Climate Change, Structural Change, and Innovation

Hyungju (Henry) Kim

A Dissertation
submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Graduate Program in Economics
York University
Toronto, Ontario

August 2022

Abstract

In this dissertation, I study the sources of economic growth and structural transformation by combining a quantitative framework with micro data. As the main theme of the dissertation, I focus on the impact of climate change on the agricultural sector of developing countries and spatially analyze its impact on the agricultural productivity and long-run distribution of farming activity. Then, I move onto the non-agricultural sector of a developed country by analyzing the source of economic growth through product innovation.

In **chapter 1**, I study the impact of climate change on farm size and agricultural productivity in developing countries. I combine pixel-level climate data from the Global Agro-Ecological Zones (GAEZ) project with rich household-level data on agricultural production from the World Bank's LSMS for Uganda. I assess the implications of anticipated climate change on aggregate productivity through the lens of a two-sector model with endogenous farm size and crop choice, featuring heterogeneity in both land quality and farmer ability. Calibrating the model to current conditions, I dis-entangle farmer ability from land quality at the farm-level. Feeding in projected land quality for the climate change to the year 2050, I find that agricultural productivity would increase 23% and farm size would increase 27%, with a switch towards food crops. While the direct effect of climate change accounts for nearly half of these changes, the other half is due to the indirect effects on structural change induced by the climate shock. The effect of climate change however is not uniform within Uganda, implying a widening of regional productivity differences.

In **chapter 2**, I study the interaction of climate change with transport frictions to determine the long-run distribution of farming activity in Ethiopia. I assess the implications of anticipated climate change through the lens of a spatial model featuring locational heterogeneity in land quality, transportation costs, and agricultural production, by incorporating a quantitative spatial framework to rich micro-data combined from different sources. I calibrate the model to the benchmark economy with baseline (1961-1990) geographical conditions, then change geographical conditions alone to the future level (year 2050). Findings suggest that the overall increase in land quality data for Ethiopia lead to higher agricultural production, but is associated with larger inequality between districts. Districts that had higher initial yield and transportation cost experienced larger percentage gains in land quality and yields.

Where does the product innovation come from? From entering plants or incumbents? From existing products or brand new products? In **chapter 3**, I answer these questions by quantifying the sources of innovation in South Korea over the years 2001-2011. To account for the sources of innovation, I combine unique Korean data on the universe of non-farm private sector establishments and the growth framework of Garcia-Macia et al. (2019), which infers the sources of innovation from job creation and job reallocation flows among incumbent and entrant firms. I find that over 4/5 of the innovation in Korea is accounted for by incumbents, in particular through their own variety improvements. I also find that the incumbent innovation in US is of smaller overall magnitude (2/3). Among entrants, most of the innovation comes through creative destruction in Korea, while through new varieties in US.

Acknowledgements

I wouldn't be where I am if it wasn't for my supervisor Professor Tasso Adamopoulos. I owe much of my growth and development over the last five years to his extraordinary supervision. He would always make time to enlighten and encourage me, even when I knew his schedule was tight. With every meeting, he has allowed me to leave with clearer visions and perspectives on my next steps as a researcher and as an economist by providing effective problem-solving advice whenever I needed it. Over the years, he has come to be more than my supervisor; he is my life mentor. Thank you for everything sir, I will forever be grateful.

I am indebted for enormous support from my committee members Professor Chaoran Chen and Professor Matias Cortes. Their insightful comments and advice most definitely pushed me further forward beyond my reach. From research analysis to presenting my paper in front of an audience - you have expanded and enriched my views on every aspect. Thank you for stimulating me to progress and thrive throughout this process.

I would like to acknowledge Professor Ahmet Akyol and Professor Wai-ming Ho for their help in preparing for my job market interviews; Professor Nippe Lagerlof, Professor Laura Salisbury and Professor Uros Petronijevic for every valuable comments throughout the Econ7000 seminar. I would also like to thank Dr. Pyo Han Hyung for his unconditional help in acquisition of the Korean data in Chapter 1 of the thesis. I truly appreciate all the administrative assistance from Regina, Lisa, Rex and Priyanka. I am thankful to the macro development group, including Nicole for all the helpful comments on my papers. Thank you everyone for supporting me through the biggest journey of my life yet.

Contents

List of Tables

List of Figures

Chapter 1

Climate Change, Farm Size and Agricultural Productivity

1.1 Introduction

There is mounting evidence of climate change around the globe, with increases in mean temperatures, CO_2 emissions, and extreme weather phenomena, such as heat waves, heavy rain, droughts, and tropical cyclones, which are expected to rise further over the next decades (IPCC, 2021). Two key questions are, what is the economic impact of climate change and how will economies adapt to it? These questions are of particular interest given the evidence linking temperatures to aggregate economic performance (Nordhaus, 2006; Sachs, 2003; Rodrik et al., 2004; Acemoglu et al., 2002; Dell et al., 2009, 2012, 2014).

At the micro-level, temperature has been linked to the allocation of time (Graff Zivin and Neidell, 2014), crop yields (Schlenker and Roberts, 2009), human conflict (Hsiang et al., 2013), mortality (Barreca et al., 2016), farm input adjustments (Aragón et al., 2021), sectoral labour re-allocations (Colmer, 2021), and internal migration (Feng et al., 2012), among others. At the macro-level, climate has been incorporated in quantitative micro-founded general equilibrium models, allowing for welfare analysis and policy evaluation (Golosov et al., 2014; Hassler and Krusell, 2018; Hassler et al., 2016).

Perhaps no other sector is impacted more directly by climate change than the agricultural sector that uses location-specific inputs, a big part of which are climatic conditions such as temperature and moisture. The effect of climate change on agricultural productivity is particularly important for lower income countries given that they tend to have lower productivity than developed economies, and agriculture accounts for a large part of their economy in terms of output, employment, and exports (Gollin et al., 2002; Restuccia et al., 2008; Caselli, 2005; Blair et al., 2005). Agricultural productivity is important not only for food security and welfare but also for the process of structural change.

In this paper, I study the effect of climate change on aggregate agricultural productivity and structural change in developing countries. To conduct my research, I combine high resolution geographic data and farm-level production data for Uganda, and I build on a general-equilibrium two-sector model of Adamopoulos and Restuccia (2014) featuring endogenous farm size by including land quality and crop choice. Using the micro data, and without using distributional assumptions, I am able to decompose farm productivity into a pure land quality component and a residual farmer productivity component. Climate change, as expected by natural scientists, manifests itself in the model through a change in land quality at the farm-level. Through the lens of the model I am then able to study the endogenous adaptation of farmers in terms of the size of their farm and the crops they produce.

There are two main contributions of this paper: one is from the model, the land quality parameter in the production function is separated from the unobserved residual of farmer ability and its changes represent climate shocks in the model. Another is from the data, given that climate and its projected change vary within countries and interacts with farmer heterogeneity, it is important to use micro-level data and a micro-level framework in order to assess the aggregate impact of climate change. Therefore, I exploit extremely rich micro-level data on agricultural productivity and harvested land size by combining data from two of the following sources: Global Agro-Ecological Zones (GAEZ), and Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA). Supported by the spatially combined rich micro-data from GAEZ and LSMS for each parcel of land identified by the GPS coordinates, I use a parsimonious two-sector economy model with crop selection to identify how land quality governed by the climate change impacts aggregate productivity.

High resolution geographic data I use in this research is from Global Agro-Ecological Zones (GAEZ). GAEZ divides the world into 5-arc minute resolution pixels and for each cell, provides potential yields which are projected crop outputs even when the crops are not actually produced within the cell. After choosing water supply and cultivation inputs, potential yield for each crop is generated by combining pixel specific land quality attributes with crop specific agronomic models. Therefore, using the GAEZ potential yield, land quality attributes representing the impact of climate change can be analyzed at a level much finer than just countries or provinces. When estimating future potential yields, GAEZ accounts for CO_2 concentrations. This is important because plants use CO_2 for photosynthesis, possibly facilitating the plant growth process and raising yields. Thus CO_2 fertilization may offset the negative effects of rising temperatures.

My research share a common ground with Costinot et al. (2016) in a sense that rich micro-level data from GAEZ is implemented into a general-equilibrium model to assess the impact of climate change on aggregate agricultural production. However, Costinot et al. (2016) focus on how changes in spatial frictions or location specific productivity shocks, such as climate change, can globally affect aggregate productivity in equilibrium models of Ricardian trade. I study the effect of climate change at the micro-level on farm size and farmer crop choices within a developing country.

An important aspect of this research is that I focus on Uganda. This paper is not about the effect of climate change on world agricultural production. Rather it studies the effects of climate change within a specific developing country, taking into account spatial heterogeneity and the endogenous responses of farmers on what they produce and the size of their farm. There are three main reasons why Uganda is the main focus of my counter-factual experiments. First, the agricultural sector in Uganda accounts for 70% of the work force and 25% of GDP. Second, the Living Standards Measurement Survey - Integrated Survey on Agriculture (LSMS-ISA), in collaboration with Uganda National Panel Survey (UNPS), provides very detailed farm-level micro-data on Uganda's agricultural sector. Lastly, the mean annual temperature of Uganda in year 2000 is at the high end of optimal growth temperature for most crops (23 °C) and the difference of mean annual temperature between the hottest and the coldest plot of land in Uganda is around 21.5 °C. This range is very high compared to other LSMS-ISA Sub-Saharan African countries such as Burkina Faso (3.8 °C), Mali (4.7 °C), Niger (9 °C), and Nigeria (11 °C). Considering the fact that the temperature of soil is a significant parameter when it comes to farming and most of the staple crops have an optimal growth temperature of 15 - 25 °C, high mean annual temperature and high temperature range will lead to explicit results between the locations within Uganda when conducting an experiment surrounding spatial heterogeneity in presence of climate change.

I find that overall, climate change will raise aggregate agricultural productivity and will induce further structural change with larger farm sizes and a reallocation of labour from agriculture to non-agriculture, and an increase in agricultural productivity. This is driven, partly by the direct – overall positive – impact of climate change on crop yields in Uganda, and partly through the endogenous adjustments that farmers make, with a shift towards larger farms and an increase in cash crop production. The overall positive effect of climate shock on aggregate agricultural productivity however, shows a non-monotonic trend overtime. I observe a larger increase in agricultural productivity from the benchmark period to year 2050, when compared to the benchmark period to year 2080, indicating a decrease in productivity from year 2050 onwards. These results are consistent with the agronomic literature that finds non-linear effects of temperature on crop yields, whereby yields gradually increase up to a threshold of temperature, and fall sharply beyond that (Schlenker and Roberts, 2009). Similarly, Burke et al. (2015) also find non-linearities in the relationship between temperature and economic production, including agriculture. My model suggests that the climate change is going to speed up the process of structural transformation in the near future but will reverse over time as land quality shows a steep decrease further into the future. Most importantly, these effects are not uniform across Uganda and shows spatial variation in the effect of temperature on crop yields. There are some regions, with currently higher temperatures that will experience a drop in their agricultural productivity. I find that in the short-run, warming temperatures will be a net gain in Uganda's agricultural production. This however, confounds within country heterogeneity which would see some regions benefit while others suffer from warming temperatures.

The rest of the paper proceeds as follows. Section II illustrates the background information of Uganda including regional average temperature heterogeneity. Section III explains two of the main data sources used in the paper and the process of combining them to create rich-micro data used for experiments. Section IV describes a simple two-sector model of agriculture and non-agriculture featuring land quality. Section V presents the details on the calibration of the model. Section VI analyzes the results for experiments surrounding climate change scenarios. Section VII breakdown the channels that generate overall quantitative effects. Section VIII concludes.

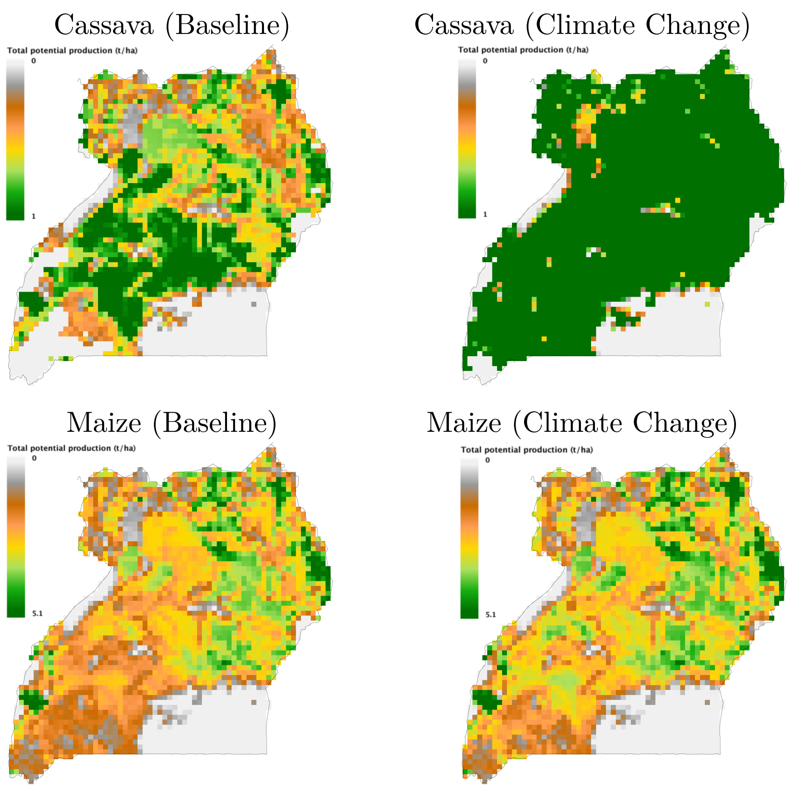
1.2 Background

Uganda is a landlocked country in the Great Lakes region of East Africa. Uganda has diverse landscapes encompassing snow-capped mountains. According to the World Bank in 2018, Uganda is approximately 241,038 km^2 in size with 34.4% of arable land and around 72% of the total land dedicated to agricultural production. The population of Uganda is 45.74 million with around 70% of employment in the agricultural sector.

1.2.1 Farming in Uganda

Figure 1.1 illustrates the evidence of spatial and crop type heterogeneity by comparing the potential yields of each plot within Uganda before and after the climate change. From Figure 1.1, the climate change scenario provided by GAEZ will lead to improvements in yields for some crop, like cassava, but hardly any effect for others like maize. With maize, potential yields generally improved after climate change, but there were some pixels that got worse. While these are the anticipated changes in potential yields by meteorologists and climatologists, I am interested in their implications for farm structure, structural change, and aggregate productivity.

Figure 1.1: Effect of Climate Change on Crop Yields in Uganda



Notes: The baseline period reflects average climatic conditions for the period 1961-1990. Climate change scenarios are based on year 2050 Hadley A1FI with CO_2 fertilization. Mapping data from GAEZ are at the 5-arc minute resolution.

There are four regions in Uganda and as shown from Table 1.1, each regions are distinguished by clear differences in crops being produced. Central region which includes the capital city of Kampala, mainly produces plantains, maize, and roots, but does not really specialize in producing any specific crop compared to other regions. Eastern region produces most millet, maize, rice, sweet potatoes, and cassava, especially the productivity (t/ha) of rice and maize is extremely high compared to other regions. Northern region has the highest mean annual temperature in Uganda and produces most of sorghum, field peas, pigeon peas, ground nuts, soya beans, and simsim. Almost all of the pigeon peas production in Uganda are from the north and have very low productivity in millet. Western region has the lowest mean annual temperature and produces 67% of total plantain and almost 88% of the irish potatoes which has lower optimal growth temperature.

Table 1.1: Crop Area and Production by Region, UCA 2008/09

Region	Plantain	Millet	Maize	Sorghum	Rice	Sweet Potatoes	Irish Potatoes
Area (Ha)							
Central (21.7 °C)	326,082	5,832	189,135	2,261	2,637	98,054	4,798
Eastern (22.8 °C)	69,504	86,911	388,762	101,645	36,033	159,948	1,271
Northern (23.9 °C)	9,195	105,656	247,780	249,330	25,912	60,573	594
Western (20.6 °C)	511,096	51,588	188,583	46,016	10,504	121,681	26,096
Production (t)							
Central (21.7 °C)	1,039,837	13,734	449,859	2,678	2,173	312,402	13,290
Eastern (22.8 °C)	342,234	106,838	1,108,554	133,313	128,195	847,140	4,624
Northern (23.9 °C)	31,626	78,572	305,798	177,088	43,719	292,932	1,311
Western (20.6 °C)	2,883,648	77,784	497,745	62,716	16,649	366,295	135,210

Region	Cassava	Beans	Field peas	Cow peas	Pigeon peas	Ground nuts	Soya bean	Simsim
Area (Ha)								
Central (21.7 °C)	127,788	120,798	470	1,135	0	26,504	750	590
Eastern (22.8 °C)	342,387	108,107	8,014	12,976	876	122,404	7,279	15,316
Northern (23.9 °C)	269,886	146,702	29,067	9,352	28,786	136,893	26,195	158,763
Western (20.6 °C)	131,328	241,915	6,286	354	139	59,431	2,220	928
Production (t)								
Central (21.7 °C)	409,812	167,276	302	281	0	32,757	208	127
Eastern (22.8 °C)	1,061,186	98,834	3,233	7,086	219	77,247	5,801	6,774
Northern (23.9 °C)	983,124	251,221	10,428	3,429	11,031	83,182	15,727	93,562
Western (20.6 °C)	440,189	411,945	2,489	261	80	51,497	1,887	565

Source: UBOS, Statistical Abstract 2013; Uganda Census of Agriculture (UCA) 2008/09.
Note: Mean annual temperature in each region (right of the region name) is the average for the baseline period (1961-1990).

Government of Uganda acknowledges the possible challenges that would arise from climate change in the future, thus in 2015, Uganda launched a National Climate Change Policy (NCCP) to cope with climate change. NCCP seeks to reduce carbon emissions and pollution, but at the same time work on mitigating the damage and adapting to the changing climate. In addition to the fact that each regions specialize in different crops, results reported later from this research suggests that segmented policy implications for different regions can be a key feature in developing a coping strategy for a climate shock.

1.2.2 Regional Heterogeneity

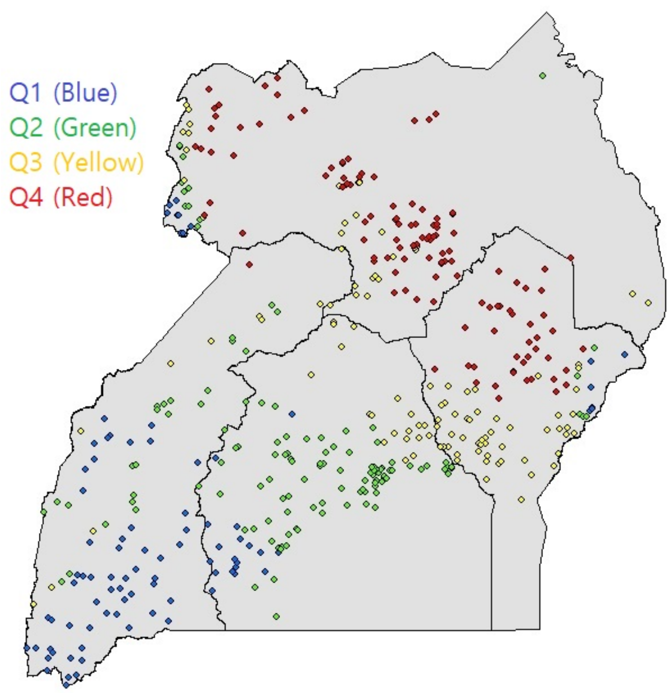
The key to coping with climate change in Uganda lies in spatial heterogeneity. In order to test spatial heterogeneity for the impact of climate change in Uganda, LSMS sample farms are divided into four different quantiles based on their mean annual temperature. Table 1.2 shows summary statistics of these quartiles. Mean annual temperature of a farm rises as they move up in number, so farms in Q4 are those that are in the hottest region within Uganda. Therefore, by comparing the simulation results between different quartile groups, I will analyze how the impact of climate change differs between the regions within Uganda based on their current temperature.

Table 1.2: Mean Annual Temperature of each Quartiles

	# of Farms	Mean Temperature	Temperature Range
Q1	546	19.1°C	13.79°C ~ 20.81°C
Q2	547	21.5°C	20.82°C ~ 22.21°C
Q3	545	22.7°C	22.22°C ~ 23.21°C
Q4	556	23.7°C	23.22°C ~ 25.87°C

Source: Q1, Q2, Q3, Q4 each describes quantiles divided by their mean annual temperature in year 2000 (data provided by GAEZ). Those that are in the upper quantile (Q3, Q4) are farms that are in hotter regions within Uganda.

Figure 1.2: Regional Average Temperature Heterogeneity within Uganda



Note: Quantiles are divided using mean annual temperature from GAEZ. Each of the dots represent sample farm operating households from Uganda National Panel Survey (UNPS) conducted by Living Standard Measurement Survey (LSMS).

Each of the dots in Figure 1.2 represent individual farms in Uganda as in the sample data summarized in Table 1.2. Temperature range between highest and lowest farm is not as large as the whole country's 21.5°C, but still shows a high range of around 12°C. From Figure 1.2, we can observe clustering patterns for temperature heterogeneity: high mean annual temperature farms are located at the north-eastern region and lower temperature farms are located at the south-western region. This pattern opens up a possibility for an experiment to see whether the impact of climate change in agriculture is significantly different between the clustered regions.

1.3 Data

Micro data used in the analysis comes from two different sources: Global Agro Ecological Zones (GAEZ) and Living Standards Measurements Survey - Integrated Surveys on Agriculture (LSMS-ISA). GAEZ is a detailed assessment on the Agro-Ecological Zones developed by Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA). Focus of the GAEZ project is to provide information on optimal crop choice in any given location on earth, that is, to help farmers to know how productive they would be at crops they are not currently growing.

In order to achieve this goal, GAEZ divides the world into just over 2.2 million pixels of land where each of the 5 arc-minute cells (pixels) are 100 km^2 along the equator, and is slightly smaller when further away due to Mercator projection. Each of the pixels are identified by GPS coordinates and provides potential yields for each crop. Potential yields are calculated as maximum possible output per hectare that can be attained given the crop's production requirements and cell-specific land quality characteristics, and it includes crops that are not actually produced in a particular pixel. GAEZ is a standardized framework for the characterization of climate, soil, and terrain conditions relevant for agricultural production because the yield estimates are computed after controlling each of the field characteristics (such as soil types and elevation) and climatic variables (such as rainfall, temperature, humidity, wind speed, sun exposure, etc). In addition to the natural aspect in which farmers have relatively little control, farmers' decisions about how to grow their crops with different levels of complementary inputs such as fertilizers, machinery and labour, and types of water supply (rain-fed, irrigated) are considered in the calculation of potential yields.

One important aspect of GAEZ potential yields is the fact that it provides data on future predictions of the potential yields. There are four main scenarios (A1, A2, B1, B2) of future predictions on pixel data that corresponds to a different "narrative storyline". Each "storyline" differs on potential evolution of demographic, social, economic, technological, and environmental variables and incorporates these changes in calculating the soil and climatic conditions for the future. For example, let's assume that there are two different scenarios A and B. In A, the use of fossil-fuels greatly increase and climate change is intensified, in B, renewable energy and sustainable technology is widely used causing climate change to slow down by a significant margin. Then, A and B would have different climatic conditions and GAEZ calculates potential yields based on these different scenarios.

