efficiency of the estimators and is determined by serial correlation in the idiosyncratic errors, u_{it}. When the u_{it} is serially uncorrelated, FE is more efficient than FD, hence, in literature FE is generally more common than the FD estimator. Nevertheless, we must remember that this assumption can be false, and that in many applications, we can expect unobserved factors that change over time to be serially correlated. On the other hand, when T is large, and N is small, as in this thesis, we must show caution in using fixed estimators.

Although distributional results hold under FE assumptions, inference can be sensitive to violations. Thus, inference with FE is potentially more sensitive to non-normality, measurement error, heteroskedasticity, and serial correlation in the idiosyncratic errors. Hence, as the FD model has the advantage of turning an integrated time series process into a weakly dependent process, it is less sensitive to violations of assumptions and thus, a more appropriate estimator (Wooldridge, 2015, p. 447).

Nevertheless, differencing is not free of difficulties either. If one or more of the explanatory variables is subject to measurement error, especially the classical errors-in-variables model, the FD estimator can be a worse option than pooled OLS. Additionally, the model causes potential problems when key explanatory variables do not vary much over time (Wooldridge, 2015). However, in the process of choosing an estimator, the main factor to evaluate is the number of T and N in the dataset. Because the dataset contains large T and small N, POLS with FD becomes the preferred estimator compared to panel data. (Wooldridge, 2015).

Statistical tests

Until now, we have reviewed the most relevant econometric models for panel data, POLS, FE, RE and FD, and ended up favoring FD because of the benefits when a dataset contains large T and small N. In our regression analysis, I therefore had to test whether both POLS and FD are consistent and thus appropriate estimators. As mentioned in the previous sub-chapter, we generate the variables for Hypotheses 2 into log-log variables before comparing the models. There are four reasons for this: First, to improve model fit. For instance, if the residuals are not normally distributed, then taking the log of a skewed variable may improve the fit by altering the scale and making the variable more "normally" distributed. This is especially important for our model where crop production in large countries may diverge the direction of the analysis. Second, better interpretation. If you log both your dependent (Y) and independent (X) variables, the regression coefficients (beta) will be elasticities and interpretation would go as follows: a 1% increase in X would lead to a ceteris paribus beta percentage increase in Y (on average). Third, taking the log of the dependent and/or independent variables may eliminate the heteroskedasticity. Forth, transforms a non-linear model into a linear model (Wooldridge, 2015).

To find the best linear regression model, I compare three POLS models with different standard errors: ordinary POLS with normal standard errors, POLS with robust standard errors and POLS with cluster robust standard errors. First, I control if the ordinary POLS with standard errors satisfy the assumptions by running tests of heteroskedasticity, multicollinearity, normality and functional form. An issue with POLS with ordinary standard errors is the vulnerability for heteroskedasticity and serial correlation. Hence, it is often less preferred than both robust standard errors and cluster robust standard errors. POLS with ordinary (robust) standard errors is presumed appropriate for this thesis because errors for a given individual, and country, are almost certainly positively correlated over time. Hence, if the functional form of ordinary POLS is rejected, cluster robust standard errors is the preferred estimator as it corrects for clusters (panel) and is robust against any type of heteroskedasticity and serial correlation (Wooldridge, 2015).

Next, we need to test for individual effects using first –order autoregression, written AR(1). This model tests if we can predict current year by using previous measurements. The last regression in the POLS analysis is an f-test for individual effects/ serial correlation. If individual effects exist, then the clusters are correlated over time. One solution is to omit these effects, however as they will go into the error term, issues of endogeneity may occur (Williams, 2015). Thus, if strong evidence of individual effects/ serial correlation exists, there is still something in our model that POLS did not solve, and which needs to be explored by adding the FD estimator.

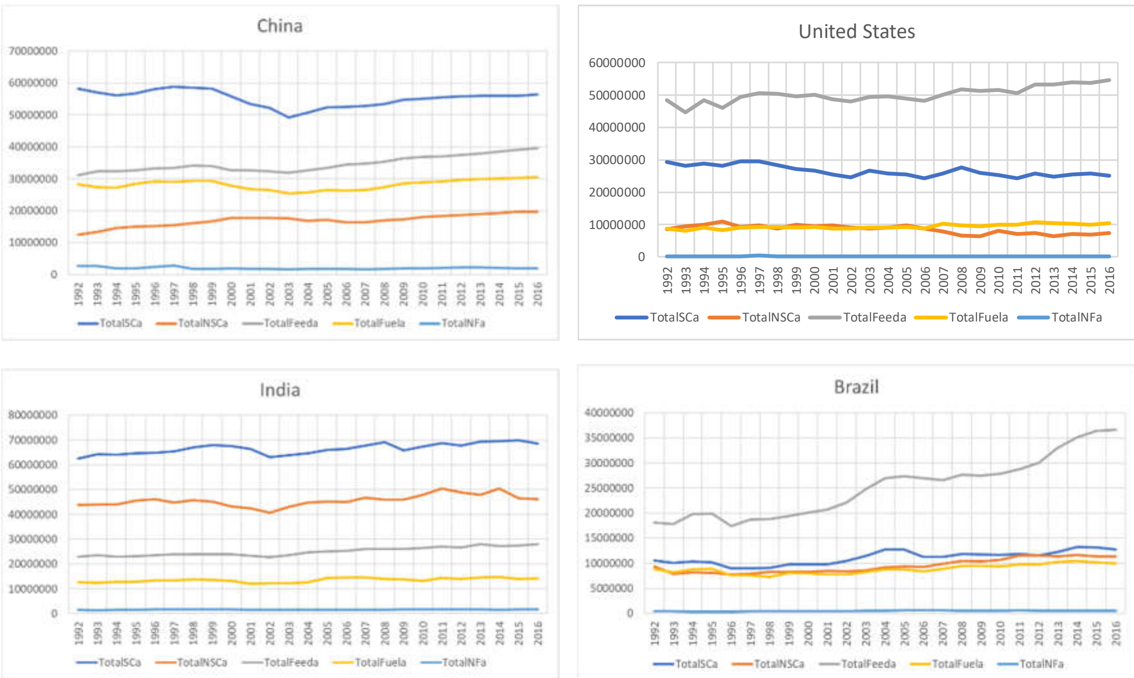
5 Results and Discussion

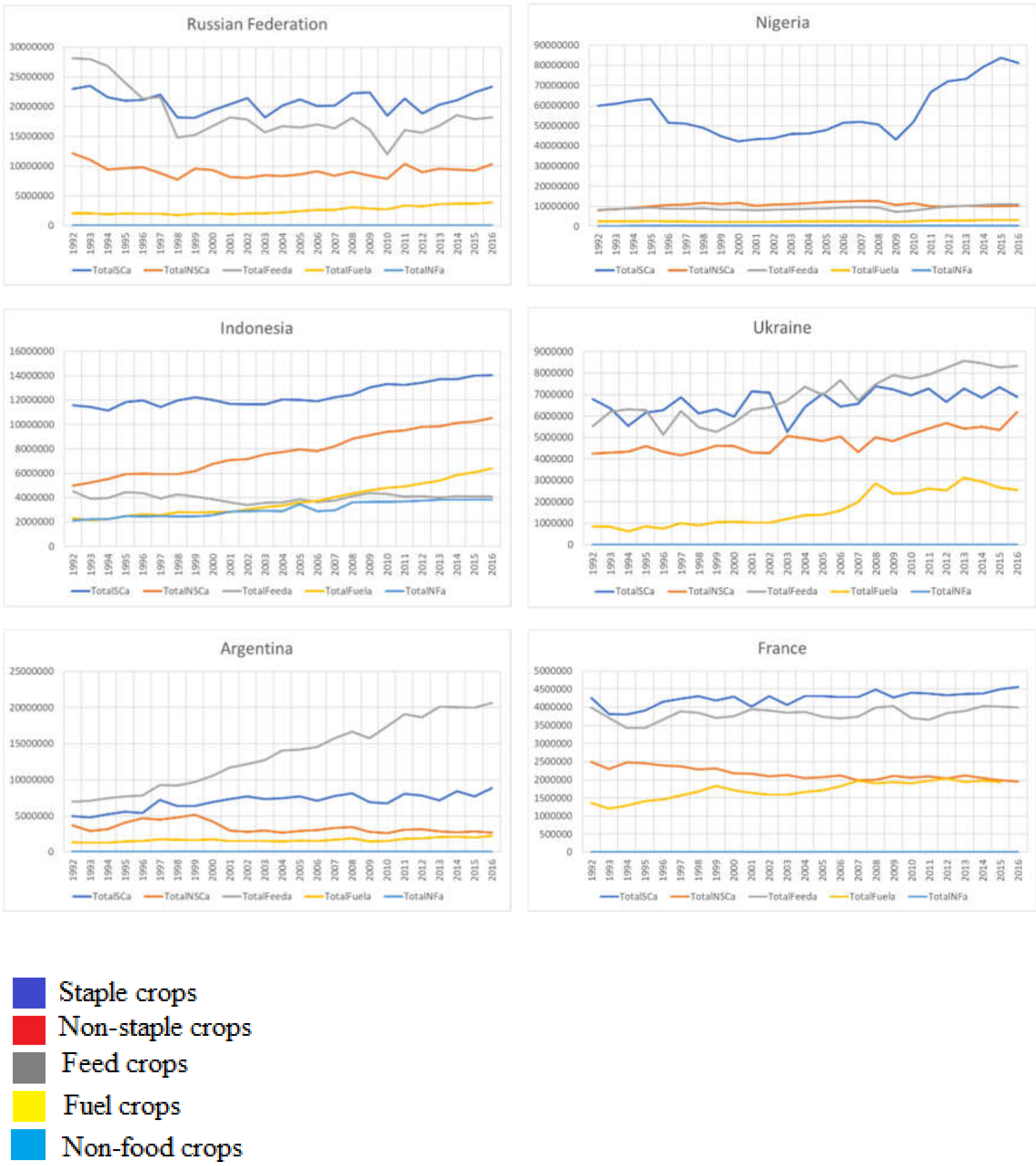
This section contains four main parts. The first part presents descriptive statistics for ten selected countries and compare national agricultural area trends. This section tests Hypothesis 1 and the trends between crop groups and total agricultural area. Hence, it also gives a better understanding of the results of the following test and in the discussion. The second part reviews price trends between 1992 and 2017. The third part presents the econometric results, followed by a discussion on the correlations between yield and agricultural area. Lastly, I point to limitations of the analysis and suggest further work.

5.1 Descriptive statistics

The lines Figure 5.1 represent the total production area (ha) of each crop group divided into the selected countries. The same trends are presented in numbers in Table 5.1. The aim is to give overview of the trends of each crop group over the last 25 years.

Figure 5.1: Change in are by crops (in hectare)





One should note that the trend lines are linked (correlated) because 18 out of 43 crops have multiple areas of usage. Thus, an increase in area of one crop group, may result in increased area of any of the other variables. This is a weakness in the method used and is discussed further in Chapter 5.6.

In many of the countries, non-staple food, feed and fuel has had the largest expansion of agricultural area between 1992 and 2016. This is especially the case for China, United States, Brazil, Indonesia and Argentina where crops such as barley, maize, soybeans, rapeseed, and palm oil are important cash crops. On the other hand, for the same countries, staple crops and non-food crops have decreased or remained the same throughout this period. Hence, several countries seem to have shifted production from food crops to non-food crops (or at least area). However, Indonesia, Nigeria and India – the three low income countries in our sample - are still experiencing growth in staple-crop production.

Table 5.1: Domestic land use changes in the selected countries between 1992 and 2016, measured in hectare (ha). The numbers in prentices shows percentage changes of agricultural area.

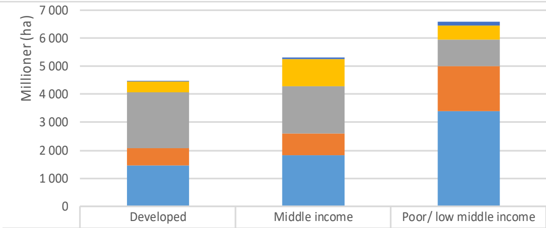
	China	*United States*	*India*	*Brazil*	*Russia*	*Nigeria*	*Indonesia*	*Ukraine*	*Argentina*	*France*	*Total*	*% of total change*
Ag. Land 1992	132 853 402	97 973 833	143 601 762	47 283 865	65 423 771	79 262 682	25 527 117	17 435 940	17 131 141	12 113 911	638 607 423	
Ag. Land 2016	148 347 369	99 651 871	158 740 345	71 232 433	55 690 675	105 709 838	38 917 378	23 993 733	34 508 181	12 448 652	749 240 476	
Tot. Increase (ha)	15 493 968	1 678 038	15 138 583	23 948 569	-9 733 096	26 447 156	13 390 261	6 557 793	17 377 040	334 741	110 633 053	110 633 053
Tot. Increase (%)	(12%)	(2%)	(11%)	(51%)	(-15%)	(-15%)	(52,%)	(38%)	(101%)	(31%)	(17%)	
Change in land area (ha) from 1992 to 2016												
Staple crop	-1 856 048 (-3%)	-4 292 929 (-15%)	6 067 382 (10%)	2 181 531 (21%)	330 964 (1%)	21 250 461 (35%)	2 473 637 (21%)	108 849 (2%)	3 896 403 (78%)	302 916 (7%)	30 463 167 (11%)	*(28%)*
Non-staple crop	7 161 075 (57%)	-1 242 108 (-14%)	2 453 679 (6%)	2 000 766 (21%)	-1 839 543 (-15%)	2 002 612 (24%)	5 555 999 (122%)	1 923 408 (45%)	-1 007 347 (-27%)	-535 176 (-21%)	16 473 364 (15%)	*(15%)*
Feed crop	8 539 157 (27%)	6 240 478 (13%)	4 990 296 (22%)	18 500 532 (102%)	-9 988 474 (-35%)	2 432 088 (29%)	-460 020 (-10%)	2 816 859 (51%)	13 617 520 (194%)	-2 316 (0%)	46 686 120 (26%)	*(41%)*
Fuel crop	2 339 736 (8%)	1 662 550 (19%)	1 435 328 (11%)	1 126 668 (13%)	1 767 377 (83%)	664 780 (27%)	4 108 063 (178%)	1 716 844 (202%)	897 486 (66%)	574 816 (42%)	16 293 647 (24%)	*(15%)*
Non-food crop	-689 952 (-26%)	-689 952 (-26%)	191 897 (13%)	139 072 (36%)	-3 420 (-100%)	97 215 (34%)	1 712 582 (80%)	-8 166 (95%)	-27 021 (-38%)	-5 499 (-62%)	716 756 (7%)	*(1%)*

In Table 5.1, the trends presented in Figure 5.1 becomes clearer. As total aggregate agricultural areas have increased 17 % between 1992 and 2016, most of the increase has large been for feed crops (26 %) and fuel crops (24 %), whereas staple crops area only increased by 11 %. The "% of total change" gives the share of the changes in total area caused by the different crop groups. Most notably is feed crops which represent 41 % of total agricultural area increase from 1992 to 2016, while staple crops represent 28 %. In connection with the column on the right, "Total", we see that, 11% increase in staple crops area from 1992 – 2016 represents 28% increase on total agricultural area. Hence, between 1992 and 2016, 28% of the crop area growth is represented by staple crops, while feed crop area and fuel crop area has had the highest growth rate in the same period. This is because staple crops represent 40 % of total area crops, hence changes in the staple agricultural sector generates a significant absolute change in total crop area.

