Palacky University Olomouc
University of Clermont Auvergne
University of Pavia

MASTER THESIS

Dibekulu Mulu Birhan

Supervisor: Professor Maria Sassi

GLODEP 2021

The Cost of Misclassification of Seeds: Evidence from Improved Maize Variety Adopters in Ethiopia

Dibekulu Mulu Birhan

Supervisor: Professor Maria Sassi

Declaration

I declare that this thesis entitled "The Cost of Misclassification of Seeds: Evidence from Improved Maize Variety Adopters in Ethiopia" is my original work representing my ideas in my own words submitted only to the GLODEP Consortium 2021. Others' ideas and words included are adequately cited and referenced. I also declare that I have adhered to all academic honesty and integrity principles to not misrepresent or fabricate or falsify any idea/data/fact/source in my thesis. I understand that any violation of the above is subject to disciplinary action by the consortium. It can also evoke penal action from the sources that are not properly cited or from whom proper permission is not taken when needed.

Dibekulu Mulu Birhan

Date: 31 May 2021

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2020/2021

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Dibekulu Mulu BIRHAN**
Osobní číslo:	**R190705**
Studijní program:	**N1301 Geography**
Studijní obor:	**International Development Studies**
Téma práce:	**Improved seed adoption: misclassification and suboptimal decision**
Zadávající katedra:	**Přírodovědecká fakulta**

Zásady pro vypracování

Growth in crop production comes from three sources: expanding the crop field, increasing the cropping frequency, and improving the crop variety (Bruinsma, 2003, p.125). Given the first two options approaching their limits, it seems that the hope we have to boost farm production is on improved seed adoption. This is vital for developing countries particularly of sub-Saharan Africa where farming is a backbone of the economy and a major source of access to food.

In the literature, improved seed adoption in LDC’s agriculture is well discussed however most using self-declared data from households. Recent breakthrough to objectively measure the adoption status, DNA fingerprinting identification, shows substantial gap from the self-reported which makes one question the validity of previous empirical analyses in the sector which relied on self-declared input (Wineman et al, 2020; Floro et al, 2018).

Experiencing considerable gap, misclassification, is not surprising given the poor literacy, poor seed certification system, and dominant seed recycling practice by the farming community (Kosmowski et al, 2020; Wossen et al, 2019). Wrong perception and reporting of the type of seed planted i.e. improved or local version, can have a direct potential adverse impact on crop productivity through influencing households’ allocation of complementary inputs that go together with the seed like fertilizer and labour. Hence, this thesis is aimed to analyze misclassification and its potential repercussion, suboptimal input allocation, which in turn pays off households with suboptimal output using the latest wave of (2018/19) Ethiopian socioeconomic survey data from the World Bank’s LSMS which for the first time included sample barcodes of some crop varieties.

Rozsah pracovní zprávy:	**20-25 tisíc slov**
Rozsah grafických prací:	**dle potřeby**
Forma zpracování diplomové práce:	**tištěná**
Jazyk zpracování:	**Angličtina**

Seznam doporučené literatury:

1. Bruinsma, Jelle (ed.) (2003). *World Agriculture: Towards 2015/30, an FAO Perspective*, London: Earthscan and Rome: FAO.
2. Floro, Victorino O., et al. (2018). *Household Determinants of the Adoption of Improved Cassava Varieties Using DNA Fingerprinting to Identify Varieties in Farmer Fields: A Case Study in Colombia.* Journal of Agricultural Economics, vol. 69, no. 2, pp. 518–36. DOI.org (Crossref), doi:10.1111/1477-9552.12247.
3. Frederic, Kosmowski, et al. (2020). *Shining a Brighter Light: Comprehensive Evidence on Adoption and Diffusion of CGIAR-related Innovations in Ethiopia*. SPIA synthesis report.
4. Poets, Ana et al. (2020). *DNA Fingerprinting for Crop Varietal Identification: Fit-for-Purpose Protocols, their Costs and Analytical Implications.* SPIA and GEMS report.
5. Wineman, Ayala, et al. (2020). *A Case of Mistaken Identity? Measuring Rates of Improved Seed Adoption in Tanzania Using DNA Fingerprinting.* Journal of Agricultural Economics, vol. 71, no. 3, pp. 719–41. DOI.org (Crossref), doi:10.1111/1477-9552.12368.
6. Wossen, Tesfamicheal, et al. (2019). *Agricultural Technology Adoption and Household Welfare: Measurement and Evidence.* Food Policy, vol. 87, p. 101742. DOI.org (Crossref), doi:10.1016/j.foodpol.2019.101742.

Vedoucí diplomové práce:	**prof. Maria Sassi** University of Pavia
Datum zadání diplomové práce:	**29. ledna 2021**
Termín odevzdání diplomové práce:	**31. května 2021**

Acknowledgement

Various people immensely contributed to the thesis. First and foremost, I am indebted to my family for their unconditional support and love. Then, I would like to thank my supervisor, Professor Maria Sassi, who has always been happily available, to guide me throughout the work. Without her constructive comments, suggestions, and creative insights, the completion of this thesis would not be possible. I also want to extend my heartfelt appreciation and respectful gratitude to the GLODEP administrative staff in general who have been very helpful since our first day of arrival. Your support to ease complications to our academic life, given our study converged with the global COVID pandemic outbreak, has been priceless. Last but not least, my amazing GLODEP friends know that you will always be in my heart.

Abstract

Improved seed adopters in developing countries are likely to misclassify the adopted variety as a landrace due to challenges such as dominant informal seed market, poor seed certification system, and widespread seed recycling practice that surround the agriculture sector. Improved seeds need differential treatment in terms of the supply of farm inputs for optimal farm production. Misclassification error could lead to suboptimal allocation of the inputs and thus cause a loss of potential yield. The purpose of this study is to estimate the loss of the yield attributable to the misclassification measurement error. Employing a combination of DNA fingerprinting and self-reporting cross-sectional plot-level data, the study analyzed the yield loss in the 2018/19 cropping season by Ethiopian farmers who adopted improved maize varieties using the propensity score matching technique. The seed adopted on 61% of the plots is found misclassified (false-negative adopters). The misclassification borne yield loss found is considerable. The average maize yield from false-negative plots is less in the range of 609 to 776 kgs per hectare than from true-positive counterfactuals. Although this study sheds light on the causal impact of misclassification on yield, given it is the first empirical investigation on the topic, further studies are needed to verify the robustness and generalizability of the findings.

Keywords: Misclassification, Improved seed, Propensity score matching, DNA fingerprinting, Self-reporting, False-negative, True-positive

Contents

List of Tables

List of Figures

Acronyms

AGRA	Alliance for a Green Revolution in Africa
ASE	Amhara Seed Enterprise
CGIAR	The Consultative Group on International Agricultural Research
CIA	Conditional Independence Assumption
CIMMYT	The International Maize and Wheat Improvement Center
DarTseq	Diversity Arrays Technology
DNA	Deoxyribonucleic Acid
EA	Enumeration Area
EIAR	The Ethiopian Institute of Agricultural Research
ESE	Ethiopian Seed Enterprise
ESS4	Ethiopian Socioeconomic Survey Wave 4
FAO	The Food and Agricultural Organization of the United Nations
GDP	Gross Domestic Product
GMOs	Genetically Modified Seeds
GNP	Gross National Product
GPS	Global Positioning System
IFPRI	The International Food Policy Research Institute
ILRI	The International Livestock Research Institute
LSMS	The Living Standards Measurement Study
NGO	Nongovernmental Organizations
OPV	Open Pollinated Variety
OSE	Oromia Seed Enterprise
SSE	Southern Seed Enterprise
DTMA	Drought Tolerant Maize for Africa
MT	Metric Ton
WTO	The World Trade Organization

Chapter 1 - Introduction

1.1 Background

Adopting improved seed, expanding cropping area, and increasing cropping intensity are the sources that growth in crop production can be derived (Bruinsma, 2017). The arable land expansion has been the main source of boosting crop production in Ethiopia (Seyoum Taffesse et al., 2013). However, land holdings in the country are now fragmented into small parcels, and it is likely for the average farm size of less than one hectare to decline further (FAO, 2018). Bacha et al. (2001) asserted that the fragmentation is caused by the mounting pressure from the growing population. Thus, unlike decades ago, production increase from cropping area expansion is less feasible, making yield growth from the use of the cropping area already available necessary to produce enough for consumption and sell.

Farming in Ethiopia depends on the summer rain. Moreover, most farmers employ poor technological equipment. Preparing the land consumes a substantial portion of the farmers' time due to the poor equipment employed. Workneh et al. (2021) indicated that land preparation for a September harvest starts in March and runs until the end of July. The intensity of crop production is thus constrained to only once in a year. Adopting improved seeds could be a feasible pathway to enhance the productivity and stability of crop production in the country. The Ethiopian Ministry of Agriculture defined improved seed as any seed that has been tested and evaluated for its superiority over existing varieties (Ethiopian Ministry of Agriculture, 2013 cited in Kosmowski et al., 2019).

According to Neill (2018), there are three types of improved seeds depending on how they are created. These are hybrid seeds, open pollinated varieties (OPVs), and genetically modified seeds (GMOs). The hybrid seeds are created by crossbreeding of two plants of desirable quality. The genetically modified seeds are produced by transferring a gene from an organism to a plant's genome. As the name would imply, the open pollinated varieties are created by natural pollination of plants. Adopting improved seeds has been the focus of the Ethiopian government in the past two decades. Nevertheless, the adoption of improved seeds by Ethiopian farmers, particularly small-scale farmers, is still very low. As reported by AGRA (2019), the national intake of the hybrid seeds is only 10%.

The improved seed adoption rate of the country is very low even when compared to countries in eastern Africa. For instance, as stated by AGRA (2019), the neighbouring country Kenya has a national hybrid seed adoption rate of 60%. Neill (2018) noted that the low adoption rate of improved seeds in Ethiopia is due to supply constraints. It is indicated that the demand for the improved seeds of five of the countries' major cereals (i.e., maize, teff, wheat, sorghum, barley) exceeded the supply by 72% in 2008. The supply challenge is partly related to the limited engagement of the private sector in the breeding and dissemination of improved seeds. Seed breeding is a costly business for firms hence they have several conditions to join the market.

For example, the government should have a strong intellectual property rights protection policy, a promise very difficult to make for a country like Ethiopia that is not yet a member of the World Trade Organization (WTO). Furthermore, some improved seeds such as genetically modified seeds and open pollinated varieties can be used for more than one seed cycle without losing their quality traits. Farmers do not need to buy the improved seeds every cropping season, and this disincentivizes the engagement of the private firms in the seed breeding business. The Ethiopian seed market is thus controlled by the Ethiopian Seed Enterprise (ESE), a state-owned entity. The distribution of the adoption of improved seeds is socioeconomically skewed in the country.

As stated by Tura et al. (2010) and Abadi et al. (2015), institutional variables such as access to credit are strong predictors of who is the improved seed adopter in Ethiopia. Hasen Ahmed et al. (2017) showed the presence of complementarity in the adoption of improved seeds and other agricultural technologies. The improved seeds are usually adopted complementarily with inorganic fertilizer and manure. Furthermore, Mulesa et al. (2021) noted a differential access to preferred seed varieties across gender and sex. Even though the adoption of improved seeds in the country is limited, lots of researches have been carried out to investigate the yield gain from the use of improved seeds.

Ample evidences exist in support of the view that improved seeds bring more yield per unit area than traditional seeds, hence they have land augmenting effects (Tesfaye et al., 2016; Abel et al., 2014; Abate et al., 2015). However, according to Meughoyi (2018), Shiva (1991), Negi (1994), and Press (1996), the improved seeds need differential treatment to outperform the traditional seeds and provide more yield. The authors pointed out that the productivity of the improved seeds does not depend only on the productive quality of the improved seeds. It also depends on the application of other packages of technology. For instance, Ceteris paribus, the improved variety needs augmentation with greater levels of purchased inputs such as fertilizer and pesticide than the traditional variety to provide optimal yield.

Thus, knowing the status (i.e. improved or traditional) of a seed adopted is vital to differentially treat the seed based on its specific needs. Recent studies revealed that farmers in Ethiopia and other developing countries substantially misclassify the adopted seed (Wossen, Abdoulaye, et al., 2019; Wineman et al., 2020; Jaleta et al., 2020; Kosmowski et al., 2019; Floro et al., 2018; Wossen, Alene, et al., 2019). Seed misclassification is a measurement error that occurs when the farmers wrongly perceive the status of the seed they are using. It thus occurs in two situations; when an improved seed adopted is classified as a traditional, and when a traditional seed adopted is classified as an improved by the farmers.