The LSMS-ISA project is a statistical survey conducted at the household-level, organized by the World Bank's Development Data Group. LSMS-ISA collaborates with national statistical offices in Sub-Saharan African countries to provide nationally-representative panel survey with a strong focus on agriculture. Each observation in the sample is identified by the Household IDs (HHID) and provides extremely detailed data on agriculture such as farm size, output, input, etc. In the case of Uganda, the LSMS-ISA implements the agricultural content of the Uganda National Panel Survey (UNPS) to create a panel data set.

I link the two data sets spatially, using the GPS coordinates of households in the LSMS with the GPS coordinates of pixels in GAEZ. This allows for a combination of household-level data from LSMS with the geographic data of GAEZ to create an extremely rich micro data set necessary for analysing the impact of climate change on agricultural productivity in the model. To combine the data, I separately compute the output yield (y) and harvested land (ℓ) for each chosen crops in LSMS. Then, I use GPS coordinates with ArcGIS to combine potential yield (z) from GAEZ with the LSMS data on actual yields and harvested land for each crop type.

There are two different crop types in the model, so Table A.1 in Appendix A shows a summary of the crops used in this research after dividing them broadly into cash or food crops. For uncertain crops, their type is decided based on the consumption trend in Uganda; for example, groundnuts and soybeans used to be a staple food crop in Uganda, but has been widely regarded as a cash crop by FAO prior to year 2009, so they are classified as a cash crop. In deciding what crops to be used in the research, the major cash crops (banana, coffee, and tea) have been added to the 18 main crops listed in Adamopoulos and Restuccia (2018)[1]. Although cocoa is one of the main cash crops, it was omitted because GAEZ doesn't provide data on future potential yield of Cocoa for most future scenarios. As for bananas, UNPS (LSMS) divides banana into three different divisions based on their uses: food, beer, sweet; however, GAEZ only reports a single price, so I have used the same price for all banana products.

Crop potential yields from GAEZ are measured in tonnes/ha, while yield units for LSMS are in kilograms. I use FAO prices by crop in Geary-Khamis dollars (GK\$/t) to compete real output under both data sets[2]. Potential yield data from GAEZ based on average historical climate conditions on years 1961-1990 (scenario Hadley CM3 A1FI) is used as the proxy for land quality in the benchmark economy, as this would help to resolve year-to-year idiosyncratic weather shocks. For the future reference periods, I use main scenarios (A1, A2, B1, B2) on years 2050 and 2080 provided by GAEZ to analyze the impact of climate change on agricultural productivity[3].

1.4 Model

This is a two-sector model of agriculture and non-agriculture featuring heterogeneity in both land quality and farmer ability. There are two types of goods being produced: agricultural (a) and non-agricultural (n) goods. The agricultural good (a) is then divided into two crop types: cash crops (c) and food crops (f). In the model, benchmark economy is initially endowed with a fixed amount of total farm

[1] 18 main crops are: wheat, rice, maize, sorghum, millet, other cereals (barley, rye, oat and other minor cereals), tubers (white potato, sweet potato), roots (cassava, yam and cocoyam), sugarcane, sugarbeets, pulses (chickpea, cowpea, dry pea, grams, pigeon pea), soybean, sunflower, rapeseed, groundnut, oilpalm, olive, and cotton. In year 2000, these crops cover 83 percent of the world production in terms of acreage, and 60 percent in terms of value of production.

[2] Price (GK\$/t) is taken from p.48 of GAEZ user guide, and average price is used for crops that provide range of prices (ex: pulses 235-500). For cotton, the average price of cotton seed and lint was used.

[3] LSMS data on actual yields and size of harvested lands are based on UNPS year 2009 survey; although the years do not align between LSMS and GAEZ, it would not pose too much problem because the main focus of this paper is to study the effect of future climate scenarios

land L and is populated by a continuum of household members of measure 1, all of whom are endowed with one unit of time inelastically supplied to the labour market. Household heads then decide on what fraction of their members work as cash crop farmers, food crop farmers, or non-agricultural workers in order to maximize the utility subject to a household-level budget constraint. There are two notable aspects of this model: first, the land quality variable z[4] is distinct from the unobserved heterogeneity of farmer ability. Second, instead of making distributional assumptions on the farmer ability and land quality, I use actual data from two different sources (GAEZ and LSMS) to infer the farmer ability from the model, conditional on observed land quality.

1.4.1 Technology

Non-agricultural sector Non-agricultural production function is very simple with its good being produced by a stand-in firm with constant returns to scale (CRS) technology that only requires labour N_n as an input.

$$Y_n = A_n N_n, \tag{1}$$

Y_n is the total amount of non-agricultural output produced and N_n is the total amount of labour input in non-agriculture, where A_n is a non-agricultural sector specific productivity parameter.

Households, firms, and farms behave competitively in factor/output markets and non-agricultural good is the numeraire, with its price normalized to one. Therefore, the representative firm takes the wage rate w as given and chooses its demand for labour to maximize profits:

$$\max_{N_n} \ \{A_n N_n - w N_n\},$$

which implies $w = A_n$.

Agricultural sector Each agricultural good is produced by a farm with a decreasing returns to scale (DRS) technology that requires the inputs of a farm operator of ability s as well as land ℓ with land augmented quality variable z. There are two different types of agricultural goods j, denoted as $j \in \{c, f\}$, where c is a cash crop and f is a food crop. The agricultural production function of farmer i producing crop j is:

$$y_j^i = (As^i)^{1-\gamma}(z_j^i \ell_j^i)^{\gamma}, \tag{2}$$

where y_j is the output of a farm for each crop type, ℓ_j is the amount of land input, and parameter $\gamma \in (0, 1)$ is a span-of-control parameter that govern returns to scale at the farm level. The three sources of productivity are: sector-specific productivity A (common to all farmers and all crops), farmer's idiosyncratic productivity s (common to both crops by a same farmer) and farm and crop-specific land productivity z_j. Climate

[4] Land quality variable z is backed by potential yield from GAEZ, and is later used to measure the impact of climate change.

change shock is represented by the change in the value of the idiosyncratic land quality variable z which is measured by the GAEZ potential yields capturing soil quality, climate, and topography and changes over time.

A farmer with ability s^i, producing crop j maximizes profits by demanding land ℓ^i_j with rental price q, and relative prices for cash and food crops p_c and p_f as given:

$$\pi_j(s,z) = \max_{\ell}\{p_j y - q\ell - p_j F_j\}$$

where, F_j is a fixed cost of operating a farm with the corresponding crop type j. Fixed cost of operating a food crop farm F_f is normalized at zero meaning that fixed costs incur only to cash crop farmers as F_c. Solving for the farmer's problem gives:

$$\begin{aligned}
\ell^*_j &= (\frac{\gamma p_j}{q})^{\frac{1}{1-\gamma}} z^{\frac{\gamma}{1-\gamma}} As \\
y^*_j &= (\frac{\gamma p_j}{q})^{\frac{\gamma}{1-\gamma}} z^{\frac{\gamma}{1-\gamma}} As \\
\ell^*_j &= y^*_j(\frac{\gamma p_j}{q}) \\
\pi^*_j &= p_j y^*_j - q\ell^*_j - p_j F_j \\
\pi^*_j &= (1-\gamma)p_j y^*_j - p_j F_j \\
\pi^*_j &= (1-\gamma)p_j^{\frac{1}{1-\gamma}} (\frac{\gamma}{q})^{\frac{\gamma}{1-\gamma}} z^{\frac{\gamma}{1-\gamma}} As - p_j F_j
\end{aligned}$$

The optimality conditions show that the farm size (harvested land), output, and profit positively depend on farmer ability s and land quality z. This implies that farmers with higher ability s^i and better land quality z^i_j will have larger farms with more output and eventually higher profits. Also, the input demand functions are linear in s and z, so farmer ability and land quality can be combined to be denoted as $g \equiv z^{\frac{\gamma}{1-\gamma}} s$.

$$\pi^*_j = (1-\gamma)p_j^{\frac{1}{1-\gamma}} (\frac{\gamma}{q})^{\frac{\gamma}{1-\gamma}} Ag - p_j F_j$$

1.4.2 Preferences and Endowments

Stand-in household has preferences over two types of goods following the utility function:

$$U = \phi\ log\ (c_a - \bar{a}) + (1-\phi)\ log\ (c_n), \tag{3}$$

where $\bar{a} > 0$ is a subsistence constraint for agricultural consumption, and ϕ is a preference weight for the agricultural good. c_n denotes non-agricultural goods consumption. The agricultural good is a composite of the cash and food crops given by the cobb-douglas function $c_a = c_c^{\beta} c_f^{1-\beta}$. Each member of the household is equally productive in the non-agricultural sector but is heterogeneous in their ability and crop-specific quality of land endowments in the agricultural sector. These heterogeneity in crop-specific land quality are defined and calculated directly for current climate conditions and is inferred from the GAEZ data. The household decides on the fraction of its members working to produce each of the crop types in the agricultural sector as well as in the non-agricultural sector.

The stand-in household maximizes utility by choosing household consumption across goods and the allocation of labour across sectors and crops given prices subject to the following budget constraint:

$$I = (1 - N_c - N_f)w(1 - \varepsilon) + \eta N_c \int_{s,z} \pi_c(s, z)dF(s, z) + N_f \int_{s,z} \pi_f(s, z)dF(s, z) + qL$$

where I is the income of the household given by:

$$I = p_c c_c + p_f c_f + c_n$$

In the budget constraint, it is assumed that working in non-agriculture is subject to a sector labour mobility barrier $\varepsilon \in (0, 1)$, and a cash crop specific wedge η. Each of the barriers ε and η are introduced for purely quantitative purposes to reconcile the sector and crop specific value productivity ratio between the model and the data. The first order conditions with respect to the share of farmers in each crops (N_c, N_f) implies the separate no-arbitrage conditions for cash and food crops:

$$w(1 - \varepsilon) = \eta \int_{s,z} \pi_c(s, z)dF(s, z)$$

$$w(1 - \varepsilon) = \int_{s,z} \pi_f(s, z)dF(s, z)$$

Which imply that the household allocates workers across sectors and crops until the average expected returns are equalized. The no-arbitrage conditions imply that income is:

$$I \equiv w(1 - \varepsilon) + qL$$

The problem of the representative household maximizing utility subject to their income is:

$$\max_{c_c, c_f, c_n} U = \{ \phi \; log \; (c_c^{\beta} c_f^{1-\beta} - \bar{a}) + (1 - \phi) \; log \; (c_n) \}$$

$$subject \;\; to : \qquad I = p_c c_c + p_f c_f + c_n$$

$$\mathcal{L} = \phi \; log \; (c_c^{\beta} c_f^{1-\beta} - \bar{a}) + (1 - \phi) \; log(c_n) + \lambda(I - p_c c_c - p_f c_f - c_n)$$

The first order condition to the household's consumption maximization problem shows that:

$$c_c = \frac{I\phi\beta}{p_c} + \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{1-\beta} \tag{4}$$

$$c_f = \frac{I\phi(1-\beta)}{p_f} + \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{-\beta} \tag{5}$$

$$c_n = I(1-\phi) + \frac{\bar{a}(1-\phi)p_f}{1-\beta}(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{-\beta} \tag{6}$$

1.4.3 Equilibrium

Households, firms, and farms behave competitively in factor and output markets. The price of non-agricultural goods is normalized to 1. The crop prices (p_j) are relative to the numeraire non-agricultural good.

Labour market The labour market clearing condition is simply noted as the sum of agricultural and non-agricultural labour with agricultural labour divided into labour spent on producing each of the different crop types:

$$N_a + N_n = 1$$
$$N_c + N_f = N_a$$

Land market Total land used for agricultural production is composed of cash and food crops:

$$L = N_c \int_{s,z} \ell_c(s,z)dF(s,z) + N_f \int_{s,z} \ell_f(s,z)dF(s,z)$$

Goods markets The cash crop market clearing condition:

$$c_c = N_c \int_{s,z} y_c(s,z)dF(s,z) - N_c F_c$$

The food crop market clearing condition with the fixed cost being zero:

$$c_f = N_f \int_{s,z} y_f(s,z)dF(s,z)$$

A *competitive equilibrium* is set of allocations for the household $\{c_c, c_f, c_n, N_c, N_f\}$, firms in the non-agricultural sector $\{N_n\}$, farmers $\{[\ell_j, y_j]_{s,z}\}$ and a set of prices $\{p_j, q, w\}$ such that: (i) given prices, allocation of the household solve the household's problem; (ii) given prices, the allocations of firms in non-agriculture and farmers in agriculture solve their problems; and (iii) markets clear.

1.5 Calibration

I calibrate the benchmark pre-climate change economy to LSMS (2009) data. For the GAEZ data I use baseline climate conditions 1961-1990. The parameters to be calibrated are: preference parameters $\{\bar{a},\phi,\beta\}$, technological parameters $\{A,A_n,F_c,\gamma\}$, barriers $\{\varepsilon,\eta\}$, triplets $\{z_c^i,z_f^i,s^i\}$ and the land endowment L. The calibration steps are listed in Appendix A, but I briefly outline the calibration strategy here. Using GAEZ and LSMS data, farmer ability distribution is directly computed from the production function. Using the calibrated parameters with the farmer ability and land quality from data, I solve for the model equilibrium matching prices to clear the market. The representative household allocates individuals into occupations and crop types in order to maximize utility subject to their income.

One of the unique aspects of this paper is that there are no distributional assumptions used to pin down crop and farm specific land quality and farmer ability. Instead of using parametric assumptions about their joint distribution, I directly use micro-data from GAEZ and LSMS to calculate land qualities and farmer ability for each of the farms in the sample. First, I use the GPS coordinates in the LSMS data to identify the spatial pixel at the 5-arc minute resolution, where the household-farm falls in. Second, I use the data from GAEZ on potential yields by crop (under baseline climate conditions) to assign a land quality measure to each farm, for both cash and food crops[5]. Third, given the land quality measures, I back-out farmer ability from the production function. In order to calculate farmer ability parameter 's', I use combination of micro data (actual yield as $\frac{y}{\ell}$, potential yield as z, and harvested land size as ℓ) implemented in the agricultural production function. Actual yield (GK$/ha) is measured from LSMS by taking actual output converted into international dollars (GK$) and dividing it by harvested land size (ha). Potential yield is taken from GAEZ measured by tons per hectare and the unit of measurement is converted to GK$/ha by multiplying prices of each crops. Since GAEZ provides potential yields of different crops for each plot of land identified by the GPS coordinates, and LSMS provides data on output and harvested land size identified by household ID (HHID) and GPS coordinates, combination of these data with the agricultural production function gives farmer ability 's_c' and 's_f' for each of the household. Then, for each farm, farmer ability is computed as the arithmetic average of s_c and s_f which prevents returning zero for some observations when other measures, such as the geometric average, are taken[6]. The model parameters along with their targets and calibrated values are provided in Table 1.3.

Without loss of generality, sector-specific productivity parameters A, A_n are normalized to 1. Given this, the relative prices of each crops (p_c and p_f) to non-agricultural goods are solved by the model matching the no-arbitrage conditions of cash and food crops. The span-of-control parameter γ is set to 0.38 to closely represent the land income share suggested for Malawi in Restuccia and Santaeulalia-Llopis (2017) because Uganda and Malawi have much similarity in their economic characteristics. The subsistence constraint

[5] Note that the potential yields are available in GAEZ regardless of whether the crop is actual being produced in that pixel.

[6] Farmer ability s is calculated directly from $y = (As)^{1-\gamma}(z\ell)^{\gamma}$, which can be re-arranged as: $s_c = \left(\frac{y_c}{\ell_c}\right)^{\frac{1}{1-\gamma}} \left[\left(\frac{A}{\ell_c}\right)^{1-\gamma} {z_c}^{\gamma}\right]^{\frac{1}{\gamma-1}}$; $s_f = \left(\frac{y_f}{\ell_f}\right)^{\frac{1}{1-\gamma}} \left[\left(\frac{A}{\ell_f}\right)^{1-\gamma} {z_f}^{\gamma}\right]^{\frac{1}{\gamma-1}}$.
For each household-farm, $s = \frac{(s_c+s_f)}{2}$.

Table 1.3: Parameterization

Parameters	Value	Target
Technological Parameters		
A	1	Normalization
A_n	1	Normalization
γ	0.38	Agricultural land income share
F_c	0.31	Share of cash crop farms
Preference Parameters		
$\bar{a}$	0.16	Share of food crop farms
ϕ	0.05	Long-run employment share in agriculture
β	0.45	Cash to food crop consumption ratio
Land Endowment		
L	2.28	Average harvested land size
Barrier		
ε	0.96	Sector value productivity ratio
η	0.4	Crop type value productivity ratio

$\bar{a}$ is computed to match the food crop market clearing condition. The cash crop specific fixed costs F_c is computed, given, to match the cash crop market clearing condition while F_f is fixed at zero[7]. For the benchmark economy, a share of labour in agriculture N_a of 0.7 is targeted, with a split between the share of cash crop farmers N_c and food crop farmers N_f of 0.2975 and 0.4025 respectively to represent real data values for Uganda. ϕ is chosen to be 0.05 to assume a long-run share of employment in agriculture of 5%. Land (L) is calculated, to match an average harvested land size in the sample of 3.25 hectares. β is set at 0.4475 such that the consumption ratio of cash crops to food crops in the model is 0.81, as estimated from the LSMS survey data[8].

I introduce two barriers in the model to reconcile the value productivity ratios of different sectors and crop types between the model and the data: ε is a barrier to labour mobility between the two sectors and η is a barrier to labour allocation between the two crop types. The barrier to labour mobility ε, is chosen such that the value of productivity disparity between non-agriculture and agriculture in the model (0.149) is as close as possible to data (0.143). This represents the 7-fold labour productivity ratio difference between the two sectors in Uganda.

$$\frac{\left(\frac{N_c}{N_a} p_c y_c + \frac{N_f}{N_a} p_f y_f\right)}{\frac{Y_n}{N_n}} = \frac{1}{7}.$$

[7]Appendix A contains the details on calculations of $\bar{a}$ and F_c.

[8]Compute β from the cobb-douglas property and match it to cash crop to food crop consumption ratio from data estimated from LSMS Uganda sample (0.81).

$$\frac{\frac{dc_a}{dc_c}}{\frac{dc_a}{dc_f}} = \frac{\beta}{1-\beta}\frac{c_f}{c_c} = \frac{p_c}{p_f}, \qquad \frac{p_c c_c}{p_f c_f} = \frac{\beta}{1-\beta} = 0.81$$

The cash crop specific wedge η is chosen such that the value productivity ratio between producing food crops and cash crops in the model (0.525) matches closely to the data (0.5):

$$\frac{p_f Y_f}{N_f} \frac{N_c}{p_c Y_c} = \frac{1}{2}.$$

1.6 Quantitative Analysis

The main goal of this paper is to analyze the impact of climate change on aggregate productivity through the lens of the model. In the benchmark economy, the land quality across farms is reflected in the potential yields from GAEZ for current climatic conditions. To implement climate change in the model, the main quantitative experiment feeds in expected future potential yields for each pixel in Uganda based on projected climatic changes by meteorologists. This section reports results on the counter-factual experiment regarding future climate change with "rain-fed" water supply and "intermediate inputs" to match the conditions of Uganda. The main quantitative experiment analyzes the impact of climate change on future aggregate values from the benchmark economy calibrated to baseline year (1961 ~ 1990) Uganda. To do so, I change the value of the land quality variable z_j for each crop $j \in \{c, f\}$, which is represented by the change in GAEZ potential yield data before and after the climate change scenario. There is no clear "benchmark" or a "central" scenario, due to substantial amount of uncertainty in various emission scenarios (Nakicenovic et al., 2000). Thus, for my benchmark climate change scenario by GAEZ, I use those obtained from the Hadley CM3 A1FI model as used by Costinot et al. (2016).

Table 1.4 shows that climate change overall has a positive impact on agricultural productivity for Uganda in both years 2050 and 2080. This result is mostly backed up by the fact that the aggregate potential yield from GAEZ (z), which is an exogenous parameter representing climate change, has increased in future scenarios. Relative to the benchmark economy in column 1, year 2050 A1 scenario with CO_2 fertilization at column 2 shows around 15% decrease in the share of employment in agriculture with around 4.5% decrease in the share of cash crop farmers and 10.3% decrease in the share of food crop farmers. This change in farmer shares with adjusted prices, result in relative farm size increase by 27% with a 23% increase in agricultural productivity, mostly coming from a 34% increase in food crop productivity followed by a 12% increase in cash crop productivity. GDP is less sensitive to climate change given that it includes non-agriculture which here is not affected by climate change directly. Nevertheless, GDP per capita increases by 12%.

Another important result from Table 1.4 is that, although years 2050 and 2080 all show positive results, the increase is non-monotonic. With CO_2 fertilization, year 2050 would result in a 23% increase in agricultural productivity which is much higher than the 13% increase in year 2080. This means that there is an increase in productivity from the baseline period to year 2050, but declines in the future showing that climate change will eventually become a threat for Uganda. One of the reasons for this trend would be that the agro-climatic conditions become extreme enough, that it starts to show negative impacts for aggregate productivities. For example, if the optimal conditions for growing most of the crops have hit its maximum

Table 1.4: Impact of Climate Change

	BE	2050 C	2050 N	2080 C	2080 N
Relative Land Quality (z)					
z	1	1.15	1.09	1.03	0.94
z_c	1	0.99	0.93	0.74	0.68
z_f	1	1.32	1.25	1.31	1.20
Employment share (%)					
N_a	70	55.14	57.77	60.43	65.39
N_c	29.75	25.20	25.96	26.21	27.33
N_f	40.25	29.94	31.82	34.22	38.06
Relative Labour Productivity					
$\frac{Y_a}{N_a}$	1	1.23	1.18	1.13	1.06
$\frac{Y_c}{N_c}$	1	1.12	1.08	1.03	0.98
$\frac{Y_f}{N_f}$	1	1.34	1.28	1.27	1.17
Relative Average Farm Size	1	1.27	1.21	1.16	1.07
Relative GDP	1	1.12	1.10	1.08	1.04

Note: BE denotes benchmark economy; results are relative to the BE. Land quality is the exogenous parameter from GAEZ potential yields. The counter-factual experiment is based on the predictions of the Hadley CM3 A1FI model of GAEZ for year 2050 with 'Carbon dioxide fertilization'.

point around year 2050, as time goes on, agro-climatic conditions become too harsh so that it becomes difficult to grow more and more crops, even those that can withstand warmer and disastrous environments. This non-monotonic trend of farmer labour productivity resulting from my general equilibrium model is consistent with the agronomic literature whereby yields gradually increase up to a threshold of temperature, and fall sharply beyond that (Schlenker and Roberts, 2009; Burke et al., 2015). More importantly, in between years 2050 and 2080, the potential yields for Uganda is decreasing much faster for cash crops compared to food crops. This is causing the agricultural productivity difference between cash crops and food crops to become larger over time, adding another concern for farmers in Uganda.

Although emission of carbon dioxide into the atmosphere causes global warming with many experts warning about its negative impact on food production, CO_2 fertilization caused by increased concentration of carbon dioxide increases the rate of photosynthesis which contributes to faster growing of crops. Column 2 shows that considering CO_2 fertilization would result in a 23% increase in agricultural productivity, while not accounting for CO_2 fertilization would result in a 18% increase. Therefore, the CO_2 fertilization has a positive effect on agricultural production with a significant contribution of over 20% of agricultural growth for the case of year 2050 Hadley CM3 A1FI scenario.