Hypothesis 1 states that *the increase in total agricultural area is driven by non-food and fuel crops, not staples*. As just observed, this is partly confirmed by our descriptive statistics.

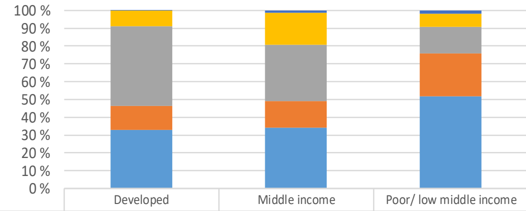
To further explore when and where production of different crop categories has changed from 1992-2016. Figures 5.2, 5.3, and 5.4, and Table 5.2, the countries were divided into geographical categories; developed countries (USA, Russia, Ukraine, France), middle-income countries (China, Brazil, Argentina) and poor/ low middle-income countries (India, Nigeria, Indonesia). The reason for this is to visualize if the Prebisch-Singer hypothesis holds, i.e. if rich countries mainly produces non-food crops for profits, whereas poorer countries mainly produce food crops.

a)



	Developed	Middle income	Poor/ low middle income
TotalNFa	5 102 137	63 514 714	125 471 016
TotalFuela	384 626 150	966 242 492	498 632 408
TotalFeeda	1 987 536 329	1 668 237 967	953 524 274
TotalNSCa	618 719 281	782 463 164	1 603 327 143
TotalSCa	1 457 248 063	1 823 490 549	3 390 742 708

b)



	Developed	Middle income	Poor/ low middle income
TotalNFa	5 102 137	63 514 714	125 471 016
TotalFuela	384 626 150	966 242 492	498 632 408
TotalFeeda	1 987 536 329	1 668 237 967	953 524 274
TotalNSCa	618 719 281	782 463 164	1 603 327 143
TotalSCa	1 457 248 063	1 823 490 549	3 390 742 708

Figure 5.2: Change of crop area between the different geographical categories from 1992-2016, measured in percent (a) and in millions of hectare (b).

In Figure 5.2, Table a) shows that the largest production area of staple crop –and non-staple crop production has occurred in the selected poor/ low middle-income countries. In fact, the total agricultural area of food production in the selected poor/ low middle-income countries is larger than total agricultural area for all crop groups in the selected developed countries, and almost equally as much as total agricultural area of all crop groups in the selected middle-income countries. The distribution of crop group production has been similar between the selected developed– and middle-income countries, however developed countries seem to produce more feed, while middle-income countries produce more fuel (biofuel in Brazil). The table on the right shows the same results, however, crop area is measured in percentage of total agricultural area in each of the geographical categories. For the selected poor/ low middle-income countries, 75 % of agricultural area has been cultivated for staple –and non-staple food crops, whereas for developed and middle-income countries, less than 50 % of agricultural area has been cultivated for food crops. Hence, for richer countries it seems that feed and fuel has been the focus of production the last 25 years.

Table 5.2 shows numerically changes in agricultural area from 1992 – 2016 in the different geographical categories. It shows that the largest area transformation has occurred in middle-income countries (46 %), where feed and non-staple food crops has experienced the largest increase. The selected poor/low middle-income countries have experienced major transformations for all crop groups, where largest increases have happened in the fuel and non-food sectors. The selected developed countries, which according to Figure 5.2 has produced the most feed crops (measured in ha) the past 25 years has experienced an overall decrease in area of production. The same trend is seen for non-staple food crops.

Table 5.2: Changes in agricultural areas from 1992 – 2016 divided by geographical categories.

	Developed	Middle-income	Poor/ low middle-income
Ag. Land 1992	190 585 344	197 573 207	248 391 561
Ag. Land 2016	189 924 683	243 573 536	303 367 561
Tot. Increase (ha)	-660 661	46 000 329	54 976 000
Tot. Increase (%)	-0,3 %	23 %	22 %
Change in land area from 1992 - 2016			
Staple crop	-3 550 200 (6 %)	434 333 (1 %)	29 791 480 (22 %)
Non-staple crop	-1 693 420 (-6 %)	11 085 248 (42 %)	10 012 290 (18 %)
Feed crop	-933 453 -1 %	29 856 547 (55 %)	6 962 364 (19 %)
Fuel crop	5 721 586 (43 %)	5 183 247 (14 %)	6 208 171 (35 %)
Non-food crop	-205 175 (61 %)	-559 046 (-18 %)	2 001 694 (51 %)

Furthermore, disregarding geographical location, Figure 5.2 displays changes in agricultural areas measured in millions of hectare and percentage change.

a) *b)*

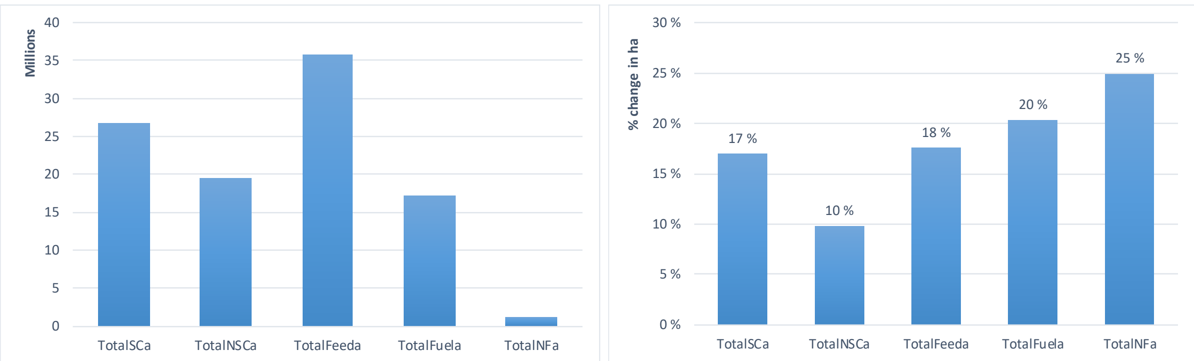
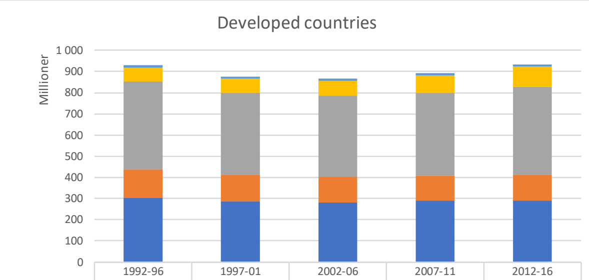


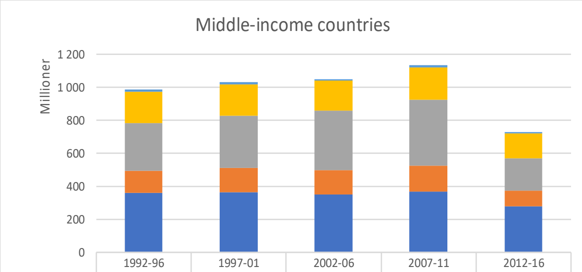
Figure 5.3: Changes in agricultural areas measured in millions of hectare (a) and percentage change (b).

Regarding Hypothesis 1, the graphs shows that, even though the larges changes has happened in the non-food sectors (Figure 5.3b), staple crops and non-staple food crops areas has increased more than fuel –and non-food crops (Figure 5.3a). Further, the largest increase, per hectare, is feed crops, which has increased by 18 % since 1992. Hence, because staple crops have experienced the second largest increase in agricultural area from 1992-2016, we can reject our hypothesis, i.e., non-food– and fuel crops, not staple crops, drive the increase in total agricultural area. However, analyzing Figure 5.2b, non-food crops has had the largest relative increase during the same period, indicating the direction of current agricultural trends.

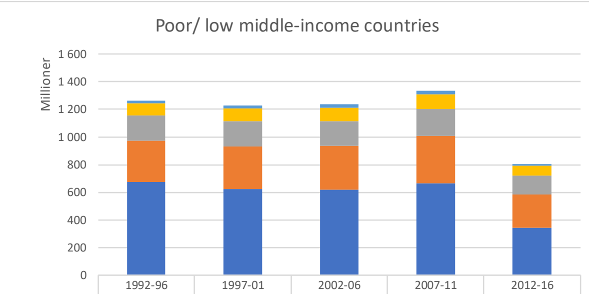
Furthermore, it is interesting to explore the more specific time periods in which agricultural trends changes. I split the 1992-2016 period into five periods. Figure 5.4 shows that total agricultural area decreased slightly between 1996 and 2016, before the agricultural area increased back to estimates for 1992-96. The largest decrease occurred for feed crops, which might be because of increased yield. During the same period, both middle-income –and poor/ low middle-income countries experiences a large drop in agricultural area production after 2011, which can be explained by more productive agricultural production.



	1992-96	1997-01	2002-06	2007-11	2012-16
TotalNFa	11775867	10265476	8965613	9481395	10764635
TotalFuela	64240912	69046811,21	71374114,52	86443275,49	93521036,3
TotalFeeda	413078887	383611610,1	382296051,8	390914321,1	417635459,3
TotalNSCa	134330318,8	124976756	122700337,8	115378417,6	121333451
TotalSCa	305380793,8	289327552,1	281581224,6	290786617,5	290171875,4



	1992-96	1997-01	2002-06	2007-11	2012-16
TotalNFa	13919773	12623980	12061343	12713150	10763714
TotalFuela	189598885,5	189673340,7	180084110,1	195956008	150409668,4
TotalFeeda	291775751	314948413,5	360968058	403368672	192721297,7
TotalNSCa	130786424,1	146751664,9	145200883,1	155656202,5	96410163,51
TotalSCa	362172698,1	366483185,8	353107164,1	367663258,1	279933748,1



	1992-96	1997-01	2002-06	2007-11	2012-16
TotalNFa	20560251	22882033	24920672	27759162	8376031
TotalFuela	89104961,26	91916252,79	95608628,73	106035183,9	71540630
TotalFeeda	181750445,2	181580562	183582210,2	195775763,8	137426878
TotalNSCa	297913240,9	309488687,7	315183808,3	339825727,7	240138336,9
TotalSCa	676376885,2	624302245,8	618918557,4	667727792,2	345134752,1

Figure 5.4: Change in crop group areas over three time periods. Countries are divided into developed countries (a), middle-income countries (b) and poor/ low middle-income countries (c).

5.1.1 Food price trends 1988 – 2017

Throughout history, high food prices have been of high concern, especially among poor households. Since 2000, the world has experienced two major food crises, where high food prices have caused social unrest (2008 and 2010) (Bellemare, 2015). These events are clearly seen in the FAO Food Price Index (FPI), marking the importance of stable food markets. The FAO FPI is a measure of the international prices of food and consists of the average of five commodity group price indices (Cereal, Vegetable oil, Dairy, Meat and Sugar), weighted with the average export shares of each of the groups (FAO, 2018).

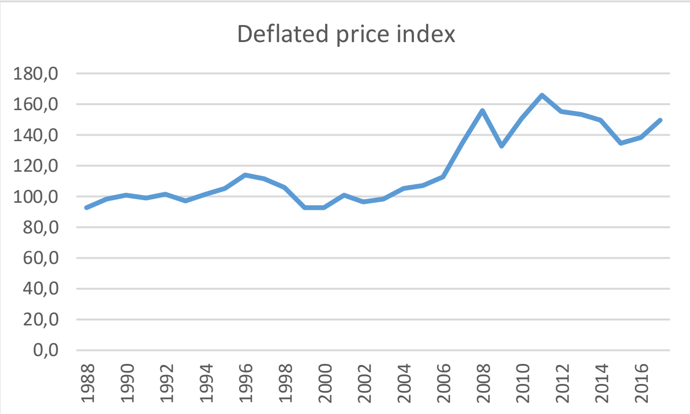


Figure 5.5: FAO Food Price Index. Source: (FAO, 2018)

This next section consists of figures 5.6, 5.7 and 5.8 and illustrates price trends for the different crops from 1988-2017 (Index Mundi, 2018).

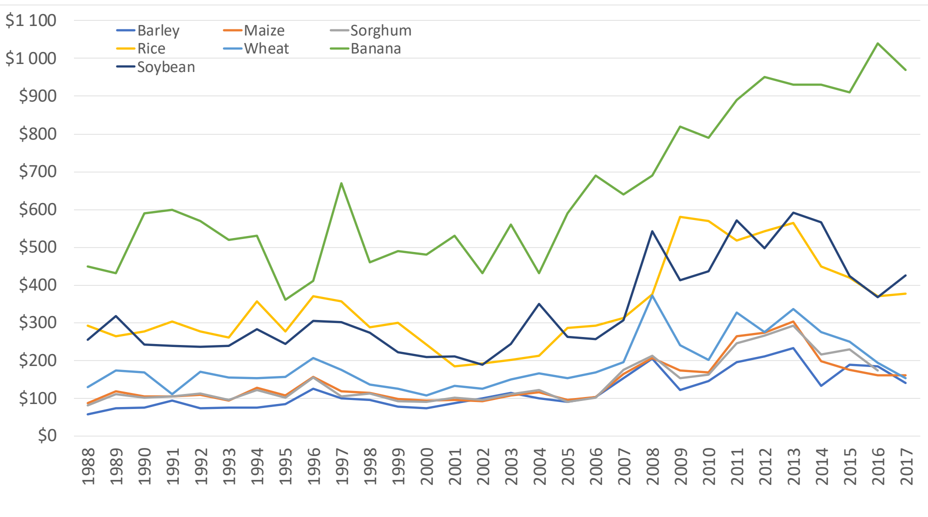


Figure 5.6: Price trends for staple crops, 1988-2017 (current US$). Source: (Index Mundi, 2018)

In figure 5.6, barley, wheat and sorghum are relatively clustered together throughout the entire period. Soybean and rice are on average higher priced and experienced a price hike between 2009 and 2014. Banana is the most valuable staple crop and has had an increasing price trend since 2004.