There are several factors that could drive farmers to seed misclassification error. Imperfect seed market is one source of the misclassification error (Wossen, Abdoulaye, et al., 2019). For example, a farmer out to a market to purchase an improved seed could get easily deceived by a seed adulterated by the seller. It is difficult for the farmer to visually identify the status of the seed while purchasing. Wineman et al. (2020) mentioned that informal seed sourcing, seed recycling, and self-seed breeding by the farmers are also potential causes of the misclassification error. Farmer -to-farmer seed exchanges in the form of purchases and borrowing are widespread practices. The exchange process is likely to be accompanied by either willful deception or losing track of the status of the variety by the seller.

Kosmowski et al. (2019) stated that the quality of an improved variety deteriorates overtime. It is highly likely for the improved seed to be misclassified as a traditional when its quality drops. Mixed cropping practice on the plot, a common risk management strategy by the farming community, can also bring the misclassification error (Spielman et al., 2017 cited in Wineman et al., 2020). Inconsistent naming of the improved seeds across sources (for instance, different naming of the same variety in different villages) is another explanation for the misclassification error indicated by Floro et al. (2018).

The presence of seed misclassification error could lead farmers to suboptimally treat the adopted improved seed. For instance, as outlined above, the improved seed often needs a greater level of inorganic fertilizer relative to the traditional seed. However, if the improved seed is misclassified and perceived as a traditional seed by the farmer, the farmer could supply a level of fertilizer that is optimal for the traditional seed. Thus, the productivity of the improved seed adopted would be lower. Therefore, because of the misclassification error, farmers that adopt the improved seeds could lose yield that would otherwise be gained.

1.2 Motivation

Agriculture plays a critical role in the lives and livelihood of peoples in most developing countries. In the sub-Saharan Africa region, the agriculture sector employs 75% of the population, and contributes for 35% of the gross national product (GNP) and 40% of the foreign exchange earnings (Workneh et al., 2021; Urgessa, 2014). In Ethiopia, agriculture is a backbone of the economy supplying 84% of the exports, employing 72% of the population, and contributing for 37% of the gross domestic product (GDP) (FAO, 2018). Several evidences indicated that adopting improved seeds is crucial to boost farm production and improve the welfare of the farmers (Mugisha & Diiro, 2010; Tesfaye et al., 2016; Meughoyi, 2018). The evidences suggest developing countries to adopt the improved seeds.

Some developing countries such as India and Mexico managed to boost agricultural production by using the improved seeds. The use of improved seeds is now the focus of the agriculture sector in most developing countries. As evidenced by (Negi, 1994; Press, 1996), the improved seeds need differential treatment to provide the supposed productivity gain. It is not only the supply and distribution of the improved seeds that is important to increase the yield on the farm but also the farmers correct classification and perception of the improved seed as an improved seed. If the improved seeds are misclassified, farmers could fail to meet the specific needs of the improved seeds, and thus they would get a suboptimal yield from the use of the improved seeds.

The adoption actors in developing countries (researchers, governments, crop institutes) have tended to focus on creating high yielding improved seed varieties and increasing the adoption of the improved varieties. Knowledge is lacking about the loss of potential yield that could arise after the improved seed is adopted but misclassified. Smallholder farmers dominate the agriculture sector in developing countries. The land they own is very limited and optimally using it is a matter of survival for them. For example, as indicated by FAO (2018), around 90 % of the area cropped and agricultural produce in Ethiopia is by smallholder farmers, and the average land size in the country is less than one hectare. It is thus essential to provide evidence for the misclassification borne yield loss.

Up to my knowledge, there is no previous empirical work dedicated to examining the causal impact of the misclassification measurement error on yield loss. This thesis will thus be a novel addition to the broader adoption and varietal identification literature. It is also the first to exploit the nationally representative crop DNA fingerprinting identification data, the 2018/19 wave of the Ethiopian socioeconomic survey from the World Bank's LSMS which got released December 2020. This is the first dataset in the LSMS repository to include barcodes of crops.

1.3 Objective & Hypothesis

Based on empirical evidences from previous varietal identification studies that the vast majority of improved seed adopters in developing countries misclassify the improved seed as a landrace, and based on the knowledge that the productivity of improved seeds depends on other agricultural inputs supplied, the aim of this thesis is to estimate the loss of potential yield caused by the misclassification error. The following are set as guiding research questions to meet the aim.

a) Does misclassification error exist in the study area?

b) Do misclassifiers and correct classifiers differentially treat the improved seed adopted?

c) Does the average yield of misclassifiers and correct classifiers significantly differ?

From these guiding questions, the following hypotheses are made.

H1: Adopters in Ethiopia misclassify the improved seed as a traditional seed

H2: Misclassification of the adopted improved seed has a negative and significant impact on yield

The thesis focuses on improved maize variety adopters in Ethiopia to meet the outlined research objective.

1.4 Ethiopia's Maize Profile

Maize is an introduced crop for Ethiopia. Unlike other cereals, Ethiopia is neither the origin nor the centre of diversity for maize (Lipper et al., 2005). It is believed that maize arrived in Ethiopia in the seventeenth century, one century later from the time it first entered to Africa (Huffnagel, 1961 cited in Abate et al., 2015). Even though maize arrived late in Ethiopia compared to most countries in Africa, Ethiopia is now one of the major maize producers in the continent. As of 2015, Ethiopia was the third largest producer of maize in the sub-Saharan Africa region following South Africa and Nigeria (Abate et al., 2015).

Cereals dominate Ethiopian crop production. Seyoum Taffesse et al. (2013) indicated that on average, cereals were grown on 73.4% of the total area cultivated by a total of 11.2 million farmers that produced a yearly average of 12 million tons of cereals which was 68% of the total agricultural production over the period 2004 to 2007. Together with the other four major cereals (teff, wheat, sorghum, and barley), maize dominates the crop production in Ethiopia. The authors mentioned that the five majors alone accounted for three-quarters of the total area cultivated, and 29% of the agriculture GDP in 2005. Maize is produced in diverse agroecological zones in the country.

These are dry lowlands (<1000 m above sea level, <700 mm rainfall), dry mid-altitude (1000–1600 m above sea level; 650–900 mm rainfall), moist lowlands (<900 m above sea level, 900–1200 mm rainfall), moist lower mid-altitude (900–1500 m above sea level; 900–1200 mm rainfall), moist and semimoist (1700–2000 m above sea level; 1000–1200 mm rainfall), and moist upper mid-altitudes (2000– 2400 m above sea level; >1200 mm rainfall) (Abate et al., 2015). The moist and semimoist midaltitude zone covers around 75% of the country's maize area. Major maize producing regions of western and northern western Amhara, western and southwestern Oromia, and parts of Benishangul Gumuz and Southern Nations Nationalities and People region are in this zone. The dry ecology covers the remaining 25% of the country's maize area.

Maize is the top of the cereals in Ethiopia in terms of gross production and productivity. Ifpri (2010) indicated that maize production was 4.2 million tons, 40% higher than teff, 56% higher than sorghum, and 75% higher than wheat production in 2007. The article also indicated that over the period 1995 to 2008, the average maize, wheat, and sorghum yield was 1.74, 1.39, and 1.36 tons per hectare, respectively. Furthermore, it is mentioned that more than any other crop, around eight million smallholders were involved in maize production during the 2008 production season, compared to 5.8 million for teff, and 4.5 million for sorghum, the second and third most cultivated crops in Ethiopia. According to Abate et al. (2015), the number of households producing maize grew to 9 million as of 2013.

The rate of growth of the maize producers over years is higher than the rate of growth of the other major cereals producers. Between the period 2004 and 2013, the average annual growth of the maize producers in the country was 3.5% each year compared to 1.8% for barley, 2.1% for wheat, 3.0% for sorghum, and 3.1% for teff (Abate et al., 2015). Moreover, maize is dominant in terms of the total area allocated for production. The above-mentioned study indicated that the maize area in the country doubled in the past two decades from 1 million to more than 2 million. As of 2013, maize covered 22% of the total cereal acreage of Ethiopia.

The increase in maize area is partly attributed to production shifts away from the other major cereals (Abate et al., 2015). Production destruction related to weaver bird invasion in the 1980's pushed sorghum farmers in the Rift Valley to resort to maize production. Farmers in the major teff producing regions also switched to maize to exploit the considerable productivity gains new maize varieties provided. Between 1981 and 1983, as indicated by the above study, maize occupied roughly 16% of the total cereal area compared to 30% for teff, 20% for sorghum, 14% for wheat, and 19% for barley. Over the period 2001 to 2003, the area covered changed to 24% for maize, 31% for teff, 17% for sorghum, 15% for wheat, and 13% for barley. Cereals in general contribute the most to daily household diets in Ethiopia. The five major cereals together constituted 64% of the calories consumed in 2005 (Ifpri, 2010).

Maize is the major contributor among the cereals for household diets in Ethiopia. In 2003, it alone contributed 20% of the average Ethiopian daily caloric intake (i.e., 1858 kilocalories per day) (Ifpri, 2010). Maize is also the lowest cost caloric source among the major cereals. Ifpri (2010) mentioned that the unit cost of calories per US dollar for maize is one-and-a-half and two times lower than wheat and teff, respectively. The article also mentioned that maize is a low-cost source of protein compared to the other cereals. Maize provides 0.2 kg of protein per USD, compared to 0.1 kg of protein per USD from teff and 0.2 kg of protein from wheat and sorghum. Maize is the largest cereal in terms of the adoption of improved seeds. From 1997 to 2015, the percentage of maize area covered with improved seeds compared to other cereals grew from 5% in 1997 to 46.38% in 2015 (Byerlee et al., 2007 cited in Hasen Ahmed et al., 2017).

There is a disparity in the rate of adoption of improved maize varieties amongst maize regions in Ethiopia. The regions such as Oromia, Amhara, and SNNP employ improved maize varieties the most in the country. Over the period 2010-2012, the percentage of the maize area covered with improved maize varieties was 49% in Oromia, 33% in Amhara, and 18% in SNNP, respectively (Abate et al., 2015). Maize production in the country has witnessed a significant boost in the past few decades despite the fact that the agricultural sector in general has been challenged by the vagaries of weather shock, low technological adoption, and environmental problems related to soil degradation and deforestation as indicated by (FAO, 2018). Fufa & Hassan (2006) underlined that the increase in the productivity of the maize sector was mainly because of extension services. They noted that the extension services raised the use of inorganic fertilizer and contributed to the productivity of the maize yield.

Abate et al. (2015) indicated that the average growth of maize productivity between 2004 and 2013 was 6.3% per annum. As a result, the national maize production doubled from about 1.50 million tonnes per hectare in the 1990's to 3.23 million tonnes per hectare in 2013. The study also indicated that the improvement in the sector is the result of a combination of factors. One of the factors mentioned is the increase in the supply and wider adaptability of improved varieties. According to Frederic Kosmowski et al. (2020), the total number of improved maize varieties released from 1970 to 2019 reached more than 60. The other factors indicated that contributed to the improvement in the maize sector are better research and extension work provision, growing market demand for consumption, improvements in production conditions, and better market access for the farmers.

1.5 Organization

The thesis is organized as follows. Chapter two presents the literature review. It discusses definitions and concepts related to improved seeds, the role of improved seed adoption for farm productivity, determinants of improved seed adoption, and seed misclassification error. Chapter three presents the methodology of the study. It introduces the data source, the definition of the variables used, and the theoretical model of the study. This chapter also illustrates the estimation strategy used. Chapter four is for discussion of results. The final chapter provides conclusions and policy recommendations based on the findings.

Chapter 2 - Literature Review

2.1 Definition & Concept

Improved seed generally designates a seed that has been tested and evaluated for its superiority over existing varieties (Ethiopian Ministry of Agriculture, 2013 cited in Kosmowski et al., 2019). Several traits have been added to improved seeds to supersede the benefits offered by traditional seeds. According to Wainaina et al. (2014), the traits that have been added to the improved seeds include quality improvement (improving nutritional value, flavor, and beauty), increasing yield, increasing tolerance to environmental pressures (temperature, salinity, and drought), resistance to diseases (fungi, viruses, and bacteria), increasing tolerance to insects and herbicides, and longer storage capacity. There are three categories of improved seeds; hybrid seeds (developed by cross-breeding of two plants of desirable quality, and they retain their quality trait only for one seed cycle), open pollinated varieties (created by selecting and saving naturally reproduced seeds), and genetically modified seeds (created by inserting a desirable gene from an organism to a plants genome) (Neill, 2018).

Most adoption studies set their own specification of what an improved seed means in their context. Often, the specifications don't conform to the above formal and commonly agreed definition adopted by the Ministry of Agriculture. For instance, Floro et al. (2018) considered only cultivars developed by seed breeding companies or agricultural institutes in their specification of what an improved seed means. However, the improved seeds are not always bred by a formal breeding institute or company. The farmers themselves can breed their preferred cultivars locally. Natural events such as wind can also create the improved seeds. Based on the above definition adopted by the Ministry of Agriculture, seeds bred by farmers and wind are improved if they have a superior attribute over the existing varieties. The authors acknowledged that their specification may misdefine some of the improved varieties as a traditional seed.