One noticeable aspect from Table 1.4 is that the potential yields for each crop type show very different trends. In response to climate change, the potential yield of food crops (z_f) increases while the potential yield of cash crops (z_c) decreases, and as a result, the increase in food crop productivity is much higher than the increase in cash crop productivity. Note that cash crop productivity increases even though there

Table 1.5: No crop type barrier ($\eta = 0$)

	BE	2050 C	2050 N	2080 C	2080 N
Employment share (%)					
N_a	70	57.95	60.01	61.39	64.93
N_c	29.75	24.81	25.64	26.14	27.55
N_f	40.25	33.15	34.37	35.25	37.38
Relative Labour Productivity					
$\frac{Y_a}{N_a}$	1	1.21	1.17	1.14	1.08
$\frac{Y_c}{N_c}$	1	1.13	1.09	1.03	0.98
$\frac{Y_f}{N_f}$	1	1.28	1.23	1.25	1.18
Relative Average Farm Size	1	1.21	1.17	1.14	1.08
Relative GDP	1	1.09	1.07	1.07	1.04

Note: The agricultural productivity difference is similar regardless of the crop type barrier η (Result between Tables 1.4 and 1.5). So, most of the changes are coming from the differences in land quality arising from climate change.

is a decrease in z_c. This is mainly caused by a decrease in the share of cash crop farmers driven by the change in prices. It seems rather clear that the difference in productivity change between the two crop types is caused by the change in potential yields.

In addition, the cash crop specific wedge η from the model restricts the production of cash crops from being more productive, resulting in much larger increase in food crops in response to climate change. Comparing the results from Table 1.4 and Table 1.5, eliminating the crop type barrier η decreases the productivity gap between the two crop types. However, there is still a significant disparity between cash and food crops, suggesting that the productivity difference between the crop types is mainly caused by differences in land quality z.

Figure 1.3 illustrates the farm size distribution before and after the climate shock. Farm size distribution for each period is based on the model values and we can see that the agricultural sector of Uganda is dominated by small-scale farms. Comparing before and after the climate change, Figure 1.3 shows that the entire distribution of farm sizes shifts right. This encompasses that in year 2050, there is a higher proportion of large-scale farms as well as the increase in average farm size and agricultural labour productivity.

Quantitative experiments so far show rather positive results up until year 2080, but there are other prediction scenarios as well. GAEZ estimates are available under 4 scenarios (A1, A2, B1, B2) and 4 general circulation model (GCM) combinations including Hadley CM3. Each of the GCM is developed independently by different climatologists from different countries and are then combined with the emission scenarios from the IPCC program. Therefore, comparing the results from different general circulation models could serve as a great tool for robustness checks.

The A1FI scenario features rapid economic growth and introduction of new and more efficient fossil-intensive technologies, with global population peaking in mid-century and declining thereafter. Compared to

Figure 1.3: Farm size Distribution

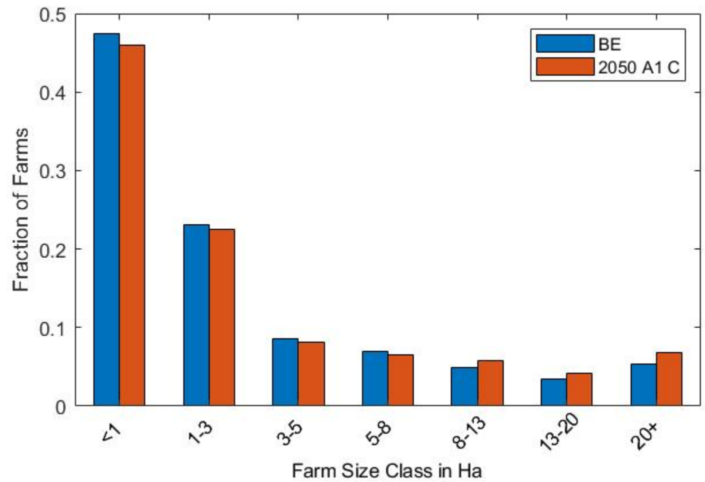


Table 1.6: Different Scenarios (Robustness Check)

	BE	2080 A1	2080 A2	2080 B1	2080 B2
Relative Land Quality (z)					
z	1	1.03	1.03	1.14	1.18
z_c	1	0.74	0.84	0.98	1.04
z_f	1	1.31	1.23	1.29	1.32
Employment Share (%)					
N_a	70	60.43	60.02	55.90	55.49
N_c	29.75	26.21	26.39	25.42	25.34
N_f	40.25	34.22	33.63	30.48	30.15
Relative Labour Productivity					
$\frac{Y_a}{N_a}$	1	1.13	1.14	1.21	1.22
$\frac{Y_c}{N_c}$	1	1.03	1.05	1.11	1.11
$\frac{Y_f}{N_f}$	1	1.27	1.25	1.33	1.33
Relative Average Farm Size	1	1.16	1.17	1.25	1.26
Relative GDP	1	1.08	1.08	1.11	1.11

Note: The columns represent different scenarios provided by GAEZ. All scenarios allow for theCO_2 fertilization.

Column 2 : year 2080, Hadley CM3 A1FI. **Column 3 :** year 2080, CSIRO Mk2 A2.
Column 4 : year 2080, Hadley CM3 B1. **Column 5 :** year 2080, CCCma CGCM2 B2.

scenario A1 from GAEZ, scenarios A2 and B2 feature a very heterogeneous world that focuses on regionally oriented economic development and technological change. Table 1.6 shows results for a completely different

GCM-SRES each corresponding to a unique "scenario", that would provide different results. Columns 2 and 3 (scenario A) report close similarities in magnitudes while columns 4 and 5 (scenario B) show more of an increase when compared to the results from scenario A. The main reasoning behind this disparity is because of the fact that scenario A depicts the world with rapid economic growth, while scenario B emphasizes the world focused on environmental sustainability. Although there are differences in magnitudes, Table 1.6 reports that different scenarios show consistent pattern of increasing productivity after climate change.

Table 1.7: Regional (Temperature) Heterogeneity (year 2050)

	Year 2050 (with CO_2)				Year 2050 (w/o CO_2)			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Relative Land Quality (z)								
z	1.47	1.19	1.08	1.01	1.39	1.12	1.02	0.96
z_c	1.30	1.08	0.90	0.75	1.22	1.01	0.85	0.71
z_f	1.67	1.31	1.24	1.22	1.59	1.24	1.17	1.15
Employment Share (%)								
$N_a(\%)$	46.08	59.05	68.52	75.76	47.92	62.24	72.26	79.64
$N_c(\%)$	22.15	26.49	28.45	29.25	22.81	27.33	29.16	29.76
$N_f(\%)$	23.87	32.56	40.07	46.51	25.11	34.91	43.11	49.88
Relative Labour Productivity								
$\frac{Y_a}{N_a}$	1.45	1.16	1.01	0.93	1.39	1.11	0.97	0.89
$\frac{Y_c}{N_c}$	1.26	1.08	0.97	0.91	1.22	1.04	0.95	0.89
$\frac{Y_f}{N_f}$	1.64	1.24	1.08	1.00	1.57	1.17	1.03	0.95
Relative Average Farm Size	1.52	1.19	1.02	0.92	1.46	1.12	0.97	0.88
Relative GDP	1.19	1.09	1.01	0.96	1.18	1.06	0.98	0.93

Note: Divide the sample into quantiles based on their baseline (1961-1990) mean annual temperature.

Q1 = 13.79°C - 20.81°C; **Q2** = 20.82°C - 22.21°C;
Q3 = 22.22°C - 23.21°C; **Q4** = 23.22°C - 25.87°C.

It is crucial to acknowledge that the changing climate does not have homogeneous impact on all regions of the globe. To assess the heterogeneity on the impact of climate change, I divide sample farms in Uganda into four quartiles based on their mean annual temperature reported in GAEZ during 1961-1990. As was shown in Figure 1.2, dividing the LSMS sample into quartiles based on their mean annual temperature leads to locational clustering patterns. Therefore, by examining the quartiles separately, identifying the impact of climate change for different locations within Uganda becomes possible. Results in Table 1.7 show that the overall productivity increase in year 2050 was mostly driven by the colder regions (Q1, Q2) while farms in Q3 barely contribute to productivity changes and farms in Q4 experience a decline in agricultural productivity for both crops alike. This means that there will be an immediate impact of climate change in the clustered areas with higher mean annual temperature, while relatively colder regions will be given some time before experiencing negative impacts. The first quartile, which shows a large increase in productivity for year 2050, forms a geographic clustering pattern in south-western region of Uganda; thus, concentrating the number of farms and possibly forming an agricultural zone in the region, while improving the non-agricultural sector on the other regions could serve as a coping strategy taken by Uganda in response to the changing climate.

1.7 Channels of Productivity

I decompose the mechanisms through the lens of the model that generate the quantitative effects in response to climate change. Climate change can increase agricultural productivity by directly improving the quality of the land operated by farmers. This is the direct effect and it improves the agricultural output of a farmer with the same ability and endowments even when the prices are not changing. However, conditional on this direct effect, there are also indirect or a general equilibrium effects. The indirect effect of climate change is associated with changes in prices, which in turn have an impact on choices of farm size and crop mix by farmers, that would induce productivity changes.

Table 1.8: Channels of productivity (Direct effect vs. Indirect effect)

	2080 A1 C
$N_a(\%)$	60.43
Rel. AFS	1.16
Rel. $\frac{Y_a}{N_a}$	+13.1%
Rel. $\frac{Y_a}{N_a}$ Direct	+ 5.8%
Rel. $\frac{Y_a}{N_a}$ Indirect	+ 7.3%
Rel. $\frac{Y_c}{N_c}$	+ 2.6%
Rel. $\frac{Y_c}{N_c}$ Direct	− 2.8%
Rel. $\frac{Y_c}{N_c}$ Indirect	+ 5.4%
Rel. $\frac{Y_f}{N_f}$	+26.7%
Rel. $\frac{Y_f}{N_f}$ Direct	+18.0%
Rel. $\frac{Y_f}{N_f}$ Indirect	+ 8.7%

Note: Results are based on the climate change experiment of year 2080 Hadley A1FI scenario with CO_2 fertilization. Recorded direct effects of productivity in row 4 is calculated using the production function: $y = (As)^{1-\gamma}(z\ell_{BE})^{\gamma}$.

In order to test the magnitude of the direct and indirect effects, the experiment results for year 2080 A1 C are shown in Table 1.8 with a decomposition of the contributions for each channel. The direct effect is calculated by fixing the prices, labour shares, crop choices, farmer ability and input used in production, so when computing the agricultural productivity, only variable that varies is the land quality z. The indirect effect is simply the difference between the total effect and the direct effect. When looking at the impacts of direct and indirect effects, the direct effect accounts for around 44% of the total effect. This means that there is a 56% effect due to the general equilibrium changes, which amplifies the productivity increase causing higher productivity than directly impacted by the climate change. This indirect effect is mainly caused by structural changes arising from climate change such as farmers moving into the non-agricultural sector and increase in average farm sizes.

One thing to note is that the indirect effects are positive for both crop types even when there is a negative direct effect for the cash crop[9]. The reason behind this is because the climate shock will still increase the combined potential yields of cash and food crops (z). Initially, this will lead to a higher supply of agricultural goods relative to demand. And as a result, price of agricultural goods will decrease and the representative household will allocate less workers to the agricultural sector. This would increase the average farm size for all operators causing an increase in labour productivity for both crop types. As shown in Table 1.8, this indirect effect or the structural transformation caused by the climate change significantly increases the labour productivity in Uganda.

1.8 Conclusion

This paper studies the impact of climate change on aggregate productivity using a quantitative model and micro-level geographic and agricultural production data. The model features both farmer heterogeneity and land quality heterogeneity, and my framework allows me to decompose the two. In the model, conditional on ability and land quality, farmers endogenously choose farm size and the type of crop to cultivate. In order to implement the effect of climate change in the model, data on land quality represented by potential yield from GAEZ are used to combine with rich-micro data from LSMS at the farm-level. By comparing the benchmark economy to a predicted future scenario, five main results are reported: (1) CO_2 fertilization has a positive impact on agricultural productivity; (2) impact of climate change on aggregate productivity shows a non-monotonic trend; (3) farms located in colder region showed high increase in productivity while farms in hotter region experienced decline in productivity when exposed to climate shock; (4) both direct and indirect effect of climate change has significant impacts on agricultural productivity; (5) productivity gap between skilled and unskilled farmers widens in response to climate change.

As a result, the overall productivity increased by 23% between the baseline year and year 2050, and increased by 13% in between the baseline year and year 2080, but the productivity increase was mostly dictated by the farms in Q1 located at the south-western region of Uganda where the average temperature was lower than other regions. The results from this paper suggest that the farms located outside of the south-western region would experience more and more decline in their productivity as time goes on, while farms in Q1 will be enjoying increased productivity from climate change longer than other regions. As this situation intensifies, this could potentially result in the sudden migration of farmers in other areas (Q2, Q3, Q4) into south-western region of Uganda where farming is still productive. This would result in a decrease in average harvested land size and agricultural productivity in Q1 due to over-saturation of farmers in small bits of land. The over-saturation would happen at the same time as the productivity gap widens between more and less able farmers, then this could push out the unskilled farmers but will also leave skilled farmers hindered in the short run.

[9]Negative direct effect is coming from the fact that the potential yields after the climate change in year 2080 from GAEZ, decreased compared to the benchmark economy. (z_c decreased from 1 to 0.74)

Uganda is a country with low productivity yet high dependency on the agricultural sector. And the result suggests that the climate shock will overall be beneficial to Ugandan agriculture in the short run. There are regions that will flourish because of global warming, but there also exist locations that will suffer drastically as well. Therefore, it would be interesting to extend the analysis to different countries to examine the extent to which the thermal climatic heterogeneity within a country is the main driver of the impact of climate change. Also, because the model does not allow for regional migration, regional heterogeneity in this paper analyzes the same group of farmers with different land quality induced by climate change so that there only exists sectoral migration from agriculture to non-agriculture. But it could be interesting to model migration between the different locations (quantiles) to further study the impact of structural changes in response to climate shock.

Chapter 2

Climate Change and the Distribution of Agricultural Production across Space

2.1 Introduction

While climate change and global warming are already being felt, they are expected to intensify over the next decades (IPCC, 2021). However, climate change is expected to be heterogeneous across space, with some locations experiencing particularly adverse effects while other locations less so or even improvements. How will economic activity shift across space as economies adapt to idiosyncratic environmental changes, in the face of realistic geographic barriers or institutional frictions? Linking the evidence of climate change to aggregate economic performance has spurred a large body of economic research(Nordhaus, 2006; Sachs, 2003; Acemoglu et al., 2002; Dell et al., 2014). One of the key challenges in the literature is figuring out an optimal way to adapt to changing climate.

Adaptation to climate change is particularly important to the agricultural sector, that uses location-specific inputs, a big part of which are climatic conditions. The issue is especially relevant for developing countries given that they have disproportionately low agricultural productivity and large agricultural sectors when compared to developed economies (Gollin et al., 2002; Caselli, 2005; Restuccia et al., 2008). Another characteristic of low income countries is that they have poor transport infrastructure with high internal costs (Adamopoulos, 2011). The transport friction serves as a major obstacle for farmers to generate profits and further hinders agricultural productivity. The type of frictions I consider here, transportation costs and labor mobility frictions, are both prevalent in the developing world, in countries that also face large agricultural economies. There is a large literature emphasizing the role of transportation costs for development Adamopoulos (2011, 2020); Gollin and Rogerson (2014); Sotelo (2020) and migration Asher and Novosad (2020). Implicit or explicit barriers impeding the mobility of labor across space are also well documented for developing countries (De Janvry et al., 2015; Munshi and Rosenzweig, 2016; Caliendo et al., 2019; Adamopoulos et al., 2022).

Of the adaptation techniques, migration has frequently been described in other literature as an effective form of adaptation to climate change(Black et al., 2011; Kniveton et al., 2012; Gemenne and Blocher, 2017). At the micro-level, climate change linked with internal migration has been studied by economists including Feng et al. (2012). And at the macro-level, quantitative micro-founded general equilibrium models have been incorporated with climate change for welfare analysis and policy evaluation(Golosov et al., 2014; Hassler et al., 2016; Hassler and Krusell, 2018). However, less is known for how the interaction between climate change with transportation friction has an impact on the long-run distribution of farming activity. Climate change might push farmers to move to districts that would have better land quality after climate change, but that location might be restricted by transportation infrastructure. In this research, I analyze the impact of climate change on each district (or woreda) of Ethiopia using a spatial model featuring heterogeneity in transportation costs and land quality.

The main feature of this paper is to infer spatial reallocation of agricultural production after climate change. In the long run, it is expected that the climate change will lead to farmers' relocation patterns, and transportation cost also has a tremendous effect on the relocation of farming activity(Adamopoulos, 2020). For example, let's assume that location A has very good land quality and its geographical conditions will get better after climate change, but is located in a mountainous region where transportation is very poor. Location B has poor land quality that will further hinder agricultural production after climate change, but is located next to a highway close to grain markets. In this setting, climate change will have a negative impact on location B but positive impact on A, and farmers in A would be producing more than farmers in B in the future. However, farmers in B may make more profit even after climate change, due to poor transportation infrastructure in A. Climate change is pushing farmers to move from B to A, but transportation cost is obstructing the farmer migration in this setting. This example shows that poor transportation infrastructure in productive locations can be an obstacle in adapting to climate change by limiting labor mobility and causing misallocation. Therefore, transportation infrastructure is of ample importance when analyzing the impact of climate change on internal migration of farmers.

Ethiopia is a rugged, landlocked country in the Horn of Africa. The capital of Ethiopia is Addis Ababa and with about 115 million people in 2020, Ethiopia is the second most populous nation in Africa after Nigeria and still the fastest growing economy in the region with 6.3 percent growth in 2020/21[1]. However, it is also one of the poorest, with a per capita gross national income of $890. By total area, Ethiopia is as large as France and Spain combined, with about 34 percent of the country's total land area dedicated to agricultural production.

Ethiopia is a country that went through a major comprehensive road expansion program over 1997 to 2014. Research on the outcomes of the road expansion has been studied by Adamopoulos (2020) using a quantitative spatial model combined with district-level panel data. This paper further builds on this research by including climate change to analyze the interaction of climate change with transportation frictions on the spatial distribution of agricultural production. Ethiopia also serves as a great candidate when it comes to spatial research on climate change; high temperature gap across locations result in high variance in land quality, and this allows for a close analysis on farmer migration patterns within the country.

In this chapter I study the migration of farmers across districts in Ethiopia as an adaptation mechanism to long-term climate change, in the presence of heterogeneous labor mobility barriers impeding their migration and heterogeneous transportation costs for delivering goods to market. To do so, I combine granular geographic data on land quality, as well as production and population data with a general equilibrium spatial model featuring an endogenous distribution of farmers across locations. In my model the distribution of the farmers across space is shaped by: (i) the inherent quality of the land (a big part of which is climate) in each location; (ii) location-specific transportation costs for delivering farm output to

[1]Source: The World Bank

market; and (iii) location-specific frictions impeding the mobility of labor across space. I calibrate my model to the current district-level data for Ethiopia. Taking as given how climate change impacts yields across districts heterogeneously, I find that at the aggregate level agricultural productivity will increase. This is due to the exogenous increase in average land quality as well as the reallocation of farmers towards districts that have experienced an increase in their productivity as a result of climate change. This aggregate result confounds an increased regional inequality across locations, with some benefiting (those around the center of the country) and others hurt (more remote areas) by climate change. If transport and labor mobility frictions were to be removed, labor would further reallocate across space more according to districts' natural comparative advantages.

The rest of the paper proceeds as follows. Section II explains the main data sources and how they are combined to be used in the analysis of the research. Section III describes the spatial framework featuring land quality and transportation costs. Section IV presents the details on the calibration of the model. Section V reports the aggregate and distributional results from the quantitative experiments. Section VI concludes.

2.2 Data

There are three main sources of data used in the analysis for Ethiopia: land quality data from Global Agro-Ecological Zones (GAEZ) project; transportation cost data from the Ethiopian Roads Authority (ERA) and the Regional Roads Authorities in Ethiopia; and agricultural production data from the Ethiopian Agricultural Sample Survey (AgSS). Aside from these three sources, I also use district (or woreda) level population data from IPUMS.

Land Quality Data Land quality (or location productivity) in my analysis is represented by GAEZ potential yields. GAEZ divides the world into GPS identified 5 arc-minute cells that are roughly 10 by 10 km in size. For each of these pixels, GAEZ reports data on location-specific geographic attributes that are important for agricultural production. These geographic attributes include soil quality (depth, fertility, chemical composition), climate conditions (temperature, precipitation, wind speed), and topography (elevation, slope). But most importantly, GAEZ uses these geographic attributes to estimate potential yields for different crops in each pixel[2]. Potential yields are computed as maximum possible output per hectare that can be attained given the crop's production requirements and cell-specific land quality characteristics, including the crops that are not actually produced in a particular cell. GAEZ is a standardized framework for the characterization of climate, soil, and terrain conditions relevant for agricultural production because the yield estimates are computed after controlling each of the field characteristics (such as soil types and elevation) and climatic variables (such as rainfall, temperature, humidity, wind speed, sun exposure, etc). Another important aspect of GAEZ potential yields is that it provides estimates on future predictions

[2]For analysis I use the following 30 crops from GAEZ: cereals (wheat, wetland rice, dryland rice, maize, barley, sorghum, rye, pearl millet, foxtail millet, oat, buckwheat); roots and tubers (white potato, sweet potato, cassava, yam and cocoyam); sugar crops (sugarcane, sugarbeet); pulses (phaseolus bean, chickpea, cowpea, dry pea, gram, pigeonpea); oilcrops (soybean, sunflower, rapeseed, groundnut, oilpalm, olive); and cotton.

calculated by natural scientists. Future predicted potential yields are provided for different climate scenarios, with each corresponding to a different scenario based on demographic, social, economic, technological, and environmental shift[3]. Therefore, GAEZ potential yields serve as a great proxy for change in land quality distribution in response to changing climate.

Roads Data The transportation costs I use in the analysis are computed using the travel time in minutes between the district centroids and the nearest destination grain market[4]. The destination markets are Ethiopia's 33 major wholesale grain markets obtained from Ethiopian Grain Trading Enterprise which are spread throughout Ethiopia. Starting in 1997, Ethiopia has embarked on an extensive road development program as a pilar of its growth strategy, resulting in a 3-fold increase in the volume of the total road network from 1997 to 2014 (Adamopoulos, 2020). I use the transportation costs estimated in Adamopoulos (2020) for the year 2014 to reconcile the recent changes in the transportation infrastructure due to the road expansion program. In Adamopoulos (2020), travel time from district centroids to the nearest grain market is estimated from actual road network data with high resolution geographic data on elevation and land use, along with the GPS coordinates of the district centroids and the destination grain markets.

Agricultural Production Data Household-level data from the Ethiopian Agricultural Sample Survey (AgSS) contains information at the field level on what crops are produced, what quantity is produced, and how much of the land is allocated to the production of the crop. Given that the AgSS data do not necessarily follow the same households over time, I use pooled household data from AgSS for years 2012, 2013, and 2014 to increase the sample size and representativeness of the survey data across Ethiopia. AgSS do not contain GPS information on the location of individual households, so I aggregate the households at the district (or woreda) level, the lowest level of spatial disaggregation for which a reliable dataset could be constructed.