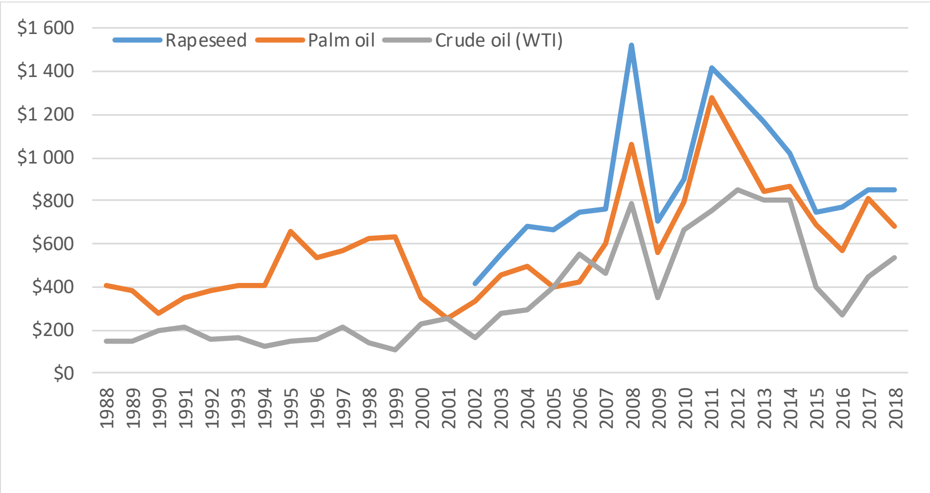


Figure 5.7: Price trends for fuel crops, 1988-2017 (current US$). Source: (Index Mundi, 2018)

In figure 5.7, prices of rapeseed and palm oil has similar price trend throughout the entire period. For illustration, crude oil (West Texas Intermediate) prices per ton is included. It shows that crude oil has similar trend as rapeseed and palm oil with price hikes in 2008 and 2011, and price dumps in 2009 in 2015-2016.

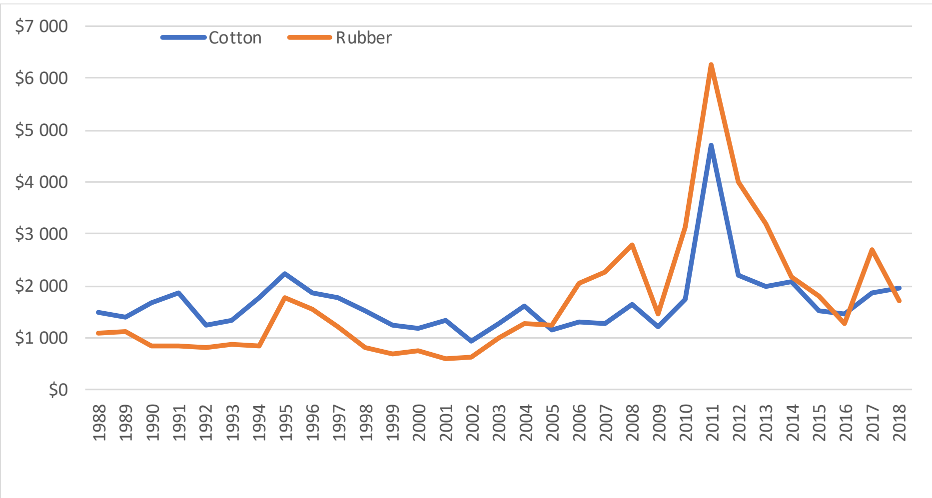


Figure 5.8: Price trends for non-food crops, 1988-2017 (current US$). Source: (Index Mundi, 2018)

Figure 5.8 shows that cotton and rubber have similar price trends throughout the entire period, with rubber being the most valuable of the two. Additionally, prices of non-food crops have similar trends as fuel crops with price hike in 2011 and price dump in 2016. These prices can explain farmer's choice of crops as well as country's agricultural policies.

5.2 Econometric results

5.2.1 Selecting the regression model

This section contains the econometric results for Hypothesis 2.

- H2: The yield area relationship differs between crop types. For staple food, higher yield results in lower crop area; for non-food and fuel crops, higher yield results in larger crop area

The hypothesis is split into four sub-hypotheses, where staple-crops (SC), fuel crops, feed crops and non-food crops (NF) are estimated in separate regression models.

Looking at various factors influencing our dependent variables, I began analyzing data by running a POLS regression with ordinary –and robust standard errors on both the original and main model. This will confirm whether POLS is consistent, especially with respect to standard errors. Because heteroskedasticity and serial correlation are main concerns for using panel data, robust standard errors are often applied as a remedy since they are robust to these concerns (Wooldridge, 2015). Another alternative is to use cluster-robust standard errors, however, because of few clusters (10), it would lead low statistical power and is therefore excluded from the analysis. To test whether robust standard errors were necessary, I ran the White's test for heteroskedasticity and Wooldridge test for autocorrelation for POLS with ordinary standard errors, with the assumption that normally distributed errors has been relaxed.

The null hypotheses of were rejected at significance level of 0.05, indicating that the POLS model suffers from heteroskedasticity and serial correlation. The tests were therefore repeated with robust standard errors. However, the same hypotheses were rejected, indicating that robust standard errors did not help to solve these issues.

Because individual effects/serial correlation is identified in all the POLS estimates, we need test if panel data estimators are more suitable for the analysis. To test for this, I ran the Breusch-Pagan test for random effects. The Breusch and Pagan Lagrangian multiplier test reject that $var(u) = 0$, which indicate that $var(u) > 0$. There is evidence of substantial individual heterogeneity. Hence, POLS is not an appropriate model. To test for model fit, the Ramsey RESET test for functional misspecification was generated, where $P > 0,05$ is considered a strong model and $0.05 > p > 0.001$ is considered a weak model. The Ramsey for POLS rejected any evidence of linearity ($P<0.001$), i.e., the model does not properly account for the relationship between the dependent and explanatory variables. Hence, a panel data estimator might be a better option. For robustness, two additional regressions were generated. First, since autocorrelation in the genuine errors seems to be present, POLS with an AR(1) disturbance and years as dummy variables was tested. Using an F-test, years as dummy variables were not significant, and POLS with an AR(1) disturbance were not considered for further testing, other than comparison between the ordinary and main model. Second, I added lagged variables to the robust standard errors model, however, adding a lagged value did not improve the model.

Further, to check for multicollinearity, i.e., control for correlation between the independent variables, we ran a VIF test. I observed some multicollinearity between population, fuel crop area and non-staple crops, however the mean multicollinearity is below the acceptable limit (VIF < 10), hence with these variables in mind, we can continue with the analysis. The analysis of POLS is shown in Appendix B.2 (Table B.2.1)

In this sub-section, we have tested if POLS with either ordinary, robust, AR(1) or lagged robust standard errors is appropriate for our model. Because of issues with heteroskedasticity, serial correlation and functional form, we concluded that other regression models must be carried out. Recalling from chapter 4.5 that RE and FE are inappropriate models for our analysis, it is suggested to obtain the First Difference model with quadratic terms (Wooldridge, 2015, p. 279). The discussions in the next chapters are therefore mainly based on the FD model.

5.2.1 Testing the models

Hypothesis 2.1 aimed to explore the correlations between change in $D.lnSCy$ and staple crop area $D.lnSCa$. The DF_ord and DF_r generated similar results with no serial correlation, and strong evidence of homoskedasticity and linearity. The DF_lr provided a slightly better fit of the data than the latter, i.e. higher R^2 and smaller rmse, but, the regression suffers from serial correlation and is therefore excluded. I also tested for individual effects using an AR(1) test and added a dummy variable for each year. The F-test generated p-value > 0.05, hence, years from 1993 to 2016 does not have significant impact on the dependent variable, and we exclude years as dummy variables from further analysis. Comparing the results from DF_ord and DF_r, variables are equally significant, however, DF_r has a stronger functional form, which indicates a better model. Hence, further discussions of Hypothesis 2.1 will focus on results generated by FD_r.

In the second regression, with $D.lnNSCa$ as dependent variable, tests for heteroskedasticity and serial correlation was rejected DF_ord and DF_r. The DF_lr does not suffer from serial correlation and provided a slightly higher R^2, however many observations were dropped, which is unfavorable as we originally do not have many observations. The F-test rejected any time-specific effects, while the Ramsey RESET test estimated a solid functional form for all correlations. However, DF_r presents higher values, indicating a better model. Hence, further discussions will focus on results generated by FD_r

Hypothesis 2.2 aimed to explore the correlation between non-food crop areas and non-food crop yields. Hence, three separate regressions with $D.lnFeeda$, $D.lnFuela$ and $D.lnNFa$ as dependent variables were performed, respectively.

With $D.lnFuela$ as dependent variable, tests of heteroskedasticity and serial correlation were rejected for FD_ord and FD_r and accepted for FD_lr with one lag. The VIF test showed a mean multicollinearity below the recommended limit (VIF > 10), i.e. low level of correlation between independent variables. Hence, results generated by DF_lr was used in our analysis.

In the fourth regression, using $D.lnFeeda$ as dependent variable, none of the models generated a strong functional form, and rejected tests of heteroskedasticity and serial correlation. Additionally, the VIF test showed a mean multicollinearity below the recommended limit. To solve for issues heteroskedasticity or serial correlation, removing control variables was attempted, without providing a better model. Keeping in mind that we have issues of heteroskedasticity and serial correlation, results generated by the DF_r model was used for our analysis as it provided the strongest model (p > 0.001). Further, the R^2 is below 0.20, which means that the explanatory variables only explains up to 20% of changes in the dependent variable. Hence the covariates do a poor job explaining and/ or predicting the response values.

In the fifth regression, with $D.lnNFa$ as dependent variable, the tests for functional form/ misspecification indicated an unfit model (p = 0.000), i.e., the combination of explanatory variables does not have any power in explaining the response variable. Tests of heteroskedasticity and serial correlation also indicated larger issues with the regression model, even after attempting to remove control variables. Nevertheless, the model generated statistically significant negative results for $D.lnNFy$ and $D.lngdp$, and significant positive results for $D.lnFeedy$ and $D.lnpopa$.

5.2.2 Model results

This sub-chapter is divided into three parts, each presenting the econometric results of the hypotheses, as well as comparing expected and actual signs.

Hypothesis 2

Hypothesis 2 states that *the yield area relationship differs between crop types. For staple food, higher yield results in lower crop area; for non-food and fuel crops, higher yield results in larger crop area.*

Table 5.3 presents the results of the FD regression for Hypothesis 2, and is divided into two parts, H2.1 and H2.2. The dependent variable for H2.1 is $D.lnSCa$ and $D.lnNSCa$, while $D.lnFeeda$, $D.lnFuela$, $D.lnNFa$, and $D.lnFuela$ are dependent variables for H2.2, in separate regressions. The main independent variables for the hypothesis were $D.lnSCy$, $D.lnNSCy$, $DlnFeedy$, $DlnFuely$, and $DlnNFy$, whereas the remaining variables were used as control variables. The VIF test of multicollinearity obtained adequate mean correlation between the independent variables (<10) for both H2.1 and H2.2. Hence, I continued with our variables and ran the tests for first differences DF_ord, DF_r and DF_lr.

For H2.1, I hypothesized that an increase in staple crop yield reduces staple crop area. In Table 5.3, $D.lnSCy$, $D.$lnNSCy, $D.lnFeedy$, and $D.lnforest$ are highly significant. Staple crop yield is negatively correlated, while non-staple food crop- and feed crop yield, and forest area is positively correlated. In Chapter 4.3 we predicted the sign of staple crop yield to be negative, signs of fuel and feed crop yield to be positive and forest to be unknown.

According to Table 5.3, when staple crop yield increases with one %, staple crop area decreases by 0.43 %. Staple crop area is expected to increase with 0.13 %, and 0.28 %, when non-food staple crop- and feed crop yield increases with one tones/ha, respectively. In relation to theory (Chapter 3), the Borlaug hypothesis vs. the Jevons paradox discussion, these results are in line with both Borlaug and Jevons argued, i.e. improved productivity reduces and enhances agricultural expansion.

For non-staple food crops, $D.lnSCy$, $D.lnNSCy$ and $D.lngdp$ are positively statistical significant. The table indicates that an increase in non-staple food crop yield with one % decreases crop area by 0.32 %, which is in line with our hypothesis. However, an increase in staple crop yield has the opposite effect, namely an increase in non-staple food crop area by 0.31 %. The reason for this might be because many non-staple food crops are also feed and fuel crops, such as sugar cane, which might change the expected signs. Although, there is a high significance and the effect is consistent with results from regressing staple crops.

Table 5.3: Estimation results for Hypothesis 2.1 and 2.2, with staple crop -, fuel crop -, feed crop -, and non-food crop area as dependent variables.