For instance, their specification excludes open pollinated varieties and hybrid seeds that are bred locally by the farmers from the improved seed category. Wineman et al. (2020) considered all released varieties with a superior characteristic attribute as improved seed irrespective of the breeding sources. It also included recycled seeds that maintain their superior attribute as improved seed. Wossen et al. (2019) provided a more nuanced specification. The authors used DNA identification to identify which seed is improved and which one is a landrace. Their definition counted both officially released and unreleased varieties as improved seed. Officially released varieties include seeds developed by formal breeding companies, and seeds developed by natural mechanisms and the farmers. Unlike the specification by Floro et al. (2018), this definition counted informally bred varieties as improved seed.

Moreover, the definition by Wossen et al. (2019) departs from the definitions by Floro et al. (2018) and Wineman et al. (2020) by noting the need to consider unreleased varieties. It is pointed out that unreleased varieties (i.e., varieties that are not officially out from the research centers) have to be included in the definition due to the possibility that adopters can acquire them through backchannels, leakages, and spillovers from the research centers. The outlined differences between studies in the definition of the improved seed they adopted could be related to the varietal identification method the studies used. For instance, the use of a survey (self-reporting by the adopters) variety identification data could make the researcher to worry about the classification of the improved seed as a landrace wrongly.

Thus, the researcher could set a definition that best reduces the classification error. For instance, the researcher could consider only seeds bred by formal sources as improved seed. DNA fingerprinting identification, used by Wossen et al. (2019), is almost perfect to identify the variety type. It frees the researcher from concerns related to the misclassification error. In this case, a definition that counts all conceivable improved varieties in the study area can be adopted. This thesis employs DNA fingerprinting identification data. Thus, improved seed refers to any seed that has a superior attribute over the existing varieties. It covers open pollinated varieties, hybrid varieties, and genetically modified varieties. Recycled varieties that maintain their superiority trait are also counted as improved seeds.

2.2 Role of Adoption for Farm Productivity

The green revolution was aimed at transforming agriculture and food production in the developing world. High yielding varieties of several cereals such as wheat, maize, and rice produced by a plant breeding program led by Dr Norman Borlaug played a key role in catalyzing the revolution. The adoption of the high yielding varieties transformed crop production in many countries. The two showcases can be Mexico and India. Wheat production couldn't meet the domestic need in Mexico in 1943. The adoption of Borlaug's high-yielding semi-dwarf improved wheat varieties made Mexico a wheat net exporter amounting 500, 000 tons a year in 1965. The total production in 1965 was 6 times greater than the production in 1943 (University of Minnesota, n.d.).

India was on the brink of mass famine in the 1960's. The adoption of IR8, a semi-dwarf improved rice often dubbed as the miracle rice, increased the rice production from two tons per hectare in the 1940's to 6 tons in the 1990's. It also reduced the cost of a ton of rice from $550 in the

1970's to under $200 in 2001(Barta, 2007). India is now one of the most successful rice exporters in the world. The success story of the high yielding varieties of the green revolution was repeated in several developing countries such as Philippines, Pakistan, and China. Dr Norman Borlaug is considered as the father of the Green revolution, and he also won the Nobel Peace Prize in 1970 for his tireless contribution to feed the world's hungry people. Improved seeds are now used by farmers in all countries of the world.

The worlds' population size is now shooting. Further expansion of the arable land is very limited if not impossible. Vandana Shiva, speaking to the New Yorker considered by many as a crusade against genetically modified seeds, noted that "for most of the past ten thousand years, feeding more people simply meant farming more land. That option no longer exists; nearly every arable patch of ground has been cultivated" (Specter, 2014, para.5). Thus, intensive use of the cropping area is essential. Eliazer Nelson et al. (2019) mentioned that between 1950 and 1990, the global cereal production increased by 174% while the global population increased by 110%. The authors also mentioned that the total productivity (i.e., yield per hectare) and total food production in the developing world doubled over the period 1960 and 1985. Specifically, between the period 1960 and 2000, the adoption of high yielding varieties enabled an increase in the average production in the developing world by 157% for maize, 208% for wheat, and 109% for rice, respectively (Pingali, 2012).

Several adoption studies provided evidences that improved seeds increase crop productivity. Datta (1968) investigated the impact of the adoption of the IR8. They found that the yield from the use of the improved rice variety was progressively higher compared to the yield from the use of the traditional rice variety. Tesfaye et al. (2016) studied the impact of improved wheat variety adoption on wheat productivity of households in Arsi, Ethiopia. They found that adopters of the improved wheat variety produced about 1 to 1.1 tons per hectare over non-adopters. A Productivity gap decomposition analysis by Meughoyi (2018) in Cameroon showed that the average yield of the improved maize seed adopters outweighed the non-adopters by about 0.351kg/hectare, which is 1.42 times higher.

In a study in Uganda, Mugisha & Diiro (2010) found that the average yield from the use of improved maize varieties was higher than that of local varieties by a margin of 2941.5 to 1694. kg/ha per season. Findings from Abel et al. (2014) on rice varieties in Ghana also confirmed the dominance of improved seeds over traditional seeds in boosting crop productivity. Although high yielding varieties transformed the global agricultural production during the green revolution and after, there are studies that are still doubtful about the role of improved seeds for crop productivity. Their central argument is that improved seeds perform well only if they are provided with higher amounts of other inputs.

The implication from the view of the studies which doubt the productivity of improved seeds is that traditional seeds can outperform the improved seeds if they are treated with equal inputs as the improved seeds. Press (1996) noted that improved seeds perform well in environments with good irrigation and water control. It is also noted that improved seeds concomitantly need augmentation with high levels of purchased inputs such as fertilizer and pesticide. Igbozurike (1978) disregarded the notion of the superiority of improved seeds by comparing the monoculture system of improved seeds with the polyculture of traditional seeds. The study was done in the Obukba region, a densely populated area in eastern Nigeria. In the study area, polyculture was a condemned and monoculture was a lauded cropping practice. The findings of the comparative analysis concluded that polyculture is economically more productive than monoculture.

The study by Igbozurike (1978) relied on the cash income reported by the farmers from a hectare of crop for the economic analysis. However, the self-reported income figure is very likely for understatement thus it weakens the results reported. Negi (1994) is another study that paid little attention to the superiority of improved seeds over traditional seeds in terms of productivity. The study analysed the impact of two high yielding and one traditional wheat varieties on yield when cultivated traditionally by farmers in the Himalayan mountains, India. Farmers who were willing to grow the improved wheat varieties, HS-86 and VL-404, were supplied with the varieties for free. The varieties except for VL-404 were grown in both rain feed and irrigation conditions on the plot in separate terraces but in the same tiny clusters.

The varieties were exposed to similar soil and environmental conditions. Whether to grow wheat varieties under the rainfed or irrigation condition as well as the application (type and quantity) of the other inputs (fertilizer, labour, manure etc) were completely farmers' decision. The results found show that in the rainfed fields, the local variety produced more yield. The improved varieties performed well only on irrigated fields. Furthermore, Meughoyi (2018), although they found improved seeds produce more yield than traditional seeds, pointed out that the yield they found from the use of improved seeds is quite below what is theoretically expected. They noted that the reason could be the failure of the adopters to comply with the specific guidelines the improved seeds need. They mentioned that the guidelines include the use, period, and method of fertilizer, and the application of herbicide on the farm.

2.3 Determinants of Adoption

Even though adopting improved seeds plays a substantial role for crop productivity, the adoption of the improved seeds in sub-Saharan Africa agriculture is still quite low. For instance, the adoption of improved seeds in Ethiopia is historically sparse. The country's national hybrid seed intake rate is only 10% (AGRA, 2019). According to Neill (2018), the low adoption of the improved seeds in Ethiopia is mainly due to supply constraints. It is indicated that the demand of the farmers for the improved varieties of five of the country's major cereals exceeded the supply by 72% in 2008. The constraint in the supply of improved seeds is partly due to limited engagement of the private sector in the seed production and dissemination process.

The limited engagement of the private sector in the seed market could arise from several reasons. For instance, the cost for varietal research and development efforts is large. It is thus vital to have a strong intellectual property rights protection to attract the private firms to the seed market. However, Ethiopia is not yet member of the WTO, the responsible organization for the protection of intellectual property rights. Some types of the improved seeds such as open pollinated varieties and genetically modified seeds can be effectively recycled for more than one season without losing their quality attributes. The farmers thus do not need to buy the improved seeds every cropping season which is not to the interest of the profit oriented private firms. Therefore, the production and distribution of improved seeds in Ethiopia is dominated by a government entity, Ethiopian Seed Enterprise.

The farmers' decision to adopt improved seeds correlates with several characteristics. Based on the literature reviewed, these characteristics can be broadly categorized into three categories: the adopters' socioeconomic condition, the institutional capacity in the adoption area, and the characteristics of the improved seeds. The factors related to the characteristics of the adopters include education level, off-farm work, distance from the market, human capital (such as supply of adult workers and experience level), the adoption of other inputs (such as fertilizer and manure), and asset endowment (such as land size and livestock). Tura et al. (2010) investigated the adoption and continued use of improved maize varieties in central Ethiopia using a bivariate probit model. They found that human capital and asset endowment of the adopters are key factors influencing the adopters' decision to adopt.

Abadi et al. (2015) also studied the factors that determine the adoption of improved varieties in central Ethiopia. The results confirmed that socioeconomic conditions of the adopters play a significant role in influencing the farmers decision to use the improved seeds. This study and the study by Tura et al. (2010) are totally redundant to each other. The study area of both studies is central Ethiopia. The empirical approach both studies used is the same (i.e., bivariate model). The results from both studies showed that the adopters' literacy, family size, livestock wealth, and access to credit are positively correlated with improved seed adoption. On the other hand, the adopters' distance from the main market negatively affected the adoption of the improved seeds.

The institutional capacity in the adoption area covers factors related to the availability and quality of institutions concerned with the distribution of agricultural technologies. The factors include access to extension programs, access to improved seed suppliers (membership to agricultural cooperatives, and to farmer associations), provision of credit services, and access to the other farm inputs (inputs such as fertilizer). Tura et al. (2010) showed the institutional variables, specifically access to credit and membership to cooperatives are positively correlated with the decision to adopt the improved seeds. Abadi et al. (2015) found that farmer associations negatively influence the adoption decision. The result from this study about the role of credit contradicts with the result from Tura et al. (2010). While Tura et al. (2010) found a positive correlation between the improved seed adoption and credit service, Abadi et al. (2015) showed that credit for fertilizer is negatively correlated with the decision to adopt improved seeds.

There is evidence provided for joint adoption of improved seeds and other agricultural technologies. Hasen Ahmed et al. (2017) explored cropping system diversification, improved seed, manure, and inorganic fertilizer use by farmers in eastern Ethiopia. They found complementarity in the adoption of the farm technologies. The complementarity found is between the adoption of improved seeds and fertilizer, manure, and inorganic fertilizer. Kaliba et al. (2000) investigated the determinants of the adoption of improved maize varieties in Tanzania. Their results indicated that the adoption of improved maize varieties depends on weather conditions such as the availability of adequate rainfall, and the characteristics of the maize variety itself.

Kaliba et al. (2000) noted that farmers preferred maize varieties not to maximize the yield, rather to minimize the loss of yield. Moreover, according to Sánchez-Toledano et al. (2018), the improved seed adoption decision is conditional on the adopters' attitude towards risk. They underlined that young adopters are relatively risk-takers, and they are open to give a try to agricultural innovations such as improved seeds.

2.4 Seed Misclassification

The problem surrounding improved seed adoption in developing countries is not limited to the poor adoption rate. The problem, as revealed by recent studies, includes substantial misclassification of the improved seed used as a traditional seed. Most studies on improved seed adoption relay on self-reported adoption data, which assumes that adopters correctly identify the type of seed they adopted. The reliance on self-reported data is a critical limitation of the majority of the studies on seed adoption. Seed misclassification is a measurement error that occurs when the improved seed adopted is wrongly perceived and reported as a traditional seed or when the traditional seed adopted is reported as an improved seed by the farmers. Thus, we have two groups of misclassifiers; false-negative adopters, and false-positive adopters (Wossen, Abdoulaye, et al., 2019; Wineman et al., 2020).

False-negative adopters are those who perceive the improved seed they use as a traditional seed. False-positive adopters represent those who classify the traditional seed they use as an improved seed. Kosmowski et al. (2019) and Jaleta et al. (2020) investigated the misclassification of sweet potato and wheat in Ethiopia, respectively. Kosmowski et al. (2019) used three types of household-based methods. These are asking the households questions about the variety adopted, showing the households visual aid protocol to elicitate about the phenotypic attributes of the variety, and recording of the phenotypic attributes by the enumerator visiting the field with the help of the aid protocols. The elicitated information was then compared against the DNA identification benchmark. They found that 20% of the observations identified the local potato variety as an improved variety, and 19% of the observations identified the improved potato variety as a local variety.