I combine this production data (inputs and outputs) with gridded geographical data from the Global Agro-Ecological Zones (GAEZ) project and the road data from the Ethiopian Road Authority (ERA) to generate a district-level dataset containing production input and output, with land quality and transportation costs. In order to combine data at the district-level, I homogenize the coding across all years using the 2007 IPUMS zonal and district boundaries and identifiers to avoid problems arising from re-districting of zones and woredas in Ethiopia. Then, pixel-level potential yields from GAEZ on 30 crops are aggregated to each district organized by IPUMS for the year 2007; and subsequently merged with the district level data from AgSS and ERA. Then, the combined district-level dataset is merged with the population data for each district in year 2007 taken from IPUMS to represent the farmer share.

[3] In my analysis, the benchmark economy uses the baseline climate conditions (1961-1990), and the future scenario I study uses climate conditions for the year 2050.

[4] Calculation of iceberg transport cost using travel time is described in the calibration section.

2.3 Model

I develop a spatial equilibrium model of agriculture to assess the interaction of climate change with transport and labor mobility frictions on the long-run distribution of farming activity in Ethiopia. The model features heterogeneity in land quality (or location productivity), farmer ability, transportation costs and mobility barriers across locations which together pin down the allocation of land within locations and the distribution of farmers across space.

2.3.1 Environment

Consider an economy with a finite set of locations, indexed $i \in \mathcal{M} \equiv \{1, 2, ...M\}$ where $\mathcal{M}$ is a discrete set. Location $i \in M$ is endowed with a fixed factor (land) with a density equal to $\bar{L}(i)$. The economy is populated by $\bar{N}$ agents, which are farmers in the agricultural sector. Agents are subject to a mobility barrier that is heterogeneous in each location when migrating for production purposes. Production can take place in any location, producing the same good. All goods produced at each location have to be delivered to a nearest grain market, subject to transportation costs.

Agents are *ex-ante* identical, and can choose the location in which they will engage in production. After choosing a location i, agents become *ex-post* heterogeneous with respect to their managerial ability in operating a farm. In particular, conditional upon locating at $i \in M$, each agent is assigned a productivity s which is heterogeneous across locations. The ability s of a farmer in location i becomes the idiosyncratic productivity of the farm. Besides the productivity of the farm operator, production also requires some of the available fixed factor (land) in that location.

Production Function A farm operator of ability s at location $i \in \mathcal{M}$ produces the agricultural good using own ability and land as inputs according to a decreasing returns to scale technology of the Cobb-Douglas form.

$$y = A(i)s^{1-\gamma}\ell^{\gamma},$$

where ℓ is the amount of land input under the operator's control and $\gamma \in (0, 1)$ is the span-of-control parameter that captures the importance of land in agricultural production. Locations are heterogeneous with respect to their inherent productivity in producing the single good, denoted as $A(i)$ for location $i \in \mathcal{M}$. This term captures differences in land quality, such as soil and terrain characteristics, climate, rainfall etc. Shocks to land quality is represented as a change in the distribution of $A(i)$ across space.

Transport Costs After production, producers have to deliver their output to the nearest grain market for consumption, subject to iceberg type trade costs. In particular, a producer s in location i has to ship an amount of output $T(i) \geq 1$ from location $i \in \mathcal{M}$ in order for one unit to arrive at the wholesale market. Thus, the overall iceberg cost for a producer s in location i is $T(i)$ which operates similar to an output tax.

Labor Mobility Frictions A farmer in each location faces a labor mobility barrier $\varepsilon(i)$ that is heterogeneous of each location when trying to migrate to a different location. For quantitative purposes, mobility friction is chosen to match the spatial distribution of farmers.

2.3.2 Analysis

In each location the market for the fixed factor is perfectly competitive. Let $q(i)$ be the rental price of land in each location, the profit maximization problem for a producer of ability s at location i is:

$$\max_{\ell(s,i)} \left\{ \frac{A(i)}{T(i)} s^{1-\gamma} \ell(s,i)^{\gamma} - q(i)\ell(s,i) \right\}$$

which implies a demand for land,

$$\ell(s,i) = \left[\frac{\gamma}{q(i)} \frac{A(i)}{T(i)} \right]^{\frac{1}{1-\gamma}} s \tag{1}$$

and output supply,

$$y(s,i) = A(i)^{\frac{1}{1-\gamma}} \left[\frac{\gamma}{q(i)T(i)} \right]^{\frac{\gamma}{1-\gamma}} s \tag{2}$$

The corresponding profits for farmer s in location i are,

$$\pi(s,i) = (1-\gamma) \left[\frac{A(i)}{T(i)} \right]^{\frac{1}{1-\gamma}} \left[\frac{\gamma}{q(i)} \right]^{\frac{\gamma}{1-\gamma}} s \tag{3}$$

The optimality conditions show that the farm size, output, and profit depends on farmer ability s, location specific land quality $A(i)$, and location specific transportation costs $T(i)$. Farms that are operated by skilled farmers, located in a district with better overall land quality and transportation infrastructure will produce more output and generate higher profits.

Consider two individuals in the same location i with abilities s and z respectively. Then, the ratio of land input demands for these two farmers would be,

$$\frac{\ell(s,i)}{\ell(z,i)} = \frac{s}{z}$$

This determines the allocation of resources across producers within locations. Transport costs and land quality do not affect the allocation of resources within locations because they are common to all producers.

Land Market Clearing Condition At each location i the land market must clear,

$$\bar{L}(i) = N(i) \int_{\underline{s}}^{\bar{s}} \ell(s,i) dF(s,i) \tag{4}$$

where $N(i)$ is the total number of farmers locating at i. Average farm size in location i is then,

$$AFS(i) \equiv \frac{\bar{L}(i)}{N(i)} = \int_{\underline{s}}^{\bar{s}} \ell(s,i) dF(s,i)$$

Labor Market Clearing Condition The market clearing condition for farmers for the whole economy is,

$$N = \int_i N(i) di \tag{5}$$

No-Arbitrage Condition Farmers must expect the same profit at each location i, given the mobility barrier ε that is heterogeneous across locations.

$$\bar{\pi} = \varepsilon(i) \int_{\underline{s}}^{\bar{s}} \pi(s,i) dF(s,i) \tag{6}$$

Aggregate output at location i is,

$$Y(i) = N(i) \int_{\underline{s}}^{\bar{s}} y(s,i) dF(s,i)$$

Combining the land market clearing condition in location i, (4), and individual demand for land, (1), gives:

$$\bar{L}(i) = N(i) \left[\frac{\gamma A(i)}{T(i) q(i)} \right]^{\frac{1}{1-\gamma}} \int_{\underline{s}}^{\bar{s}} s dF(s,i) \tag{7}$$

From the cross-space no-arbitrage condition (6) and the individual farmer profit in location i (3),

$$\bar{\pi} = \varepsilon(i)(1-\gamma) \left[\frac{A(i)}{T(i)} \right]^{\frac{1}{1-\gamma}} \left[\frac{\gamma}{q(i)} \right]^{\frac{\gamma}{1-\gamma}} \int_{\underline{s}}^{\bar{s}} s dF(s,i) \tag{8}$$

From (7) and (8),

$$q(i) = \frac{\gamma}{\varepsilon(i)(1-\gamma)} \frac{N(i)}{\bar{L}(i)} \bar{\pi} \tag{9}$$

Plug (9) into (8) to solve for the number of farmers locating in i, $N(i)$ as a function of $\bar{\pi}$,

$$N(i) = \frac{\varepsilon(i)^{\frac{1}{\gamma}} (1-\gamma)^{\frac{1}{\gamma}} \left[\frac{A(i)}{T(i)}\right]^{\frac{1}{\gamma}} \bar{L}(i)}{\bar{\pi}^{\frac{1}{\gamma}}} \left[\int_{\underline{s}}^{\bar{s}} s dF(s,i) \right]^{\frac{1-\gamma}{\gamma}} \tag{10}$$

Plugging (10) in the economy-wide labor market clearing condition (5), the average profit is solved for each location i,

$$\bar{\pi} = \frac{\varepsilon(i)(1-\gamma)}{N^{\gamma}} \Big\{ \int_i \Big[\frac{A(i)}{T(i)}\Big]^{\frac{1}{\gamma}} \bar{L}(i) \Big[\int_{\underline{s}}^{\bar{s}} s dF(s,i) \Big]^{\frac{1-\gamma}{\gamma}} di \Big\}^{\gamma} \tag{11}$$

Spatial Distribution of Farmers Plugging $\bar{\pi}$ from (11) back into (10) and solving for the share of farmers is location i gives,

$$\frac{N(i)}{N} = \frac{\varepsilon(i)^{\frac{1}{\gamma}} \Big[\frac{A(i)}{T(i)}\Big]^{\frac{1}{\gamma}} \bar{L}(i) \Big[\int_{\underline{s}}^{\bar{s}} s dF(s,i) \Big]^{\frac{1-\gamma}{\gamma}}}{\int_m \varepsilon(m)^{\frac{1}{\gamma}} \Big[\frac{A(m)}{T(m)}\Big]^{\frac{1}{\gamma}} \bar{L}(m) \Big[\int_{\underline{s}}^{\bar{s}} s dF(s,m) \Big]^{\frac{1-\gamma}{\gamma}} dm} \tag{12}$$

Equation (12) describes the distribution of farmers across locations (space), i.e. the share of producers allocated to location i in equilibrium. Where farmers locate will depend on the transport costs of those locations, the frictions impeding labor mobility, and their land quality (relative to all other locations). In other words, climate change that impacts heterogeneously different locations will affect the distribution of farms across space.

Definition of equilibrium A *competitive equilibrium* is a set of allocations of each farm $\{[\ell_i, y_i]_s\}$, an allocation of producers across locations and land rental price $q(i)$ in location i, such that (i) the production allocation for each farm in location i maximizes profits given price, transportation cost $\{T(i)\}$, land quality, $A(i)$ and land $\bar{L}$; (ii) the land market clears in each location $\bar{L}(i) = N(i) \int_{\underline{s}}^{\bar{s}} \ell(s,i) dF(s,i)$; (iii) the economy-wide labor market clears $\int_i N(i) di = N$.

2.4 Calibration

The spatial unit of observation in the model is a district (or woreda) in the Ethiopian data. This is the most disaggregate level for which a reliable panel of agricultural production and geographic data could be constructed. My main strategy is to calibrate the benchmark economy to the pre-climate change economy[5] at the Ethiopian district-level. Parameters to be calibrated to match the spatial agricultural production structure of the Ethiopian economy are: the vector of farmer productivity $\{\int_{\underline{s}}^{\bar{s}} s dF(s,i)\}_{i=1}^{M}$; the technological parameter γ; vector of transportation costs $\{T(i)\}_{i=1}^{M}$; labor mobility barriers $\{\varepsilon(i)\}_{i=1}^{M}$; location specific productivity $\{A(i)\}_{i=1}^{M}$; total agricultural land $\{\bar{L}(i)\}_{i=1}^{M}$ and labor N.

Calibration of this paper does not rely on parametric assumptions about the distributions from which productivities and distortions are drawn. Instead, land quality before and after climate change is directly represented by geographic measures of GAEZ potential yields, which are then combined with location-specific agricultural production data, labor data, and transport costs. Using the merged data, I directly estimate the unobserved farmer ability for each household and aggregate them at the district-level

[5]Pre-climate change economy is calibrated using the baseline year (1961-1990) from GAEZ and pooled household data from AgSS.

to solve for the model. In the data M = 542, which includes the districts for which agricultural production data, transport cost data, and land quality data are available. Then the model equilibrium is calibrated at the district-level using these aggregated data at the district-level.

Farmer Productivity In order to calculate individual farmer ability s, I use combination of micro data (output as y, district land quality as $A(i)$, and total agricultural land as ℓ) implemented in the agricultural production function. Output y and total agricultural land ℓ are pooled from AgSS for years 2012, 2013 and 2014, and the land quality $A(i)$ is represented by GAEZ potential yields aggregated at the district-level using ArcGIS. The span-of-control parameter γ is set to 0.38 to closely represent the land income share suggested for Malawi in Restuccia and Santaeulalia-Llopis (2017) because Ethiopia and Malawi have much similarity in their economic characteristics. Directly combining these data with the agricultural production function gives farmer ability s for each of the households[6]. Then, for each district, I take the average of abilities of farmers in that district to estimate the district-level TFP.

Agricultural land by location/total labor The total amount of land for each location $\bar{L}(i)$ is taken directly from the Ethiopan Agricultural Sample Survey (AgSS) as a sum of total agricultural land across all households. Total amount of labor in the economy N is normalized to one for simplicity, as whether $N(i)$ is a share or a number does not have an impact on other parameter values.

Labor Mobility Barrier Labor mobility barrier $\varepsilon(i)$ in each district is chosen such that the share of farmers locating to district i in the model equation (12) exactly matches the year 2007 population shares data from IPUMS. Appendix B shows the calibration of $\varepsilon(i)$.

Transportation costs by location The transportation cost is estimated with travel times from district centroid to the nearest grain market within Ethiopia through the existing road infrastructure network measured from the geographic analysis for 2014. Since the model has iceberg transport costs, I map the travel times associated with transportation into iceberg transport costs following Adamopoulos (2020). The time costs are mapped into goods costs via the following function:

$$T(i) = 1 + \psi \cdot (tt_i)^{\eta},$$

where $\psi > 0$ is a scale parameter regulating the level of the implied iceberg transport costs, tt_i is the travel time (in minutes) from location i to the nearest grain market, and η captures the sensitivity of iceberg transport costs with respect to travel time tt_i. I use the elasticity η estimated at 0.8 by Combes and Lafourcade (2005) across French districts in 1998. For the scale parameter ψ, I use the scale parameter for cereals estimated by Adamopoulos (2020) such that the amount of resources devoted to transport matches the share of transportation costs in the data for Ethiopia.

Location Productivity/Land Quality Location specific productivity $A(i)$ is directly represented by GAEZ potential yields aggregated at the Ethiopian district (or woreda) level. GAEZ provides potential

[6]Farmer ability s is directly estimated from $y = A(i)s^{1-\gamma}\ell^{\gamma}$, which can be re-arranged as a function of known variables: $s = \left(\frac{A(i)\ell^{\gamma}}{y}\right)^{\frac{1}{1-\gamma}}$.

yields of different crops for each plot of land identified by the GPS coordinates. In the benchmark economy, I aggregate the baseline potential yield at the pixel-level for 30 crops to the IPUMS 2007 woreda borders, by taking the average of the land productivity measured by Geary-Khamis dollars per hectare of land. Table 2.1 reports the descriptions and values of the economy-wide parameters that are common across all locations. Table 2.2 shows the parameters that are mapped into actual district-level data as well as their descriptions and data targets.

Table 2.1: Common Parameters

Parameter	Description	Value
γ	Agricultural land income share	0.38
η	transport costs to travel time	0.8
ψ	transport cost scale parameter	0.00258
N	Normalization	1

Table 2.2: District-specific Parameters

Parameter	Description	Target Data
$A(i)$	Location-specific Productivity	GAEZ potential yields
$T(i)$	Iceberg transport cost	travel time to nearest grain market
$\bar{L}(i)$	Total agricultural land	total land from AgSS Data
$\varepsilon(i)$	Labor mobility barrier	IPUMS 2007 district population shares
$s(i)$	Farmer productivity	Production & Transport cost data

2.5 Quantitative Experiments

My objective is to study the impact of heterogeneous climate change across space in the presence of transportation cost frictions and labor mobility barriers. The main quantitative experiment involves "shocking" districts in Ethiopia with heterogeneous changes in climate predicted by natural scientists, as manifested by GAEZ's predicted potential yields for 2050. In this experiment, to isolate the effects of climate change, all other parameters including frictions, are held constant to the calibrated economy before climate change. This allows us to observe the aggregate and spatial micro-level effects on the Ethiopian economy as the only change between now and the future would be coming from differences in land quality. I do this by replacing the distribution of land quality for each district in Ethiopia $\{A(i)_{baseline}\}$ to future potential yields from GAEZ $\{A(i)_{2050}\}$ based on projected climate change by climatologists[7]. While keeping the labor mobility barrier ε at the benchmark economy level, the share of farmers $N(i)$[8] and price is allowed to vary in response to the change in land quality parameter $A(i)$, in order to track the change in the distribution of farmers across space.

[7]Change in land quality is directly represented by the change in GAEZ potential yield aggregated at the Ethiopian district-level.

[8]$N(i)$ represents the share of farmers in each district because the total number of farmers N is normalized to one.

This section reports the results on this counter-factual experiment regarding future climate change at year 2050 with rain-fed water supply and intermediate inputs to closely match the agriculture of Ethiopia. According to Nakicenovic et al. (2000), emission scenarios are associated with extreme uncertainty; therefore, there is no clear "benchmark" scenario for future predictions. Therefore, similar to Costinot et al. (2016), I choose Hadley year 2050 CM3 A1FI model with carbon dioxide fertilization as the future scenario, as it provides potential yields that are in the average range of other scenarios.

Table 2.3: Aggregate Effects of Climate Change

	All Crops (%)	Cereals (%)	Non - Cereals(%)
	Aggregate Statistics		
Land Quality (A)	47.9	9.3	46.6
Real Yield (Y/L)	62.3	14.8	53.7
Labor Productivity (Y/N)	53.2	12.6	63.7
Profit (π)	53.1	12.6	63.6

Note: Potential yield is changed from baseline (1961-1990) to year 2050 Hadley CM3 A1FI scenario allowing for CO_2 fertilization. The numbers represent percentage increase from the benchmark economy to the counter-factual experiment.

Aggregate Effects In Table 2.3, I can see that from the baseline (1961-1990) to year 2050, there is an increase in aggregate GAEZ potential yield A measured by climatologists. This exogenous improvement in overall land quality, feeds into an increase in overall yield and profit through the lens of the model. Result shows that the productivity increases more than the land quality; this is because of the reallocation of farmers towards the locations where productivity has gone up due to climate change. Also, from running the quantitative experiment three times: for all crops, for cereals only, and for non-cereals only, I observe that the most of the productivity increase was dictated by non-cereals. This is most likely because cereals generally tend to do better in moderate temperatures, where some non-cereals including oil crops and cotton requires high temperature and bright sunshine for its growth.[9] The aggregate results shows a large difference in productivity increase between the crop types in response to climate change; therefore, it would be interesting to add farmer's crop-selection into the model for stronger analysis.

Spatial Effects The main focus on this research is the impact of climate change on the spatial distribution of agricultural production activity. Table 2.4 shows the effect of climate change on the district-level statistics. An important take away from Table 2.4 is that climate change leads to a higher dispersion of yield and land quality across the districts. Increase in standard deviation of yield and land quality over the few decades indicates that there will be more agricultural productivity inequality between locations. This mainly happens due to the large temperature disparity between locations[10], where higher temperature regions experience decrease in land quality while relatively lower temperature regions experience increase in land quality.

[9] From the baseline year to year 2050, mean annual temperature of Ethiopia has increased for over 2.5°C providing better growth condition for crops that can endure hotter temperatures.

[10] Ethiopia is a country with high temperature dispersion between the locations. Mean annual temperature between the highest and lowest temperature is around 20°C.

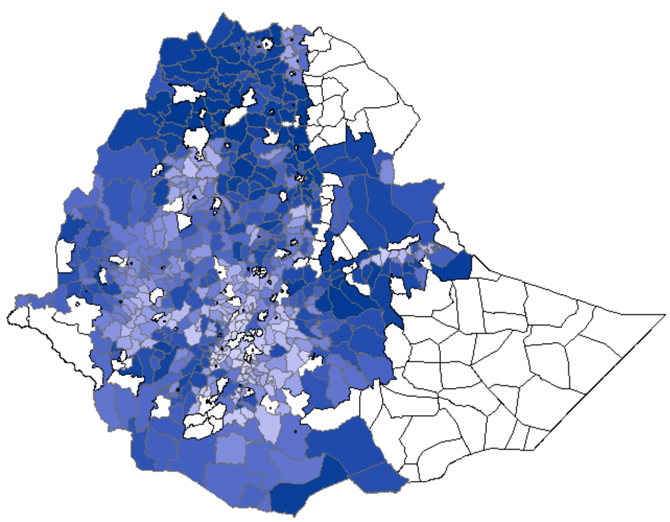
Table 2.4: District-level Effects of Climate Change

	Benchmark Economy	Year 2050 Land Quality
District-level Statistics		
STD of log-Yield	1.12	1.45
STD of log-Land Quality	0.69	0.78
CORR of log-(Yield, Land Quality)	0.07	0.36
CORR of log-(Yield, Transport Costs)	-0.3	-0.39

Note: Results are for all crops (cereal + non-cereal).

The correlation of log yield and log land quality is close to zero in the benchmark economy, but there is a large increase in correlation between yield and land quality after climate change. This is because from the model, farmers migrate to districts that have high land quality in response to climate change resulting in increased yields[11]. As expected, the lower transportation cost is correlated with higher yields in the benchmark economy, but the negative correlation gets stronger after climate change.

Figure 2.1: Change in Farmer Distribution across space



Notes: Above map shows the change in farmer shares in each district after climate change. Climate change scenario is based on year 2050 Hadley A1FI with CO_2 fertilization.

The main channel through which adaptation to climate change occurs is labor mobility across districts. Figure 2.1 helps visualize the farmer distribution in Ethiopia before and after climate change. Blue gets darker when farmers migrate into the district, and gets lighter as farmers move away from the district. Hollow (or white) districts are those that were omitted from the experiment due to data availability. Figure 2.1 clearly shows that there is farmer migration in response to climate change. Farmer shares are highly correlated to land quality and transportation costs, so the districts that gain farmer shares (darker blue)

[11]Correlation coefficient between $\Delta N(i)$ and $A(i)_{2050}$ is also extremely high at 0.94.

are those exposed to higher relative gains in land quality after climate change.

The standard deviation from Table 2.4 shows higher dispersion of yield and land quality after climate change. To further analyze spatial inequality, I divide the districts into quintiles based on district-specific characteristics. In first column of Table 2.5, Panel A, I order districts according to their iceberg transportation costs and group them into quintiles. The second column shows the percentage change in land quality from the aggregated GAEZ potential yields data, when the time period changes from the baseline to year 2050. The third column shows the percentage change in yield from the model in response to the climate change. Panel A shows that there is a larger increase in land quality for district that have better transport infrastructure (Q1) and the increase in yield for these districts are far higher than the districts in other quintiles. The districts that have poor transport infrastructure (Q5) even show decrease in yield after climate change, because farmers move away from these districts into those that have better land quality and transport infrastructure. The farmer migration due to climate change promote further inequality in agricultural production, but could serve as a viable adaptation technique if more productive districts can be supported to be specialized in agriculture.