Hypothesis	H2.1	H2.2	H2.2	H2.2	H2.2
Dep. Variable	SCa	NSCa	Feeda	Fuela	NFa
Variable name	FD_r	FD_r	FD_r	FD_lr	FD_lr
D.lnSCy	**-0.434***** (0.0984)	**0.305***** (0.0878)	-0.0193 (0.0924)	-0.0515 (0.117)	-0.455 (0.275)
D.lnNSCy	**0.125*** (0.0525)	**-0.315***** (0.0898)	**0.156**** (0.0589)	-0.0251 (0.0788)	-0.104 (0.287)
D.lnFeedy	**0.281***** (0.0702)	-0.0230 (0.0629)	**0.143*** (0.0695)	0.0988 (0.0747)	**0.665*** (0.295)
D.lnFuely	-0.0212 (0.0610)	0.0892 (0.0688)	-0.0239 (0.0665)	**0.212*** (0.0901)	-0.371 (0.287)
D.lnNFy	-0.00345 (0.0155)	-0.000445 (0.0108)	0.00530 (0.0205)	0.00252 (0.0121)	**-0.635***** (0.129)
D.lnforest	**0.0179*** (0.00843)	-0.00420 (0.00935)	**-0.0348***** (0.00790)	**0.0178*** (0.00890)	0.0399 (0.0321)
D.lnpop	0.472 (0.569)	0.283 (0.582)	0.178 (0.628)	-0.981 (0.703)	**8.247*** (3.225)
D.lnUnem	-0.0261 (0.0286)	-0.0583 (0.0421)	**-0.0807**** (0.0300)	-0.0705 (0.0362)	-0.134 (0.100)
D.lngdp	0.0280 (0.0286)	**0.0708*** (0.0303)	0.0146 (0.0317)	0.0641 (0.0347)	**-0.333*** (0.142)
L.u				-1.95e-09 (1.68e-09)	-6.80e-08 (6.25e-08)
_cons	-0.00180 (0.00736)	-0.00247 (0.00833)	0.00369 (0.00829)	**0.0207*** (0.0104)	**-0.101*** (0.0487)
N	237	237	237	237	237
R^2	0.244	0.204	0.187	0.194	0.385
adj. R-sq	0.214	0.173	0.155	0.159	0.357
rmse	0.0618	0.0691	0.0618	0.0713	0.291

Notes: * p<0.05, ** p<0.01, *** p<0.001; p-values in parenthesis

In Table 5.4, $D.lnFuely$ is weakly positive significant. We hypothesize that higher fuel crop yield, increases fuel crop area, and ran the regression with the same control variables as in hypothesis 2.1. The regression results in Table 5.4 shows that one % increase in fuel crop, increases fuel crop area with 0.21 %. This result correspond with findings in (Angelsen and Kaimowitz, 2001) and are in line with the Jevons paradox.

In contrast to our expectations, $D.lnforest$ is weakly positive significant (p-value < 0.05) with fuel crop area. An increase in forest area with one % increases fuel crop area 0.02 %. We expected that an increase in forest area would either lead to decrease in fuel crop area because they are expected to be contrasting variables or leave more forest area available for extraction. However, from our results we observe the opposite, i.e. that an increase in forest area also increases fuel crop area. As explained in Chapter 4.4, one of the major issues with data from FAO is that there might be noise in the data as countries use different methods in compiling data (they report). Increase in cultivation of biofuel crops like palm oil might therefore be counted as increase in forest cover, even though FAO has specified that no agroforests are included in the dataset.

According to Table 5.4, an increase in GDP increases fuel crop area, while an increase in population decreases fuel crop area. The reasons might be that as wealth increases, the country focuses production on non-food crops, i.e., in accordance with Prebisch-Singer hypothesis. Another possible interpretation is that, as fuel crops also provides food and feed, hence, an increase in GDP per capita corresponds with increased consumption of food, feed and fuel, which results in increased agricultural area. However, the variables are not significant due to low correlation between GDP, population and fuel crop area[4], hence the control variables do not have much explanatory power. Further, it is important to remember that our analysis suffers from heteroskedasticity and serial correlation. The R^2 is 0.19 and rmse is 0.071, which tells us that our explanatory variables only explain 19,4 % of variances in our model, with low model fit.

I hypothesized that an increase in feed crop yield, increases feed crop area. Table 5.2 indicates that $D.lnFeedy$ and $D.lnNSCy$ is positive statistically significant, while $D.lnforest$ and $D.lnUnem$ is negative and statistically significant. Thus, when feed crop yield increases with one %, feed crop area increases with 0.14 %. These results correspond with our findings in the latter hypothesis, namely the Jevons paradox. The second significant variable, non-food crops

[4] Correlation matrix for hypothesis 2 in Appendix C B.1 (Table B.1.2),

yield, unexpectedly increases feed crop area with 0.016 % per yield percentage increase. This might be due to collinearity, i.e., crops that are qualified as feed crops are also large food crops. Hence, an increase in feed crop area will also increase non-staple food crop area. Furthermore, as expected, forest area is negatively and significant correlated with feed crop area. This can also confirm some of hypothesis two where we ask which crops that impacts forest cover. The regression shows that when forest cover increases with one %, feed crop area decreases with 0.03 %. From the last significant variable, unemployment rate, suggests that feed crop area decreases if unemployment increases with one %. This might be the result because, as agriculture is a labor-intensive sector, increase in unemployment means that the sector is becoming more productive. And as we can observe from the results, increase in feed crop yield also increases feed crop area. Once again, we must keep in mind the issues of heteroskedasticy, serial correlation and multicollinearity when evaluating the results. The adjusted R^2 is 0.187 and rmse is 0.0618, which tells us that our explanatory variables only explain 18.7 % of variances in our model, with low model fit.

If non-food crop yield increases with one %, non-food crop area decreases with 0.635 %. This is in accordance with Borlaug hypothesis and in contrast to Hypothesis 2.2. Further, an increase in feed yield corresponds with increase in non-food crop area, indicating that we might have some collinearity in our mode. However, because the model suffers from misspecification, the results cannot be concluded upon. The reason for the weak functional form might be that because the non-food crop group is the smallest of all the crop groups. Hence, its explanatory power might be absorbed by other crop groups, and variables important for the model is omitted or measurement errors.

One of the main issues in the regression analysis was misspecified functional forms, i.e., unreliable models due to non-linearity. To solve this, it often makes sense to add squared variables, which captures diminishing or increasing effects on the dependent variable. If the squared variables are significant, they can be added to the model (Wooldridge 2015, p.279). Because issues of functional form were detected in the hypothesis, we regressed them a second time adding squared variables to area and yield variables in POLS with FD. This is because area and yield are nonlinear, i.e. the independent variables are expected to increase/decrease with dependent variable until a certain point before they change direction. Hence, the quadratic function solves the issues of non-linearity and lets us calculate tipping points in area and yield production. Including quadratic variables in our regression analysis provided a stronger model (functional form and R^2), as expected, and more statistically significant variables. However, larger issues with heteroscedasticity and serial correlation were also detected[5]. This might be because quadratic terms can be symptomatic of other functional form problems, such as using the level of a variable when the logarithm is more appropriate, or vice versa. Additionally, even though squaring variables can solve issues with functional form, the interpretation is very complicated and requires calculating maximum and minimum values to find turning point. Therefore, the method was not included as a main model for regressions in this thesis.

5.3 Discussion

Changing demand for different agricultural products causes shifts in land use and new markets within the agricultural sector has increased pressure on production areas. However, one process could reduce this pressure, at least in part, by increasing agricultural yields rather than expanding crop land. In Chapter 3, we introduced theories in relation to these questions. Norman Borlaug argued that improved agricultural practices decrease agricultural area, while William Jevons argued the opposite. The findings from the current analysis may improve our knowledge about how changes in yield in practice influences agricultural areas and correlations between crop groups. It may also give better knowledge about the extent of which forest cover changes with agricultural land use trend, as well as the impacts of changes in unemployment rate, GDP per capita and population growth.

Growth in agricultural output has previously been obtained by bringing new land in to production. Today, new technologies have made it possible to increase output without necessarily increasing area of production, but in practice, this is not always the case.

In our regression analysis, we observed the effects of increased yield on crop area for the different crop groups. The results show that the land-sparing effects of increased crop yield was uneven. Increasing yield in staple crops has strong positive land-sparring effects, reducing staple crop area by -0.43 % per one % increase in yield. The reason for this may be that market prices of food crops, agricultural rents, are low compared feed, fuel, forest and other market commodities. Hence, the additional benefits of increasing staple crop area may not exist, according to our analysis. Besides market effects, these patterns might also occur because of political factors, that favor feed- and fuel crops with low yield, at the expense of food

[5] Regression results and a short discussion in Appendix A

production. Thus, the results in our regression, where staple crop area decreases as yield increases, may also be explained by political policy decisions, and not exclusively by more productive land use. Nonetheless, we can confirm the Borlaug hypothesis with high significance regarding staple crops, i.e., increasing staple crop yield do not trigger area expansion. Further, increasing feed crop- and non-food staple crop yield, and forest area increases staple crop area with statistical confidence. Hence, when non-staple crops become more productive, more agricultural land becomes available for staple crop production.

From H2.2, increase in feed crop yield by 1 % increases feed crop area with 0.67 %. Besides normal market mechanisms, there might be government policies favoring production of feed crops, especially for exports. According to Table 5.2, it is the developed countries who are the main producers of feed crops, hence, according to Prebisch-Singer thesis, wealthier countries has focused production on non-food crops rather than food crops.

The same reasoning is valid for fuel crop yield, where, as fuel crop production becomes more productive, fuel crop area significantly increases by 0.22 %. The effects of higher feed crop yield and food crop yield on staple crop area are in line with the theory of Jevons paradox and some of the findings of Angelsen and Kaimowitz (2001). It is also interesting to notice that GDP per capita has is positively correlated with feed and fuel crop. Hence, the regression analysis suggests that when wealth increases, agricultural area expands. However, the variables are not statistically significant and might be impacted by endogeneity and multicollinearity, and one should be careful when making any conclusions. The last regression for H2.2, testing for non-food crop area indicated that increased productivity significantly reduces area by 0.64 %. This result contradicts the hypothesis and supports Borlaug hypothesis. Additionally, the non-food crop area significantly decreases as GDP per capita increases. Hence, we can accept Hypothesis 2, that *higher yield of staple crops reduces total agricultural area, while higher yield for feed-, fuel- and non-food crops enhance total agricultural area,* with high statistical confidence for staple-, feed-, and fuel crops, while it must be rejected for non-food crops.

To sum up, prior to the 1900s, the most common pattern in agricultural land use change involved simultaneous increases in agricultural yields and cultivated areas (Rudel et al., 2009). However, improved technologies have made it possible to increase output without increasing area. The regressions for H2.1 shows that higher staple crop yield reduces area for staple crops, with high statistical confidence. On the other hand, the regressions in H2.2 showed that increase in feed crop- and fuel crop yield enhances negative land sparing effects due to higher market prices from increasing demand due to change in diets and high energy prices. According to the

Prebisch-Singer hypothesis, countries should focus their production on non-food goods instead of non-food crops to enhance economic growth. The regression shows there might also be some political incentives to invest in feed and fuel crops. However, as higher yield also enhances agricultural expansion for feed -and fuel crops, yield improvements should be incentivized in places where fuel and feed crop area are at its maximum stage because it will have the least impact on agricultural area expansion. Nevertheless, explained by changes in demand for agricultural products, our analysis confirms both the Borlaug hypothesis and the Jevons paradox with high significance.

5.4 Limitations of the analysis

Originally, the dataset included 20 countries over a 55-year period and several control variables, but because of missing values, our dataset became unbalanced and we had to settle with ten countries over a 25-year period. Narrowing down the dataset also limited our options of regression analysis. Hence, computing panel data regression through random effects or fixed effects became inappropriate.

Furthermore, the analysis confirmed the limits mentioned in chapter 4.4 where three estimation issues were suggested. First, our dataset is biased, i.e. our dataset was not collected randomly. Second, there might be some issues of endogeneity as we might have omitted unknown variables that are essential for our study. This issue became evident when we encountered issues with our functional form, which could not be fixed by excluding variables from our regression. Hence, there might exist some variables in our error term with large explanatory power that should have been included in our analysis. And third, there are large measurement errors due to methods of data collection by FAO which might have caused area noise in our dataset. Hence, some of our variables behaved unexpectedly.

In the regression, I also encountered some issues with multicollinearity, serial correlation and heteroskedasticity. These are major weaknesses of our model and can be explained by the nature of our dataset. The variances between our panels were very different as it included both major agricultural economies and small ones. Additionally, because of the different usage areas of crops included in the analysis, several explanatory variables are dependent on each other. Hence, we also had problems with multicollinearity. Even so, the regressions presented in this analysis did not have many issues with misspecification and we could continuous with our regression, keeping in mind that we have issues with heteroskedasticity and serial correlation.

5.5 Further research

To get a better understanding of driving forces behind agricultural land use changes, more countries over longer time periods are suggested. Additionally, more control variables such as market supply and demand, agricultural subsidy, daily per capita food consumption (kcal), export and import share of crops, prices of all crops for each year, are suggested to solve issues of endogeneity, although including all these variables might create new endogeneity ad multicollinearity issues. Another possible approach is to divide the population within each country in to urban and rural population. As mentioned in Chapter 3, the Prebisch-Singer hypothesis suggests that countries should focus production on non-food crops rather than food crops to enhance economic growth. By dividing the population in to rural and urban population and compare it with changes in GDP per capita, it would be possible to investigate if this hypothesis is valid for all countries. It would also explain why some countries are focusing their production on feed and fuel crops, rather than food.

Furthermore, one of the major limitation of this analysis was multicollinearity. It might therefore be suggested to either establish clearer differences between crop groups or include more crops in the analysis. Although better data by year and country on the allocation of "multi-functional" crops on the different groups will allow for a more precise analysis.

6 Conclusion

Changing demand motivate shifts in agricultural land use. Until recently, food production was the main agricultural practice for many countries. Recently, however, new technology has opened market opportunities to enhance economic growth through cultivating crops for unconventional purposes. By using panel data with a sample of crops from the top ten staple-crop countries, divided into five crop groups, we studies land use changes in the periods 1992-2016. We estimated the driving forces mainly by using Pooled OLS with First Differences estimator and ran both an original model and main model to find the preferred model. The basic premise is that land use changes are driven by non-food and fuel crops, not by staple-crops. The first question to ask when assessing trends in the agricultural sector is: *What are the trends in production of non-food crops and staple crops?* Subsequently, the thesis investigates the debate of Borlaug hypothesis vs. Jevons paradox, and asked: *Does higher yield of staple crops and non-food crops enhance or reduce total agricultural area?*

The thesis shows that while there is high correlation between production of staple crops and non-food crop, descriptive statistics suggests that feed crops are the main contributor of agricultural area expansion, which has increased by 18% since 1992. Staple crops have experienced the second highest increase in agricultural area during the same period, hence, it is also a large contributor to agricultural expansion, especially in poor/ low middle-income countries as confirmed by the Prebisch-Singer hypothesis. Further, feed-, fuel- and non-food crops has experienced the largest relative growth, indicating the direction of trends in agricultural production.