Jaleta et al. (2020) found that only 28-34% of the farmers were able to identify the type of the wheat seed adopted correctly. Similarly, Wossen, Abdoulaye, et al. (2019), and Floro et al. (2018) studied the misclassification of Cassava varieties in Nigeria and Columbia, respectively. The findings from Wossen, Abdoulaye, et al. (2019) showed that 35% of the cassava variety was misclassified. Around 25% of the farmers reported the improved cassava variety as a local variety. The remaining 10% classified the local cassava variety as an improved variety. The results from Floro et al. (2018) also indicated that farmers substantially misclassify and overestimate their use of improved seeds.

Furthermore, Wineman et al. (2020) studied misclassification of Maize seeds in Tanzania. They found that 30% of the observations wrongly identified the status of the maize seeds adopted. While 16% of the observations perceived the traditional maize variety they grew as improved, 14% perceived the improved maize variety as a traditional variety. Seed misclassification can arise from several reasons. The source the seed is obtained is one key correlate of the misclassification error. The seed market in developing countries is imperfect. For instance, the seller might adulterate the seed before it comes to the market (Wossen, Abdoulaye, et al., 2019). Informal seed sourcing, seed recycling, and self-seed breeding by farmers are other sources of misclassification (Wineman et al., 2020). Farmer - to-farmer seed exchanges in the form of purchases or borrowings are widespread practices by the rural agricultural community. The exchanges are likely to be accompanied by willful deception or losing track of the type of variety by the seller.

Due to factors related to limited supply and the high cost of the new varieties, the use of seeds from previous harvests is widespread. Farmers also breed seeds by crossing their preferred cultivars locally. When the seed recycling practice and the self-seed breeding are repeated over a period of time, adopters could become unsure of the genetic identity of the variety they are adopting. Kosmowski et al. (2019) mentioned that overtime, the quality of an improved variety deteriorates. The farmers could then perceive it as a traditional variety since low quality is mostly associated with the traditional seeds. Jaleta et al. (2020) examined the correlates of misclassification. The results showed that missclassification is positively correlated with seed recycling and negatively correlated with the purchase of seeds from trusted sources such as agricultural cooperatives and known farmers. Multiple cropping practice, a risk management strategy, can also bring misclassification (Spielman et al., 2017 cited in Wineman et al., 2020).

Inconsistent naming of the varieties across sources (for instance, different naming of the same variety in different villages) is another potential explanation indicated by Floro et al. (2018). The misclassification error is correlated with the socioeconomic characteristics of the adopters (responders). Wineman et al. (2020) mentioned that the major sources of heterogeneity that explain the probability of misclassification are the level of education, access to information, and location of the adopters. Wossen, Abdoulaye, et al. (2019) found that the access to information (i.e., mobile phone ownership, access to extension service, membership to cooperatives), and the regional location of the adopters are correlated with the misclassification of the improved seed as a landrace seed.

Moreover, Jaleta et al. (2020) showed that the number of years since the release of the variety is positively correlated with the misclassification status. Seed misclassification distorts the estimates of the impact of adoption. Wossen, Abdoulaye, et al. (2019) analysed the productivity impact of adoption considering misclassification. They approached the concept by looking into the deviation of the yield estimates when the self-reported and DNA data are used. They found that misclassification is endogenous to the household's characteristics and biases the productivity impact of adoption by around 22 percentage points. They showed that when behavioral adjustments (i.e., farm management decisions that surround the perceived seed adopted) are ignored, the bias gets lower to 13 percentage points. The authors used the instrumental variable model to derive the casual impact of adoption. The instrument the authors chose is self-reported improved seed adoption status by the neighbours. However, as the seed from the self-reporting by the sample observations is misclassified, there is no reason to assume that the seed from the self-reporting by the neighbours will not be misclassified. The robustness of their results is not confirmed using another instrument or estimation strategy.

As Meughoyi (2018) pointed out, the optimal yield from improved seed adoption requires adopters to comply with specific guidelines in terms of the supply of complementary inputs and other farm decisions. The misclassification of the improved seed as a traditional seed would not allow the adopters to provide the differential treatment the improved seeds need for optimal production. Wineman et al. (2020) provided several correlation exercises related to the misclassification error. One of their exercises showed that the average yield of the improved seed identified correctly is almost 700 kilograms greater than the improved seed misclassified. The authors constructed the yield proxy using the amount of yield harvested and area cropped self-reported by the farmers.

However, variables such as harvest and plot area are often subject to deliberate misestimation by the farmers in times of self-reporting (Carletto et al., 2015 & Desiere and Jolliffe, 2018 cited in Wineman et al., 2020). For instance, the amount harvested is often underestimated for large plots and overestimated for small plots. The area of the plots is often underestimated for large plots and overestimated for small plots. The result from Wineman et al. (2020) doesn't imply causality, but it clearly hints that improved seed adopting farmers are likely to lose a potential yield because of the misclassification error. This thesis tries to drive the causal impact of the misclassification error on yield loss.

The latest wave of the Ethiopian Socioeconomic Survey data, the 2018/19, from the World Bank's LSMS is used. As discussed above, the harvest and plot data self-reported by the farmers are subject to misestimation. To minimize the systematic misestimation of the yield variable, a harvest data directly collected and measured by enumerators is used. The enumerators collected both the fresh and dry weights of the harvest from a randomly selected 16 m^2 section of the plot as part of the DNA fingerprinting identification process of the variety. This figure is not expected to be vulnerable to misestimation as to the response to asking a farmer how much he/she harvested months/year ago from the plot. For the plot size, the GPS measure is recorded, and it sufficiently overcomes the misestimation from the self-reporting. The propensity score matching technique adopted enables to derive the causal impact of the misclassification error.

Chapter 3 - Methodology

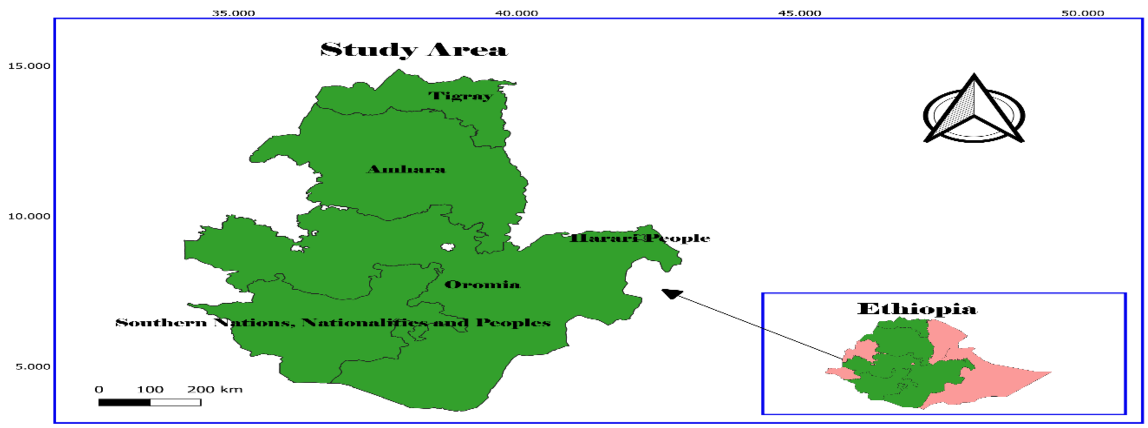
3.1 Data Source

The data used, the Ethiopian socioeconomic survey 2018/19 wave (ESS4), is obtained from the World Bank's Living Standard Measurement Study (LSMS). This is the first wave of the new panel of households started. Unlike previous versions, ESS4 extends the samples to regions that were previously aggregated as one (i.e., Afar, Harari, Somali, Dire Dawa, Gambela, and Benishangul-Gumuz). To ensure representativeness rural-urban wise, all regions in the country including the two administrative cities are covered by this new panel. The survey was administered in three visits that occurred between October 2018 and August 2019. During the first visit, administered between September and October 2018, the post-planting agriculture was collected. The second visit was to collect the post-harvest agriculture, and it was administered between February and March 2019. The final visit was to collect the household and community questionnaire, and it took place between June and August 2019.

This study exploits the post-planting module of the ESS4. ESS4 consists of barcodes of some crops planted such as maize, sorghum, and barley. This makes it the first survey from all countries in the LSMS repository for integrating objective measures of crop varietal identification in a large, institutionalized, and nationally representative survey. The survey consists of a maize varietal identification data from six major maize producing regions. The regions are Amhara, Oromia, Southern Nations, Nationalities, and Peoples, Tigray, Harar, and Dire Dawa. The sampling procedure followed a two-stage stratified probability approach. The primary sampling units (i.e., enumeration areas (EAs)) were first selected. Then, the secondary sampling units (i.e., households) were selected from the enumeration areas.

While a total of 10 agricultural and 2 nonagricultural households were selected from each rural EAs, 15 households were taken from urban EAs. A maximum of 10 fields by crop were selected using systematic random sampling from a frame of all plots of the sampled households. This procedure provided a total of 374 monocrop field observations from five major maize producing regions in the country from a total of 341 households. The sample observations from Dire Dawa are excluded by the data cleaning procedures (for example, selection of only monocrop fields) implemented for the strength of the analysis. Figure (2) shows the maize regions where the sample plots used in this study are picked.

Figure 1: Sample Coverage



3.2 DNA Extraction

DNA fingerprinting identification, which is a very recent breakthrough and according to Wossen, Alene, et al. (2019), which is a gold standard varietal identification technique, is all about matching the genetic identity of a crop variety with that of the genetic material of the same crop in the reference library. Thus, DNA extraction from all varieties that could conceivably be found in the landscape in question is an integral part of the identification process. The reference library lists the DNA extracted from all released varieties from all breeders. A comprehensive reference library is vital for DNA fingerprinting identification (Jaleta et al., 2020). For maize, the reference library for Ethiopia was compiled under a DNA fingerprinting research project conducted by the International Maize and Wheat Improvement Center (CIMMYT) and the Ethiopian Institute of Agricultural Research (EIAR) funded by the Bill and Melinda Gates Foundation (Kosmowski et al., 2020). From the 1970s to 2019, as presented in Table (1), 60 improved maize varieties were released in Ethiopia from various crop research centers (Kosmowski et al., 2020).

Table 1: Maize Varieties Released in Ethiopia since the 1970's

Variety	Year of Release	Type	Breeder
A-511	1973	Hybrid	Awassa ARC
Alemaya Composite	1973	Hybrid	HU
Katumani	1974	OPV	Bako ARC
Abo-Bako	1986	Hybrid	IITA/EIAR
BH 140	1988	Hybrid	Bako ARC
Gutto	1988	OPV	Bako ARC
BH 660	1993	Hybrid	Bako ARC
Kuleni	1995	Hybrid	Bako ARC
BH540	1995	Hybrid	Bako ARC
Jabi	1995	Hybrid	Pioneer
Tesfa	1996	OPV	ACA
Fetene	1996	OPV	ACA
Rare-1	1998	Hybrid	HU
Gibe Comp-1	2001	OPV	Bako ARC
Gambela Comp1	2001	OPV	EIAR
Melkassa-1	2001	Hybrid	Melkassa ARC

Table 1: continued

Tabor	2001	Hybrid	Pioneer
Shindi	2001	OPV	Pioneer
BH-670	2002	OPV	Bako ARC
BH-QP-542	2002	OPV	Bako ARC
BH-541	2002	Hybrid	Bako ARC
Melkassa -2	2004	OPV	Melkassa ARC
Mekassa-3	2004	OPV	Melkassa ARC
Hora	2005	OPV	Ambo ARC
AMH-800	2005	Hybrid	Ambo ARC
BH-543	2005	Hybrid	Bako ARC
Toga	2005	Hybrid	ESE
SC715	2005	OPV	SEED-Co
SC713	2005	Hybrid	SEED-Co
BH-544	2006	OPV	Bako ARC
Melkassa-4	2006	OPV	Melkassa ARC
Welel	2006	OPV	Pioneer
Shone	2006	Hybrid	Pioneer
Aba raya	2006	OPV	SEED-Co
BHQPY-545	2008	Hybrid	Bako ARC
Morka	2008	OPV	EIAR
AMH-850	2008	Hybrid	Ambo ARC
Melkassa-7	2008	Hybrid	Melkassa ARC
Melkassa-6Q	2008	Hybrid	Melkassa ARC
Mekassa-5	2008	Hybrid	Melkassa ARC
Agar	2008	Hybrid	Hybrid
AMH-851	2009	Hybrid	Ambo ARC
Gibe-2	2011	Hybrid	Bako ARC
BH 661	2011	Hybrid	Bako ARC
AMH760Q	2012	Hybrid	APRC/EIAR
Hawassa-1	2012	Hybrid	ESE
MHQ138	2012	Hybrid	Melkassa ARC
MH130	2012	Hybrid	Melkassa ARC
Limmu	2012	Hybrid	Pioneer
BH547	2013	Hybrid	Bako ARC
BH546	2013	Hybrid	Bako ARC
MH140	2013	Hybrid	Melkassa ARC
Melkasa-1Q	2013	Hybrid	Melkassa ARC
SPRH1	2015	Hybrid	Bako ARC
SBRH1	2015	Hybrid	Bako ARC
BHQP548	2015	OPV	Bako ARC
Damote	2015	Hybrid	Pioneer
AMH853	2016	Hybrid	Ambo ARC
AMH852Q	2016	OPV	Ambo ARC
Kortu	2017	Hybrid	Pioneer

Source: The report by the Standing Panel on Impact Assessment (SPIA), 2020. Shinning a brighter light: comprehensive evidence on adoption and diffusion of CGIAR related innovations in Ethiopia.
Note: ACA and HU stand for Awassa College of Agriculture and Haramaya University, respectively.