Table 2.5: Spatial Inequality from Climate Change

Panel A : Ordered according to $T(i)$			
Quintile	Transportation Cost	% Change in Land Quality	Model % Change in Yield
Q1	1.06	72.8	142.2
Q2	1.11	54.5	75.3
Q3	1.16	41.0	35.3
Q4	1.22	27.4	5.7
Q5	1.43	24.3	-2.1
Panel B : Ordered according to $Y(i)$			
Quintile	Initial Yield	% Change in Land Quality	Model % Change in Yield
Q1	3.8	33.3	31.4
Q2	10.1	43.1	55.4
Q3	17.7	41.0	35.0
Q4	29.3	51.0	65.3
Q5	67.7	52.0	67.2
Panel C : Ordered according to $N(i)$			
Quintile	Initial Farmer Share(%)	% Change in Land Quality	Model % Change in Farmer Share
Q1	0.07	35.5	-14.5
Q2	0.12	46.2	0.9
Q3	0.17	46.5	4.6
Q4	0.23	46.7	1.8
Q5	0.33	45.4	-2.9

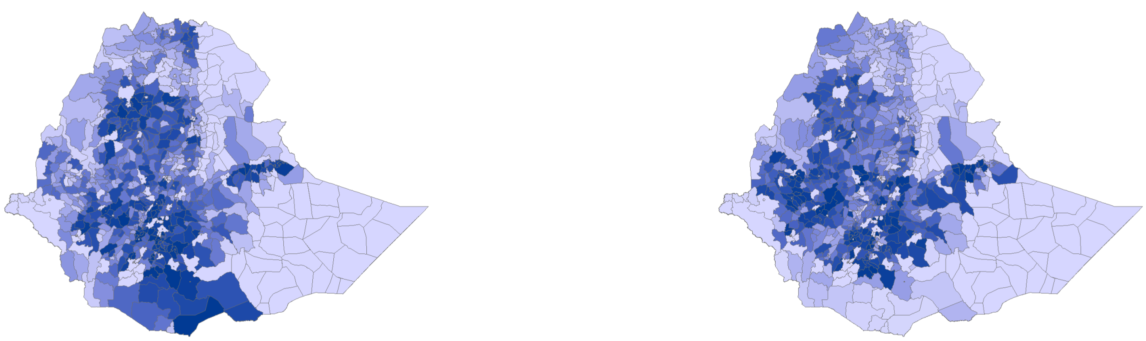
Note: The results in each Panel order districts into quintiles based on (A) transportation costs, (B) initial (before climate change) model yield, and (C) initial (before climate change) farmer share.

Panel B of Table 2.5, order districts according to their initial yield in the benchmark economy (1961-1990), and columns 2 and 3 show % change in land quality and yield when the land quality is changed to year 2050 level. Although the process is not entirely monotonic, land quality and yield generally show higher percentage increase for the districts that already had higher yields in the benchmark economy. Again, this shows an increasing wealth gap between the districts, since those that were doing well in years 1961-1990 (Q5) experience larger percentage increase in year 2050. Panel C orders districts according to the farmer share in the benchmark economy, and the results do not show any clustering migration pattern for districts that already had high share of farmers (Q5).

Figure 2.2: Farmer Distribution without Frictions

Equal Transportation Costs

Equal Mobility Barriers



Notes: Equal transportation cost shows farmer share in each district when transportation dispersion between the districts are removed. Equal mobility barriers shows farmer share in each district when the dispersion of mobility frictions are removed after climate change.

Figure 2.2 shows the farmer distribution in Ethiopia after removing the dispersion of transportation and mobility frictions in each districts. There are clear differences in the distribution of agricultural production when we remove the dispersion in transportation costs and mobility barriers. The standard deviation of log farmer distribution between the districts was around 0.56 for the benchmark economy. After the climate change, the standard deviation increased to 0.93, showing further deviation between the districts. Removing the dispersion of transportation costs after climate change decreased the standard deviation from 0.93 to 0.85; and removing the dispersion of labor mobility barriers increased the standard deviation from 0.93 to 2.86. As expected, lack of dispersion in mobility frictions caused farmers to migrate away from a less productive to a more productive district causing higher disparity in share of farmers between locations within Ethiopia.

2.6 Conclusion

Migration is one of the most effective adaptation methods to reduce the climate change risks, especially to developing countries (Vinke et al., 2022). Climate change pushes farmers to migrate to locations with higher land quality, but those locations are often linked with heterogeneous transportation costs and labor mobility barriers. Therefore, this paper studies the interaction of climate change with transport frictions to determine the long-run distribution of farming activity in Ethiopia. To measure the impact of climate change, I combined a quantitative spatial framework with novel data on land quality, agricultural production and transportation costs.

The aggregate increase in land quality in Ethiopia predicted by climatologists is expected to lead an even larger increase in output yields in Ethiopia due to the reallocation of farmers across space. Most of those increases were from non-cereals which are more resilient to higher temperature than the cereals showing a high disparity on the crop-specific impact of climate change. AgSS provides household production data for individual crops and GAEZ also provides potential yield data for individual crops. Therefore, quantification of the crop-specific impact of climate change with a choice in crop production interacted with location-specific transportation costs would be an interesting addition to this research. When I zoom in to the district-level, I found that the inequality between the districts intensify after climate change. This was mainly due to the fact that the percentage increase in land quality were generally higher for the districts that were already producing more agricultural yields. And I found that the regions with low transportation costs experienced higher increase in land quality and yields also further promoting agricultural production inequality between the districts.

Climate change is an on-going process which is predicted to have a large impact on the agricultural sector. Therefore, as a country that is heavily skewed towards agriculture, it is crucial for Ethiopia to understand the mechanism behind climate change and its impact on farmer migration. Climate change will push farmers to migrate to locations that have better land quality and low transportation costs. Another interesting aspect regarding climate would be to analyze climate shocks and its impact on farmer distribution in addition to long-term change in land quality.

Chapter 3

Sources of Innovation: A Quantitative Analysis with Korean Data

3.1 Introduction

Productivity is important for understanding economic disparities across countries, across sectors, and across firms; and innovation is one of the crucial factors that determine productivity growth (Syverson, 2011). Studies surrounding innovation depict it as having many different forms; but there are three major types of innovation that are studied the most. First, innovating firms can improve on existing products made by other firms; such "creative destruction", pioneered by Schumpeter et al. (1939), led to other researches conducted by Stokey (1988); Aghion and Howitt (1990); Klette and Kortum (2004). Second form of innovation is shaped by firms improving on existing products made by their own firms; this form of 'own-innovation' was emphasized by Krusell (1998) and Lucas Jr and Moll (2014). Innovation can also come from brand new varieties; this form called 'new-variety innovation' was mainly researched by Romer (1990); Acemoglu (2003); Jones (2016). All of these previous researches that depict innovation as either one of three different forms have different implications for innovation policies. However, their rich structure presents an obstacle in trying to determine the intensity on how each of the different sources of innovation contribute to economic growth.

A model from "how destructive is innovation", by Garcia-Macia et al. (2019) (henceforth GHK), has an advantage over previous researches by providing a parsimonious model that includes innovation as an exogenous parameter. By classifying innovation rates exogenously, inferred parameter values from calibration allows a comparison of intensity between each sources of innovation. This process is executed through matching each of the population data moments to the model driven moments which allows the calibration for each of the innovation arrival rates as parameters classified as: own-variety innovation by incumbents, creative-destruction by incumbents, creative-destruction by entrants, new product innovation by incumbents and new product innovation by entrants. The empirical identification of the different sources of innovation exploits the fact that the different sources of innovation have different implications for job reallocations across firms at the micro-level. Through the model, the sources of innovation are backed out from observed patterns of job creation and job destruction.

Relatively recent researches including Klette and Kortum (2004) or Atkeson and Burstein (2019), models innovation by endogenizing investment in research and development (henceforth, R&D). The innovation models from these papers assume that spending on R&D increase the chance of innovation arrivals and constitutes for one of the main factors in economic growth. Although GHK disregards endogenizing the role of spending on R&D in their model, mainly to follow parsimonious modelling method in determining each innovations' contributions on economic growth, there are numerous researches surrounding the importance of investments in R&D to bring innovation within firms.

The focus of the GHK study is the universe of firms in the United States; thus, extending the analysis of identifying sources of innovation to individual sectors, other time periods, and countries, would serve as an interesting topic. In this paper I use the same quantitative framework as GHK, but with unique micro-level data from Korea, a country with substantial productivity growth and deliberate efforts towards innovation. I also compare the results in Korea to those in the United States. The results show that while most of the innovation is accounted by incumbents and own-variety improvements in both countries, these sources are more pronounced in Korea. Among entrants, most of the innovation comes through creative destruction in Korea, while through new varieties in the US. Also, creative destruction accounts for the same of overall innovation in the two countries. More results will be explained in detail on section 5.

The rest of the paper proceeds as follows. Section 2 explains why this study is conducted around Korean data. Section 3 lays out the parsimonious exogenous growth model originally by GHK. Section 4 compares the data moments from the U.S. LBD and Census on Establishments from Statistics Korea. Section 5 presents the details on the calibration of the model with inferred parameter values for U.S and Korea. Section 6 concludes.

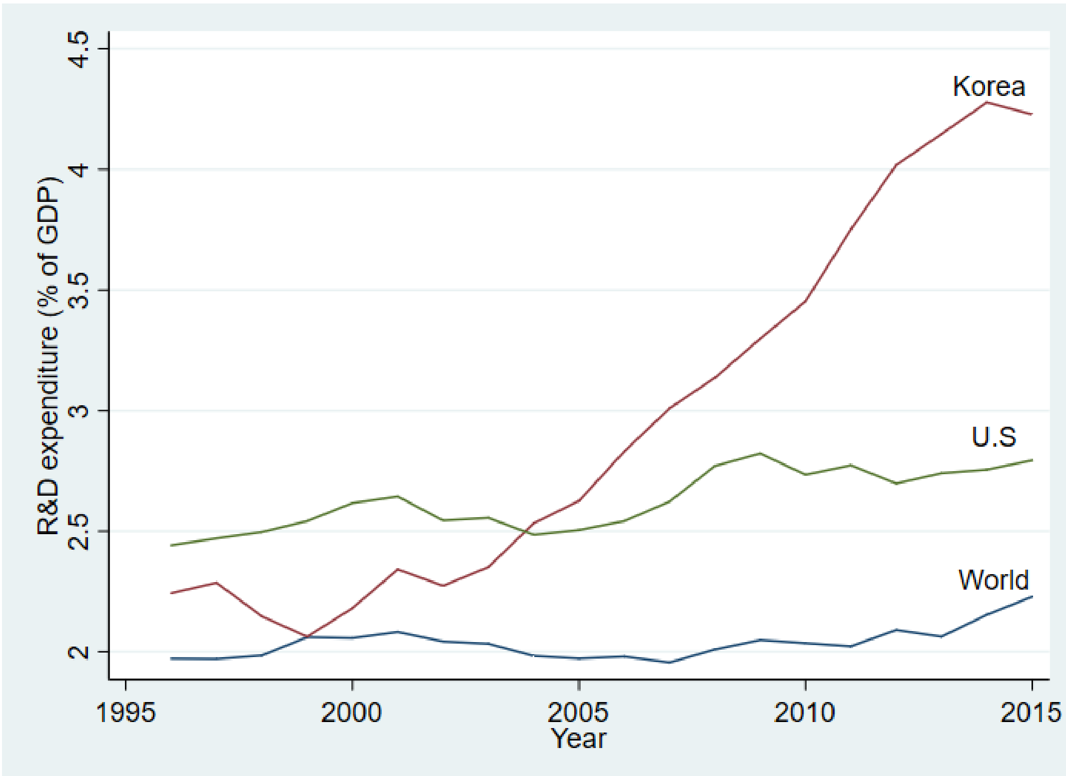
3.2 Why Korea?

Figure 3.1 shows that since the late 1990s, R&D expenditure as a share of GDP has more than doubled in South Korea, while the US and world average share have remained roughly constant. Given the assumptions from endogenous innovation models that investments in research brings innovation, it seems essential for Korea to correctly identify the sources of each innovation to maximize efficiency from various policy implications.

South Korea is a great example of a country that has achieved sustained improvements in labour productivity – a key factor behind escaping the middle-income trap. Korea's GDP which was just $2.7 billion in 1962 took a leap and attained a high value of $230 billion in 1989 at an enormous growth rate of over 8 percent per year. Furthermore, data from GGDC 10-sector database shows that since the early 1960s, real value added per worker (in constant 2005 prices) in manufacturing has increased 22-fold, with an average annual growth rate of 5.1% between 2000 and 2010. How has Korea achieved such phenomenal growth in just over a four decades? While earlier growth accounting shows that South Korea's remarkable growth in the post-war period was driven by inputs (Young, 1994), recent work indicates that since the late 1990s, South Korea's growth is increasingly accounted for by TFP (Pyo et al., 2008). In addition, Bartzokas (2008) have pointed out to a strong Korean national innovation system (NIS) as one of the important factors of Korea's economic growth. He states that private industries and government-sponsored research institutes (GRIs) have played important roles in Korea's economic development and have emerged as strategic partners in the local R&D "policy mix". One of the key reasons to how a small country with limited inputs into production has become prosperous over the recent years is believed to be from high productivity coming from constant technological innovation.

Chung (2011) argues that science, technology, and innovation has been cited as the key factors behind

Figure 3.1: R&D expenditure trend from 1996-2015



Source: The data is taken from The World Bank for years 1996-2015 on Korea, U.S, and World. Y axis is the percent spending of GDP on research and development for each subjects.

the economic success of Korea. By making continuous and massive investments in innovative R&D, Korea has succeeded in building a unique innovation system that supports sustainable economic growth. Figure 3.1 supports this statement by showing a steep increase in research expenditure. Since spending on R&D is on a steep increasing trend during the target period of my research (2001-2012), identifying Korea's sources of innovation would provide an interesting comparison against United States.

3.3 Model

The aggregate growth model of this paper follows a same model from "How destructive is innovation" by GHK. In the model, the static equilibrium derived from profit maximization of a CES output function, determines the equilibrium values necessary to derive productivity in the model. Then, the aggregate growth rate of the wage and output per worker (productivity) is presented as a function of exogenous innovation parameters.

There are four main assumptions underlying the model, first and second assumptions are that there is a constant exogenous arrival rate for each type of innovation, and that innovation arrives proportional to the number of products owned by a firm. Third assumption is that the quality of an improved product through innovation is built on the existing quality level of the products, and is drawn from a Pareto distribution with shape parameter θ and scale parameter equal to the existing quality level. This means that the average improvement in quality from innovation on existing products is: $s_q = \left(\frac{\theta}{\theta-(\sigma-1)}\right)^{\frac{1}{\sigma-1}} \geq 1$, where σ is the elasticity of demand for quality. The average quality of a brand new product s_κ, can be better or worse than the quality level of the existing products, therefore, s_κ could be greater or less than 1. Lastly, entrants will have one product that is either taken from another firm or created as a new variety, but incumbent firms can have multiple products.

3.3.1 Static Equilibrium

The only production unit in this growth model is a firm that requires labour as only source of input. The aggregate output is a CES combination of quality-weighted varieties:

$$Y = \left[\sum_{j=1}^{M}(q_j y_j)^{1-\frac{1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

here, subscript j represents individual variety, σ is elasticity of demand for quality and M represents aggregate number of varieties. q_j denotes the quality of variety j and y_j the quantity. Output of variety j is given by $y_j = l_j$, where l_j is amount of labour used to produce variety j.

Bertrand competition allows the firm with a highest quality product to charge a mark-up of $\frac{\sigma}{\sigma-1}$ and each firms may produce multiple varieties. Therefore, under market structure for interim goods with monopolistic competition subject to a competitive fringe,

$$P_j = \mu MC_j = \mu W, \qquad \mu = min\left\{\mu^2, \frac{\sigma}{\sigma-1}\right\} \tag{2}$$

where, μ becomes the optimal mark-up under Bertrand competition and a competitive fringe of firms.

After normalizing the price of aggregate output to 1 and solving for profit maximization problem of a firm, a total employment of a firm L_f is given by:

$$L_f \equiv \sum_{j=1}^{M_f} l_j = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} LW^{1-\sigma}\sum_{j=1}^{M_f} q_j^{\sigma-1} \tag{3}$$

where, W is the wage, L is the aggregate employment and M_f is the number of varieties produced by firm f. From equation (3), it is important to note that firm employment is proportional to $\sum_{j=1}^{M_f} q_j^{\sigma-1}$, which means that larger firms produce more varieties of products with better quality. However, the above statement is satisfied if and only if the demand elasticity for quality, σ is above 1. In the case of Klette and Kortum (2004), where product quality is not an issue, σ=1 and the level of quality would not be a factor in the equation. Therefore, to account for product quality in the model, it is important to allow $\sigma > 1$; in

Table 3.1: Innovation Parameters

Channel	**Probability**
Own-variety improvements by incumbents	λ_i
Creative destruction by entrants	δ_e
Creative destruction by incumbents	δ_i
New varieties from entrants	κ_e
New varieties from incumbents	κ_i

Note: The average step size for quality improvements for own innovation and creative destruction, weighted by employment, is $s_q = \left(\frac{\theta}{\theta-(\sigma-1)}\right)^{\frac{1}{\sigma-1}} \geq 1$. The quality of new variety is drawn from the quality distribution of existing products multiplied by s_κ.

which the number of varieties produced by the firm and quality of those varieties will determine the level of employment hired by a firm.

Estimating for the wage level W, and rearranging yields:

$$W \equiv \frac{\sigma-1}{\sigma}\frac{Y}{L} = \frac{\sigma-1}{\sigma}M^{\frac{1}{\sigma-1}}\left[\sum_{j=1}^{M}\frac{q_j^{\sigma-1}}{M}\right]^{\frac{1}{\sigma-1}} \tag{4}$$

Since the inverse of the mark-up is a fixed constant, the aggregate growth in the model is equivalent to the wage rate and comes from the creation of new varieties, $M^{\frac{1}{\sigma-1}}$ and growth in average quality, $\left[\sum_{j=1}^{M}\frac{q_j^{\sigma-1}}{M}\right]^{\frac{1}{\sigma-1}}$.

3.3.2 Innovation Parameters

To calibrate the static equilibrium levels, each type of innovation rates are treated as exogenous parameters and are defined as the probability of each innovation arriving. Time is discrete and each type of innovation arrival rates increases as the number of products owned by a firm increases. Assigned notations for each type of innovations are shown in Table 3.1 as follows: λ_i is the probability that an incumbent innovates on their own product. So, if incumbents would fail to innovate on their own product, then they would be subject to creative destruction by other firms. δ_i is the probability that a product is innovated by other incumbents conditional on not being innovated by the owner of the product (creative destruction by other incumbents). δ_e is the probability that a product is innovated by an entrant conditional on not being innovated by the owner of the product (creative destruction by entrants). κ_i is the rate at which new varieties are created by incumbents and κ_e is the rate of new variety creation by entrants.

In addition to these five channels of innovation, $\tilde{\delta_i} \equiv \delta_i(1-\lambda_i)$ is the unconditional probability of innovation by another incumbent, which shows a probability that an incumbent will creatively destroy other firm's product when the owner of the product fails to innovate their own product. $\tilde{\delta_e} \equiv \delta_e(1-\delta_i)(1-\lambda_i)$ is the unconditional probability of innovation by an entrant and shows a probability that an entrant creatively destroys other incumbents' product when the owner of the product fails to innovate their own product and other incumbents fail to creatively destroy the owner's product. Therefore, the probability of improvement of an existing product by any firm is $\lambda_i + \tilde{\delta}_i + \tilde{\delta}_e$.

It was mentioned that the quality of the products are drawn from a Pareto distribution and is improved upon the existing quality of the product. The average improvement in quality of an existing product is $s_q > 1$ and the quality of a brand new product is s_κ multiplied by the quality distribution of existing products, s_κ could be greater or less than 1. The overhead labour costs are the last parameters introduced in the model. Derived from the value function, the overhead cost determines the cut-off quality of varieties which determines the threshold where firms decide to exit if there are present discounted value of profits. Therefore, regarding overhead costs would result in the net growth rate of varieties to be $\kappa_e + \kappa_i - \delta_o$.

3.3.3 Sources of Growth

Appendix C provides detail on the derivation of a growth function from the static model. Using the static equilibrium and the parameters described above, the growth rate of the wage and output per worker can be expressed as:

$$\mathbb{E}\big[(1+g)^{\sigma-1}\big] = 1 + \underbrace{s_\kappa\big(\kappa_e + \kappa_i\big)}_{\text{new varieties}} + \underbrace{\big(s_q^{\sigma-1} - 1\big)\lambda_i}_{\text{own innovation}} + \underbrace{\big(s_q^{\sigma-1} - 1\big)\big(\tilde{\delta}_e + \tilde{\delta}_i\big)}_{\text{creative destruction}} - \delta_o\psi \tag{5}$$

here, the first term, shows that the innovation rates of new varieties paired with the quality change depicts the contribution of new variety innovation in aggregate growth. The second term, captures the contribution of own-product innovation paired with the improvement in average quality and the last term, similarly captures the contribution of creative destruction.

Rearranging equation (5) would give:

$$\mathbb{E}\big[(1+g)^{\sigma-1}\big] = 1 + \underbrace{s_\kappa\kappa_e + \big(s_q^{\sigma-1} - 1\big)\tilde{\delta}_e}_{\text{entrants}} + \underbrace{s_\kappa\kappa_i + \big(s_q^{\sigma-1} - 1\big)\Big(\lambda_i + \tilde{\delta}_i\Big)}_{\text{incumbents}} - \delta_o\psi \tag{6}$$

now, it is possible to express growth in terms of entrants vs. incumbents rather than a function of three types of innovations. The notational interpretation is similar to equation (5). Therefore, the innovation probabilities $(\kappa_i, \kappa_e, \delta_i, \delta_e, \lambda_i)$ and quality steps (θ, s_κ) determine the aggregate growth rate either as a function of each sources of innovation or as entrants vs. incumbents. By using patterns in the U.S. LBD and statistics Korea census on establishments micro data, each of the parameter values will be estimated. Then the parameters will be used to determine the share of growth for three types of innovation as well as the contribution of incumbents vs. entrants.

3.3.4 Dynamics

The model's static equilibrium points out that a firm's employment depends on the number of varieties created by a firm and the qualities associated with each of those varieties. Therefore, I can examine a firm's change in employment by analysing the transition dynamics of quality and quantity of the products produced by a firm. As mentioned in the previous subsection, innovation arrival rates dictate the transition dynamics of varieties and those determine the changes in a firm's level of employment. Entrants can enter the market by either creatively destroying an existing variety by an incumbent or by creating a brand new variety to the market. Incumbents that experience a successful innovation either by creatively destroying other firms' product, improving upon its own product, or creating a brand new variety, will grow in employment. Incumbents that lose their product to another incumbent or an entrant, will decline in employment. Incumbents will exit when all of their varieties are lost.

The polar models in GHK with only one source of growth, provide traits for each of the different cases to identify the moments necessary for determining exogenous innovation parameters. The polar models show that: the distribution of job creation and destruction dictates the change in number of varieties across firms, growth in firm employment is driven by the accumulation of varieties, exit rates of a firm decline sharply with size, and the employment share of entrants is positive and is pinned down by the rate of creative destruction by entrants. To infer new variety creation from the data moments, growth in the number of varieties can be interpreted from growth of total employment. The quality of new varieties and of whom the principal producer is, would be determined from average employment between incumbents and entrants.

Taking reference to GHK's polar model cases and setting the minimum employment of an establishment to be equal to 1, moments that need to be satisfied are:

1. Aggregate Total Factor Productivity (TFP) growth rate
2. Standard Deviation of log employment across plants
3. Job creation rate
4. Job destruction rate
5. Share of job creation due to plant employment growth ≤ 1
6. The employment share of entrants
7. Exit rate of large/small establishments
8. Average employment for incumbents relative to entrants

3.4 Data

One thing different about this paper in relation to GHK is that instead of using firm-level data, this research is conducted at the establishment-level. Although business stealing can be an issue when the research is focused at the establishment-level, there are results including Pyo et al. (2016) suggesting that establishment level and firm level data do not show statistically significant differences in relation between job creation and sizes. Therefore, as for the population of U.S establishments, establishment-level data on employment from the U.S. Census Longitudinal Business Database (LBD) is used. LBD is a comprehensive database of all employment records of U.S. business establishments which provides annual data of business variables such as job creation, job destruction, exit rates, entry rates, etc. However, two moments: standard deviation of log employment across firms, and share of job creation due to firm employment growth < 1, is taken from GHK's organized firm-level data, since these cannot be computed using the publicly available data and would not show much difference between firm-level and establishment-level.