The Borlaug hypothesis and Jevons paradox theories proposes two opposing outcomes of increased agricultural yield. To tests for this, we introduced the following hypothesis:

i. For staple crops: Higher yield *reduces* crop area (Borlaug hypothesis)
ii. For non-food crops: Higher yield *increases* crop area (Jevons paradox)

The first main finding indicates that increase in staple crop yield reduces staple crop area. This is in accordance with the Borlaug hypothesis. Hence, higher staple crop yield does not encourage agricultural expansion and deforestation. The second main finding indicates that increase in feed- and fuel crops yields increases cultivation area. However, higher non-food crop yield decreases agricultural area. Hence, we can only somewhat confirm the Jevons paradox: higher yield of fuel- and feed crops do not enhance land sparring, but instead encourage agricultural expansion, while the relationship does not hold for non-food crops. Hence, a major implication from the analysis of this thesis is to better distinguish between different types of crops in the debate on the impact of yield on crop area expansion.

7 References

Achterbosch, T. J., van Berkum, S., Meijerink, G. W., Asbreuk, H. and Oudendag, D. (2014) *Cash crops and food security: Contributions to income, livelihood risk and agricultural innovation.* LEI Wageningen UR.

Akar, T., Avci, M. and Dusunceli, F. (2004) 'Barley: Post harvest operations', *Food and Agriculture Organization (FAO) of the United Nations, The Central Research Institute for Field Crops, Ankara, Turkey,* 64.

Alcott, B. (2005) 'Jevons' paradox', *Ecological economics,* 54(1), pp. 9-21.

Amido, T. (2015) *The Latest Hot Crop: Sorghum: This gluten-free whole grain has so many nutritional benefits.* . Foodnetwork.com: Foodnetwork.com. Available at: https://www.foodnetwork.com/fn-dish/news/2011/03/the-latest-from-food-network-stars (Accessed: 20.04 2018).

Anderson, W., You, L., Wood, S., Wood-Sichra, U. and Wu, W. (2014) 'A comparative analysis of global cropping systems models and maps'.

Angelsen, A. (1999) 'Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights', *Journal of development economics,* 58(1), pp. 185-218.

Angelsen, A. (2007) *Forest cover change in space and time: combining the von Thünen and forest transition theories.* World Bank Publications.

Angelsen, A. (2010) 'Policies for reduced deforestation and their impact on agricultural production', *Proceedings of the National Academy of Sciences,* 107(46), pp. 19639-19644.

Angelsen, A. 2017. Synergies and trade-offs between forestland management and food system. *Climate change, land use and food security.* FAO.

Angelsen, A. and Kaimowitz, D. (1999) 'Rethinking the causes of deforestation: lessons from economic models', *The world bank research observer,* 14(1), pp. 73-98.

Angelsen, A. and Kaimowitz, D. (2001) *Agricultural technologies and tropical deforestation.* CABi, p. 402.

Baffes, J. and Etienne, X. L. 'Reconciling high food prices with Engel and Prebisch-Singer'. *International Conference on Food Price Volatility: Causes and Consequences, Rabat, Morocco*: Citeseer.

Balmford, A., Green, R. and Scharlemann, J. P. (2005) 'Sparing land for nature: exploring the potential impact of changes in agricultural yield on the area needed for crop production', *Global Change Biology,* 11(10), pp. 1594-1605.

Barbier, E. B. (1989) 'Cash crops, food crops, and sustainability: the case of Indonesia', *World development,* 17(6), pp. 879-895.

Bellemare, M. F. (2015) 'Rising food prices, food price volatility, and social unrest', *American Journal of Agricultural Economics,* 97(1), pp. 1-21.

Benhin, J. K. (2006) 'Agriculture and deforestation in the tropics: a critical theoretical and empirical review', *AMBIO: A Journal of the Human Environment,* 35(1), pp. 9-16.

Benoit, K. (2011) 'Linear regression models with logarithmic transformations', *London School of Economics, London,* 22(1), pp. 23-36.

Borlaug, N. (2007) 'Feeding a hungry world', *Science magazine,* 318.

Buttler, R. (2013) *Europe importing more palm oil for biofuels, raising risks for rainforests*. mongabay.com. Available at: https://news.mongabay.com/2013/09/europe-importing-more-palm-oil-for-biofuels-raising-risks-for-rainforests/ (Accessed: 21.04 2018).

Byerlee, D., Stevenson, J. and Villoria, N. (2014) 'Does intensification slow crop land expansion or encourage deforestation?', *Global Food Security,* 3(2), pp. 92-98.

CEPR (2010) *Food Prices and Rural Poverty.* Centre for Economic Policy Research.

CGIAR (2016) *Cowpea*. Available at: https://www.cgiar.org/our-strategy/crop-factsheets/cowpea (Accessed: 20.03 2018).

Cuddington, J. T., Ludema, R. and Jayasuriya, S. A. (2002) *Prebisch-Singer Redux.* Central Bank of Chile.
De la Torre Ugarte, D. 'Opportunities and challenges of biofuels for the agricultural sector and food security of developing countries'. *Expert Meeting on Participation of Developing countries in New Dynamic Sectors for World Trade, Geneva*.
Edgerton, M. D. (2009) 'Increasing crop productivity to meet global needs for feed, food, and fuel', *Plant physiology,* 149(1), pp. 7-13.
Eicher, C. K. and Staatz, J. M. (1998) *International agricultural development.* JHU Press, p. 156.
Evenson, R. and Rosegrant, M. (2003) 'The Economic Consequences 23 of Crop Genetic Improvement Programmes', *Crop variety improvement and its effect on productivity: the impact of international agricultural research*, pp. 473.
Ewers, R. M., Scharlemann, J. P., Balmford, A. and Green, R. E. (2009) 'Do increases in agricultural yield spare land for nature?', *Global Change Biology,* 15(7), pp. 1716-1726.
FAO (1996) *The world sorghum and millet economies: facts, trends and outlook.* Food & Agriculture Org.
FAO (2000) *The World Cassava Economy: Facts, Trends and Outlook.* International Fund for Agricultural Development.
FAO (2017) *The future of food and agriculture: Trends and challenges*: United Nations.
FAO (2018) 'FAO Food Price Index'. Available at: http://www.fao.org/worldfoodsituation/foodpricesindex/en/ (Accessed.
FAOSTAT (2013) 'Food and agriculture data' (Accessed.
FAOSTAT (2018a) 'DEFINITION AND CLASSIFICATION OF COMMODITIES'. Available at: {FAOSTAT, 2018 #226} (Accessed.
FAOSTAT (2018b) 'Food and agriculture data' (Accessed.
Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Tóth, G., Van Velthuizen, H., Verelst, L. and Wiberg, D. (2012) 'Global Agro-ecological Zones (GAEZ v3. 0)-Model Documentation'.
Foley, J. (2016) *A five-step plant to feed the world*. Feeding 9 billion: National Geographic. Available at: https://www.nationalgeographic.com/foodfeatures/feeding-9-billion/ (Accessed: 22.01 2018).
Francois, J., Stringer, R. and Sarris, A. (2005) *Agricultural trade and poverty: can trade work for the poor?*: Food and Agriculture Organization.
Grujcic, D., Hansen, T. H., Husted, S., Drinic, M. and Singh, B. R. (2018) 'Effect of nitrogen and zinc fertilization on zinc and iron bioavailability and chemical speciation in maize silage', *Journal of Trace Elements in Medicine and Biology*.
Harvey, F. (2013) *Growth in crop yields inadequate to feed the world by 2050 – research*. Environment: The Guardian. Available at: https://www.theguardian.com/environment/2013/jun/20/crop-yeilds-world-population (Accessed: 13.02 2018).
Hertel, T. W. (2012) 'Implications of agricultural productivity for global cropland use and GHG emissions: Borlaug vs. Jevons', *Global Trade Analysis Project (GTAP) Working Paper,* 69.
IDRC (2010) *Facts & Figures on Food and Biodiversity*: International Development Research Center. Available at: https://www.idrc.ca/en/article/facts-figures-food-and-biodiversity (Accessed: 17.02 2018).
IGC (2014) *Five-year global supply and demand projections*, http://www.igc.int: International Grains Council. Available at: http://www.igc.int/en/downloads/grainsupdate/igc_5yrprojections2014.pdf.
Index Mundi (2018) *Crops - Montly Price - Commodity Prices - Price Charts, Data, and News* https://www.indexmundi.com/search.html: Index Mundi. Available at: https://www.indexmundi.com/search.html (Accessed: 20.04 2018).
Islas-Rubio, A. and Higuera-Ciapara, I. (2002) 'Soybeans: post-harvest operations', *FAO, United Nations*.

Jimoh, A., Abdulkareem, A., Afolabi, A., Odigure, J. and Odili, U. (2012) 'Production and characterization of biofuel from refined groundnut oil', *Energy Conservation*: InTech.
Kaimowitz, D. and Angelsen, A. (2008) 'Will livestock intensification help save Latin America's tropical forests?', *Journal of Sustainable Forestry,* 27(1-2), pp. 6-24.
Kindleberger, C. P. (1943) 'Planning for foreign investment', *The American Economic Review,* 33(1), pp. 347-354.
Kraft, M. 2018. Engels law. Wikipedia: Wikipedia.
Laitner, J. (2000) 'Structural change and economic growth', *The Review of Economic Studies,* 67(3), pp. 545-561.
Lam, E., Shine, J., Da Silva, J., Lawton, M., Bonos, S., Calvino, M., Carrer, H., SILVA-FILHO, M. C., Glynn, N. and Helsel, Z. (2009) 'Improving sugarcane for biofuel: engineering for an even better feedstock', *Gcb Bioenergy,* 1(3), pp. 251-255.
Leff, B., Ramankutty, N. and Foley, J. A. (2004) 'Geographic distribution of major crops across the world', *Global Biogeochemical Cycles,* 18(1).
Maxwell, S. and Fernando, A. (1989) 'Cash crops in developing countries: the issues, the facts, the policies', *World development,* 17(11), pp. 1677-1708.
McNally, R., Enright, A. and Smit, H. (2014) 'Finding the Right Balance: Exploring Forest and Agriculture Landscapes'.
Monfreda, C., Ramankutty, N. and Foley, J. A. (2008) 'Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000', *Global biogeochemical cycles,* 22(1).
Mula, M. and Saxena, K. (2010) *Lifting the level of awareness on pigeonpea-a global perspective.* International Crops Research Institute for the Semi-Arid Tropics, p. 107.
Nellis, M. D., Price, K. P. and Rundquist, D. (2009) 'Remote sensing of cropland agriculture', *The SAGE handbook of remote sensing,* 1, pp. 368-380.
Nelsen, A. (2016) *Leaked figures show spike in palm oil use for biodiesel in Europe*. theguradian.com. Available at: https://www.theguardian.com/environment/2016/jun/01/leaked-figures-show-spike-in-palm-oil-use-for-biodiesel-in-europe (Accessed: 21.04 2018).
Nepstad, D. C. and Stickler, C. M. (2008) 'Managing the tropical agriculture revolution', *Journal of Sustainable Forestry,* 27(1-2), pp. 43-56.
Nguyen, V. (2002) 'Rice production, consumption and nutrition. Chapter I'.
NOAA (2017) *What is remote sensing?* Available at: https://oceanservice.noaa.gov/facts/remotesensing.html (Accessed: 20.04 2018).
Nuss, E. T. and Tanumihardjo, S. A. (2010) 'Maize: a paramount staple crop in the context of global nutrition', *Comprehensive reviews in food science and food safety,* 9(4), pp. 417-436.
O'Connor, S. (2014) *Staple Food Crops of the World*: National Geographic. Available at: https://www.nationalgeographic.org/maps/wbt-staple-food-crops-world/ (Accessed: 17.02 2018).
OECD (2008) *Consensus document on compositional considerations for new varieties of barley (Hordeum vulgare L.): key food and feed nutrients and anti-nutrients. Series on the Safety of Novel Foods and Feeds*: OECD.
Pavelescu, F.-M. (2011) 'Some aspects of the translog production function estimation', *Romanian Journal of Economics,* 32(1), pp. 41.
Portmann, F. T., Siebert, S. and 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', *Global Biogeochemical Cycles,* 24(1).
Ramankutty, N., Evan, A. T., Monfreda, C. and Foley, J. A. (2008) 'Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000', *Global Biogeochemical Cycles,* 22(1).
Ray, D. K., Mueller, N. D., West, P. C. and Foley, J. A. (2013) 'Yield trends are insufficient to double global crop production by 2050', *PloS one,* 8(6), pp. e66428.
Roser, M. (2018) 'Employment in Agriculture'.

Rudel, T. K., Schneider, L., Uriarte, M., Turner, B. L., DeFries, R., Lawrence, D., Geoghegan, J., Hecht, S., Ickowitz, A. and Lambin, E. F. (2009) 'Agricultural intensification and changes in cultivated areas, 1970–2005', *Proceedings of the National Academy of Sciences,* 106(49), pp. 20675-20680.

Ruf, F. (2001) 'Tree crops as deforestation and reforestation agents: the case of cocoa in Côte d'Ivoire and Sulawesi', *Agricultural technologies and tropical deforestation*, pp. 291-315.

Rulli, M. C., Saviori, A. and D'Odorico, P. (2013) 'Global land and water grabbing', *Proceedings of the National Academy of Sciences,* 110(3), pp. 892-897.

Ruttan, V. W. and Hayami, Y. (1972) *Strategies for agricultural development.* Food Research Institute, Stanford University.

Rye and Health (2013) *Statistics and Usage*. http://www.ryeandhealth.org/: GRAINITYproject. Available at: http://www.ryeandhealth.org/statistics-a-usage (Accessed: 21.04 2018).