From the maize sample plots in the ESS4 selected for varietal identification, barcoded crop cuts from a randomly selected 16 m^2 quadrant of the plot were taken for DNA processing. The barcoded and dried field samples were transported to the International Livestock Research Institute (ILRI) campus in Addis Ababa, where they were dried further and then ground to obtain 50 grams of flour. The DNA from the dried cut was extracted in Addis Ababa, Ethiopia, using Qiagen DNeasy plant mini kits. Plates containing the DNA samples were then shipped to the Diversity Arrays laboratory in Canberra, Australia, for genotyping by sequencing using Diversity Arrays Technology Sequencing (DarTSeq) platform.

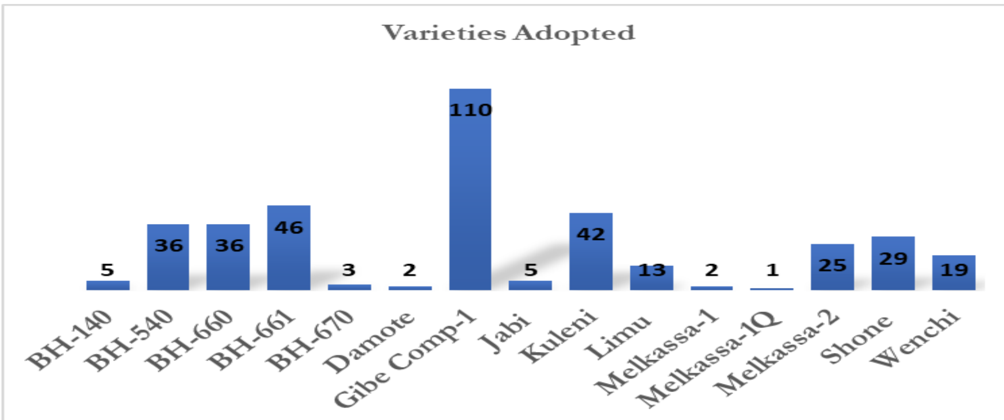
As described by Kilian et al. (2012), the DarTSeq platform uses a combination of a proprietary complexity reduction method and a next-generation sequencing platform. For each sample, approximately 200,000 fragments of DNA are sequenced, while matching relies on 20,000 polymorphic markers. The matching identifies the name of the crop variety to which each sample variety is matched, together with the level of purity of the sample. Additional outputs for analysis include the genetic separation between reference library samples as well as the sequenced genomic data. The report used in this study, matching the genetic material of the sample maize varieties in the ESS4 with the closest genetic match in the reference library, is obtained from the Consultative Group on International Agricultural Research (CGIAR).

As can be seen from Table (1), around 2/3 of the maize varieties released are hybrids and 78% of them are released after 2000. It is also evident that some breeding centers played a major role in the production of the maize varieties. For example, Bako ARC alone produced 1/3 of the varieties. Except for the hybrids Shone and Agar, the hybrids that came into production between 2005 and 2008 are limited in amount (Abate et al., 2015). The authors mentioned that BH661, promoted under the auspices of the Drought Tolerant Maize for Africa (DTMA) project, is essential due to its drought tolerance, resistance to major diseases, higher yield potential and wide adaptability. The demand by the seed companies for this hybrid was growing, thus it was expected to replace the two dominant hybrids in the Ethiopian seed market, BH660 and BH540.

Only one year after its release, in 2012, ESE produced 6 metric tons (MT) of certified BH661seed. In 2014, five companies, including Amhara Seed Enterprise (ASE), Avallo, ESE, Oromia Seed Enterprise (OSE), and Southern Seed Enterprise (SSE) produced nearly 2,900 MT of BH661 seed (Abate et al., 2015). As mentioned by the authors, among OPVs released, only four of them are widespread. These are Melkassa2, Melkassa4, Melkassa6, and Gibe2. However, their use is restricted to drought-prone areas such as the Rift Valley. Melkassa2 and Melkassa4 have been used extensively over several years. The relatively new releases Melkassa6 and Gibe2 are expected to take away their market share before they themselves are replaced by other new releases.

Only 15 of out of the 60 released improved maize varieties are adopted on the 374 sample plots used for this study. Figure (2) lists the name and frequency of adoption of the varieties.

Figure 2: Maize Varieties Adopted in the Study Area



A substantial gap in the adopters' preference for the varieties can be observed from Figure (2). The OPV Gibe Comp-1, planted on 29 % of the plots (i.e., used on 110 total plots), is the most favored by the farmers. At the same time, the hybrid Melkassa-1Q is the least preferred (i.e., adopted only on a single plot). The other varieties preferred by the adopters are BH-661 planted on 46 plots, Kuleni planted on 42 plots, and BH-540 and BH-660 planted on 36 plots each. Looking at the release year of some of the preferred varieties, we can witness the adopters' less initiative to adopt newly released varieties. For instance, Kuleni, BH-540, and BH-660 are released in the early 1990's. They are relatively old varieties, but they together covered around 30% of the sample plots. We can't attribute this adoption gap amongst varieties strictly to adopters' preference of one variety over the other since other players such as supply and price can also contribute.

3.3 Theoretical Model

Maize is an introduced crop to Ethiopia (Lipper et al., 2005). No farmer in the country adopts a landrace variety (i.e., all farmers use a genetically improved variety although the varieties can be at a different degree of purity). The maize farmers thus fall into two groups; false-negative adopters (who misclassify the adopted improved maize wrongly as a landrace), and true-positive adopters (who identify the adopted improved maize correctly as improved). The bivariate selection function into false-negative and true-positive adopter is as follows:

$$S = \beta 0 + \beta 1\ X + \epsilon \ldots\ldots\ldots\ldots\ldots (1)$$

Where, S denotes adopter type, and X captures a vector of explanatory variables.

Avoiding the misclassification of the seed could entail incurring additional costs over the actual price of the seed. For instance, as discussed in section 2.4 of the literature review, getting the seed from trusted sources such as agricultural cooperatives overcome the error. However, these sources are not as accessible as the less trusted sources such as friends and neighbours. Consider an adopter J, encountered with the decision to misclassify improved seed M, adopted on plot P.

$$Y^*_{JPM} = U^*_{MC} - U_{CC} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where, $Y^*_{JPM,}$ U^*_{MC}, and U_{CC} denote the net gain from adoption, gain from adoption in the presence of misclassification, and gain from adoption in case of correct classification, respectively. The latent variable Y^*_{JPM} could be measured in the form of utility, thus it is not observable to the researcher. What the researcher can observe is the misclassification status and conditional on it, the amount of yield from the plot. Following Hasen Ahmed et al. (2017), adopters misclassify the improved seed when Y^*_{JPM} is greater than zero. The selection variable, S, can thus be modelled in a random utility model.

$$S = \begin{cases} 1, \text{ if } Y^*_{JPM} > O \\ 0, \text{ otherwise} \end{cases} \ldots\ldots\ldots\ldots (3)$$

A treatment-control framework is adopted for the study. The exposure to misclassification error is the measure for treatment status. Thus, false-negative adopters who misclassified the improved seed adopted are the treated group. True-positive adopters are the control group since they are not exposed to the misclassification error. Means are the most commonly used evaluation parameters (Heckman et al., 1997). Following Heckman et al. (1997), the mean effect of the misclassification error, the treatment, on the treated is given by:

$$E\ (Y1 - Y0 \mid X, S = 1) = E\ (\Delta \mid X, S = 1) \ldots\ldots (4)$$

Where, Y1 and Y0 are the outcome variables for the treated observations with and without treatment, respectively.

However, we observe the observations only in one state. For instance, we know only the outcome values of the treated observations after treatment. We don't know the outcome values of the treated groups without the treatment. Likewise, we don't know the outcome value of the control group with treatment. We know only the values without the treatment. Thus, only Y1 is observable for the treated observations. This means equation (4) is not directly estimable.

As described by Heckman et al. (1997), the evaluation problem thus comes to missing data problem. Comparing the mean outcome of the treated group with that of the control group does not address selection bias. In fact, the whole sample (i.e., the treatment and control group together) is chosen randomly from the population. Thus, it can be fairly assumed that there is no bias in observing this sample. Nevertheless, self-grouping of the observations to the status of S=1 and S = 0 may not be random, thus the observations in the two groups can be already systematically different on variables that are important in determining the outcome variable, Y.

According to Heckman et al. (1997), matching reduces the selection bias. Rosenbaum & Rubin (1983) suggested to matching observations by a single score variable, in our case S(X) as specified in equation (1), instead of matching along each covariate. The suggestion is to reduce the dimensionality problem that could arise in the presence of large selection variables. An important assumption of the technique is the conditional independence assumption (CIA). It assumes that the selection into S is assumed entirely to be on the observables, and the potential outcome Y is assumed independent of the treatment assignment. With this assumption, the treated observations' counterfactual outcome had not they receive the treatment can be identified by the control observations' outcome.

Imposing the common support condition ensures the observations from the two groups to be compared are comparable in the first place (Heckman et al., 1997). With the CIA, equation (4) can be rewritten as

$$E\ (Y1 \mid S(X), S= 1) - E\ (Y0 \mid S\ (X, S = 0) \ \ldots\ldots\ldots\ (5).$$

This final equation captures the average loss of yield by the false-negative adopters because of the misclassification error.

3.4 Estimation Technique

The estimation strategy follows the theoretical model discussed above. As provided in the literature review (Chapter 2, Section 2.4), selection to the misclassification error depends on various adopter and plot specific characteristics. Based on the review, the variables chosen are responder gender, responder age, holder education level, age of the variety, source the variety is obtained, area of the plot the variety is adopted, number of plots on the parcel, location region, and access to advisory service. Thus, the bivariate selection function is specified as follows.

$$S = \beta_0 + \beta_1\ Age + \beta_2\ Gen + \beta_3\ Educ + \beta_4\ AdvS + \beta_5\ VarYear + \beta_6\ VarSour + \beta_7\ Area + \beta_8\ Plot + \beta_9\ VarReg + \epsilon_I \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (6)$$

The model, S, is set to a standard probit model, and it can be estimated using the maximum likelihood estimation.

3.5 Definition of Variables

3.5.1 Adopter Type

This is the main variable of interest, the treatment variable. The variable is measured by the match between the DNA fingerprinting identification and the self-reported adoption. The variable takes a value of 1 if the plot is false-negative (i.e., the seed planted on the plot is identified as improved by the DNA identification but reported as landrace by the adopters self-reporting), and 0 if the plot is true-positive, (i.e., both the DNA and self-reporting identified the seed as improved).

3.5.2 Maize Yield

This is the outcome variable that captures the cost (the loss of yield) of the misclassification error. Unlike the plot type variable, there is no DNA measure for this variable. However, unlike most studies that used recall data (asking farmers the yield they harvested from the plot), this study uses the yield collected and weighted from the plot by the enumerators themselves. This can significantly reduce the misestimation of the harvest by the farmers in self-reporting. As part of the genetic identification process, a fresh and dry weight of the maize crop was collected from a randomly chosen $16m^2$ segment of the plot. The yield figure is transformed to a hectare equivalent, a unit of measure most studies use (e.g., Ahmed et al., 2016; Wineman et al., 2020; Mugisha & Diiro, 2010). The conversion to hectare is done using the meter - hectare relationship (1 hectare = 10000 m^2). So, the yield from the $16m^2$ plot is multiplied by 625 to get the yield from the hectare equivalent. Here, it is assumed that each $16m^2$ segment of the hectare is equally productive.

3.5.3 Other Socioeconomic Variables

The age of the responder is measured in in number of years. Holders educational level is a dummy that captures the completion of primary schooling which in Ethiopia is till grade 8. To capture the age of the varieties, the varieties are categorized into three categories; varieties released before 2000, varieties released between 2000 and 2010, and varieties released after 2010. The source the variety is obtained is also aggregated into three categories based on their credibility in revealing the identity of the improved seed. The categories are markets, agencies (cooperatives, government agencies, farmer-based associations, and NGOs), and other sources (i.e., neighbors, relatives, friends, and seeds non-purchased such as recycled varieties).