Census on Establishments (henceforth, CE) from Statistics Korea provide the Korean data for the research. Although it does not contain annual longitudinal data as in U.S. LBD, they provide raw data for census on all the individual establishments with industrial classification, established year and month, number of employees, etc. In order to connect each establishments by year, each establishment's ID is used; upon request, this data is only accessible in the private server under supervision of Statistics Korea. Using establishment ID and level of employment data provided by CE, the data is organized into a panel to estimate each of the data moments that were mentioned in the previous section. Data provided by Statistics Korea has fewer time periods; thus throughout my research, I have only used the census data on years 2001 to 2012 for both Korea CE and US LBD. Establishments in the public, educational, agricultural/fishery/forestry, and mining sectors are dropped.

In order to distinguish industries, KSIC 9th standard industry classification from Statistic Korea is used. Since 9th standard industry classification was implemented in 2006, any industry classification codes prior to the change were standardized into 9th classification and were matched accordingly. Raw data from Census on Establishment contains some minor errors such as missing/duplicate IDs, recording errors, etc. Therefore, observations with duplicate/missing IDs, establishments that were missing data in between years, plants that had been established more than 200 years ago were dropped. At the end, I have retained about 90% of the total number of establishments, so there weren't much loss of observations even after getting rid of the outliers. After restricting the industries and matching the establishments using establishment IDs, replication was done for all of the annual variables in LBD format using Korean data.

The main problem of comparing the two countries: U.S. and Korea, arises from the differences in data collection process of two institutions. During the data collection process, U.S. LBD excludes employee types of: self-employed, domestic service workers, most government employees, employees in foreign countries, etc, and Korea CE excludes: self-employed in agriculture and fishery industry, domestic service workers, street vendors, military personnel, employees in foreign countries, etc. The two exclusions have most in common, but the problem lied in the exclusion differences for self-employed workers. Since Korea CE only excludes self-employed workers for agriculture and fishery industry, there was a tendency to over-estimate number of establishments and under-estimate number of average employment per establishment compared to data provided from U.S. LBD in the other sectors. Therefore, I have dropped the self-employed workers in all other sectors as well so the exclusion criteria matched between U.S. and Korea.

When calculating establishment age from data, LBD measures age starting from the year an establishment enters the market; however, limited number of time periods in CE becomes on obstacle when following the same method. Therefore, for Korean establishments, I calculate their age by subtracting each observations' established year from the current year of data collection. This age measurement difference seems to be negligible since entrants and incumbents in both countries behave the similar way and aggregate rates show similar pattern between entrants and incumbent of both countries.

The research is done on an annual basis due to the short time period of the given data. Therefore, entrants are classified as an establishment that are less than one year old and incumbents are those that have existed in the market for at least one year. The raw data for both of the institutions are collected during February to March every year, so in order to maintain continuity in the data, growth rates are calculated by dividing the change over the average of the present and previous years' values.

Table 3.2: Summary Statistics for U.S and Korea

	Labour	Plants	Average Labour	SD of log Labour	Labour Growth	TFP Growth
U.S	114.68	6.59	17.41	1.26	0.13%	0.99%
Korea	14.73	1.99	7.40	0.91	0.01%	2.73%

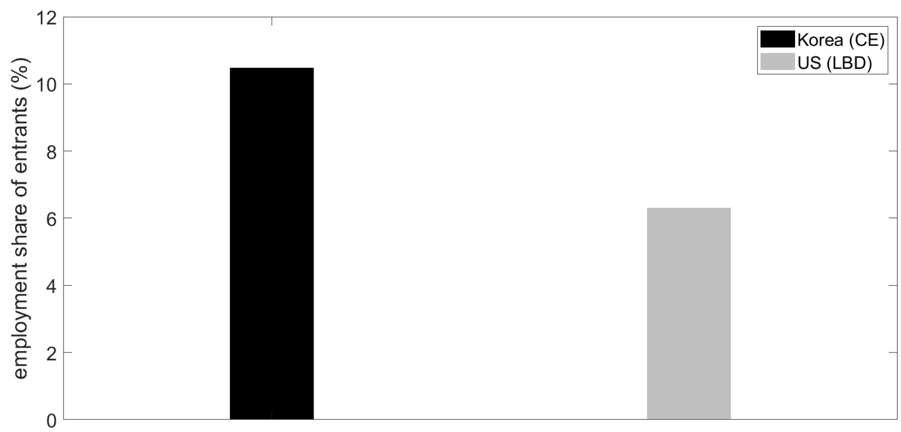
Source: U.S. Census Longitudinal Business Database (LBD) and Statistics Korea Census on Establishments (CE) for establishments in non-farm business sector. Total factor productivity (TFP) growth data is average of annual multi-factor productivity taken from OECD stats. Total employment and number of plants (establishments) are in millions and the summary shows average of each variables for 2001 to 2012 for both countries.

Table 3.2 shows summary statistics for establishments of United States and Korea after excluding public, education, agriculture/fishery/forestry, and mining sectors. The main difference between the two countries lies in the third column of average workers per establishment. Assuming that older companies generally hire more workers in average, such high difference in average number of workers in an establishment could be an indicator suggesting that mean age of the establishments are higher in US. Annual employment growth rate is calculated using job creation and job destruction rates; thus small difference in job creation/destruction rate is the reason behind Korea's near zero employment growth rate[1]. Taking account of the model where growth in aggregate number of varieties is dictated by new variety arrivals and the aggregate level of employment is determined by number of varieties and quality associated with those varieties, growth differences would be indicating that new variety arrivals are relatively more important in U.S. compared to Korea.

Figure 3.2 presents employment share of entrants which is the average of annual employment share of entrants during 2001-2012. The entrants are defined as plants that have been in the market for 12 months or less. Compared to Korea (around 10.5%), U.S (around 6.3%) has less share of entrants' employment, which in other words translates to higher incumbents' employment share. It is inferred from Figure 3.2 that entrants constitute for a small part of total employment in both countries. Also, the share of entrant establishments proportional to the total number of plants in Korea is 11.29% and for U.S is 9.7%. So for

[1] Annual employment growth rate = $(1+\mathrm{JC}-\mathrm{JD})^{\frac{1}{5}} - 1$

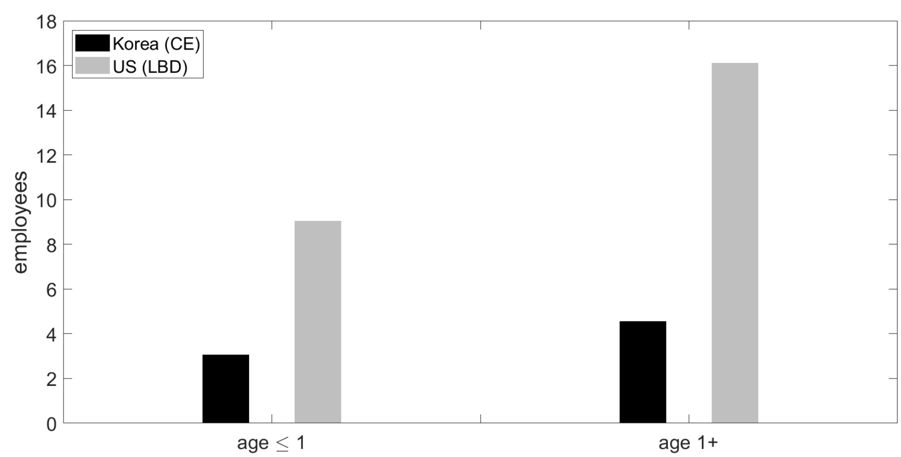
Figure 3.2: Employment Share of Entrants



Note: The employment share of entrants for U.S and Korea for years 2001-2012 is the average of annual share of entrant employment for 12 years taken from the U.S. LBD and Statistics Korea CE. Annual share of entrant employment is calculated by dividing employment by entrants from the total employment.

both countries, the data supports a common belief that entrants in average hires less workers compared to incumbents.

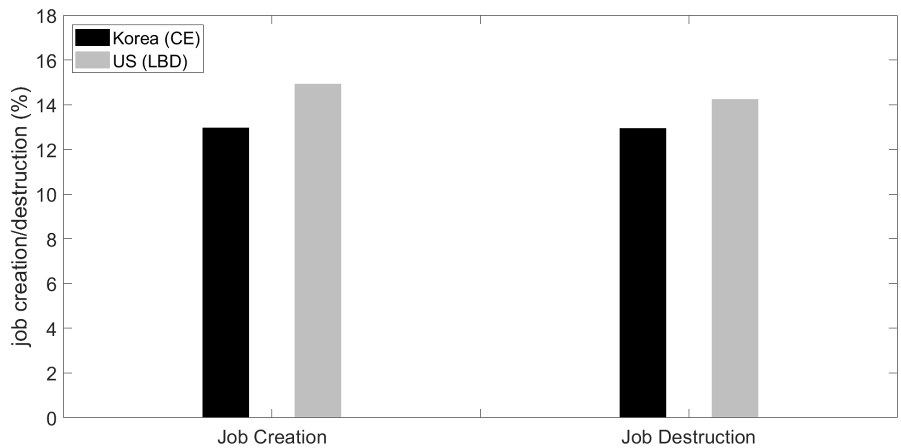
Figure 3.3: Employment per plant, Entrants vs Incumbents



Note: Left set of graphs represent entrants and right set of graphs represent incumbents. The figure shows average number of workers in each of the groups for 2001-2012.

Figure 3.3 shows that incumbents (older establishments) generally have more workers compared to entrants (young establishments) in both countries. This trend also shows in the model simulation, as companies that innovate tend to survive and grow larger and those that fail to innovate would exit the market.

Figure 3.4: Job Creation and Destruction Rates



Note: The job creation (destruction) rate is the sum of employment changes at establishments with rising (falling) employment divided by denom (average employment between time t and t-1). This figure shows average of annual job creation and destruction rates for 2001-2012 statistics taken from U.S. LBD and Statistics Korea CE.

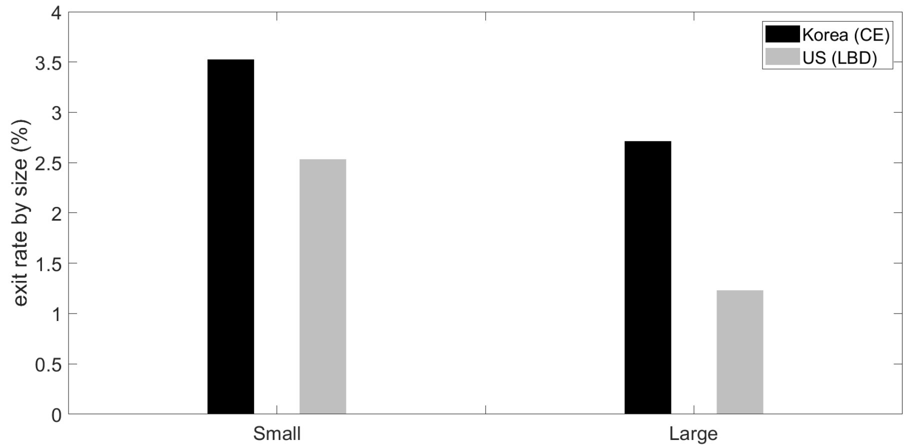
Figure 3.4 shows annual calculation of job creation and destruction rates. Job creation (destruction) rate is count of all jobs created (destroyed) within a plant over the last 12 months divided by denom[2], meaning that job creation records newly hired labour within the establishment. Job creation and destruction shows the population dynamics within establishments; therefore its values are highly influenced by all three of the innovation arrival rates as the change in level of employment in the model across time is dependent on number of varieties and qualities determined by the innovation arrival rates. The job creation rate is slightly higher than the destruction rate for both countries, but U.S shows around 0.7% higher reallocation rate (JC - JD) compared to Korea.

Exit rates for large and small establishments are compared between the two countries in Figure 3.5. Large establishments are defined as those with above average employment and small establishments have below average employment. In both countries, small plants have higher exit rates compared to large plants as the model would explain this trend since plants' employment largely depends on number of varieties and having more varieties would mean more employment and lower rate of exit, given that the exit condition is when plants lose all of their products.

Combining the information from Figure 3.4 with the fact that the fraction of employment size for incumbents/entrants are 1.49 for Korea and 1.78 for United States (older establishments are larger in size), shows that older plants would have lower exit rates compared to younger plants, which have been a robust trend in the U.S. Census of Manufacturing. Although the exit rates are lower for larger plants, the intensity

[2]Denom = Average of the values between periods t and t-1, so denom for job creation (destruction) rate would be average employment between years t and t-1.

Figure 3.5: Exit Rate, Large vs Small



Note: The exit rate is the annualized fraction of establishments that operated in year t-1 but not in year t (since the calculation is done for one year moments). The data shows percentage averages of annual exit rates from 2001 to 2012 for statistics from U.S LBD and Statistics Korea CE on non-farm business sector.

of the difference between small and large plants' exit rates differ between two countries. Figure 3.5 shows that larger plants in United States have much lower exit rate compared to Korea, while the difference in exit rates for smaller plants (1%) are lesser than the difference in exit rates for larger plants (1.5%). In other words, this could be translated to larger establishments in U.S having higher survival rates and surviving plants in U.S having faster growth. The tendency of U.S plants having a much larger size compared to some other countries as they age, is supported by Hsieh and Klenow (2014). The large gap of exit rates by size for United States compared to Korea, would also be one of the main factors that deter parameter estimations between the two countries.

3.5 Calibration and Results

Given the data moments described in the previous section, parameters are estimated by matching model simulations to the corresponding data moments. The parameters to be matched are 5 innovation rates ($\delta_i, \delta_e, \kappa_i, \kappa_e$ and λ_i), 2 quality step-size parameters (θ and s_κ), and the overhead cost. The target moments to be matched are aggregate TFP growth, minimum employment of 1, standard deviation of log employment, aggregate rates of job creation and destruction, share of job creation due to employment growth ≤ 1, employment share of entrants, exit rate of large/small establishments, and average employment for incumbents/entrants. Overhead cost pins down the exit rate of varieties due to overhead cost (δ_o) and the average quality of exiting varieties (ψ).

Conditional on the overhead cost, the 7 parameters jointly determine the productivity growth rate from equation (5) and (6), so the combination of the parameters are set such that expected TFP growth in the model exactly equals TFP growth from the data for both countries. The rest of the moments are matched to infer parameter values as in Table 3.3. Simulation algorithm for this paper follows a simulation method from GHK, and is summarized in Appendix C.

Table 3.3: Inferred Parameters Values

	Korea	U.S.
Creative destruction by incumbents δ_i	14.8%	54.3%
Creative destruction by entrants δ_e	100%	100%
New varieties from incumbents κ_i	0.0%	0.0%
New varieties from entrants κ_e	0.9%	2.1%
Own-variety improvements by incumbents λ_i	88.4%	86.0%
Exit probability from overhead costs δ_o	0.9%	1.4%
Pareto shape of quality draws θ	38.7	113.4
Relative quality of new varieties s_κ	0.52	0.41
Average quality of exiting products ψ	0.13	0.07

Table 3.3 records the inferred parameters from the simulation matching. For Korea, given θ of 38.7 and σ of 4, s_q=1.027 which represents 2.7% average improvement in quality when there is an innovation on an existing product.[3] $s_\kappa = 0.52$ shows the average quality of new varieties being introduced in the market are around 52% of the average quality of existing varieties. Given the values for the conditional creative destruction parameters, unconditional probability that a product improves from creative destruction by incumbents $\tilde{\delta_i}$ is 1.72% and by entrants $\tilde{\delta_e}$ is 9.88%[4]. The probability that an establishment will exit due to the overhead cost is 0.9% and regarding the average quality of exiting products ψ, the impact of overhead cost on the aggregate growth only accounts for about 1.1% of the total. Similar to the results from GHK, new varieties are only created by entrants, and arrive with 0.9%.

Same estimation process on parameter values are done by United States and the calculation results show that $s_q = 1.009$, $\tilde{\delta_i} = 7.60\%$ and $\tilde{\delta_e} = 6.40\%$. Compared to Korea, U.S. shows relatively higher unconditional creative destruction entrants, but the overall unconditional creative destruction rates are rather similar between the two countries. Main disparity in parameter values between the two countries are in δ_i, κ_e and s_κ, and these will be the driving force behind differences in identifying sources of growth when we compare contributions in Table 3.4.

[3] $s_q = \left(\frac{\theta}{\theta-(\sigma-1)}\right)^{\frac{1}{\sigma-1}}$

[4] For Korea 2001-2012, $\tilde{\delta_i} \equiv \delta_i(1-\lambda_i) = 1.72\%$ and $\tilde{\delta_e} \equiv \delta_e(1-\delta_i)(1-\lambda_i) = 9.88\%$

Table 3.4: Contribution to Aggregate Growth

	Entrants	Incumbents	
Korea			
Creative destruction	9.6%	1.7%	11.3%
New varieties	4.2%	0.0%	4.2%
Own-variety improvements	-	84.6%	84.6%
	13.8%	86.3%	
U.S.			
Creative destruction	5.0%	6.0%	11.0%
New varieties	22.4%	0.0%	22.4%
Own-variety improvements	-	66.7%	66.7%
	27.4%	72.6%	

Note: This table presents the percentage contribution of each source of innovation to aggregate TFP growth. Rows use equation (5) to decompose aggregate TFP into the contribution of creative destruction, new varieties, and own innovation. Columns use equation (6) to decompose aggregate TFP into the contribution of entrants and incumbents.

Table 3.4 shows contributions to growth for each innovation sources in Korea and U.S. during the years 2001-2012. Results show that incumbents are the main driving force behind aggregate growth with around 5 times the contribution of those of the entrants. Between the innovation types, own-variety improvement is the main source of growth accounting for over 80% of total growth, followed by creative destruction and new variety creation. Therefore, we can see that own-variety improvement by incumbents is the most important source of growth in Korea. Also, incumbents focus solely on own-variety improvements rather than creative destruction as creative destruction and new variety creation are mostly dictated by entrants.

Things are slightly different in U.S. compared to Korea and the most notable difference between the two countries can be seen under innovation through new variety creation. Similar to the Korean case study, incumbents are more important and own-variety improvements are the main source of growth; however, the amount of contribution to economic growth coming from new variety creation is almost a quarter of total productivity growth. Also, the importance of creative destruction coming from incumbents are much emphasized in U.S., with half of the creative destruction coming from each of the player groups. Another aspect from the results is that, entrants play a larger part in productivity growth for U.S. compared to Korea. The moment that pins down the contribution of entrants is the entrant share of employment; however, even when the entrant share of employment is higher in Korea, Table 3.4 result shows higher contribution of entrants from U.S., and this is because of high contribution coming from new variety creation.

The main results are as shown in Table 3.4, then which feature of the data is responsible in creating the differences between the two countries? There are few moments that has an impact on the contribution of new variety creation to economic growth, but most of the difference comes from the average establishment size differences. U.S. establishments are much larger in average, which means more varieties per plant, and the model determines a firm's size from number of varieties and qualities of those varieties. Ratio of average employment between incumbents and entrants also has an impact on contribution of new varieties by pinning down the quality of new varieties being introduced to the market. New varieties are of lower quality than existing ones, and if entrants are the main source of producing new varieties, this will tend to make entrants smaller than incumbents. Own variety improvements are mainly decided by share of job creation where employment growth ≤ 1, as increase in this moment would imply increase in own-variety improvement relative to the arrival rate of creative destruction. This is because own innovation only results in small changes in employment as it would increase the quality of the products but not the number of varieties being produced by the firm. On the other hand, a creative destruction would generate large changes in employment by having an impact through both quality and quantity. Creative destruction's contribution to growth is mostly determined by job creation and destruction rates. Therefore, nearly zero job re-allocation rate from Korean data becomes a reason for such small share of incumbents' creative destruction share.

Table 3.5: Comparing Firm and Establishment-level Contributions

	Entrants	Incumbents	
U.S. firm (2003-2013)			
Creative destruction	3.3%	7.6%	10.9%
New varieties	5.4%	0.0%	5.4%
Own-variety improvements	-	83.4%	83.4%
	8.8%	91.1%	
U.S. establishments (2001-2012)			
Creative destruction	5.0%	6.0%	11.0%
New varieties	22.4%	0.0%	22.4%
Own-variety improvements	-	66.7%	66.7%
	27.4%	72.6%	

Note: This Table compares the results between firm-level presented in GHK and establishment-level from this paper. General layout is same as Table 3.4, and results under U.S. firm (2003-2013) is directly from GHK with entrants specified as those that are less than or equal to one year, just like how entrants are classified in this paper.

Due to the fewer years of recorded panel data in Korea, entrants were classified as establishments that are less than one year, this is relatively tight compared to benchmark GHK where they treat entrants as firms that are 5 years or under. However, GHK also report results where they treat entrants as firms that are less than a year and those are shown in Table 3.5. The years under research are relatively similar, so I can carefully assume that the differences in Table 3.5 arises mainly due to whether the data at the firm-level or the establishment-level. The notable difference in Table 3.5 is in the new varieties and own-improvements, and this is mainly due to the differences in average size of companies by age and job creation at the lower level of growth (JC$\leq$1). Since incumbent firms that survive grow larger by diversifying themselves into multiple establishments, older companies are relatively much larger than younger ones at the firm-level (2.1) compared to the establishment-level (1.7). Combined with the fact that in the model, entrants producing new varieties will tend to make entrants smaller than incumbents, I can explain the magnitude difference in contribution of new variety creation by entrants. The moment, JC $\leq$ 1, shows a value of 32.2% at the firm-level and 26.1% at the establishment-level. As mentioned before, higher job creation at the lower level of growth would imply a high rate of own-improvement, explaining some of the difference amongst own-variety improvements in Table 3.5. However, most of the difference can be explained from the fact that job creation rates are positively related to new variety arrivals while negatively related to own-improvement arrival rates since new variety arrivals increase employment by adding varieties to a company and own-variety improvements restrict the frequency of creative destruction from happening, thus reducing the job creation and destruction rates. The job creation rates show a value of 12.1% at the firm-level and 14.9% at the establishment-level, this would explain why the contribution of new varieties is much higher and own-improvement is much lower at the establishment-level compared to the firm-level. In addition, share of employment of entrant is much higher at the establishment-level (6.7%) compared to the firm-level (3.9%) and this is shown in Table 3.5 as the entrants' contribution to growth of is much higher at the establishment-level.

Table 3.6: Model fit

	Korea		U.S.	
	Data	Model	Data	Model
Employment share of entrants	10.5%	10.3%	6.3%	6.7%
Employment growth rate	0.0%	0.0%	0.1%	0.1%
Job creation rate	13.0%	14.3%	14.9%	14.9%
Job destruction rate	12.9%	14.3%	14.2%	14.2%
Share of job creation $\leq$ 1	21.5%	20.1%	54.0%	26.1%
SD of log employment	0.91	0.96	1.26	1.07
Exit probability by size	0.77	0.81	0.49	0.47
Size by age	1.49	1.13	1.78	1.69

Note: Exit probability by size means exit rate of large establishments divided by exit rate of small establishments (large and small being greater than or less than average size). Size by age means number of employment of incumbents divided by entrants. Targeted moments for both countries data and model moments during years 2001 to 2012 is recorded in this table.

Table 3.6 shows the fitness of the model for Korea and U.S during 2001-2012 and I can say that the model calibrated moments match reasonably well with the data for both countries. However, the model largely understates the share of job creation $\leq$ 1 for U.S. and average employment comparison of incumbents and entrants for Korea. The reason that the model understates "size by age" moment is due to little heterogeneity in the number of varieties. Regardless of a company being an incumbent or an entrant, Table 3.8 shows that most of the plants in the model, especially Korea, generate only 1 product and there are very few plants producing 3 or more varieties. Judging by the fact that the size of a company in this model depends on number of varieties being produced and the quality of those varieties, having very few products in general might have such impact on understatement of average employment by age.

Table 3.7: Job Creation and Growth contribution by age

	Korea		U.S.	
	Job Creation	TFP Growth	Job Creation	TFP Growth
Age < 5	81.8%	46.0%	59.6%	52.9%
Age 5-10	7.7%	22.6%	13.0%	14.1%
Age 11-15	4.5%	13.2%	8.7%	10.0%
Age > 15	6.0%	18.2%	18.6%	23.0%

Table 3.7 shows how much each of the age groups contribute to aggregate factors. From the table results, relatively newer companies contribute much in job creation but not so much to productivity growth, and contrary to this, older companies contribute less to job creation but more to TFP growth. Although the value is quite different from GHK as explained in their paper, this discrepancy arises because entrant innovation largely takes the form of creative destruction while incumbents focus more on own innovation. Employment will grow less from own variety improvement compared to creative destruction due to acquiring more varieties even when the step size parameter is identical. In real life scenario, it is easy to understand that there will be more job reallocation if a company takes resources from another than to improve upon their own product. Therefore, even when the contribution to aggregate growth was mainly driven by incumbents, entrants tend to have more impact on job reallocation in the economy.

Table 3.8: Distribution of products per establishment

	Korea	U.S.
1 Product	93.3%	71.8%
2 Products	6.2%	18.4%
3 Products	0.5%	6.0%
>3 Products	0.1%	3.8%
Average products per plant	1.07	1.42

As the distribution in Table 3.8 supports the lack of heterogeneity, it also shows similar distribution of plants according to their size. The model implies that the level of employment is determined by the number of varieties and their qualities; therefore, if a plant produces more number of varieties, that plant is highly likely to have more employees. For both U.S and Korea, the distribution of number of establishments according to their size is heavily skewed to the lower bracket where the establishment size is small (there are much more small plants compared to large ones). However, U.S. is less skewed compared to Korea, which could be a correct indicator shown in data that the average size of the establishments in Korea are much smaller compared to their counterparts in US. Nevertheless, results from Table 3.8 showing majority of the plants having only one variety for both countries with the average product of 1.07 and 1.42 for all of the establishments in the model seems to be uncanny; but the lack of data on number of varieties being produced by an establishment makes it hard to compare and contrast on this matter.

3.6 Conclusion

The main focus of this paper is to compare each innovation type's contribution to aggregate growth between US and Korea. Although analysed in a different setting, results from Table 3.4 show consistency with GHK's results that: most of the growth comes from incumbents, most of the growth comes from improvements of existing varieties rather than creation of brand new products, and own-variety improvements appear to be the most important source of growth. However unlike GHK, role of creative destruction on contribution to aggregate growth is understated in this paper. This would emphasize the social return relative to the private return; and could have been caused by the fact that the research was conducted at the establishment-level. By comparing the calibration results for two countries, it is shown that entrants in Korea contribute to growth mainly through creative destruction, while US entrants rely heavily on creating new varieties. Also, it is notable that most of the creative destruction in Korea is executed by entrants, and new variety creation account for almost a quarter of growth for US.

The moments match well between the model and data; this implies a possibility that the results derived from this paper could potentially contribute to Korean government's business policy implications when regarding subsidies based on company types. By answering the questions such as: how much of innovation comes from each of the innovation types? or how much of innovation comes from entrants vs incumbents? I could identify the intensity for each of the sources of growth which could potentially be relevant for policy implications. There are still limitations to the simulation algorithm and the parsimonious model due to the incapability to fully replicate the dynamics or to factor in other relevant aspects to the model. However, the base comparison between Korea and U.S. in this paper could provide a framework for possible additions to the research. Simulation by adding endogenous innovation investments or counter-factual policy implications as in Atkeson and Burstein (2019) to the GHK's simulation algorithm would be such examples.

References

Acemoglu, D. (2003). Labor-and capital-augmenting technical change. *Journal of the European Economic Association*, 1(1):1–37.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly journal of economics*, 117(4):1231–1294.

Adamopoulos, T. (2008). Land inequality and the transition to modern growth. *Review of Economic Dynamics*, 11(2):257–282.

Adamopoulos, T. (2011). Transportation costs, agricultural productivity, and cross-country income differences. *International Economic Review*, 52(2):489–521.

Adamopoulos, T. (2020). Spatial integration and agricultural productivity: Quantifying the impact of new roads. Technical report, Working Paper.

Adamopoulos, T., Brandt, L., Chen, C., Restuccia, D., and Wei, X. (2022). Land security and mobility frictions. Technical report, National Bureau of Economic Research.

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–97.

Adamopoulos, T. and Restuccia, D. (2018). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. Technical report, National Bureau of Economic Research.

Aghion, P. and Howitt, P. (1990). A model of growth through creative destruction.

Aragón, F. M., Oteiza, F., and Rud, J. P. (2021). Climate change and agriculture: Subsistence farmers' response to extreme heat. *American Economic Journal: Economic Policy*, 13(1):1–35.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American economic review*, 110(3):797–823.

Atkeson, A. and Burstein, A. (2019). Aggregate implications of innovation policy. *Journal of Political Economy*, 127(6):2625–2683.

Banerjee, A. V., Gertler, P. J., and Ghatak, M. (2002). Empowerment and efficiency: Tenancy reform in west bengal. *Journal of Political Economy*, 110(2):239–280.

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., and Shapiro, J. S. (2016). Adapting to climate change: The remarkable decline in the us temperature-mortality relationship over the twentieth century. *Journal of Political Economy*, 124(1):105–159.

Bartelsman, E. J. and Doms, M. (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic literature*, 38(3):569–594.

Bartzokas, A. (2008). Country review korea. *UNU-MERIT, URL: http://ec. europa. eu/invest-in-research/pdf/download_en/korea. pdf Erişim*, 24.

Berry, R. A. and Cline, W. R. (1979). *Agrarian Structure and Productivity in Developing Countries: A study prepared for the International Labour Office within the framework of the World Employment Programme.* Johns Hopkins Univ. Press, Baltimore.

Besley, T. and Burgess, R. (2000). Land reform, poverty reduction, and growth: Evidence from india. *Quarterly Journal of Economics*, 115(2):389–430.

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., and Thomas, D. (2011). Migration and global environmental change: future challenges and opportunities.

Blair, T. et al. (2005). *Our common interest: Report of the Commission for Africa.* Commission for Africa.

Borras, S. M. (2003). Inclusion–exclusion in public policies and policy analyses: The case of philippine land reform, 1972–2002. *Journal of International Development*, 15(8):1049–1065.

Bouis, H. E. and Haddad, L. J. (1990). *Agricultural Commercialization, Nutrition, and the Rural Poor*. Lynne Rienner, Boulder, Colorado, U.S.A.

Buera, F. J., Kaboski, J. P., and Shin, Y. (2014). Macro-perspective on asset grants programs: Occupational and wealth mobility. *The American Economic Review: Papers and Proceedings*, 104(5):159–164.

Burke, M. and Emerick, K. (2016). Adaptation to climate change: Evidence from us agriculture. *American Economic Journal: Economic Policy*, 8(3):106–40.

Burke, M., Hsiang, S. M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, 87(3):741–835.

Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production. *Oxford Economic Papers*, 36:131–145.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth*, 1:679–741.

Cassman, K. G. (1999). Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences*, 96(11):5952–5959.

Chung, S. (2011). Innovation, competitiveness, and growth: Korean experiences. *ABCDE*, 17:333.

Colmer, J. (2021). Temperature, labor reallocation, and industrial production: Evidence from india. *American Economic Journal: Applied Economics*, 13(4):101–24.

Combes, P.-P. and Lafourcade, M. (2005). Transport costs: measures, determinants, and regional policy implications for france. *Journal of economic geography*, 5(3):319–349.

Cornia, G. A. (1985). Farm size, land yields and the agricultural production function: An analysis for fifteen developing countries. *World Development*, 13(4):513–534.

Costinot, A. and Donaldson, D. (2012). Ricardo's theory of comparative advantage: Old idea, new evidence. *The American Economic Review*, 102(3):453–458.

Costinot, A. and Donaldson, D. (2016). How large are the gains from economic integration? theory and evidence from us agriculture, 1880-1997. Technical report, National Bureau of Economic Research.

Costinot, A., Donaldson, D., and Smith, C. (2016). Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, 124(1):205–248.

DAR (2006). Primer on carp extension. Department of Agrarian Reform, Philippines.

De Janvry, A. (1981). The role of land reform in economic development: Policies and politics. *American Journal of Agricultural Economics*, 63(2):384–392.

De Janvry, A., Emerick, K., Gonzalez-Navarro, M., and Sadoulet, E. (2015). Delinking land rights from land use: Certification and migration in mexico. *American Economic Review*, 105(10):3125–49.

Deininger, K. and Feder, G. (2001). Land institutions and land markets. *Handbook of Agricultural Economics*, 1:288–331.

Dell, M., Jones, B. F., and Olken, B. A. (2009). Temperature and income: reconciling new cross-sectional and panel estimates. *American Economic Review*, 99(2):198–204.

Dell, M., Jones, B. F., and Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66–95.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature*, 52(3):740–98.

Di Falco, S. and Chavas, J.-P. (2009). On crop biodiversity, risk exposure, and food security in the highlands of ethiopia. *American Journal of Agricultural Economics*, 91(3):599–611.

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *The Journal of Economic Perspectives*, 30(4):171–198.

Donovan, K. (2020). The equilibrium impact of agricultural risk on intermediate inputs and aggregate productivity. *Review of Economic Studies, forthcoming.*

Doorenbos, J. and Kassam, A. (1979). Yield response to water. *Irrigation and drainage paper*, 33:257.

Easterly, W. and Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of monetary economics*, 50(1):3–39.

Espe, M. B., Cassman, K. G., Yang, H., Guilpart, N., Grassini, P., Van Wart, J., Anders, M., Beighley, D., Harrell, D., Linscombe, S., et al. (2016). Yield gap analysis of us rice production systems shows opportunities for improvement. *Field Crops Research*, 196:276–283.

FAO (2000). Yield gap and productivity decline in rice production. Food and Agricultural Organization, United Nations. See Chapter "High-yielding technologies for increasing rice production in the Philippines," S.R. Obien, E.D. Redona, S.R. Francisco, P.Alviola IV, and L.S.Sebastian.

FAOSTAT (2000). Production statistics. *Food and Agricultural Organization (FAO)*, (Value of Agricultural Production, Area Harvested). http://www.fao.org/faostat/en/#data (accessed January 20, 2021).

Feng, S., Oppenheimer, M., and Schlenker, W. (2012). Climate change, crop yields, and internal migration in the united states. Technical report, National Bureau of Economic Research.

Fischer, G., Nachtergaele, F., Prieler, S., Teixeira, E., Toth, G., Velthuizen, H., Verelst, L., and Wiberg, D. (2012). Global agro-ecological zones (gaez v3.0) - model documentation. *Food and Agricultural Organization (FAO) and International Institute for Applied Systems Analysis (IIASA)*. https://www.gaez.iiasa.ac.at/ (accessed March 11, 2018).

Foster, A. D. and Rosenzweig, M. R. (2017). Are there too many farms in the world? labor-market transaction costs, machine capacities and optimal farm size. Technical report, National Bureau of Economic Research.

Fuwa, N., Edmonds, C., and Banik, P. (2007). Are small-scale rice farmers in eastern india really inefficient? examining the effects of microtopography on technical efficiency estimates. *Agricultural Economics*, 36(3):335–346.

GAEZ (2000). Global agro-ecological zones, version 3.0. *Food and Agricultural Organization (FAO) and International Institute for Applied Systems Analysis (IIASA)*. https://www.gaez.iiasa.ac.at/ (accessed March 11, 2018).

Gallup, J. L., Sachs, J. D., and Mellinger, A. D. (1999). Geography and economic development. *International regional science review*, 22(2):179–232.

Galor, O., Moav, O., and Vollrath, D. (2009). Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. *The Review of Economic Studies*, 76(1):143–179.

Galor, O. and Özak, Ö. (2015). Land productivity and economic development: Caloric suitability vs. agricultural suitability. working paper.

Galor, O. and Özak, Ö. (2016). The agricultural origins of time preference. *American Economic Review*, 106(10):3064–3103.

Garcia-Macia, D., Hsieh, C.-T., and Klenow, P. J. (2019). How destructive is innovation? *Econometrica*, 87(5):1507–1541.

Gemenne, F. and Blocher, J. (2017). How can migration serve adaptation to climate change? challenges to fleshing out a policy ideal. *The Geographical Journal*, 183(4):336–347.

Ghatak, M. and Roy, S. (2007). Land reform and agricultural productivity in india: A review of the evidence. *Oxford Review of Economic Policy*, 23(2):251–269.

Godefroy, R. and Lewis, J. (2018). The impact of uncertainty in agriculture. working paper, Université de Montréal.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014a). Agricultural productivity differences across countries. *The American Economic Review: Papers and Proceedings*, 104(5):165–170.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014b). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.

Gollin, D., Parente, S., and Rogerson, R. (2002). The role of agriculture in development. *American economic review*, 92(2):160–164.

Gollin, D., Parente, S. L., and Rogerson, R. (2004). Farm work, home work and international productivity differences. *Review of Economic Dynamics*, 7(4):827–850.

Gollin, D. and Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, 107:38–48.

Golosov, M., Hassler, J., Krusell, P., and Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, 82(1):41–88.

Graff Zivin, J. and Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26.

Guilpart, N., Grassini, P., Sadras, V. O., Timsina, J., and Cassman, K. G. (2017). Estimating yield gaps at the cropping system level. *Field crops research*, 206:21–32.

Guner, N., Ventura, G., and Xu, Y. (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4):721–744.

Hassler, J. and Krusell, P. (2018). Environmental macroeconomics: the case of climate change. In *Handbook of Environmental Economics*, volume 4, pages 333–394. Elsevier.

Hassler, J., Krusell, P., and Smith Jr, A. A. (2016). Environmental macroeconomics. In *Handbook of macroeconomics*, volume 2, pages 1893–2008. Elsevier.

Herrendorf, B. and Schoellman, T. (2015). Why is measured productivity so low in agriculture? *Review of Economic Dynamics*, 18:1003–1022.

Heston, A., Summers, R., and Aten, B. (2009). Penn world table, version 6.3. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania*. https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt-6.3?lang=en.

Holmes, T. J. and Lee, S. (2012). Economies of density versus natural advantage: Crop choice on the back forty. *Review of Economics and Statistics*, 94(1):1–19.

Hsiang, S. M., Burke, M., and Miguel, E. (2013). Quantifying the influence of climate on human conflict. *Science*, 341(6151).

Hsieh, C.-T. and Klenow, P. J. (2009a). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2009b). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics*, 129(3):1035–1084.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, In Press.

Jaenicke, E. C. and Lengnick, L. L. (1999). A soil-quality index and its relationship to efficiency and productivity growth measures: two decompositions. *American Journal of Agricultural Economics*, 81(4):881–893.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of political Economy*, 114(3):538–575.

Jones, C. I. (2016). Life and growth. *Journal of political Economy*, 124(2):539–578.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of political economy*, 112(5):986–1018.

Kniveton, D. R., Smith, C. D., and Black, R. (2012). Emerging migration flows in a changing climate in dryland africa. *Nature Climate Change*, 2(6):444–447.

Kravchenko, A. N. and Bullock, D. G. (2000). Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, 92(1):75–83.

Krusell, P. (1998). How is r&d distributed across firms? a theoretical analysis. *unpublished paper*.

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *The American Economic Review*, 103(2):948–980.

Lahiff, E. and Li, G. (2012). Land redistribution in south africa: A critical review. WP 80874, World Bank. Washington D.C.

Levine, D. I. and Yang, D. (2014). The impact of rainfall on rice output in indonesia. Technical report, National Bureau of Economic Research.

Lipton, M. (2009). *Land Reform in Developing Countries: Property Rights and Property Wrongs*. Routledge.

Lobell, D. B., Cassman, K. G., and Field, C. B. (2009). Crop yield gaps: their importance, magnitudes, and causes. *Annual review of environment and resources*, 34:179–204.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523.

Lucas Jr, R. E. and Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, 122(1):1–51.

Mendola, M. and Simtowe, F. (2014). The welfare impact of land distribution: Evidence from a quasi-experimental initiative in malawi. Centro Studi Luca d'Agliano Development Studies Working Paper.

Midrigan, V. and Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *The American Economic Review*, 104(2):422–458.

Monfreda, C., Ramankutty, N., and Foley, J. A. (2008). Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global biogeochemical cycles*, 22(1).

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., and Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature*, 490(7419):254.

Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.

Nakicenovic, N., Alcamo, J., Davis, G., Vries, B. d., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. Y., Kram, T., et al. (2000). Special report on emissions scenarios.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences*, 103(10):3510–3517.

Nunn, N. and Qian, N. (2011). The potato's contribution to population and urbanization: Evidence from a historical experiment. *The Quarterly Journal of Economics*, 126(2):593–650.

Porteous, O. (2019). High trade costs and their consequences: an estimated dynamic model of african agricultural storage and trade. *American Economic Journal: Applied Economics*, 11(4):327–66.

Prasada Rao, D. (1993). Intercountry comparisons of agricultural output and productivity. Technical report. FAO Economic and Social Development Paper 112, Food and Agricultural Organization, Rome.

Pyo, H., Hong, S., Kim, A., et al. (2016). Firm size and job creation in korea: do small businesses create more jobs? *The Korean Economic Review*, 32(1):137–166.

Pyo, H. K., Chun, H., and Rhee, K. H. (2008). Total factor productivity by 72 industries in korea and international comparison (1970-2005). Technical report, BOK working papers.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10.

Restuccia, D. and Santaeulalia-Llopis, R. (2017). Land misallocation and productivity. Technical report, National Bureau of Economic Research.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics*, 55(2):234–250.

Rodrik, D., Subramanian, A., and Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of economic growth*, 9(2):131–165.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102.

Sachs, J. D. (2003). Institutions don't rule: direct effects of geography on per capita income. Technical report, National Bureau of Economic Research.

Saulo-Adriano, L. (1991). *A general Assessment of the Comprehensive Agrarian Reform Program*. Philippine institute for development studies.

Schils, R., Olesen, J. E., Kersebaum, K.-C., Rijk, B., Oberforster, M., Kalyada, V., Khitrykau, M., Gobin, A., Kirchev, H., Manolova, V., et al. (2018). Cereal yield gaps across europe. *European Journal of Agronomy*, 101:109–120.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy.* Harper and Brothers.

Schumpeter, J. A. et al. (1939). *Business cycles*, volume 1. Mcgraw-hill New York.

Sherlund, S. M., Barrett, C. B., and Adesina, A. A. (2002). Smallholder technical efficiency controlling for environmental production conditions. *Journal of Development Economics*, 69(1):85–101.

Sotelo, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy*, 128(7):2690–2738.

Steduto, P., Hsiao, T. C., Fereres, E., and Raes, D. (2012). *Crop yield response to water*. FAO Roma.

Stokey, N. L. (1988). Learning by doing and the introduction of new goods. *journal of Political Economy*, 96(4):701–717.

Syverson, C. (2011). What determines productivity? *Journal of Economic literature*, 49(2):326–65.

Tittonell, P. and Giller, K. E. (2013). When yield gaps are poverty traps: The paradigm of ecological intensification in african smallholder agriculture. *Field Crops Research*, 143:76–90.

TM (2008). World borders dataset. *Thematic Mapping*. http://thematicmapping.org/downloads/world_borders.php (accessed July 25, 2014).

Tombe, T. (2015). The missing food problem: Trade, agriculture, and international productivity differences. *American Economic Journal: Macroeconomics*, 7(3):226–258.

Udry, C. (2012). Misallocation, growth and financial market imperfections. Unpublished manuscript, SED plenary talk.

UNPS (2009). Uganda bureau of statistics, national panel survey (unps) 2005-2009. *World Bank*, (Ref. UGA_2005-2009_UNPS_v01_M.). https://microdata.worldbank.org/index.php/catalog/1001/get-microdata (accessed June 15, 2020).

USDA (2015). Usda national nutrient database for standard reference, release 28. *US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory*. https://www.ars.usda.gov/Services/docs.htm?docid=8964 (accessed April 5, 2018).

Van Ittersum, M. K., Van Bussel, L. G., Wolf, J., Grassini, P., Van Wart, J., Guilpart, N., Claessens, L., de Groot, H., Wiebe, K., Mason-D'Croz, D., et al. (2016). Can sub-saharan africa feed itself? *Proceedings of the National Academy of Sciences*, 113(52):14964–14969.

Vinke, K., Rottmann, S., Gornott, C., Zabre, P., Nayna Schwerdtle, P., and Sauerborn, R. (2022). Is migration an effective adaptation to climate-related agricultural distress in sub-saharan africa? *Population and Environment*, 43(3):319–345.

Young, A. (1994). Lessons from the east asian nics: A contrarian view. *European economic review*, 38(3-4):964–973.

Zhai, F., Lin, T., and Byambadorj, E. (2009). A general equilibrium analysis of the impact of climate change on agriculture in the people's republic of china. asian development review, vol. 26 (1), pp. 206-225.

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., et al. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 114(35):9326–9331.

Appendix A

Chapter 1 Appendix

A.1 Data

Calculating Actual Yields:

1. The conversion factor variable provided by LSMS, converts the unit of quantity variable from the unit to kilograms. However, the reported values are not identical to the code sheet (codes for unit of quantity) and some observations provide extremely high or low values for the conversion variable. So, I use the codes for unit of quantity to accurately portray the conversion factor variable as much as I can, and I drop all the observations that have unspecified values, for their unit of quantity.

2. After step 1, I would have the quantity harvested in terms of kilograms which I convert into tonnes. And the next step involves multiplying crop prices (GK$/t from GAEZ user guide) to each of the different crop types so that I would have quantity harvested in terms of GK$.

3. Last step would be to merge the total output (quantity harvested (GK$)) with the harvested land size variables using the household ID, and calculate actual yield by dividing output (GK$) by harvested land input (hectare).