Sali, G. 'Agricultural Land Consumption in Developed Countries'. *2012 Conference, August 18-24, 2012, Foz do Iguacu, Brazil*: International Association of Agricultural Economists.

Siebert, S., Portmann, F. T. and Döll, P. (2010) 'Global patterns of cropland use intensity', *Remote Sensing,* 2(7), pp. 1625-1643.

Stevenson, J., Byerlee, D., Villoria, N., Kelley, T. and Maredia, M. (2011) 'Agricultural technology, global land use and deforestation: A review', *CGIAR, Rome*, pp. 318.

Stevenson, J. R., Villoria, N., Byerlee, D., Kelley, T. and 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), pp. 8363-8368.

Tan, K., Lee, K., Mohamed, A. and Bhatia, S. (2009) 'Palm oil: addressing issues and towards sustainable development', *Renewable and sustainable energy reviews,* 13(2), pp. 420-427.

Tenenbaum, D. J. (2008) 'Food vs. fuel: diversion of crops could cause more hunger', *Environmental health perspectives,* 116(6), pp. A254.

Tiffin, R. and Irz, X. (2006) 'Is agriculture the engine of growth?', *Agricultural Economics,* 35(1), pp. 79-89.

Torres-Reyna, O. (2007) 'Panel data analysis fixed and random effects using Stata (v. 4.2)', *Data & Statistical Services, Priceton University*.

Ufop (2013) *Rapeseed – Opportunity or risk for the future!?* Available at: http://www.etipbioenergy.eu/images/ufop_brochure_rape_seed_2013.pdf.

Verma, S. K. (2015) 'Political Economy of Global Rush for Agricultural Land: a Tract on India's Overseas Acquisitions', *Future of Food: Journal on Food, Agriculture and Society,* 2(2), pp. 62-68.

Vogel, S. (2017) *Global Wheat Demand: Feeding the World by Milling and Feeding*. https://research.rabobank.com: Rabobank. Available at: https://research.rabobank.com/far/en/sectors/grains-oilseeds/global_wheat_demand_article_1.html (Accessed: 21.04 2018).

Von Thünen, J. H. (1966) *Isolated state.* Pergamon Press.

Waggoner, P. E. and Ausubel, J. H. (2001) 'How much will feeding more and wealthier people encroach on forests?', *Population and Development Review,* 27(2), pp. 239-257.

Weingärtner, L. (2010) *Assessment and appraisal of Foreign Direct Investments (FDI) in land in view of food security. Agriculture, fisheries and food* Eschborn: GTZ.

Wiggins, S., Henley, G. and Keats, S. (2015) 'Competitive or complementary? Industrial crops and food security in sub-Saharan Africa', *Overseas Development Institute Report, 41pp*.

Williams, R. (2015) 'Panel Data: Very Brief Overview', *University of Notre Dame*.

Wooldridge, J. M. (2014) *Introduction to Econometrics: Europe, Middle East and Africa Edition.* Cengage Learning.

Wooldridge, J. M. (2015) *Introductory econometrics: A modern approach.* Nelson Education, p. 373.

World Bank (2008) *Agriculture for Developent*, Washington DC: World Bank.

World Bank (2010) *Rising Global Interest in Farmland: Can it Yield Sustainable and Equitable Benefits,* , Washington DC.

World Bank (2018a) 'GDP per capita (current US$)'. Available at: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD (Accessed.
World Bank (2018b) 'Unemployment, total (% of total labor force) (modeled ILO estimate)'. Available at: https://data.worldbank.org/indicator/sl.uem.totl.zs (Accessed.
WWF (2018) *SOY AND ITS USES*: World Wildlife Fund. Available at: http://wwf.panda.org/what_we_do/footprint/agriculture/soy/soyreport/soy_and_its_uses/ (Accessed: 21.04 2018).
You, L., Wood, S., Wood-Sichra, U. and Wu, W. (2014) 'Generating global crop distribution maps: From census to grid', *Agricultural Systems,* 127, pp. 53-60.

Appendix A: Econometric results and discussion

A.1 Hypothesis 1

Estimation results for Hypothesis 1 with quadratic variables from the original model and main model. Variables are in log-log form and ln(totagarea) is dependent variable.

	Original model			Main model		
	POLS_ord	POLS_r	POLS_lr	FD_ord	FD_r	FD_lr
D.lnSCa	**0.510***** (0.00731)	**0.510***** (0.00805)	**0.506***** (0.00614)	**-1.332***** (0.150)	**-1.332***** (0.277)	**-1.397***** (0.273)
D.lnSCa2				**0.0516***** (0.00459)	**0.0516***** (0.00868)	**0.0534***** (0.00857)
D.lnNSCa	**0.188***** (0.0104)	**0.188***** (0.00950)	**0.190***** (0.00857)	**0.363** (0.193)	**0.363** (0.282)	**0.201** (0.258)
D.lnNSCa2				**-0.00739** (0.00617)	**-0.00739** (0.00890)	**-0.00245** (0.00813)
D.lnFuela	**0.128***** (0.00974)	**0.128***** (0.00800)	**0.121***** (0.00713)	**-0.0475** (0.127)	**-0.0475** (0.121)	**-0.00471** (0.113)
D.lnFuela2				**0.00511** (0.00434)	**0.00511** (0.00410)	**0.00357** (0.00383)
D.lnNFa	**0.0199***** (0.00106)	**0.0199***** (0.00109)	**0.0207***** (0.000880)	**-0.000605** (0.00223)	**-0.000605** (0.00170)	**-0.000372** (0.00201)
D.lnNFa2				**-0.277** (0.214)	**-0.277** (0.275)	**-0.258** (0.257)
D.lnFeeda	**0.263***** (0.00649)	**0.263***** (0.00831)	**0.266***** (0.00643)	**0.0191**** (0.00663)	**0.0191*** (0.00839)	**0.0187*** (0.00789)
D.lnFeeda2				**0.000534*** (0.000226)	**0.000534**** (0.000180)	**0.000557**** (0.000186)
D.lnforest	**0.0267***** (0.00301)	**0.0267***** (0.00276)	**0.0274***** (0.00237)	**0.808*** (0.365)	**0.808*** (0.404)	**1.000**** (0.369)
D.lnforest2				**-0.0393*** (0.0178)	**-0.0393*** (0.0197)	**-0.0561**** (0.0179)
D.lnpop	**-0.120***** (0.0118)	**-0.120***** (0.0109)	**-0.118***** (0.00915)	**0.0421** (0.0668)	**0.0421** (0.0613)	**-0.0787** (0.0674)
D.lnUnem	**-0.0257*** (0.0115)	**-0.0257** (0.0131)	**-0.0321**** (0.0108)	**0.00230** (0.00384)	**0.00230** (0.00437)	**0.00273** (0.00421)
D.lngdp	**-0.00326** (0.00318)	**-0.00326** (0.00292)	**-0.00335** (0.00278)	**0.00520** (0.00319)	**0.00520** (0.00358)	**0.00456** (0.00373)
L.u			**0.814***** (0.0776)			**-0.0869** (0.0886)
_cons	**1.733***** (0.122)	**1.733***** (0.131)	**1.814***** (0.106)	**0.00111** (0.000859)	**0.00111** (0.000706)	**0.00221**** (0.000735)
No. of observations	247	247	237	237	237	227
R^2	0.998	0.998	0.998	0.978	0.978	0.981
adj. R-sq	0.997	0.997	0.998	0.977	0.977	0.980
rmse	0.0414	0.0414	0.0347	0.00787	0.00787	0.00732

Notes: * p<0.05, ** p<0.01, *** p<0.001; standard errors in parenthesis

Staple crop area ($D.lnSCa$), feed crop area ($D.lnFeeda$) and total forest area ($D.forest$) are all highly significant in linear and quadratic order. Feed crop area is positively correlated, as expected. However, staple crop area in linear condition is negative. It means that an increase in staple crop area decreases total agricultural area. This is unexpected, bit might be true due to area noise (Chapter 4.3) or changes in production preferences, i.e., feed crop production increases more than staple crop production decreases. Hence, the total production increases. According to Table 5.2, when staple crop areas increases by one %, total agricultural area decreases by -1.33 % until it reaches a certain minimum point where it starts increasing with 0.05 %. If feed crop increases by one %, total agricultural area increases by 0.02 %. The results do not correspond to our hypothesis, which stated that changes in total agricultural area is mainly driven by non-food crops and not staple crops. There are several possible reasons for these results. The main reason might be that there are external effects affecting total agricultural area over time which might have been picked up by other variables.

To correct for potential endogeneity, the same regression was tested using price variables, however because of high collinearity they were excluded. Additionally, when it comes to the other area variables ($D.lnNSCa,\ D.lnFuela,\ D.lnNFa$), there are reasons to believe they are not significant because explanatory power might be captured by the other variables due to high correlation between variables.

Population ($D.lnpop$) and Unemployment rate ($D.lnUnem$) are positively correlated with total agricultural area, as expected. However, they are not significant because other variables might capture their explanatory power. As previously mentioned, the VIF test indicated multicollinearity and is possibly redundant. Both population and unemployment are highly correlated (corr > 50%) with many area variables in the dataset[6].

The reason GDP per capita ($D.lngdp$) has a positive sign might be because several of the countries in the analysis are emerging economies. Hence, countries focus a larger share of their economy towards developing and expanding agricultural production through areal increase. Later, according to Prebisch-Singer hypothesis, as the economy grows, the country begins focusing on producing manufacture goods and increase productivity of the agricultural sector. However, because we do not have serial correlation nor heteroskedasticity in our analysis, and a solid functional form, there are reasons to believe the entirety of our results are conclusive

[6] See correlation matrix Appendix B.1, Table B.2.

A.1.2 Hypothesis 1: Linear form

Estimation results for Hypothesis 1 from main model, with TotalSCa(left) and TotalFuela(right) as dependent variables.

	FD_ord	FD_r	FD_lr
D.TotalNSCa	-0.285	-0.285	-0.292
	(0.236)	(0.384)	(0.385)
D.TotalFuela	0.847	0.847*	0.844*
	(0.454)	(0.332)	(0.334)
D.TotalNFa	0.497	0.497	0.497
	(1.134)	(0.340)	(0.351)
D.TotalFeeda	1.050***	1.050**	1.043**
	(0.275)	(0.365)	(0.373)
D.forest	12.11	12.11*	11.66**
	(26.26)	(5.107)	(4.456)
D.pop	0.00820	0.00820	0.00608
	(0.0300)	(0.0198)	(0.0217)
D.Unem	-124766.6	-124766.6	-90219.5
	(194210.4)	(146262.8)	(150552.5)
D.gdp	153.0	153.0	148.0
	(313.6)	(248.0)	(244.5)
L.u			0.0180
			(0.0250)
_cons	-154838.6	-154838.6	-139194.1
	(268213.4)	(263448.6)	(288021.0)
N	144	144	144
R-sq	0.187	0.187	0.197
adj. R-sq	0.139	0.139	0.143
rmse	2119468.0	2119468.0	2114741.6

Standar errorr in parenteses
* p<0.05, **p<0.01, *** p<0.001

	FD_ord	FD_r	FD_lr
D.TotalSCa	0.0297	0.0297*	0.0238
	(0.0159)	(0.0140)	(0.0122)
D.TotalNSCa	0.0984*	0.0984	0.0946
	(0.0437)	(0.0546)	(0.0588)
D.TotalNFa	0.271	0.271	0.250
	(0.211)	(0.156)	(0.158)
D.TotalFeeda	0.193***	0.193**	0.218***
	(0.0516)	(0.0594)	(0.0570)
D.forest	1.733	1.733	9.654
	(4.919)	(1.398)	(34.25)
D.pop	-0.00228	-0.00228	-0.000510
	(0.00561)	(0.00723)	(0.00793)
D.Unem	-7809.3	-7809.3	-10090.1
	(36414.1)	(22795.6)	(22728.4)
D.gdp	85.17	85.17	80.48
	(58.31)	(52.16)	(52.70)
L.u			0.0646
			(0.104)
_cons	4937.6	4937.6	9069.1
	(50281.5)	(42478.4)	(51319.2)
N	144	144	138
R-sq	0.243	0.243	0.269
adj. R-sq	0.198	0.198	0.217
rmse	396857.8	396857.8	387435.9

Standar errorr in parenteses
* p<0.05, **p<0.01, *** p<0.001

A.2 Hypothesis 2

Estimation results for Hypothesis 2.1 and 2.2, with staple crop area, fuel crop area and feed crop area as dependent variable. Quadratic form is included.

Hypothesis Dep. Variable	H2.1 SCa	H2.2 Fuela	H2.2 Feeda
Variable name	FD_r	FD_r	FD_r
D.lnSCy	**-0.683***** (0.188)	**-0.266** (0.160)	**-0.123** (0.159)
D.lnSCy2	**0.150*** (0.0701)	**0.117** (0.0696)	**0.0846** (0.0605)
D.lnNSCy	**0.118** (0.106)	**-0.107** (0.119)	**0.165** (0.114)
D.lnNSCy2	**-0.000179** (0.0271)	**0.0194** (0.0294)	**-0.00412** (0.0290)
D.lnFeedy	**0.271***** (0.0644)	**0.0535** (0.0778)	**0.144*** (0.0728)
D.lnFeedy2	**-0.0586** (0.0329)	**0.0437** (0.0378)	**-0.0923** (0.0505)
D.lnFuely	**0.0512** (0.0562)	**0.347***** (0.0918)	**0.0297** (0.0775)
D.lnFuely2	**-0.0860***** (0.0200)	**-0.157***** (0.0336)	**-0.0576*** (0.0275)
D.lnNFy	**0.00149** (0.0176)	**0.00638** (0.0134)	**0.0116** (0.0194)
D.lnNFy2	**-0.000968** (0.0212)	**0.00172** (0.0137)	**0.00925** (0.0211)
D.lnforest	**0.500** (2.670)	**3.321** (2.397)	
D.lnforest2	**-0.0226** (0.130)	**-0.160** (0.117)	
D.lnpop	**0.815** (0.538)	**-0.188** (0.721)	**0.386** (0.609)
D.lnUnem	**-0.0124** (0.0281)	**-0.0511** (0.0319)	**-0.0692*** (0.0278)
D.lngdp	**0.0202** (0.0273)	**0.0637*** (0.0302)	**0.00538** (0.0284)
_cons	**-0.00190** (0.00703)	**0.0174** (0.0101)	**0.00444** (0.00803)
N	237	237	237
R^2	0.340	0.321	0.266
adj. R-sq	0.295	0.275	0.224
rmse	0.0586	0.0662	0.0593

Notes: * p<0.05, ** p<0.01, *** p<0.001; p-values in parenthesis

For hypothesis 2.1, the DF_ord and DF_r generated similar results with no serial correlation, and strong evidence of homoskedasticity and linearity. The DF_lr provided a better fit of the data than the latter, i.e. higher R^2 and smaller rmse, however, the regression suffers from serial correlation and is therefore excluded. I also tested for individual effects using an AR(1) test and added a dummy variable for each year. The F-test generated p-value > 0.05, hence, years from 1993 to 2016 does not have significant impact on the dependent variable, and we exclude years as dummy variables from further analysis. Comparing the results from DF_ord and DF_r, variables are equally significant, however, DF_r has a stronger functional form, which indicates a better model. Hence, further discussions of Hypothesis 2.1 will focus on results generated by FD_r.