The area of the plot used is measured with a GPS device. The number of plots represents the number of fields that are owned by the household and found on the same parcel with the sample plot.

Finally, regional dummies are controlled to account for the disparities in region-specific characteristics. Variety naming in Ethiopia often follows regional languages. So, the adopter's identification of the varieties can be influenced by the know-how of the regional language after which the variety adopted is named. Table (2) summarizes the variables selected.

Table 2: Summary of Variables

Variables	Definition
Dependent Variable	
S, Adopter Type	False negative = 1
Independent Variables	
Age	Age in years
Area	Plot size in m^2
Plot	Number of plots on the parcel
Gen	Male = 1
Educ	Primary school completion = 1
AdvS	Access to Advisory Service = 1
Variety Age	
VarT1	Released ≤ 2000
VarT2	Released between 2000 & 2010
VarT3	Released ≥ 2010
(Reference category)	
Variety Source	
Sour1	Agency (cooperatives, government agencies, farmer-based organizations, NGOs)
Sour2	Markets (Local and main markets)
Sour3	Others
(Reference category)	
Variety Region	
Tigray	
Amhara	
Oromia	
SNNP	
Harar	
(Reference category)	

Chapter 4 - Results & Discussions

4.1 Misclassification Error

4.1.1 Extent of the Error

Table (3) presents the misclassification error observed. The DNA identification revealed that all plots are covered with improved maize variety (i.e., all are true-positive plots). This is due to the fact that Ethiopia doesn't have a maize landrace. All maize varieties in the country are genetically improved introduced varieties. However, from the self-reporting, only 39% of the plots are identified as true-positive plots. This leaves a substantial number of the plots, 61%, to be false-negative plots (perceived by adopters as if they are covered with a traditional maize variety). This is a large departure between the DNA fingerprinting and self-reporting identification.

Table 3: Misclassification Error

	Self-Reporting	
DNA Fingerprinting	False-negative	True-positive
Adopter	61.23%	38.77%
N	229	145

The misclassification error, 61%, is comparable to Jaleta et al. (2020)'s misclassification finding on wheat identification in Ethiopia. They found that only 28-34 % of the farmers were able to identify the improved wheat variety they adopted correctly. The misclassification error in Ethiopia is greater when compared to the misclassification error found by varietal identification studies in other countries. For instance, Wineman et al. (2020) found that only 14% of the maize adopters in Tanzania were false-negative. Wossen, Abdoulaye, et al. (2019) showed that only 25% of the cassava growers in Nigeria were false-negative.

A fair argument one can raise here can be instead of attributing the mismatch between the DNA fingerprinting and self-reporting identification entirely to the misclassification error by the farmers, farmers can have their definition of what an improved seed mean. For instance, in our case, the improved maize seed on 61% of the plots identified as landrace might have to be improved to a greater extent to be considered as improved by the farmers. Most farmers do not know the genetic and technical details of the variety while acquiring the seed from the sources they purchase. The perception of the status of the variety by the farmers is likely to be influenced by factors such as local naming of the variety, the price paid to purchase the variety, and yield potential of the variety when compared to other previously experienced varieties by the farmers.

For instance, a new traditional maize variety purchased could be misclassified as improved by the farmers if it provided more yield on the first trial than the improved varieties experienced previously. Parallelly, the quality of the improved variety purchased deteriorates overtime. After some years, the quality can reach to a similar or lower level that farmers associate with traditional varieties. Then, the improved variety could be classified as a traditional variety in times of self-reporting. Therefore, further tightening of what an improved variety exactly means by specifications such as the year of release of the variety and the purity level might switch some of the plots from false-negative to true-positive category.

Table (4) shows the distribution of the false-negative and true-positive plots when the definition of the improved seed is jointly conditioned on the improved status from the DNA identification and the release year of the variety. For example, we can see from the table that specifying the definition of the improved variety to improved status from the DNA identification and a year of release after 2010 reduces the percentage of the false-negative adopters down to 27.4% from the original 61%. Conversely, the percentage share of the true-positive adopters rises to 72.6% from the original 39%.

Table 4: Misclassification when the Definition of the Improved Seeds is further Tightened

Adopter	VR ≤ 2000	2000< VR ≤ 2010	VR > 2010	Total Plots
False Negative	57.2%	75%	27.4%	229
True Positive	42.8%	25%	72.6%	145
N	124	188	62	374

4.1.2 Sources of the Error

The source where the seed is obtained is one major explanation for the misclassification error to arise. The seed during the cropping season can be purchased or obtained from non-purchasing sources. The common sources for a purchased seed include local markets, agricultural cooperatives, government agencies, and farmer-based organizations. The sources that adopters can get seed without purchasing during the cropping season can include saving from previous season purchases, recycling from the previous harvest, and in-kind borrowings or gifts from relatives and neighbours, which are widespread practices in times of shortage by the rural agricultural community. Table (5) is a groupwise comparative statistics of the sources, purchased/non-purchased, the improved maize seed is obtained from.

Table 5: Purchased/Non-purchased Status of the Varieties

Adopter	Purchased Seed Plots	Total Plots
False Negative	20.52%	229
True Positive	38.62%	145

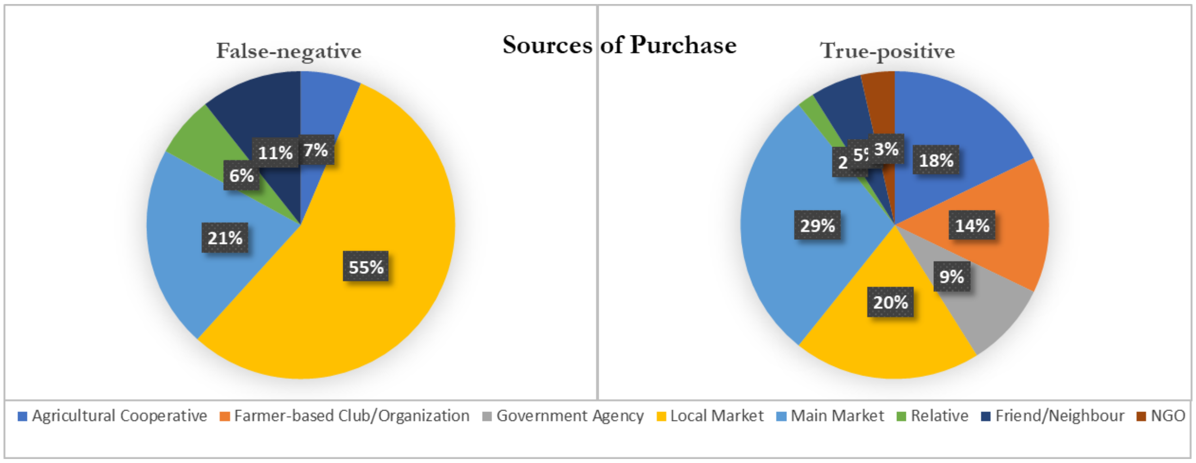
Purchased Seed Plots: plots covered with a purchased maize variety (i.e., the adopters bought the seed used) during the 2018/19 cropping season.

Table (5) reveals that only the seeds on one-fifth of the false-negative plots (i.e., around 20%) that are sourced from purchases during the 2018/19 cropping season. The remaining plots, around 80% of the group, are covered with varieties obtained through non-purchasing mechanisms. The maize seed on true-positive plots relative to false-negative plots is more acquired from purchases. The seed on 39% of the true-positive plots is purchased during the cropping season. It can be fairly assumed that getting the seed from purchases relatively eases the identification of the type of seed. The amount of money spent to buy the seed accounts a substantial portion of most farmers annual spending in developing countries. The type of the seed, improved vs landrace, is also a major determinant of the price of the seed. Thus, the adopters would better remember the identity of seed when the seed is purchased.

The observed difference from Table (5) that most false-negative plots did not get the seed from formal purchasing sources substantiates the claim that informal non-purchasing sources expose more to misclassification. In fact, purchasing the seed does not provide an equal guarantee to be free from the misclassification error since the credibility of the purchasing sources in revealing the identity of the seed is not the same. Breaking down by the specific sources from where the purchase of the purchased maize seed is made across the groups can provide further insight into the misclassification error.

Figure (3) displays the various sources of purchase of the purchased seed. Relatively agricultural cooperatives, farmer-based clubs, and government agencies are expected to be better sources in revealing the type of the seed. Others, such as traditional markets and neighbours, are more susceptible to willful deceptions that involve selling the seed, mixing the traditional with the improved. Thus, purchases from the latter group of sources are expected to expose more to the misclassification error.

Figure 3: Sources of Purchase by Group



From Figure (3), we can see that seeds purchased from sources assumed to be more credible are applied more on true-positive plots. 18% of the true-positive plots got the seed from agricultural cooperatives, a highly trusted source, compared to 7% of the false-negative plots. Government agency, another source relatively deception is expected to be less, provided 9% for true-positive plots while no false-negative plot was covered with a variety from this source. Likewise, no false-negative plot sourced maize from farmer-based club/organization, while around 14% of the true-positive plots got maize variety from this source. NGOs didn't provide seeds for false-negative plots, but they supplied around 3% of the seed for true positive plots. Main markets contributed relatively a high percentage share for both groups, but still true-positive plots got more.

The figure also reveals the other part of the story, which is relatively deceptive sources provided more maize to the false-negative plots than the true-positive plots. Local markets where local dealers play, usually with no or little regulatory checks from concerned authorities, contributed about 55% of the maize on the false-negative plots while they supplied only 20% on the true-positive plots. Friends and neighbors account for 11% of the false-negative plots, while they account only for around 5% of the true positive plots. Furthermore, while around 6% of the seed on the false-negative plots is obtained from relatives, less than 2% of the true positive plots obtained from it.

The differences observed in the sources of purchase of the seed indicate why the plots in the false-negative group came to be misclassifiers. The year of release of the varieties adopted is also another source correlated with misclassification. Keeping all other things constant, older maize varieties are expected to get misclassified as a traditional variety than recently released varieties. Table (6) presents the year of release of the maize varieties planted on the false-negative and true-positive plots.

Table 6: Varieties Adopted by Year of Release

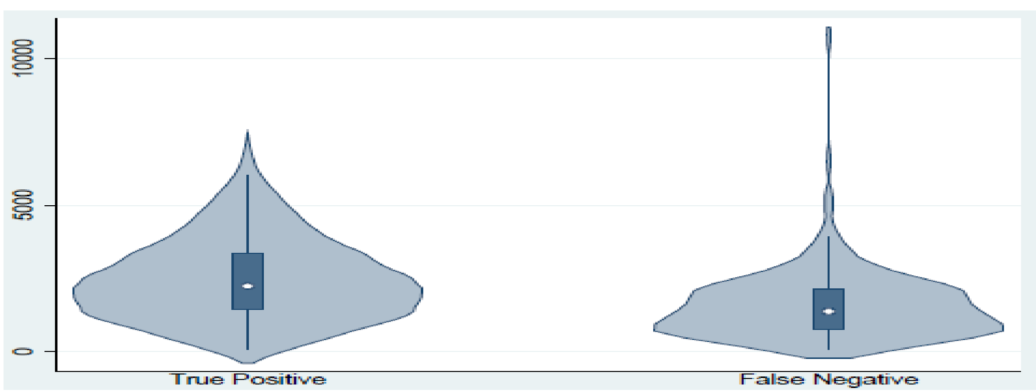
Adopter	VR ≤ 2000	2000< VR ≤ 2010	VR > 2010	Total Plots
False Negative	31%	61.57%	7.43%	229
True Positive	36.55%	32.45%	31%	145
VR, varieties released.				

Table (6), as expected, shows that older varieties are more adopted on the false-negative plots. Maize varieties released before 2010 account for 93% of the false-negative plots compared to 69% of the true-positive plots. While recent maize varieties, those released after 2010, covered only 7% of the false-negative plots, they represent 31% of the true-positive plots. The figures from Table (5) and Table (6) are in line with Jaleta et al. (2020)'s finding that misclassification is indirectly correlated to trusted sources such as cooperatives and directly related to informal sources such as recycling and the number of years since the release of the specific variety.

4.1.3 Adopters' Behavioural Adjustment

Now let's turn to investigate the input-output gaps amongst the false-negative and true-positive groups. Figure (4) shows the average maize yield per hectare by the two groups.

Figure 4: Mean Maize Yield Violin Plot, after Treatment



The Violin plot shows that the average yield per hectare by the false-negative plots (i.e., the treated group) is quite lower than that of the true-positive plots (i.e., the control group).

Table (7) present presents the Welch's t-test to check if the average yield by the two groups differs significantly.

Table 7: Test for Maize Yield Difference by the Groups, (after Treatment)

Yield	False Negative		True Positive		Mean Diff.
	Mean	SD	Mean	SD	
	1583.87	1169.07	2513.70	1428.81	-929.83***
Welch's t-test, *** significant at 1% Mean Diff., mean difference					

The average maize harvest from the false-negative plots after treatment is 1584 kilograms of maize per hectare compared to the average 2514 kilograms per hectare from the true-positive plots. The Welch's t-test conducted shows that the mean difference is statistically significant at 1%.

Table (8) presents the average differences between the groups in the inputs supplied and some other variables that are critical determinant of the yield from the plot.

Table 8: Test for Input Supply Differences by the Groups, (after Treatment)

	False Negative		True Positive		
Factor	Mean	SD	Mean	SD	Mean Diff.
Fertilizer Use	885.81	12464.10	327.34	467.10	558.47
Labour Supply	2.42	1.22	2.84	1.44	-.41***
Pesticide Use	353.54	409.44	18.78	36.18	-334.75
Seed Intensity	264.60	2598.37	64.63	187.20	199.97
Plot Size	956.98	1344.86	1617.82	1674.88	-661***
	False Negative		True Positive		
Factor	Yes	No	Yes	No	Mean Diff.
Manure	100	129	48	97	(**)
Irrigation Access	5	224	2	143	(##)
	False Negative		True Positive		
Factor	Traditional	Modern	Traditional	Modern	Mean Diff.
Plough Equipment	205	24	122	23	

Note: for pesticide, only 55% of the pesticide users considered in the calculation due to unit of measure discrepancies
Welch's t-test & Chi2 test (for the continuous & categorical variables respectively); ** 5%, *** 1%
(##), observations not enough for meaningful statistical comparison
Mean Diff., mean difference

From Table (8), the two groups after treatment are significantly different in some production inputs applied. Labour supply is negatively correlated with false-negative status. The average number of days per week the household allocated on the false-negative plots is around two and a half days compared to three days on the true-positive plots. Contrary to this, the use of manure is positively correlated with false-negative status. Around 44% of the false-negative plots used manure compared to 33% of the true-positive plots.

This gap in manure application somewhat corresponds to the finding by Wineman et al. (2020), which shows that the false-negatives' manure application was more than twice (i.e., 11% vs 26%) higher than that of true-negatives' application. This can be linked to the use of fertilizer. It might be the case that the wrong perception of the plot as false-negative switched the farmer to the use of the relatively less expensive and locally available manure than getting the relatively costly inorganic fertilizer.

Differences about the use of fertilizer and seeding intensity are unexpectedly not significantly different among the groups. The insignificant difference in fertilizer use and seeding intensity between the groups does not conform to Wineman et al. (2020)'s belief that the intensity of the use of fertilizer is higher when the adopters correctly or incorrectly identify the variety adopted as an improved variety, and the intensity of seeding of a local seed is higher than the seeding intesnity of an improved seed classifieid correctly or incorrectly. We can observe a significant difference in the conditions of the two groups of plots. While the average size of the false negative plots is 957 meters square, it is 1618 meters square for the true positive plots. Due to the very few observations both gross and groupwise about access to irrigation, it is difficult to make a meaningful comparison about this variable across the groups.

The differences in the allocation of the inputs observed above can be an indicator of the behavioural adjustment of the farmers, differential treatment of the improved seed depending on how (i.e., improved vs traditional) they perceived it.

4.2 Estimation Results

4.2.1 Selection to Treatment Model

The Common support region from the estimation of the propensity score of the bivariate selection function lies in the range of 0.06901996 to 0.9512712. Figure (5) depicts the distribution of the propensity score across the treated and the control group. It can be seen from the figure that the control observations are relatively evenly distributed across the region while the treated observations are densely populated on the right side of the score distribution.

Figure 5: Propensity Score Distribution

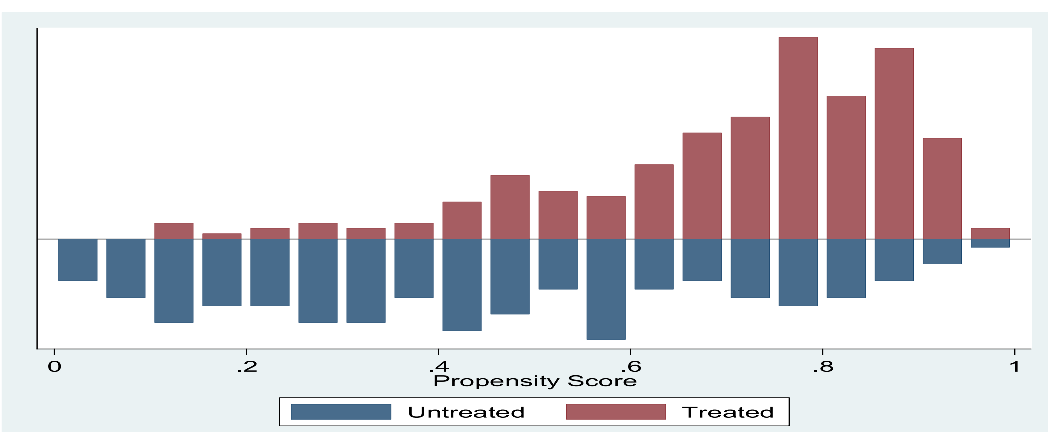


Table (9) reports the results from the estimation of the probit regression model. As can be seen, there are significant variables that cause the misclassification error. The size of the plot where the improved seed is adopted and access to advisory services are found to have a negative impact on the misclassification error. Jaleta et al. (2020) also found that the area of the plot has a negative impact on false-negative status. The age and gender of the responder are not significant correlates of the misclassification status. Likewise, the regional location and educational level (completion of primary schooling) of the adopters are not significant. Wineman et al. (2020) & Jaleta et al. (2020) also found that the age, gender, and education status of the adopters are not significant correlates of false-negative status.

Interestingly, obtaining the seed from trusted sources (cooperative, government agency, and NGOs) is found to have a negative impact on misclassification in comparison to the reference category of obtaining the seed from sources such as friends, neighbours, relatives, and recycled seed. As expected, the age of the varieties adopted has a positive impact on misclassification. Adopting varieties released before 2000 and released between 2000 and 2010 is positively correlated with misclassification error compared to using the reference category of varieties released after 2010. This confirms Jaleta et al. (2020) finding that getting the seed from trusted sources and the number of years since the release of the variety affects false-negative status negatively and positively, respectively.

Table 9: Estimation of the Probit Selection Model

Variables	Coeff.	Stand. err.	P > Z
Adopter Specific Characteristics			
Age	.003	.003	0.424
Gender (Male =1)	.130	.155	0.400
Education (Prim. school com.= 1)	.228	.281	0.417
Advisory Service	-.525***	.158	0.001
Seed from Agency	-1.416***	.365	0.000
Seed from Market	-.108	.203	0.595
Variety Released ≤ 2000	.660***	.224	0.003
Variety Released 2001 to 2010	.989***	.224	0.000
Plot Specific Characteristics			
Plot Area	-.0001***	.00005	0.004
Plots on Parcel	.029	.022	0.199
Tigray Region	.116	.405	0.774
Amhara Region	-.122	.386	0.751
Oromia Region 3	-.639	.396	0.107
SNNP Region 4	-.132	.432	0.759
_cons	.044	.421	0.916
***; significant at 1%			

4.2.2 Treatment Effect

Nearest neighbour matching with repetition is first used to estimate the treatment impact. No observation from the treatment group is discarded in the process. Other matching estimators are also employed as a test for the robustness of the result across the estimators. Table (10) summarizes the average treatment effect on the treated. Model (1) is from the use of the nearest neighbour matching. Models from 2 to 4 are from the use of radius matching, kernel matching, and stratified matching, respectively. The outcome variable is a maize yield obtained from a hectare of a plot.

Table 10: Average Treatment Effect on Treated

	(1)		(2)		(3)		(4)	
Outcome Variable	Nearest Neighbour Matching		Radius Matching		Kernel Matching		Stratified Matching	
	ATT	SE	ATT	SE	ATT	SE	ATT	SE
Yield	-628.57 (***)	273.05	-775.54 (***)	240.47	-671.42 (***)	204.29	-609.45 (***)	189.57

***, significant at 1%
ATT and SE; average treatment effect on treated, and standard error, respectively.
SEs are bootstrapped
Number of matched treated observations 224

The estimated result from the nearest neighbour matching reveals that because of the misclassification error, the average amount of maize yield per hectare by the false-negative adopters is 629 kilograms less than that of the true-positive counterfactuals. The result is robust to alternative matching methods. Using radius matching, the yield gap amounts to 776 kilograms of maize per hectare. The kernel and block stratified estimators estimated the gap to 671 kilograms and 609 kilograms per hectare, respectively. Maize plays a central role in the daily diets of households in Ethiopia. A 100kgs of maize is expected to feed a household of 4 members for more than two months in Ethiopia. The average land the farmers in Ethiopia own to grow crops is very small. According to FAO (2018), the current average land is less than one hectare, and it is expected to decline further due to the population pressure. In the presence of such factors, the misclassification borne yield loss estimated (i.e., 609 – 776 kgs per hectare) is disastrous to the food security of the maize farmers in Ethiopia.

Chapter 5 - Conclusions & Recommendations

5.1 Conclusions

The thesis aimed to investigate the loss of potential yield by improved seed adopters that is caused by misclassification of the improved seed as a landrace seed. The positive role the improved seeds play to boost farm productivity is widely accepted. This thesis argues that the performance of the improved seeds in increasing farm productivity is highly conditional on the supply of other inputs. The improved seeds need differential treatment such as greater supply of inorganic fertilizer and water than the traditional seeds. The adopter thus needs to supply the improved seeds with an amount of input optimal for their specific need. Farmers first need to correctly classify the improved seed they are cropping to differentially treat the improved seed. However, farmers in developing countries live with several challenges and practices that can expose them to misclassification error. For instance, the seed certification system is poor, seed recycling practice is dominant, and informal seed suppliers control the seed market.

Thus, in the presence of misclassification error, farmers could not meet the differential needs of the improved seeds. Therefore, there would be suboptimality in the allocation of farm inputs which are crucial for the productivity of the improved seed and the total amount of the yield. The focus of the adoption actors such as research institutes and governments has been mainly on increasing the rate of adoption of the improved seeds due to the considerable contribution of the improved seeds to increase farm production. Knowledge is totally lacking on the misclassification-borne output loss. The results presented in chapter 4 show that the specific objectives outlined in chapter 1 have been met. The main conclusions are:

I. Seed misclassification error is substantial in Ethiopia. Around 61% of the plots where improved maize varieties were adopted are found misclassified. The source where the improved seed is obtained from is a major correlate of the misclassification error. The descriptive statistics showed that the false-negative adopters got the improved seed more from informal and less-trusted sources compared to the true-positive adopters. Furthermore, on average, the varieties adopted on the false-negative plots are older compared to those on the rue-positive plots.

II. A gap in the allocation of some critical agricultural inputs is observed between the false-negative and true-positive adopters. The true-positive plots were supplied with a greater amount of inputs such as labour and manure compared to the false-negative counterparts.

This could be an indicator of the farmers' differential treatment of the improved and traditional seed.

III. Misclassification error causes yield loss. The average maize yield per hectare from the false-negative plots is significantly lower than that of the true-positive counterfactuals. The result from the various matching estimators shows that it is less in a range of 609 to 776 kgs per hectare.

Reliance only on English language literature was one constraint of the thesis. Adoption studies in other languages were not explored for this study's gap identification, the methodology employed, and comparative assessment. The relatively small sample size found available, fair to expect given DNA fingerprinting identification is costly, and Ethiopia doesn't have the technology (always sending to Australia), was another constraint.

5.2 Recommendations

The results presented indicated that misclassification error is substantial and causes loss of yield. The following recommendations are thus suggested.

I. The role of informal sources in the seed market should be reduced. Reliable seed sources such as agricultural cooperatives and government agencies should scale up their efforts in the distribution of the seeds to reach out more adopters.

II. Crop breeding institutes engaged in the production of improved crop varieties should also give attention to the provision of adequate information to the adopters about the qualities and characteristics of the improved varieties they are producing. The focus should not be only on the quality and quantity of the varieties being produced.

III. Reliance on self-reported adoption data, because of the misclassification error, can bias adoption estimates. For instance, it can bias the welfare impact of improved seed adoption. It can also mislead interventions aimed at targeting and improving the adoption of improved seeds. Accurate varietal identification methods such as DNA fingerprinting are suggested.

IV. This is the first study that investigated the causal impact of misclassification error on yield loss. Similar studies in other countries are strongly suggested to verify, amplify, and negate the findings of this study.