Table A.1: Crop types by crop

Crop Name	Crop Type
Wheat	Food Crop
Rice	Food Crop
Maize	Food Crop
Millet	Food Crop
Sorghum	Food Crop
Beans	Food Crop
Cow peas	Food Crop
Pigeon peas	Food Crop
Chick peas	Food Crop
Groundnuts	Cash Crop
Soybeans	Cash Crop
Sunflower	Cash Crop
Rapeseed(sesame)	Cash Crop
Sugarcane	Cash Crop
Cotton	Cash Crop
Potatoes	Food Crop
Sweet Potatoes	Food Crop
Cassava	Food Crop
Yam	Food Crop
Cocoyam	Food Crop
Banana	Cash Crop
Coffee	Cash Crop
Tea	Cash Crop

ANNEX 7: CODES FOR UNIT OF QUANTITY

Sr. No.	UNIT	CODE
1	Kilogram (kg)	01
2	Gram	02
3	Litre	03
4	Small cup with handle (Akendo)	04
5	Metre	05
6	Square metre	06
7	Yard	07
8	Millilitre	08
9	Sack (120 kgs)	09
10	Sack (100 kgs)	10
11	Sack (80 kgs)	11
12	Sack (50 kgs)	12
13	Sack (unspecified)	13
14	Jerrican (20 lts)	14
15	Jerrican (10 lts)	15
16	Jerrican (5 lts)	16
17	Jerrican (3 lts)	17
18	Jerrican (2 lts)	18
19	Jerrican (1 lt)	19
20	Tin (20 lts)	20
21	Tin (5 lts)	21
22	Plastic Basin (15 lts)	22
23	Bottle (750 ml)	23
24	Bottle (500 ml)	24
25	Bottle (350 ml)	25
26	Bottle (300 ml)	26
27	Bottle (250 ml)	27
28	Bottle (150 ml)	28
29	Kimbo/Cowboy/Blueband Tin (2	29
30	Kimbo/Cowboy/Blueband Tin (1	30
31	Kimbo/Cowboy/Blueband Tin (0.5	31
32	Cup/Mug (0.5 lt)	32
33	Glass (0.25 lt)	33
34	Ladle (100 g)	34
35	Table spoon	35
36	Tea spoon	36
37	Basket (20 kg)	37
38	Basket (10 kg)	38
39	Basket (5 kg)	39
40	Basket (2 kg)	40
41	Loaf (1 kg)	41
42	Loaf (500 g)	42
43	Buns (200 g)	43
44	Buns (100 g)	44

Sr. No.	UNIT	CODE
44	Buns (100 g)	44
45	Buns (50 g)	45
46	Bathing soap (Tablet)	46
47	Washing soap (Bar)	47
48	Washing soap (Tablet)	48
49	Packet (2 kg)	49
50	Packet (1 kg)	50
51	Packet (500 g)	51
52	Packet (250 g)	52
53	Packet (100 g)	53
54	Packet (Unspecified)	54
55	Fish – Whole (Up to 1 kg)	55
56	Fish – Whole (1 - 2 kg)	56
57	Fish – Whole (Above 2 kg)	57
58	Fish - Cut piece (Up to 1 kg)	58
59	Fish - Cut piece (1 - 2 kg)	59
60	Fish - Cut piece (Above 2 kg)	60
61	Tray of 30 eggs	61
62	Ream	62
63	Crate	63
64	Heap (Unspecified)	64
65	Dozen	65
66	Bundle (Unspecified)	66
67	Bunch (Big)	67
68	Bunch (Medium)	68
69	Bunch (Small)	69
70	Cluster (Unspecified)	70
71	Gourd (1 – 5 lts)	71
72	Gourd (5 – 10 lts)	72
73	Gourd (Above 10 lts)	73
74	Gologolo (4 - 5 lts)	74
75	Calabash (1 - 5 lts)	75
76	Calabash (Above 5 lts)	76
77	Jug (2 lts)	77
78	Jug (1.5 lts)	78
79	Jug (1 lt)	79
80	Tot (50 ml)	80
81	Tot (sachet)	81
82	Tot (Unspecified)	82
83	Tobacco leaf (Number)	83
84	Pair	84
85	Number of Units (General)	85
86	Acre	86
87	Other Units (Specify)	99

Potential Yield Summary:

Table A.2: Summary of Potential Yields

Scenarios	Mean	Std.Dev	# of Obs.
Baseline (1961-1990)	5047.864	2080.984	2193
Hadley CM3 2050 A1 C	5828.248	2223.048	2193
Hadley CM3 2050 A1 N	5504.822	2078.958	2193
Hadley CM3 2080 A1 C	5195.536	1890.251	2193
Hadley CM3 2080 A1 N	4750.036	1701.083	2193
Hadley CM3 2080 B1 C	5755.420	2181.548	2193
CSIRO Mk2 2080 A2 C	5218.295	1854.442	2193
CCCma CGCM2 2080 B2 C	5978.067	2266.280	2193

Hadley CM3 A1 scenarios are showing that in year 2050, the average potential yields are much higher than year 2080, while the standard deviation is lower. This is interesting because the agricultural productivity gap widens between different regions in Uganda as time goes on, and the standard deviation is showing an opposite movement. One possible explanation for this could be that the land quality gap is narrowing between farmers in the same quartile (region) or could just be because of the fact that there are more factors other than potential yields that needs to be considered when analyzing agricultural productivity. Compared to the baseline period, Hadley 2080 A1 N is the only scenario where the average potential yield is lower. However, since the standard deviation is also lower and due to change in prices with potential yield of food crop being higher than the baseline period, Table 3.4 shows the result that there is still an increase in agricultural productivity.

A.2 Algorithm

1. Given the LSMS data on actual yields and harvested farm land size with GAEZ data on potential yields, calculate farmer ability s_c, s_f, s using the production function.

2. Using the calibrated parameters and prices (p_c, p_f, q), solve for the model equilibrium which clears the land market as well as the cash crop and food crop goods market.

3. Fix the parameters, $F_c, \beta, \eta, \bar{a}, \gamma, \varepsilon$, and the farmer ability s to the benchmark economy. Change the value of potential yields z to the corresponding scenario of climate change and solve the model letting the prices (p_c, p_f, q) and share of farmers (N_c, N_f) vary to match five market clearing conditions: no-arbitrage conditions for cash and food crops, land market, cash and food crop goods market.

4. After solving for the model equilibrium, compare the aggregate values of the economy between the benchmark economy and the experiments representing climate change.

A.3 Crop type consumption Preference

Let the preference of agricultural crops be cobb-douglas: Then,

$$U = \phi \; log \; (c_a - \bar{a}) + (1-\phi) \; log \; (c_n),$$

where, $c_a = c_c^{\beta} c_f^{1-\beta}$ is a cobb-douglas consumption measure of two types of agricultural goods. Income of the household:

$$I \equiv (1 - N_c - N_f) w (1-\varepsilon) + \eta N_c \int_{s,z} \pi_c(s,z) dF(s,z) + N_f \int_{s,z} \pi_f(s,z) dF(s,z) + qL$$
$$I \equiv p_c c_c + p_f c_f + c_n$$

The problem of the household maximizing utility subject to their income is:

$$\max_{c_c, c_f, c_n} U = \{ \; \phi \; log \; (c_c^{\beta} c_f^{1-\beta} - \bar{a}) + (1-\phi) \; log \; (c_n) \}$$
$$subject \;\; to: \qquad I = p_c c_c + p_f c_f + c_n$$
$$\mathcal{L} = \phi \; log \; (c_c^{\beta} c_f^{1-\beta} - \bar{a}) + (1-\phi) \; log (c_n) + \lambda (I - p_c c_c - p_f c_f - c_n)$$

Solving for the F.O.C to find the marginal utilities :

$$F.O.C(c_c) \Rightarrow \quad \frac{\phi \beta c_c^{\beta-1} c_f^{1-\beta}}{p_c (c_c^{\beta} c_f^{1-\beta} - \bar{a})} = \lambda \qquad (1)$$

$$F.O.C(c_f) \Rightarrow \quad \frac{\phi (1-\beta) c_c^{\beta} c_f^{-\beta}}{p_f (c_c^{\beta} c_f^{1-\beta} - \bar{a})} = \lambda \qquad (2)$$

$$F.O.C(c_n) \Rightarrow \quad \frac{1-\phi}{c_n} = \lambda \qquad (3)$$

Letting equations (1), (2), & (3) to be equal:

$$\frac{MU_{c_c}}{p_c} = \frac{MU_{c_f}}{p_f} = \frac{MU_{c_n}}{1} \Rightarrow \frac{\phi \beta c_c^{\beta-1} c_f^{1-\beta}}{p_c (c_c^{\beta} c_f^{1-\beta} - \bar{a})} = \frac{\phi (1-\beta) c_c^{\beta} c_f^{-\beta}}{p_f (c_c^{\beta} c_f^{1-\beta} - \bar{a})} = \frac{1-\phi}{c_n}$$

Using equations (1) & (2), c_c can be estimated as a function of c_f, vice versa:

$$\frac{\phi \beta c_c^{\beta-1} c_f^{1-\beta}}{p_c (c_c^{\beta} c_f^{1-\beta} - \bar{a})} = \frac{\phi (1-\beta) c_c^{\beta} c_f^{-\beta}}{p_f (c_c^{\beta} c_f^{1-\beta} - \bar{a})} \quad \Rightarrow \quad \frac{p_f}{p_c} = \frac{1-\beta}{\beta} c_c c_f^{-1}$$

$$\Rightarrow c_c = \frac{\beta}{1-\beta}\frac{p_f}{p_c}c_f \tag{4}$$

$$\Rightarrow c_f = \frac{1-\beta}{\beta}\frac{p_c}{p_f}c_c \tag{5}$$

Letting equation (2) & (3) to be equal and rearranging, we get:

$$c_n = \frac{1}{1-\beta}\frac{1-\phi}{\phi}\frac{p_f(c_c^{\beta}c_f^{1-\beta}-\bar{a})}{c_c^{\beta}c_f^{-\beta}} \tag{6}$$

Letting equation (1) & (3) to be equal and rearranging, we get:

$$c_n = \frac{1}{\beta}\frac{1-\phi}{\phi}\frac{p_c(c_c^{\beta}c_f^{1-\beta}-\bar{a})}{c_c^{\beta-1}c_f^{1-\beta}} \tag{7}$$

Plugging equations (4) & (6) into the budget constraint yields:

$$I = p_c(\frac{\beta}{1-\beta}\frac{p_f}{p_c}c_f) + p_f c_f + \frac{1}{1-\beta}\frac{1-\phi}{\phi}\frac{p_f\big((\frac{\beta}{1-\beta}\frac{p_f}{p_c}c_f)^{\beta}c_f^{1-\beta}-\bar{a}\big)}{(\frac{\beta}{1-\beta}\frac{p_f}{p_c}c_f)^{\beta}c_f^{-\beta}}$$

$$I = p_f c_f(\frac{1}{1-\beta}) + \frac{1-\phi}{\phi(1-\beta)}p_f c_f - \frac{1-\phi}{\phi(1-\beta)}\frac{\bar{a}p_f}{(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{\beta}}$$

$$c_f = \frac{\phi(1-\beta)}{p_f}\left(I + \frac{1-\phi}{\phi(1-\beta)}\frac{\bar{a}p_f}{(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{\beta}}\right)$$

$$c_f = \frac{I\phi(1-\beta)}{p_f} + \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{-\beta} \tag{8}$$

Substituting equation (8) into (4) to find c_c gives:

$$c_c = \frac{I\phi\beta}{p_c} + \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{1-\beta} \tag{9}$$

Substituting equations (8) and (9) into the budget constraint gives c_n:

$$c_n = I(1-\phi) + \frac{\bar{a}(1-\phi)p_f}{1-\beta}(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{-\beta} \tag{10}$$

Using equation (8) and (9) with $c_a = c_c^{\beta}c_f^{1-\beta}$ to calculate optimal consumption of agricultural goods gives c_a.

A.4 Computation of the subsistence constraint and cash crop fixed cost

Compute $\bar{a}$ endogenously using the food crop market clearing condition:

$$C_f = N_f \int_{s,z} y_f(s,z)dF(s,z) = \frac{I\phi(1-\beta)}{p_f} + \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{-\beta}$$

$$\Rightarrow \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{-\beta} = N_f \int_{s,z} y_f(s,z)dF(s,z) - \frac{I\phi(1-\beta)}{p_f}$$

Then, isolating $\bar{a}$ yields:

$$\bar{a} = \frac{1}{1-\phi}(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{\beta}\Big[N_f \int_{s,z} y_f(s,z)dF(s,z) - \frac{I\phi(1-\beta)}{p_f}\Big]$$

Compute F_c endogenously using the cash crop market clearing condition and $\bar{a}$ from the food crop market clearing condition:

$$C_c = N_c \int_{s,z} y_c(s,z)dF(s,z) - N_c F_c = \frac{I\phi\beta}{p_c} + \bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{1-\beta}$$

Isolate F_c:

$$F_c = \int_{s,z} y_c(s,z)dF(s,z) - \frac{I\phi\beta}{p_c N_c} - \frac{\bar{a}(1-\phi)(\frac{\beta}{1-\beta}\frac{p_f}{p_c})^{1-\beta}}{N_c}$$

Appendix B

Chapter 2 Appendix

B.1 Calibration of price and the labor mobility barrier

Direct calibration of rental price of land $q(i)$ and the labor mobility barrier $\varepsilon(i)$.

Find $q(i)$: From the land market clearing condition, we can compute $q(i)$:

$$\begin{aligned} AFS(i) = \frac{\bar{L}(i)}{N(i)} &= \left[\frac{\gamma A(i)}{T(i)q(i)}\right]^{\frac{1}{1-\gamma}} \int_{\underline{s}}^{\bar{s}} sdF(s,i) \\ q(i) &= \left[\frac{\gamma A(i)}{T(i)}\right](AFS(i))^{\gamma-1}\left[\int_{\underline{s}}^{\bar{s}} sdF(s,i)\right]^{1-\gamma} \end{aligned}$$

Whereas, $A(i)$ is represented by GAEZ potential yield; $T(i)$ is represented by the transportation costs data from Ethiopia paper; $AFS(i)$ is represented by aggregated AgSS data at the woreda-level; and the TFP is computed from the production function and AgSS data at the farm-level.

Find $\varepsilon(i)$: Using the share of farmers in each location, compute the relative $\varepsilon(i)$ as a function of known variables (without $q(i)$).

$$\frac{\varepsilon(i)}{\int_m \varepsilon(m)dm} = \left[\frac{N(i)}{N} \frac{\int_m \left[\frac{A(m)}{T(m)}\right]^{\frac{1}{\gamma}} \bar{L}(m)\left[\int_{\underline{s}}^{\bar{s}} sdF(s,m)\right]^{\frac{1-\gamma}{\gamma}} dm}{\left[\frac{A(i)}{T(i)}\right]^{\frac{1}{\gamma}} \bar{L}(i)\left[\int_{\underline{s}}^{\bar{s}} sdF(s,i)\right]^{\frac{1-\gamma}{\gamma}}}\right]^{\gamma}$$

Where, $A(i)$, $T(i)$, $\bar{L}(i)$, $N(i)$ are represented by data. Matching $\varepsilon(i)$ to real data $\frac{N(i)}{N}$ will allow $\varepsilon(i)$ to match the real data population shares as well. We normalize N to one in the calibration, so $N(i)$ will just be population shares and equal to $\frac{N(i)}{N}$.

B.2 Aggregate Production Function

Aggregate output at location i is,

$$\begin{aligned} Y(i) &= N(i)\int_{\underline{s}}^{\bar{s}} y(s,i)dF(s,i) \\ \Rightarrow Y(i) &= N(i)A(i)^{\frac{1}{1-\gamma}}\left(\frac{\gamma}{q(i)T(i)}\right)^{\frac{\gamma}{1-\gamma}} \int_{\underline{s}}^{\bar{s}} sdF(s,i) \end{aligned}$$

As estimated from the previous appendix,

$$q(i) = \left[\frac{\gamma A(i)}{T(i)}\right]\left[\frac{N(i)}{\bar{L}(i)}\int_{\underline{s}}^{\bar{s}} s dF(s,i)\right]^{1-\gamma}$$

Plugging in the above $q(i)$ to the aggregate output $Y(i)$ yields:

$$Y(i) = A(i)\bar{L}(i)^{\gamma}N(i)^{1-\gamma}\Big[\int_{\underline{s}}^{\bar{s}} s dF(s,i)\Big]^{1-\gamma}$$

Appendix C

Chapter 3 Appendix

C.1 Aggregate Growth Function

Final goods sector problem:

$$\max_{\{y_{jt}\}_{j=1}^{M_t}} Y_t$$

$$subject\ \ to: \quad Y_t = \Big[\sum_{j=1}^{M_t} (q_{jt} y_{jt})^{1-\frac{1}{\sigma}}\Big]^{\frac{\sigma}{\sigma-1}}$$

$$* \quad E_t = W_t L = \sum_{j=1}^{M_t} y_{jt} P_{jt}$$

- Equilibrium demand for intermediate input j:
  $y_{jt} = q_{jt}^{\sigma-1} \big(\frac{P_t}{P_{jt}}\big)^{\sigma} \frac{E_t}{P_t}$
- Equilibrium aggregate price index:
  $P_t = \Big(\sum_{j=1}^{M_t} \big(\frac{P_{jt}}{q_{jt}}\big)^{1-\sigma} \Big)^{\frac{1}{1-\sigma}}$
- Under optimal price setting:
  $P_t = \mu W_t \big(\sum_{j=1}^{M_t} q_{jt}^{\sigma-1} \big)^{\frac{1}{1-\sigma}}$
- Optimal pricing: $P_{jt} = \mu W_t$
- Equilibrium demand for intermediate variety:
  $y_{jt} = q_{jt}^{\sigma-1} \Big(\frac{P_t}{\mu W_t}\Big)^{\sigma} \frac{E_t}{P_t}$

  $y_{jt} q_{jt} = q_{jt}^{\sigma} \Big(\frac{P_t}{\mu W_t}\Big)^{\sigma} \frac{E_t}{P_t}$

  $Y_t = \Big[\sum_{j=1}^{M_t} (q_{jt} y_{jt})^{\frac{\sigma-1}{\sigma}} \Big]^{\frac{\sigma}{\sigma-1}} = \frac{E_t}{\mu^{\sigma} W_t^{\sigma}} \frac{1}{P_t^{1-\sigma}} \Big[\sum_{j=1}^{M_t} q_{jt}^{\sigma-1} \Big]^{\frac{\sigma}{\sigma-1}}$

  $Y_t = \frac{E_t}{P_t^{1-\sigma}} \Big\{ \mu W_t \big[\sum_{j=1}^{M_t} q_{jt}^{\sigma-1} \big]^{\frac{1}{1-\sigma}} \Big\}^{-\sigma} \Rightarrow E_t = P_t Y_t$

  $$Y_t = \frac{E_t}{P_t} = \frac{W_t L}{P_t} \Rightarrow \frac{Y_t}{L} = \frac{W_t}{P_t} \tag{1}$$

$$\frac{P_{t+1}}{P_t} = \frac{W_{t+1}\left(\sum_{j=1}^{M_{t+1}} q_{jt+1}^{\sigma-1}\right)^{\frac{1}{1-\sigma}}}{W_t\left(\sum_{j=1}^{M_t} q_{jt}^{\sigma-1}\right)^{\frac{1}{1-\sigma}}} \tag{2}$$

- Note: the term in numerator can be re-written as:

$$\sum_{j=1}^{M_{t+1}} q_{jt+1}^{\sigma-1} = \underbrace{\sum_{j=1}^{M_t} q_{jt+1}^{\sigma-1}}_{\text{A}} + (\kappa_e + \kappa_i) s_\kappa \sum_{j=1}^{M_t} q_{jt}^{\sigma-1} \tag{3}$$

- Write term A (innovation of existing varieties) as:
$\sum_{j=1}^{M_t} q_{jt+1}^{\sigma-1} = \underbrace{\sum_{j \in M_{dt}} q_{jt+1}^{\sigma-1}}_{\text{a}} + \underbrace{\sum_{j \in M_{it}} q_{jt+1}^{\sigma-1}}_{\text{b}} + \underbrace{\sum_{j \in M_{nt}} q_{jt}^{\sigma-1}}_{\text{c}} - \underbrace{\sum_{j \in M_{ot}} q_{jt+1}^{\sigma-1}}_{\text{d}}$

- a = products subject to creative destruction by other incumbent or by new entrant
  b = products by own incumbent innovation
  c = surviving products not subject to any innovation
  d = non-surviving products

- If any type of innovation occurs for existing varieties (a or b) the quality of the product improves by s_q. ($q_{jt+1} = s_q q_{jt}$)
  Therefore, re-writing term A would yield equation (4):

$$\sum_{j=1}^{M_t} q_{jt+1}^{\sigma-1} = s_q^{\sigma-1} \sum_{j \in M_{dt}} q_{jt+1}^{\sigma-1} + s_q^{\sigma-1} \sum_{j \in M_{it}} q_{jt+1}^{\sigma-1} + \sum_{j \in M_{nt}} q_{jt}^{\sigma-1} - \sum_{j \in M_{ot}} q_{jt+1}^{\sigma-1} \tag{4}$$

- The probability for each of the cases are:
  $\tilde{\delta}_e + \tilde{\delta}_i$ = arrival rate of creative destruction
  λ_i = arrival rate of own innovation
  δ_o = exit rate of existing varieties due to overhead cost
  $1 - \tilde{\delta}_e + \tilde{\delta}_i - \lambda_i$ = probability of innovation of surviving existing product

- Since probability of innovation is independent of quality, by LLN:

$$\sum_{j\in M_{dt}} q_{jt}^{\sigma-1} = (\tilde{\delta}_e + \tilde{\delta}_i)\sum_{j=1}^{M_t} q_{jt}^{\sigma-1}$$

$$\sum_{j\in M_{it}} q_{jt}^{\sigma-1} = \lambda_i \sum_{j=1}^{M_t} q_{jt}^{\sigma-1}$$

$$\sum_{j\in M_{nt}} q_{jt}^{\sigma-1} = [1-\tilde{\delta}_e + \tilde{\delta}_i - \lambda_i]\sum_{j=1}^{M_t} q_{jt}^{\sigma-1}$$

$$\sum_{j\in M_{ot}} q_{jt}^{\sigma-1} = \delta_o\psi \sum_{j=1}^{M_t} q_{jt}^{\sigma-1}$$

- Then rewrite appendix A equation (4) as:
$\sum_{j=1}^{M_t} q_{jt+1}^{\sigma-1} = \left\{ s_q^{\sigma-1}\left(\tilde{\delta}_e + \tilde{\delta}_i\right) + s_q^{\sigma-1}\lambda_i + \left(1-\tilde{\delta}_e - \tilde{\delta}_i - \lambda_i\right) - \delta_o\psi \right\} \sum_{j=1}^{M_t} q_{jt}^{\sigma-1}$

- Which can be re-written as:

$$\left(\frac{\sum_{j=1}^{M_t} q_{jt+1}}{\sum_{j=1}^{M_t} q_{jt}}\right)^{\sigma-1} = \left(1 + s_\kappa(\kappa_e+\kappa_i) + (s_q^{\sigma-1}-1)\lambda_i + (s_q^{\sigma-1}-1)(\tilde{\delta}_e + \tilde{\delta}_i) - \delta_o\psi\right)$$

- The aggregate growth function:
$$1+g = \left(1 + \underbrace{s_\kappa(\kappa_e+\kappa_i)}_{\text{new varieties}} + \underbrace{(s_q^{\sigma-1}-1)\lambda_i}_{\text{own innovation}} + \underbrace{(s_q^{\sigma-1}-1)(\tilde{\delta}_e + \tilde{\delta}_i)}_{\text{creative destruction}} - \delta_o\psi\right)^{\frac{1}{\sigma-1}}$$

C.2 Simulation Algorithm

1. Specify an initial guess for the number of varieties in the economy and the distribution of quality across varieties (the shape parameter θ of the pareto distribution is later manually specified after an initial run is complete).

2. Simulate life paths for establishments in the economy: each entrant has one initial variety, stolen from an incumbent or newly created. In every period, an establishment faces a probability of each type of innovation per variety it owns. Incumbents can either not innovate, lose a variety to other incumbents or entrants, improve upon its own variety, create a new product, or steal a variety from another incumbent. A plant's life ends when it loses all of its varieties to others or when 80 periods have passed.

3. Based on the population of simulated establishments, calculate the moments:

   (a) Standard deviation of log employment

(b) Job creation rate

(c) Job destruction rate

(d) Share of job creation where employment growth ≤ 1

(e) Employment share of entrants

(f) Employment growth rate

(g) Exit rate for large/small plants

(h) Average employment for incumbents/entrants

4. Repeat above steps, searching for parameter values to exactly match TFP growth in the data, set minimum employment to 1 and minimize the euclidean distance between the model and data moments.