According to table 5.3, when staple crop yield increases with one %, staple crop area decreases by 0.68 %, until a certain minimum point where total agricultural area cannot decrease anymore due to higher yield. From the minimum point, an increase of one % in staple crop yield increases staple crop area with 0.15%. In relation to theory (Chapter 3), the Borlaug hypothesis vs. the Jevons paradox discussion, these results are in line with what Norman Borlaug argued.

For Hypothesis 2.2, The VIF test showed a mean multicollinearity below the recommended limit, however, looking at the variables independently, there were high multicollinearity for $D.lnforest$ and $D.lnforest2$ (VIF > 10), which can explain some of our results. Because forest area was a control variable, we excluded the variable to improve our model. However, the functional form became weaker when removing forest area, and we decided to keep forest area nonetheless. Moreover, the model suffered from heteroskedasticity and serial correlation. To fix this, reducing multicollinearity and removing the remaining control variables (unemployment, GDP and population), was tested, but without success. Removing more control variables made our model weaker, which means they have some explanatory power for changes in staple crops area. Hence, all explanatory variables were kept, and results generated by DF_r was used in our analysis. Though, keep in mind that the model suffers from heteroskedasticity and serial correlation.

In the second regression, using $D.lnFeeda$ as dependent variable, none of the models generated a strong functional form, and rejected tests of heteroskedasticity or serial correlation. The VIF test showed a mean multicollinearity below the recommended limit, however, $D.lnforest$ and $D.lnforest2$ presented high values multicollinearity (VIF > 10). The variable was omitted, and after running the Ramsey RESET test for misspecification, I could confirm a stronger functional form for FD_r. However, removing forest area did not remove issues heteroskedasticity or serial correlation, and removing additional control variables made our model weaker. Hence, only forest area was omitted, and results generated by the DF_r model was used for our analysis. Further, the adjusted R^2 is below 0.30, which means that the explanatory variables only explains up to 30% of changes in the dependent variable. Hence the covariates do a poor job explaining and/ or predicting the response values. This might be because of size of the dataset

For H2.2, were fuel crop area ($D.lnFuela$) is the dependent variable, fuel crop yield ($D.lnFuely$) is highly positive significant in linear and highly negative significant in quadratic. We hypothesize that higher fuel crop yield, increases fuel crop area and ran the regression with the same control variables as in hypothesis two. These results correspond with findings in (Angelsen and Kaimowitz, 2001) and are in line with the Jevons paradox. In contrast to our expectations, GDP per capita ($D.lngdp$) is weakly positive significant (p-value < 0.05) with fuel crop area. An increase in GDP per capita with one % increases fuel crop area with 0.06%. We expected that increase in GDP per capita led to more advanced agriculture, higher yield and reduced agricultural area, in accordance with the Borlaug hypothesis. However, from our results we observe the opposite, i.e. that increase in GDP per capita results in increased fuel crop yield, and in accordance with Jevons paradox, increase in fuel crop area. Another possible interpretation is that, as fuel crops also provides food and feed, an increase in GDP per capita corresponds with increased consumption of food, feed and fuel, which results in increased agricultural area. Further, fuel crop yield in linear and quadratic form have low correlation with control variables, potentially explaining why none of the other variables are significant. There are also some unexpected signs in the regression analysis, particularly forest area and population. According to the regression, an increase in forest area increases fuel crop area, while an increase in population decreases fuel crop area. This can be explained the low correlation between forest, population and fuel crop area[7]. Further, it is important to remember that our analysis suffers from heteroskedasticity and serial correlation. The adjusted R^2 is 0.275 and rmse is 0.0662, which tells us that our explanatory variables only explain 27,5 % of variances in our model, with low model fit.

For H2.2 where feed crop area ($D.lnFeeda$) is the dependent variable, I hypothesized that an increase in feed crop yield, increases feed crop area. To improve the model's functional form,

[7] Correlation matrix for hypothesis 2 in Appendix C B.1 (Table B.1.2),

forest area was omitted due to high VIF value. Consequently, the model fit improved and I observed weak positive correlation of feed crop yield ($D.lnFeedy$) in the quadratic, and weakly negative correlation for fuel crop yield ($D.lnFuely$) and unemployment rate ($D.lnUnem$). Table 5.2 indicates that, one % increase in feed crop yield, increases feed crop area with 0,05 % until it reaches a certain maximum point, where eventually, increase in feed crop yield, reduces feed crop area at a certain rate. These results correspond with our findings in the latter hypothesis, namely the Jevons paradox. The second significant variable, fuel crop yield in quadratic, present similar results as changes in feed crop yield, i.e. feed crop area increases in line with fuel crop area, until reaches a certain limit, when one % increase in fuel crop area generate a 0.05 % decrease in feed crop area. From the last significant variable, unemployment rate, suggests that feed crop area decreases with 0.07 % if unemployment increases with one %. This might be the result because, as agriculture is a labor-intensive sector, increase in unemployment means that the sector is becoming more productive. And as we can observe from the results, increase in feed crop yield also increases feed crop area. Once again, we must keep in mind the issues of heteroskedasticy, serial correlation and multicollinearity when evaluating the results. The adjusted R^2 is 0.266 and rmse is 0.0593, which tells us that our explanatory variables only explain 26.6 % of variances in our model, with low model fit.

Lastly, a third regression with $D.lnNFa$ as dependent variable was tested, however, after repeating the processes of the latter hypotheses, the functional form still indicated a weak model. All the variables were insignificant. Reasons might be that because the non-food crop group is the smallest of all the crop groups, hence the explanatory power might be absorbed by other variables. We will therefore not include it further in our analysis.

A.3 Hypothesis 3

Table A.3.1 Estimation results for Hypothesis 3 from the main model with ln(forest) as dependent variable. Quadratic form is included.

Variable name	FD_ord	FD_r	FD_lr
D.lnSCa	0.572	1.681	0.572
	(0.460)	(1.478)	(0.460)
D.lnSCa2	-0.0182	-0.0424	-0.0182
	(0.0145)	(0.0376)	(0.0145)
D.lnNSCa	-0.187	-2.254	-0.187
	(0.306)	(2.132)	(0.306)
D.lnNSCa2	0.00602	0.0725	0.00602
	(0.00994)	(0.0686)	(0.00994)
D.lnFuela	-0.105	0.751	-0.105
	(0.189)	(1.235)	(0.189)
D.lnFuela2	0.00379	-0.0245	0.00379
	(0.00642)	(0.0403)	(0.00642)
D.lnNFa	0.00153	-0.0102	0.00153
	(0.00179)	(0.0147)	(0.00179)
D.lnNFa2	-0.000445	0.00126	-0.000445
	(0.000265)	(0.00203)	(0.000265)
D.lnFeeda	-0.100	-6.392	-0.100
	(0.380)	(6.196)	(0.380)
D.lnFeeda2	0.00351	0.184	0.00351
	(0.0115)	(0.178)	(0.0115)
D.lnUnem	0.00145	0.0511	0.00145
	(0.0109)	(0.0577)	(0.0109)
D.lngdp	-0.000723	0.0635	-0.000723
	(0.00760)	(0.0664)	(0.00760)
L3.u	0.00426		0.00426
	(0.00707)		(0.00707)
_cons	-0.00418***	-0.0162	-0.00418***
	(0.00111)	(0.0121)	(0.00111)
N	207	237	207
R-sq	0.047	0.052	0.047
adj. R-sq	-0.018	0.001	-0.018
rmse	0.0139	0.148	0.0139

Notes: * p<0.05, ** p<0.01, *** p<0.001; standard errors in parentecis

Table A.3.2 Estimation results for Hypothesis 3 from the main model with forest as dependent variable. Ten countries included in regression.

	ord	r	lr3
D.lnSCa	-0.0240	0.335	-0.0240
	(0.0158)	(0.358)	(0.0158)
D.lnNSCa	-0.000869	0.0233	-0.000869
	(0.0110)	(0.0451)	(0.0110)
D.lnFuela	0.00511	0.0227	0.00511
	(0.0110)	(0.0603)	(0.0110)
D.lnNFa	-0.00217	0.000434	-0.00217
	(0.00147)	(0.00633)	(0.00147)
D.lnFeeda	0.00369	-0.443	0.00369
	(0.0145)	(0.438)	(0.0145)
D.lnUnem	-0.00216	0.0639	-0.00216
	(0.00794)	(0.0670)	(0.00794)
D.lngdp	-0.00401	0.0731	-0.00401
	(0.00713)	(0.0757)	(0.00713)
L3.u	0.00849		0.00849
	(0.00918)		(0.00918)
_cons	-0.00366***	-0.0151	-0.00366***
	(0.00102)	(0.0112)	(0.00102)
N	227	237	227
R-sq	0.022	0.040	0.022
adj. R-sq	-0.014	0.011	-0.014
rmse	0.0136	0.148	0.0136

Standard errors in parentheses
* p<0.05, **p<0.01, *** p<0.001

Table A.3.1 presents the results of the first difference regression for Hypothesis 3, with log of forest area ($D.lnFeedy$) as dependent variable. Using the same control variables as the previous hypothesis, the VIF test of multicollinearity showed a small mean correlation between the independent variables. However, population ($D.lnpop$) was omitted to improve model fit as the VIF value was above recommended limit. In Hypothesis 3, we aimed to explore if expansion of any of the main crop categories lead to forest loss. In Table 5.4 none of the variables are of statistical significance. Reasons are that the model has strong evidence against linearity and evidence of heteroskedasticity. When interpreting R^2 (< 0.06) and mrse (>0.14), it becomes clear that this is a weak model we cannot make any conclusions from. It was attempted to omit and include other variables, as well as time lags, however the model did not improve. Correspondingly, the actual signs of variables deviated from our expectations. There appeared to be negative impacts of increases in non-staple food crop area, non-food crop area and feed crop area on forest, as expected, while staple crop area and fuel crop area have a positive correlation.

Appendix B: STATA results

B.1 Correlation matrix

Table B.1. 1: Correlation matrix for variables in a linear form. Total agricultural area as dependent variable.

	totagarea	TotalSCa	TotalNSCa	TotalFeeda	TotalFuela	TotalNFa	TotalSCy	TotalNSCy	TotalFeedy	TotalFuely	TotalNFy	forest	Unem	gdp	pop	Year
totagarea	1.0000															
TotalSCa	0.9029	1.0000														
TotalNSCa	0.8106	0.7604	1.0000													
TotalFeeda	0.6791	0.3676	0.3035	1.0000												
TotalFuela	0.8003	0.6090	0.5055	0.6325	1.0000											
TotalNFa	0.4708	0.3282	0.3407	0.3702	0.5184	1.0000										
TotalSCy	-0.0554	-0.2377	-0.0282	-0.0314	0.3897	0.4714	1.0000									
TotalNSCy	-0.1132	-0.2633	-0.2903	0.1596	0.2398	-0.0525	0.5371	1.0000								
TotalFeedy	0.1147	-0.0603	-0.2173	0.5315	0.1333	0.5684	0.1699	0.3103	1.0000							
TotalFuely	0.0461	-0.1123	0.1399	0.0543	0.2189	0.4886	0.5842	0.0472	0.1454	1.0000						
TotalNFy	0.1431	-0.1018	0.0887	0.4174	0.2412	0.0279	0.2086	0.5896	0.3149	0.0872	1.0000					
forest	0.0372	-0.1247	-0.0697	0.3580	0.0472	-0.1134	-0.0929	-0.0274	-0.0449	0.0430	0.0015	1.0000				
Unem	-0.6656	-0.6558	-0.5376	-0.3763	-0.4942	-0.4743	-0.0450	0.1092	-0.1549	0.0593	0.0628	0.0636	1.0000			
gdp	-0.0814	-0.2621	-0.3011	0.4159	-0.0365	0.0298	0.0874	0.6914	0.7125	-0.1067	0.6223	0.0396	0.0406	1.0000		
pop	0.8830	0.7599	0.7772	0.4894	0.9070	0.5216	0.2896	0.0513	-0.0377	0.1490	0.1834	-0.0712	-0.5970	-0.1873	1.0000	
Year	0.0737	0.0391	0.0406	0.1024	0.0693	0.0486	0.2018	0.1337	0.1630	0.1868	0.1107	-0.0044	-0.1589	0.2281	0.0626	1.0000

Table B.1. 2: Correlation matrix with first differences for Hypothesis 1. Log of total agricultural area as dependent variable.