References

Abadi, T., Philip, D., & Moti, J. (2015). Factors that affect the adoption of improved maize varieties by smallholder farmers in Central Oromia, Ethiopia. *Developing Country Studies*, *5*(15), 50–58. http://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=20153319804&site=ehost live%5Cnhttp://www.iiste.org/Journals/index.php/DCS/article/view/25000/25603

Abebe, G., & Alemu, A. (2017). Role of Improved Seeds Towards Improving Livelihood and Food Security At Ethiopia. *International Journal of Research -GRANTHAALAYAH*, *5*(2), 338–356. https://doi.org/10.29121/granthaalayah.v5.i2.2017.1746

Abel, K. K. B., Samuel, A. D., & Michael, A. (2014). Improved rice variety adoption and its effects on farmers output in Ghana. *Journal of Development and Agricultural Economics*, *6*(6), 242–248. https://doi.org/10.5897/jdae2013.0544

AGRA. (2019). Ethiopian Farmers Adopt Hybrid Seeds. https://agra.org/ethiopian-farmers-adopt-hybrid-seeds/

Ahmed, M. H., Mesfin, H. M., Abady, S., Mesfin, W., & Kebede, A. (2016). Adoption of improved groundnut seed and its impact on rural households' welfare in eastern Ethiopia. In *Cogent Economics and Finance* (Vol. 4, Issue 1). https://doi.org/10.1080/23322039.2016.1268747

Amare, M., Asfaw, S., & Shiferaw, B. (2012). Welfare impacts of maize-pigeonpea intensification in Tanzania. *Agricultural Economics*, *43*(1), 27–43. https://doi.org/10.1111/j.1574-0862.2011.00563.x

Asfaw, S., Shiferaw, B., Simtowe, F., & Hagos, M. (2011). Agricultural technology adoption, seed access constraints and commercialization in Ethiopia. *Journal of Development and Agricultural Economics*, *3*(9), 436–447.

Bacha, D., Aboma, G., Gemeda, A., & De Groote, H. (2001). The Determinants of Fertilizer and Manure Use in Maize Production in Western Oromiya, Ethiopia. *Seventh Eastern and Southern Africa Regional Maize Conference*, *January*, 438–441.

Barta, P. (2007, July 28). Feeding Billiona, A Grain at a Time: as development and climate change imperil rice yields, scientists seek new Green Revolution. *The Wall Street Journal*. https://www.wsj.com/articles/SB118556810848880619

Bruinsma, J. (2017). World agriculture: Towards 2015/2030: An FAO Study. *World Agriculture: Towards 2015/2030: An FAO Study*, 1–431. https://doi.org/10.4324/9781315083858

Data, S.K., Tauro, A. C., & Balaoing, S. N. (1968). Effect of Plant Type and Nitrogen Level on the Growth Characteristics and Grain Yield of Indica Rice in the Tropics. Agronomy Journal. https://web.archive.org/web/20081202142838/http:/agron.scijournals.org/cgi/content/abstract/60/6/643

Eliazer Nelson, A. R. L., Ravichandran, K., & Antony, U. (2019). The impact of the Green Revolution on indigenous crops of India. *Journal of Ethnic Foods*, *6*(1), 1–10. https://doi.org/10.1186/s42779-019-0011-9

FAO. (2018). *Ethiopia: small family farms country factsheet*. 1–2. www.fao.org/family-farming/data-sources/dataportrait/farm-size/en

Floro, V. O., Labarta, R. A., Becerra López-Lavalle, L. A., Martinez, J. M., & Ovalle, T. M. (2018). Household Determinants of the Adoption of Improved Cassava Varieties using DNA Fingerprinting to Identify Varieties in Farmer Fields: A Case Study in Colombia. *Journal of Agricultural Economics*, *69*(2), 518–536. https://doi.org/10.1111/1477-9552.12247

Fufa, B., & Hassan, R. M. (2006). Determinants of fertilizer use on maize in Eastern Ethiopia: A weighted endogenous sampling analysis of the extent and intensity of adoption. *Agrekon*, *45*(1), 38–49. https://doi.org/10.1080/03031853.2006.9523732

Hasen Ahmed, M., Mamo Geleta, K., Tazeze, A., Mesfin, H. M., & Tilahun, E. A. (n.d.). Cropping systems diversification, improved seed, manure and inorganic fertilizer adoption by maize producers of eastern Ethiopia. *Journal of Economic Structures*. https://doi.org/10.1186/s40008-017-0093-8

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, *64*(4), 605–654. https://doi.org/10.2307/2971733

Ifpri. (2010). Maize Value Chain Potential in Ethiopia. *International Food Policy Research Institute*, *April 2010*, 1–42.

Igbozurike, U. M. (1978). Polyculture and monoculture: Contrast and analysis. *GeoJournal*, *2*(5), 443–449. https://doi.org/10.1007/BF00156222

Jaleta, M., Tesfaye, K., Kilian, A., Yirga, C., Habte, E., Beyene, H., Abeyo, B., Badebo, A., & Erenstein, O. (2020). Misidentification by farmers of the crop varieties they grow: Lessons from DNA fingerprinting of wheat in Ethiopia. *PLoS ONE*, *15*(7 July), 1–16. https://doi.org/10.1371/journal.pone.0235484

Kaliba, A. R. M., Verkuijl, H., & Mwangi, W. (2000). Factors Affecting Adoption of Improved Maize Seeds and Use of Inorganic Fertilizer for Maize Production in the Intermediate and Lowland Zones of Tanzania. *Journal of Agricultural and Applied Economics*, *32*(1), 35–47. https://doi.org/10.1017/s1074070800027802

Kilian, A., Wenzl, P., Huttner. E, Carling, J., Xia, L., Blois, H., Caig, V., Uszynska, K., Jaccoud, D., Hopper, C., Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., and Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In: Walker, J. M. (ed.) *Data production and analysis in population genomics*. Methods and protocols. In *Life Sciences* (Vol. 531, Issue 1). http://books.google.com/books?id=Ku2wPAAACAAJ

Kosmowski, F., Alemu, S., Mallia, P., Stevenson, J., & Macours, K. (2020). *Shining a Brighter Light: Comprehensive Evidence on Adoption and Diffusion of CGIAR-related Innovations in Ethiopia SPIA Synthesis*. *October*. https://cas.cgiar.org/spia

Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., Yigezu, B., & Stevenson, J. (2019). Varietal identification in household surveys: Results from three household-based methods against the benchmark of DNA fingerprinting in southern Ethiopia. *Experimental Agriculture*, *55*(3), 371–385. https://doi.org/10.1017/S0014479718000030

Lipper, L., Cavatassi, R., Winters, P. C., & Winters, P. (2005). Seed Systems, Household Welfare and Crop Genetic Diversity: An Economic Methodology applied in Ethiopia Seed Systems, Household Welfare and Crop Genetic Diversity: An Applied Economic Methodology. *Organization*.

Maruod, M. E., Breima, E. E., Elkhidir, E. E., & Naim, A. M. El. (2013). *Impact of Improved Seeds on Small Farmers Productivity, Income and Livelihood in Umruwaba Locality of North*. *3*(6), 203–208. https://doi.org/10.5923/j.ijaf.20130306.01

Meughoyi, C.T. (2018). Improved Seeds and Agricultural Productivity of Family Farms in Cameroon. In: Shimeles, A., Verdier-Chouchane, A., & Boly, A. (ed.) *Building a Resilient and Sustainable Agriculture in Sub-Saharan Africa*. https://doi.org/10.1007/978-3-319-76222-7

Mulesa, T. H., Dalle, S. P., Makate, C., Haug, R., & Westengen, O. T. (2021). Pluralistic Seed System Development: A Path to Seed Security? *Agronomy*, *11*(2), 372. https://doi.org/10.3390/agronomy11020372

Mugisha, J., & Diiro, G. (2010). Explaining the Adoption of Improved Maize Varieties and its Effects on Yields among Smallholder Maize Farmers in Eastern and Central Uganda. *Middle-East Journal of Scientific Research*, *5*(1), 6–13.

Negi, G. C. S. (1994). High yielding vs. traditional crop varieties: A socio-agronomic study in a Himalayan village in India. *Mtn Res.Dev.*, *14*(3), 251–254. https://doi.org/10.2307/3673776

Neill, M. O. (2018). *Tilling the Earth: Improved Seed Use in Ethiopia and Brazil*. *9*(1), 29–44

Ouma, J. O., & De Groote, H. (2011). Determinants of improved maize seed and fertilizer adoption in Kenya. *Journal of Development and Agricultural Economics*, *3*(11), 529–536. http://www.academicjournals.org/JDAE

Pingali, P. L. (2012). Green revolution: Impacts, limits, andthe path ahead. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12302–12308. https://doi.org/10.1073/pnas.0912953109

Press, C. (1996). *Adoption of HYV Technology in Asian Countries: The Role of Concessionary Credit Revisited Author (s): Gamini Herath and Sisira Jayasuriya Published by: University of California Press Stable URL: https://www.jstor.org/stable/2645574 All use subject to*. *36*(12), 1184–1200.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Sánchez-Toledano, B. I., Kallas, Z., Rojas, O. P., & Gil, J. M. (2018). Determinant factors of the adoption of improved maize seeds in Southern Mexico: A survival analysis approach. *Sustainability (Switzerland)*, *10*(10). https://doi.org/10.3390/su10103543

Seyoum Taffesse, A., Dorosh, P., & Gemessa, S. A. (2013). Crop production in Ethiopia: Regional patterns and trends. *Food and Agriculture in Ethiopia: Progress and Policy Challenges*, *9780812208*, 53–83. https://doi.org/10.9783/9780812208610.53

Shiferaw, B., Kassie, M., Jaleta, M., & Yirga, C. (2014). Adoption of improved wheat varieties and impacts on household food security in Ethiopia. *Food Policy*, *44*, 272–284. https://doi.org/10.1016/j.foodpol.2013.09.012

Shiva, Vandana. The Violence of the Green Revolution: Third World Agriculture, Ecology, and Politics. University Press of Kentucky, 2016. JSTOR, www.jstor.org/stable/j.ctt19dzdcp. Accessed 14 May 2021.

Specter, M. (2014, Aug 18). Seeds of Doubt: An activists' controversial crusade against genetically modified crops. *The New Yorker* https://www.newyorker.com/magazine/2014/08/25/seeds-of-doubt

Tesfaye, S., Bedada, B., & Mesay, Y. (2016). Impact of improved wheat technology adoption on productivity and income in Ethiopia. *African Crop Science Journal*, *24*(1), 127. https://doi.org/10.4314/acsj.v24i1.14s

The Beginning of the Green Revolution (n.d.). https://web.archive.org/web/20041227090100/http://www.coafes.umn.edu/The_Beginning_of_the_Green_Revolution.html

Tura, M., Aredo, D., Tsegaye, W., La Rovere, R., Tesfahun, G., Mwangi, W., & Mwabu, G. (2010). Adoption and continued use of improved maize seeds: Case study of Central Ethiopia. *African Journal of Agricultural Research*, *5*(17), 2350–2358.

Urgessa, T. B. (2014). Review of challenges and prospects of agricultural production and productivity in Ethiopia. *Journal of Natural Sciences Research*, *4*(18), 70–77. http://www.iiste.org/Journals/index.php/JNSR/article/view/15857/16354

Wainaina, P., Tongruksawattana, S., & Qaim, M. (2014). Improved Seeds, Fertilizer or Natural Resource Management? Evidence From Kenya'S Smallholder Maize Farmers. *Food and Rural Inno Vations for Healthier Societies*, *August*,1–14.

Wineman, A., Njagi, T., Anderson, C. L., Reynolds, T. W., Alia, D. Y., Wainaina, P., Njue, E., Biscaye, P., & Ayieko, M. W. (2020). A Case of Mistaken Identity? Measuring Rates of Improved Seed Adoption in Tanzania Using DNA Fingerprinting. *Journal of Agricultural Economics*, *71*(3), 719–741. https://doi.org/10.1111/1477-9552.12368

Workneh, W. A., Ujiie, K., & Matsushita, S. (2021). Farmers' agricultural tractor preferences in Ethiopia: a choice experiment approach. *Discover Sustainability*, *2*(1). https://doi.org/10.1007/s43621-021-00021-2

Wossen, T., Abdoulaye, T., Alene, A., Nguimkeu, P., Feleke, S., Rabbi, I. Y., Haile, M. G., & Manyong, V. (2019). Estimating the productivity impacts of technology adoption in the presence of misclassification. *American Journal of Agricultural Economics*, *101*(1), 1–16. https://doi.org/10.1093/ajae/aay017

Wossen, T., Alene, A., Abdoulaye, T., Feleke, S., & Manyong, V. (2019). Agricultural technology adoption and household welfare: Measurement and evidence. *Food Policy*, *87*(July 2019), 101742. https://doi.org/10.1016/j.foodpol.2019.101742

Zeng, D., Alwang, J., Norton, G., Jaleta, M., Shiferaw, B., & Yirga, C. (2018). Land ownership and technology adoption revisited: Improved maize varieties in Ethiopia. *Land Use Policy*, *72*(October 2017), 270–279. https://doi.org/10.1016/j.landusepol.2017.12.047