	lnAgArea	lnSCa	lnSCa2	lnNSCa	lnNSCa2	lnFuela	lnFuela2	lnNFa	lnFeeda	lnFeeda2	lnNFa2	lnforest	lnfore~2	lnpop	lnUnem	lngdp
lnAgArea	1.0000															
lnSCa	0.9421	1.0000														
lnSCa2	0.9387	0.9997	1.0000													
lnNSCa	0.8873	0.8688	0.8675	1.0000												
lnNSCa2	0.8812	0.8640	0.8633	0.9994	1.0000											
lnFuela	0.8305	0.6848	0.6828	0.7447	0.7422	1.0000										
lnFuela2	0.8276	0.6832	0.6815	0.7418	0.7395	0.9996	1.0000									
lnNFa	0.5502	0.4810	0.4801	0.4666	0.4649	0.6161	0.6134	1.0000								
lnFeeda	0.8006	0.5915	0.5845	0.5686	0.5620	0.7323	0.7316	0.3281	1.0000							
lnFeeda2	0.7991	0.5886	0.5816	0.5648	0.5582	0.7371	0.7364	0.3360	0.9998	1.0000						
lnNFa2	0.6318	0.5607	0.5578	0.5536	0.5498	0.6994	0.6950	0.9769	0.3825	0.3908	1.0000					
lnforest	0.4506	0.2533	0.2380	0.3587	0.3447	0.5345	0.5272	0.1542	0.6267	0.6299	0.2581	1.0000				
lnforest2	0.4366	0.2416	0.2262	0.3431	0.3284	0.5104	0.5030	0.1161	0.6214	0.6242	0.2195	0.9986	1.0000			
lnpop	0.8889	0.8381	0.8362	0.8899	0.8882	0.9073	0.9083	0.6113	0.6262	0.6294	0.7085	0.4922	0.4677	1.0000		
lnUnem	-0.6905	-0.7349	-0.7369	-0.6815	-0.6845	-0.5943	-0.5964	-0.3938	-0.4030	-0.4058	-0.4820	-0.1167	-0.0964	-0.7363	1.0000	
lngdp	-0.2269	-0.3899	-0.3913	-0.4758	-0.4743	0.0384	0.0366	-0.1381	0.1962	0.2081	-0.1370	0.2308	0.2344	-0.2347	0.2372	1.0000

Table B.1. 3: Correlation matrix with first differences for Hypothesis 2. Log of staple crops area as dependent variable.

	lnSCa	lnFeeda	lnFuela	lnNFa	lnSCy	lnSCy2	lnNSCy	lnNSCy2	lnFeedy	lnFeedy2	lnFuely	lnFuely2	lnNFy	lnNFy2	lnforest	lnfore~2	lnpop	lnUnem	lngdp
lnSCa	1.0000																		
lnFeeda	0.5915	1.0000																	
lnFuela	0.6848	0.7323	1.0000																
lnNFa	0.4810	0.3281	0.6161	1.0000															
lnSCy	-0.3219	-0.0866	0.3197	0.1319	1.0000														
lnSCy2	-0.2548	-0.1888	0.3239	0.2062	0.9541	1.0000													
lnNSCy	-0.4014	0.1149	0.1976	-0.0908	0.6853	0.6227	1.0000												
lnNSCy2	-0.3334	0.0496	0.1769	-0.0603	0.5765	0.5639	0.9634	1.0000											
lnFeedy	0.0239	0.1220	0.2680	0.5239	0.2190	0.2818	0.3938	0.4212	1.0000										
lnFeedy2	0.0952	0.2897	0.2583	0.3804	0.1306	0.1065	0.3165	0.2926	0.8535	1.0000									
lnFuely	0.0375	0.0943	0.5119	0.6856	0.4858	0.5279	0.1933	0.1797	0.4130	0.1492	1.0000								
lnFuely2	-0.0756	0.0092	0.3878	0.5420	0.5761	0.6118	0.1476	0.0756	0.2222	0.0442	0.8819	1.0000							
lnNFy	-0.1860	0.3587	0.3352	0.0690	0.4289	0.2916	0.6360	0.5897	0.2309	0.2484	0.3583	0.1865	1.0000						
lnNFy2	0.0851	0.0661	-0.0251	-0.1647	-0.2391	-0.2345	0.1303	0.2206	0.1770	0.1587	-0.0872	-0.2784	0.0509	1.0000					
lnforest	0.2533	0.6267	0.5345	0.1542	0.2277	0.1397	0.2427	0.1133	0.1254	0.2725	0.2598	0.3610	0.2466	-0.0337	1.0000				
lnforest2	0.2416	0.6214	0.5104	0.1161	0.2103	0.1216	0.2426	0.1136	0.1069	0.2604	0.2233	0.3359	0.2285	-0.0252	0.9986	1.0000			
lnpop	0.8381	0.6262	0.9073	0.6113	0.2061	0.2196	-0.0466	-0.0575	0.1175	0.1793	0.3652	0.3009	0.1095	-0.0743	0.4922	0.4677	1.0000		
lnUnem	-0.7349	-0.4030	-0.5943	-0.3938	-0.0269	-0.0553	0.1509	0.1214	-0.0848	-0.1551	-0.0899	0.0276	0.0230	-0.0742	-0.1167	-0.0964	-0.7363	1.0000	
lngdp	-0.3899	0.1962	0.0384	-0.1381	0.2902	0.2078	0.7659	0.7413	0.5475	0.5173	0.1356	-0.0067	0.6238	0.3352	0.2308	0.2344	-0.2347	0.2372	1.0000

B.2 POLS

Table B2.1: H2.1 – Staple crop

Tests	*H0:*	
White's test	Homoscedasticity exist	*Reject*
	Chi2 =	14.02
	Prob > chi2 =	0.000
Wooldridge test	No first-order autocorrelation	*Reject*
	F (1, 9)	33.912
	Prob > F =	0.0003
Ramsey RESET test	Linearity exist	*Reject*
	Chi2 =	13.09
	Prob > chi2 =	0.0000

Table B2.1: H2.1 – Non-staple food crop

Tests	*H0:*	
White's test	Homoscedasticity exist	*Reject*
	Chi2 =	14.02
	Prob > chi2 =	0.000
Wooldridge test	No first-order autocorrelation	*Reject*
	F (1, 9)	33.912
	Prob > F =	0.0003
Ramsey RESET test	Linearity exist	*Reject*
	Chi2 =	13.09
	Prob > chi2 =	0.0000

Table B2.2: H2.2 – Fuel crop

Tests	*H0:*	
White's test	Homoscedasticity exist	*Reject*
	Chi2 =	173.17
	Prob > chi2 =	0.000
Wooldridge test	No first-order autocorrelation	*Reject*
	F (1, 9)	101.509
	Prob > F =	0.0003
Ramsey RESET test	Linearity exist	*Reject*
	Chi2 =	37.20
	Prob > chi2 =	0.0000

Table B2.3: H2.2 – Feed crop

Tests	*H0:*	
White's test	Homoscedasticity exist	*Reject*
	Chi2 =	216.80
	Prob > chi2 =	0.000
Wooldridge test	No first-order autocorrelation	*Reject*
	F (1, 9)	33.912
	Prob > F =	0.0003
Ramsey RESET test	Linearity exist	*Reject*
	Chi2 =	73.13
	Prob > chi2 =	0.0000

Table B2.4: H2.2 – Non-fuel crop

Tests	*H0:*	
White's test	Homoscedasticity exist	*Reject*
	Chi2 =	8.70
	Prob > chi2 =	0.000
Wooldridge test	No first-order autocorrelation	*Reject*
	F (1, 9)	33.912
	Prob > F =	0.0003
Ramsey RESET test	Linearity exist	*Reject*
	Chi2 =	285.32
	Prob > chi2 =	0.0000

B.3 Ramsey RESET test: First-Difference

Table B3.1: H2.1 – staple crop

Ramsey Reset test using power of the fitted values of D.lnSCa	
H0: linearity exists	
F(3, 223)=	3.15
Prob > F=	0.0259

Table B3.1: H2.1 – Non-staple food crop

Ramsey Reset test using power of the fitted values of D.lnNSCa	
H0: linearity exists	
F(3, 224)=	1.78
Prob > F=	0.1526

Table B3.2: H2.2 – Fuel crop

Ramsey Reset test using power of the fitted values of D.lnFuela	
H0: linearity exists	
F(3, 223)=	1.92
Prob > F=	0,1272

Table B3.3: H2.2 – non-food crop

Ramsey Reset test using power of the fitted values of D.lnNFa	
H0: linearity exists	
F(3, 223)=	9.62
Prob > F=	0.0000

Table B3.4: H2.2 – Feed crop

Ramsey Reset test using power of the fitted values of D.lnFeeda	
H0: linearity exists	
F(3, 218)=	3.90
Prob > F=	0.0096

B.4 VIF

Table B4.1: VIF test for multicollinearity of independent variables for Hypothesis 2.

Variable	VIF	1/VIF
lnFeedy D1,	1,77	0,565637
lnSCy D1,	1,54	0,648289
lnFuely D1,	1,48	0,677412
lnNSCy D1,	1,27	0,788941
lngdp D1,	1,16	0,861315
lnUnem D1,	1,11	0,90075
lnNFy D1,	1,06	0,946086
lnforest D1,	1,04	0,965641
lnpop D1,	1,02	0,981713
Mean VIF	1,27	

Appendix C: Excel

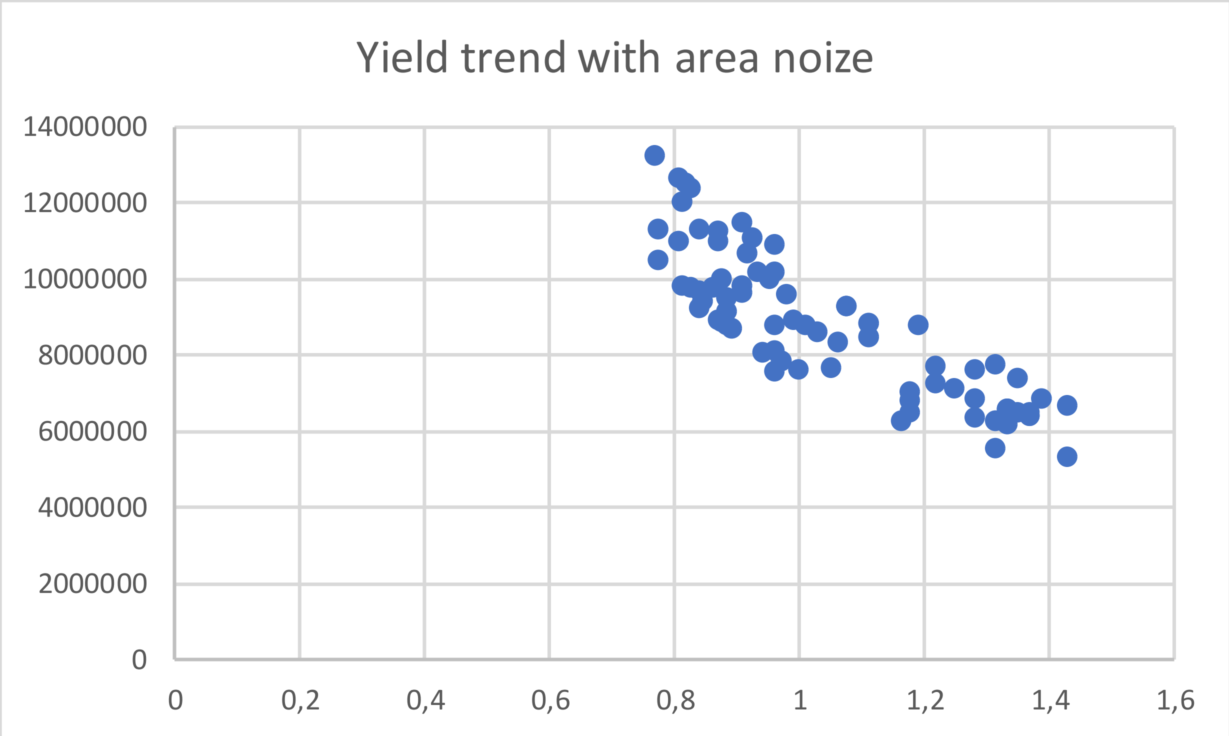


Figure C.1.1: Trends in yield due to area noise. Example from fuel crops yield in Brazil.

Appendix D: Empirical evidence

Four major global cropping systems are found to be relevant for crop trend analysis: 1) M3, 2) MIRCA2000, 3) SPAM, and 4) GAEZ.

1) *M3 dataset (Monfreda, Ramankutty, and Foley's 2008 Cropping System Model):*
   The M3 dataset provides the most complete coverage of crops (175 crops) for both harvested area and yield (Monfreda et al., 2008). It uses remote sensing products to construct a dataset for cropland and pasture circa 2000 and apply minimal modeling from subnational statistics of yield and harvested area to ease interpretation and limit complex assumptions.

2) *"Dataset of Monthly Irrigated and Rainfed Crop Areas (MIRCA) around Year 2000" (Portmann, Siebert and Döll 2010):*
   This dataset uses M3 as a starting point and relies on its input data on the spatial allocation of the total area and average yield (Portmann et al., 2010). MIRCA is mainly derived from census data and crop calendars from literature and aims to maximize consistency with subnational statistics collected by national institutions and by the FAO. (Siebert et al., 2010). Further, it downscales 26 crops and two aggregate categories of "other annual" and other "perennial" crops, which are divided into rainfed and irrigated production areas (Anderson et al., 2014). Siebert and co-authors (2010) explained that "Data derived by remote sensing were not used to produce the MIRCA2000 inventory, but to develop the datasets on global cropland and area equipped for irrigation" (Siebert et al., 2010). MIRCA therefore presents unique sets of information on irrigation, crop-specific irrigated water use, crop calendar and copping intensities. Hence, the largest differences between the M3 and MIRCA is reflected in the different downscaling methodologies and the different subnational data collections (Portmann et al., 2010).

3) *"Spatial Production Allocation Model (SPAM)" (You, Wood, Wood-Sichra and Wu 2014):*
   Relies on a different collection of subnational statistical data than that of MIRCA, focusing on covering developing countries. SPAM covers the fewest crops, just 20, but downscales the area and yield for each crop into high-input irrigated, high-input rainfed, and low-input rainfed production systems (You et al., 2014). Additionally, in different

from previously mentioned datasets, SPAM also relies on additional variables such as crop prices, population density and specific biophysical suitability to distribute subnational statistics (Anderson et al., 2014).

4) *"Global Agroecological Zones Cropping System Model (GAEZ)" (Fisher et. a., 2012):* Develops a different cropland analysis than the production systems mentioned above, which are all based on Ramankutty and co-authors (2008), and relies instead on an "extensive analysis of crop-specific agroclimatic and edaphic suitability" (Anderson et al., 2014). GAEZ include additional variables such as population density, biophysical suitability and market access as means of distributing subnational statistics (Fischer et al., 2012)

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway