CESifo
AREA
CONFERENCES 2021

Energy and Climate Economics
Munich, 4–5 March 2021

Climate Change and Migration: The Case of Africa

Bruno Conte

CESifo

Climate change and migration: the case of Africa[*]

Bruno Conte[†]

Job Market Paper

This version: January 10, 2021
[latest version here]

Abstract

This paper provides a spatial general equilibrium model to quantify the impact of climate change on the economy and migration. The model can capture the role of trade networks and agricultural suitability on the distribution of population and GDP accounting for endogenous adjustments of crop choice and trade. I use detailed geospatial data from 42 countries in sub-Saharan Africa (SSA) to simulate the impact of climate using forecasts of agricultural productivity in 2080 from FAO–GAEZ. Climate change is estimated to displace 12 percent of the SSA population and reduce real GDP by 4 percent. The capacity of switching crops, urbanizing, or trading goods reduces the impact of climate change in terms of population outflows. Finally, the adoption of modern inputs in agriculture reverses considerably the negative impacts of climate change.

Keywords: Climate change, Migration, Economic geography.
JEL Codes: O15, Q54, R12.

[†]Universitat Autònoma de Barcelona and Barcelona GSE. Address: Departament d'Economia i d'Historia Economica, Office B3-112G, 08193, Barcelona, Spain. Phone: +34 615 548 181. Email: bruno.conte@barcelonagse.eu. Web: brunoconteleite.github.io.

1 Introduction

One of the most concerning potential consequences of climate change is population displacement, recently coined as the *Great Climate Migration* (Lustgarten, 2020). Subsistence agricultural economies, like the sub-Saharan African (SSA henceforth) countries, lie at the center of this issue. These are highly agriculture–dependent economies, which are expected to be amid the fastest–growing zones, in terms of population, during the next decades (United Nations and Social Affairs, 2019). Understanding how these economies will adjust to a climate–changing world, in which crop yields will be different than today, is key to uncover where this growing population will be geographically reallocated.

Assessing which could be the decisions of economic agents when adapting to climate change is challenging, especially in an agricultural context like SSA. If facing changed agricultural yields, farmers could switch their production towards different crops, remaining however in the agricultural sector. Alternatively, these agents could switch to a non–agricultural sector, potentially moving geographically. Trade would have a crucial role in the extent to which a specialization into agricultural and non–agricultural sectors could take place. Therefore, understanding how these forces – production switching, trade, and migration – will respond to climate change is key to evaluate the effect of climate change on migration and the economy.

In this paper, I develop a quantitative spatial model that accounts for these forces and can be used to quantify how their response to climate change translates into population displacement and economic losses. The model is calibrated with a rich geographical dataset that I assemble, covering 42 countries of SSA. By simulating it for a future scenario of climate change in 2080, I find that about 12 percent of the SSA population could be displaced. The results are very heterogeneous across countries and subnational locations: the median country(location)–level population change is about -12 (-9) percent, while the 10th and 90th percentiles are -22 (-28) and 8 (3) percent, respectively. Moreover, a key finding is that urbanization, crop–switching, and access to trade are relevant margins along which the economy adjusts so to dampen the impact of climate change.

I begin my analysis by showing that the future changes in agricultural yields are expected to be spatially heterogeneous across locations and crops. As a consequence, they would consist of a shock to comparative advantages in the agricultural sector. Informed by this empirical evidence, I develop a multi-sector Ricardian spatial trade model with partial labor mobility. The model's spatial units are $1°x1°$ grid–cells, where farmers and firms can produce goods of multiple agricultural sectors (crops) and a non-agricultural sector, respectively. Differences in total factor productivity and

2

market access generate trade, shaping the spatial pattern of the sector–specialization. Relative sectoral prices determine consumers' expenditure shares, allowing for structural transformation. Migration frictions and other congestion forces counteract the agglomeration forces in the model.

The advantages of my framework are manifold. First, my general equilibrium approach considers the long–term adjustments of the economy along many dimensions as the climate changes. Second, the multi–sector feature of the model allows me to predict future sectoral expenditure shares in counterfactuals so that the results speak closely to the recent literature on urbanization and structural transformation in SSA. Third, my setup allows for the simulation of policy-relevant experiments, such as the adoption of modern inputs in agricultural production. Fourth, my quantification strategy requires only data on the distribution of the economic activity and population within countries, similar to Desmet et al. (2018). As such, it overcomes the main limitation in the migration literature, which is the need of observing within and cross-country bilateral migration flows.

In terms of empirics, I assemble a rich micro spatial (1°x1° degree) dataset on population, economic activity, transportation infrastructure, and agricultural production and suitability. The latter is a core element of this paper: climate change is modeled to impact the economy as a shock to the suitability for growing crops. In practice, I draw on the work of agronomists and climatologists from the GAEZ (IIASA and FAO, 2012) project, and use their estimates of crop-specific potential yields for several grain crops for recent, past, and future – under a climate change scenario – periods.[1] These potential yields are calculated based exclusively on natural characteristics (e.g. topographic and climatic), providing a measure of geographical natural advantages for each crop.[2] Importantly, as I focus on subsistence agriculture in SSA, I consider the main staple crops that are grown and consumed in the region: cassava, maize, millet, rice, sorghum, and wheat.[3]

A key insight obtained with the GAEZ forecasts is that agricultural suitability is not expected to change uniformly. Figure 1 illustrates this idea with the potential yields of cassava in SSA for 2000 and 2080. Several locations are expected to become much less productive as a consequence of climate change. Surprisingly perhaps, the opposite is expected for a handful of locations (e.g. in Ethiopia, Southern Kenya,

---

[1]To my knowledge, my paper is the first one to exploit the variation of the GAEZ estimates over time for past periods. Therefore, it contrasts with related research which uses the long–term averages (1960–1990) of the GAEZ estimates.

[2]Hereafter, I refer to the GAEZ potential yields as natural or fundamental productivities/advantages indistinctly.

[3]These grain crops account for about 80% of the total production, in tonnes, of the main staple and cash crops (cassava, coffee, cotton, groundnut, maize, millet, palm oil, rice, sorghum, soybean, sugarcane, and wheat; see Table C.2) and for about 50% of the caloric intake in SSA (Porteous, 2019a).

Figure 1: FAO–GAEZ Agro-climatic yields of cassava in 2000 (left) and 2080 (right).
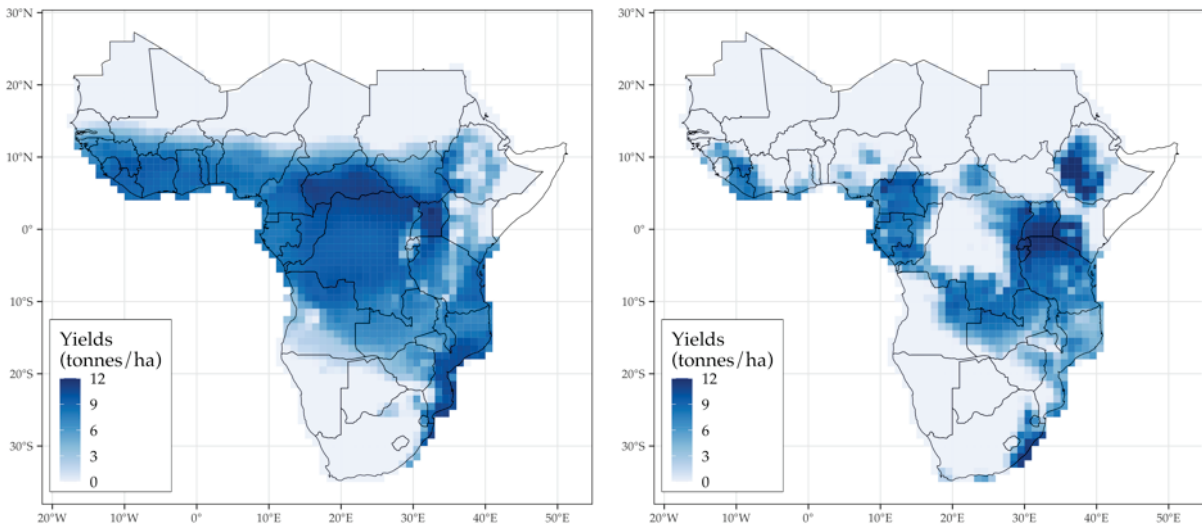
Notes: The two graphs above depict average potential yields of cassava, within $1°\text{x}1°$ grid cells, drawn from the GAEZ database for 2000 (left) and 2080 (right). The potential yields stand for the average production, in tonnes/hectares, that could be achieved in each cell (conditional on natural characteristics only). Somalia is not considered due to the unavailability of data from other sources. See Section 2 and Appendix B for details.

and Northern Tanzania), contrasting with the idea of climate change as a negative–only shock to agriculture. The same pattern is observed for other crops, though differently in terms of spatial distribution. Therefore, the presented evidence suggests that the changes in the climate will be a heterogeneous shock to the spatial degree of comparative advantages within the agricultural sector.

The model is calibrated to fit the SSA economy in 2000. To do that, I first use the GAEZ yields for 2000 as the measure of the fundamental productivities in each agricultural sector (crops). Second, I use comprehensive data on transportation infrastructure in SSA to build an optimal trade network between all locations pairs of my empirical setup. The resulting bilateral distances between location pairs are used to estimate trade frictions. Third, I quantify the unobserved fundamentals and parameters: the sets of fundamental productivities of the non–agricultural sector, location–specific productivity shifters of all sectors, and amenities. This last step requires the inversion of the spatial equilibrium so that the model achieves an exact fit of the data in terms of GDP distribution, sectoral output, and population in 2000.

I validate my calibrated model with a backcasting exercise. In particular, I simulate my calibrated model with crop suitabilities for 1975 and compare its outcomes with observable data. The model achieves a very good fit when predicting the grid cell–level changes in population between 2000 and 1975, reassuring its capacity of providing similar numbers for future periods. As an additional overidentification test, I

find that the model identifies well the degree of specialization in agriculture across countries.

My main counterfactual exercise consists of simulating a climate changed SSA in 2080. I draw the estimates for agricultural productivities in 2080 for that scenario[4,5] and simulate my model with these values, keeping all other fundamentals unchanged. The results are striking: compared to a scenario with no climate change, more than 300 million people could be displaced[6] (about 12 percent of SSA's total population), and real GDP drops by about 4 percent. As previously mentioned, the results are very heterogeneous across countries and locations.

Subsequently, I evaluate the mechanisms of the model at play by investigating the heterogeneity of my results. My findings suggest a decrease in the non–agricultural employment, on aggregate, and heterogeneous changes across countries. The median country experiences an increase in the non–agricultural labor of 1 percent, and the 10th and 90th percentiles are -3 and 5 percent, respectively. The mechanism driving this result is sectoral specialization: while the most severely hit locations (and countries) specialize out of agriculture, the opposite takes place for the least hit locations. Indeed, the capacity of moving the production towards the non–agricultural sector mitigates the impacts of climate change in terms of population losses at the grid–cell level. Along the same lines, crop–switching and access to trade are likewise important. Among the locations severely hit by climate–change, those able to reshuffle their agricultural production mix, or better connected to markets, exhibit lower rates of population outflows.

Finally, I perform a policy experiment centered on technology adoption in agriculture. In particular, I simulate a climate changed SSA where farmers exogenously adopt more modern inputs in the production, such as animal traction, mechanization, usage of high yielding varieties, and fertilizers. In such a scenario, the estimated GDP losses of climate change are considerably reversed. However, population flows remain considerably large, mainly because more productive inputs in agricultural production intensify the geographical specialization between the non-/agricultural sectors.

This paper contributes to several strands of the economic literature. First, to a large set of reduced–form studies that establish a causal relationship between weather

---

[4]The GAEZ estimates are available for different hypothetical scenarios for the future – I pick the one that compares the closest to the standards, according to climatologists, for a severe scenario: the Representative Concentration Pathway (RCP) 8.5 (see Appendix B.1).

[5]To account for the uncertainty around the estimates for the future climate, I test the robustness of all results with respect to different climate models and/or RCP scenarios used to generate the GAEZ data. That provides confidence intervals for the estimated results. See Section 6.3 for details.

[6]Population displacement is defined as the difference between the model–implied population, at the grid–cell level, of two simulations: with and without climate change in 2080. Grid–cells with positive (negative) values experience population inflows (outflows). See Section 6.1 for a careful discussion.

anomalies and migration,[7] especially in subsistence agricultural economies like SSA.[8] However, by mostly focusing on short term climate events for identification, this literature does not provide the means for assessing the resulting migration of the long–term changes entailed by climate change (Burzyński et al., 2019a). I add to this literature by using a spatial general equilibrium approach to assess migration flows caused by the climate change shock to agriculture.

As such, I also contribute to the literature that investigates the relationship between population displacement and climate change with quantitative spatial models. Within this literature, Desmet et al. (2020) quantify the reshaping of the world's economy and population upon coastal flooding with a one–sector spatial framework that embeds land losses in the spatial dynamics of the economy. Desmet and Rossi-Hansberg (2015); Conte et al. (2020); Nath (2020) embed global warming into a spatial general equilibrium setup where temperature changes dynamically shape the evolution of sectoral productivities. Unlike these papers, I focus on the heterogeneous impact of climate change within the agricultural sector, allowing for within (switching crops) and across (agriculture–urban) sector adjustments. In my results, I show that the capacity for such an adaptation is a quantitatively relevant margin that mitigates the impact of the climate shock in terms of migration at the grid–cell level.

My paper is close to the work of Costinot et al. (2016), who quantify future GDP losses due to climate change as a shock to agricultural suitability. I contribute to their work by allowing for labor mobility and quantifying the migration consequences of the climate shock in a spatial general equilibrium framework. As a consequence, my work also relates to Shayegh (2017); Burzyński et al. (2019a,b), who study climate migration in an OLG model where migrations decisions respond to temperature and sea levels. My contribution is to explicitly model the geography of the economy and the impact of climate change throughout its locations, making migration (within and across countries) a key mechanism of the long–run structural adjustment of the economy to climate change.[9]

---

[7]See Berlemann and Steinhardt (2017); Cattaneo et al. (2019) for the most recent surveys, and Baez et al. (2017); Gröger and Zylberberg (2016); Cai et al. (2016) for two examples of Latin America, Southeast Asia, and worldwide studies, respectively.

[8]The lack of technological adoption in SSA's agriculture is argued to be among the main causes of underdevelopment (Porteous, 2019b; Sheahan and Barrett, 2017) and vulnerability with respect to weather shocks (FAO, 2015). Indeed, Barrios et al. (2006); Henderson et al. (2017) show that changes in the rainfall patterns during the past decades played a determinant – and causal – role in the high urbanization rates of SSA.

[9]Indirectly, I also add to a rich and growing modern spatial economics literature on developing contexts. A non–exhaustive list include Donaldson and Hornbeck (2016); Morten and Oliveira (2018); Donaldson (2018); Pellegrina and Sotelo (2019); Ducruet et al. (2019); Balboni (2019); Sotelo (2020) on the relevance of transportation infrastructure, Desmet et al. (2018) on the dynamic effects of the spatial diffusion of ideas, Nagy (2020); Allen and Donaldson (2018) on city location and historical dependence, Allen et al. (2019) on border walls, and Pellegrina (2019); Moneke (2019) on technology adoption.

Importantly, my paper also adds to the current policy debate about potential climate migration. Policy circles have been stressing the concerns with climate refugees for several years (IPCC, 2007, 2012, 2018). More recently, several institutions have produced quantitative studies to guide policymakers in this matter. Worth noting are, among others, the World Banks' (Rigaud et al., 2018), and the Pulitzer Center's (Lustgarten, 2020) projects, which use the gravity–based spatial framework of Jones and O'Neill (2016) to estimate climate migrant flows. The closest result to the context of my paper is the estimate of about 100 million migrants by 2050 in SSA (Rigaud et al., 2018). I contribute to this debate by providing results with a richer quantitative framework. In particular, my model considers the interconnection of production, trade, and residence decisions by agents within and across countries, accounts for the heterogeneity of the climate change shock within the agricultural sector, and allows for the simulation of real–world policies.

Finally, I also contribute to the broader economics of climate change literature, based on the seminal work of William Nordhaus (1992, 2013, 2018, 2019). His DICE/RICE integrated assessment model (IAM) became a standard tool to quantify the potential economic impacts of global warming by endogenizing the global climate to economic activity. While not allowing such relation,[10] my framework is capable of accounting for the economic consequences of climate change as well as many other aspects not contemplated in IAM studies, such as production, trade, and migration decisions.[11,12]

The remaining of the paper is organized as follows. Section 2 describes the main sources of data used and Section 3 documents a number of empirical facts that illustrate the potential impact of climate change on the agricultural economy of SSA. Section 4 presents the theoretical framework. Section 5 details how the model is brought to the data, and Section 6 the results of the climate change counterfactuals, policy experiments, and several robustness checks. Section 7 concludes.

2 Data

This study builds upon several sources of geographical data. These are aggregated at $1°x1°$ degree grid cells (about 100 x 100 km at the equator), the unit of observation for this study. The set of cells covering 42 countries of SSA contains 2,032 cells. Below, I

---

[10] Africa contributes with about 3 percent of total CO_2 emissions as of 2015, which allows me to assume climate change to be exogenous to economic activity in my framework.

[11] An exception is Benveniste et al. (2020), who integrates cross–country migration within an IAM.

[12] Other studies that exploit past weather changes to evaluate the economic impacts of global warming are Burke et al. (2015); Burke and Emerick (2016); Dell et al. (2009, 2014); Schlenker et al. (2005); Schlenker and Lobell (2010).

describe the collection and aggregation of the main data sources; further details are documented in Appendix B.

GDP. I obtain data on GDP disaggregated in areas within countries from the Global Gridded Geographically Based Economic Data v4 (G-Econ, Nordhaus et al., 2006). It consists of a dataset with *gross cell product*; i.e. gross product of grid cells of 1 square degree. The data spans from 1990 to 2005 in intervals of five years.

Population. Information on the distribution of the population is gathered from several sources. First, the G-Econ database provides the population count for the periods from 1990 to 2005. I complement it with gridded population of 1975 from the Global Human Settlement Project (GHSP, Florczyk et al., 2019). Finally, I collect projections for future population, at the country level, from United Nations and Social Affairs (2019) for the period of 2021 to 2100.

Agricultural suitability. I construct a time-varying, geographically disaggregated data set of crop–specific suitabilities from the Food and Agriculture Organization's Global Agro-Ecological Zones (GAEZ, IIASA and FAO, 2012) database. This data built with a state-of-the-art agronomic model that combines fine–grained data on geographic characteristics (e.g. soil, elevation, etc.) and yearly climatic conditions to produce several agricultural-related outputs disaggregated at the 5 arc-minutes (about 0.083 degrees) resolution from 1960 to 2000. Among these, I collect and aggregate estimates of agro-climatic potential yields for the 6 crops of interest for 1975 and 2000. These potential yields, measured in tonnes/hectares, refer to the yield that a certain cell would obtain if its surface was fully devoted to a specific crop. Moreover, the GAEZ database provides this data for a climate–changed world in 2080, which I also collect. The final data is a panel, at the cell-crop level, of agro-climatic yields for 1975, 2000, and 2080.

Agricultural production. Actual crop production is obtained from two sources of data. First, GAEZ provides actual values (tonnes) of production and harvested land (in hectares) for 2000. Moreover, FAOSTAT provides crop production, in current US$, at the country level for 2000–2010.

Transportation network. In order to build up a network connecting all grid cells of SSA, I first collect the African extract of the Global Roads Open Access Data Set (gROADS v1, CIESIN, 2013), which combines the best available public domain road data by country into a global roads coverage database. The date range for the road network representations is from the 1980s to 2010 depending on the country, as their data is gathered from different sources. In order to overcome potential missing roads in some particular country, and to capture links between locations not necessarily through roads, I explore the transportation friction surface from the Accessibility to

Cities' project (Weiss et al., 2018). This high–resolution surface (0.01 degrees reso-
lution) provides the instantaneous cost of passing through a cell conditional on geo-
graphical features (e.g. type of terrain, steepness) as well as on infrastructure (whether
the cell is on a road, railroad, river, etc.).

3   Motivating Facts

This section documents two facts about the potential impacts of climate change in
SSA. I show first that these effects are expected to be strong and heterogeneous. Sec-
ond, I document that as such climate change could be determinant in the future
organization of the SSA economy. Overall, they provide empirical support for my
Ricardian approach when modelling how climate change could affect SSA.

**Fact 1: Climate change is expected to bring substantial, and spatially heteroge-
neous, changes to agricultural suitability in SSA.**

I use GAEZ estimates of agro-climatic potential yields for 2000 and 2080[13] to stress
out how severe and heterogeneous the impact of climate change is expected to be. I
define ΔA_i^k as the changes in the yields of crop k in location i between the two periods,
and ΔA_i as the change in average crop yields in every location i.

Panel A of Figure 2 shows that the average climate change shock to agricultural
yields is very heterogeneous. In terms of levels, several locations are estimated to
become less suitable to agriculture, with average yields reducing by 50% or more.
Impressively perhaps, a handfull of location are expected to become more suitable
if compared to 2000, and the magnitude of such gains are likewise substantial. This
finding goes against the general sense of climate change as a homogeneous negative
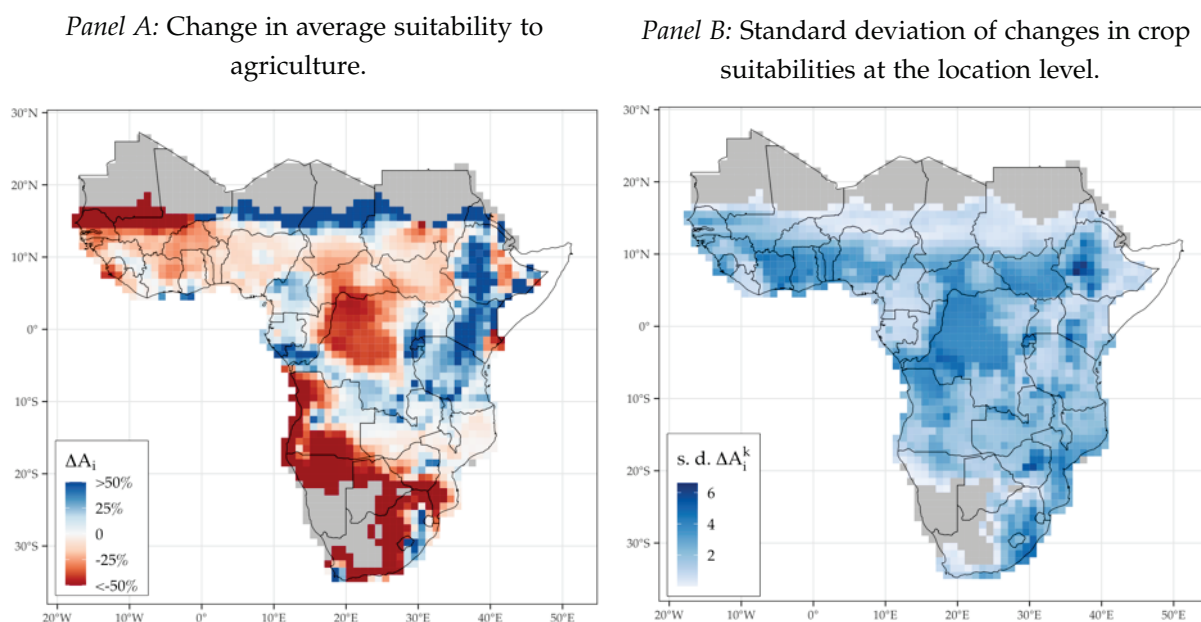shock to agriculture.

To illustrate how heterogeneous these effects are across crops, Panel B of Fig-
ure 2 documents the dispersion of the climate change effects at the location level
(in standard deviations of A_i^k at the location i level). It can be observed that the
changes in yields are not homogenous across crops so that the relative ranking of
crop–suitabilities will be differently shifted. As such, climate change will consist of a
shock to the geographical comparative advantage for each crop.

Therefore, several locations of SSA could potentially cope with the climate change
shock in terms of agricultural loss by adjusting their crop choices. Alternatively, the
economy could also reshuffle its factors to the non–agricultural sector. The degree of
natural advantages on all sectors, interacted with market access, would determine the

---

[13]The 2080 forecasts from GAEZ are calculated assuming a hypothetical scenario for the future
evolution of the world climate. Appendix B.1 describes how I choose the scenario to draw my data from
so that my numbers compare the closest possible to the standards from the climatologist community.

Figure 2: Expected impact of climate change to average crop yields (left) and standard deviation of crop–yield changes (right) in SSA between 2000 and 2080.

Panel A: Change in average suitability to agriculture.

Panel B: Standard deviation of changes in crop suitabilities at the location level.

Notes: The potential yields stand for the production, in tonnes/hectares, attained within 1°x1° grid–cells conditional on natural characteristics. Panel A documents the estimated change (truncated for easiness of visualization) in average potential yields between 2000 and 2080. Panel B shows the standard deviation of the crop–level yield changes within locations. Grey areas stand for cells in with zero potential yields for all crops in both time periods. See Section 2 and Appendix B for details.

patterns of sector–specialization upon the climate change shock through comparative advantages.

The extent to which such Ricardian economic adjustments could take place in SSA depends on the strength of comparative advantages on shaping the agricultural production and trade in the continent. The next empirical fact provides evidence for such a mechanism holding in reality and emphasize the importance of embedding it into my theoretical framework.

Fact 2: Natural suitability to grow crops explains a large degree of crop-specialization and trade in SSA.

I next document how natural advantages for growing crops explain the spatial patterns of agricultural production and trade in SSA. To do that, I first match the yields data with effective production at the grid–cell level. Second, I average out the yields at the country level and match them with bilateral crop trade data from COMTRADE.[14]

Figure 3[15] how production and trade correlate with with the GAEZ yields. In particular, Panel A plots the crop production against the yields at the location–crop level. The raw data values are first net out of location and country–crop fixed effects,

---

[14]Refer to Appendix B.2 for details on the collection and aggregation of the trade data.

[15]Appendix C.1 provides the details of the econometric models that generate the graphs of Figure 3 and several robustness checks.

Figure 3: Comparative advantage and the organization of the SSA economy: relationship of crop yields with production and trade.

Panel A: Crop production (in tonnes).   *Panel B:* Bilateral crop trade (in US$).
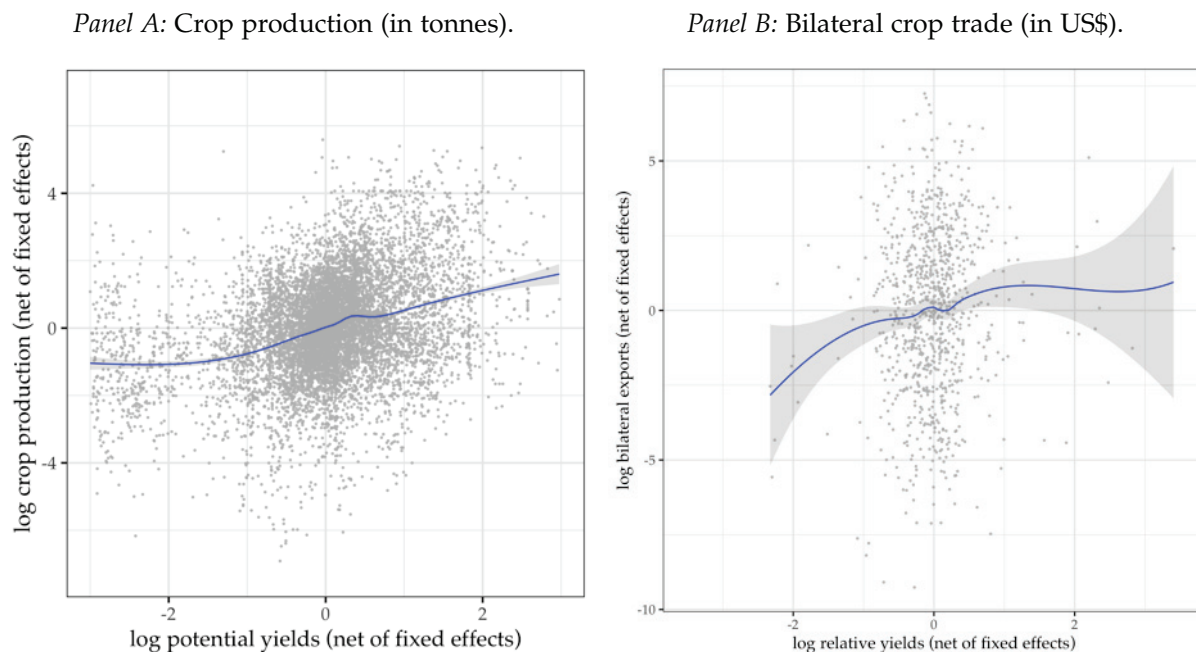
Notes: Panel A (B) plots the correlation between GAEZ potential yields and effective production (bilateral crop trade) at the location–crop (country–pair) level. The blue line stands for an estimated polynomial regression, and grey–shaded areas the 95% confidence bands. See Appendix C.1 for details.

so that confounding factors at both levels are controlled for. There is a strong correlation between the natural advantages and effective production. Analogously, Panel B shows a strong correlation between bilateral crop exports and relative yields between countries. In this case, the raw data is net out of exporter, importer, and crop fixed effects, as well as of controls at the country–pair level. These two facts are statistical significant if estimated with fixed effects regressions (see Appendix C.1).

Overall, the data conveys a sound message: crop–specialization happens both across and within countries, and country trade flows reflect that. To generate this pattern, my general equilibrium model will take the perspective of subnational units that specialize in (and trade) crops based on comparative advantage.

4   Model

This section outlines a spatial model[16] that allows for a credible quantification of the general equilibrium impact of future climate change. The model provides a tractable framework to account for the role of several dimension of heterogeneity (fundamental productivities across many sectors, market access, factor productivities, among oth-

---

[16]Further details and derivations of the model are documented in the Appendix A.

ers) across geographical locations on the spatial distribution of the economic activity.

4.1 Environment

The economy is composed by N locations $i \in S = \{1, ..., N\}$ and populated by $\mathcal{L} = \sum_{i \in S} L_i$ workers who supply their labor inelastically. There are K sectors $k \in \mathcal{K} = \{1, ..., K\}$ in the economy: $K - 1$ agricultural sectors (crops) and a non-agricultural composite K sector. Locations (can) produce a locally differentiated variety of the goods of each of these sectors. Each location has a sector-specific fundamental productivity parameter $A_i^k \in \mathcal{A} = \{A_1^1, ..., A_N^K\}$ which drives the degree of comparative advantage between locations in each sector. Moreover, each location provides an amenity value $u_i \in \{u_i\}_{i \in S} \equiv \mathcal{U}$ for workers residing in it.

Goods and labor units are mobile in S, subject to frictions. As standard in the literature, $\mathcal{T} = \{\tau_{ij}\}_{i,j \in S}$ is the bilateral trade frictions' matrix; $\tau_{ij} = \tau_{ji} \geq 1$ stand for the amount of units of the good required to ship 1 unit from location i to j. Frictions in labor mobility are instead driven by an idiosyncratic taste shock to the choice of living in a certain location i. The dispersion of the distribution of these shocks drives the extent of frictions to labor mobility.

The *geography* of the economy is the set $\mathcal{G}(S) = \{\mathcal{L}, \mathcal{A}, \mathcal{U}, \mathcal{T}\}$: the spatial fundamentals that interact with the economic forces of the model and determine the distribution of the economic activity over S. In the following, I describe how the economic component is structured.

Technology and Market Structure. In every location i, a representative firm produces goods of each sector k with labor as the unique input of the following linear production function

$$q_i^k = b_i^k A_i^k L_i^k, \tag{1}$$

where b_i^k stands for a location–sector efficiency parameter unrelated to the natural advantage of that location in producing goods of sector k (e.g. degree of technology adopted in the production). The output can be locally consumed or traded with other locations. Trade takes place in a perfectly competitive framework with full information, which implies no arbitrage in the trade between locations. Such market structure implies that the price of the sector k variety produced in i and shipped to (thus, consumed at) location j is

$$p_{ij}^k = (w_i / b_i^k A_i^k) \times \tau_{ij}, \tag{2}$$

where the first (second) subscript stands for the location of production (consumption/shipment). w_i stands for the wages in location i.

12

Preferences. Each location is populated by a continuum of heterogeneous workers. Their welfare is determined by a component related to the consumption of varieties and an amenity component. The latter is determined by their heterogeneous taste with respect to the location of living. Workers choose where to live and how much to consume so to maximize welfare; a v worker in location i has the following welfare function:[17]

$$W_i(z) = \left(\sum_{k \in \mathcal{K}} \left(C_i^k \right)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \times \varepsilon_i(v); \tag{3}$$

$\varepsilon_i(v)$ is the location taste shock the worker draws and $\eta > 1$ the elasticity of substitution between consumption of bundles of different sectors. C_i^k is the CES aggregate of the consumption of all varieties of goods from a sector k, defined as

$$C_i^k = \left(\sum_{j \in \mathcal{S}} \left(q_{ji}^k \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{4}$$

where q_{ji}^k is the per-capita quantity of the variety of sector k produced in j that is consumed in i and $\sigma > 1$ is the Armington CES.

Consumption choice. Each worker earns wage w_i, thus $\sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{K}} p_{ji}^k q_{ji}^k = w_i$ is the budget constraint for workers conditional on living in i. Welfare maximization with respect to consumption of varieties implies that the share of i's spending on j's variety of sector k is

$$\lambda_{ji}^k = \left(p_{ji}^k / P_i^k \right)^{1-\sigma}, \text{ where} \tag{5}$$

$$P_i^k = \left(\sum_{j \in \mathcal{S}} \left(p_{ji}^k \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{6}$$

is the Dixit-Stiglitz price index of sector k. An analogous result holds for the share of the expenditure on sector aggregates:

$$\mu_i^k = \left(P_i^k / P_i \right)^{1-\eta} \text{ and} \tag{7}$$

$$P_i = \left(\sum_{k \in \mathcal{K}} \left(P_i^k \right)^{1-\eta} \right)^{\frac{1}{1-\eta}} \tag{8}$$

are the share of location i's expenditure in goods from sector k and the overall price

---

[17]I set preferences with a double-nested CES structure to allow for structural transformation. Differently from assuming preferences with a lower-tier CES and an upper-tier Cobb-Douglas aggregate, my set up allows for sector shares to be endogenous, rather than fixed and set by the Cobb-Douglas shares. Moreover, it provides an empirical advantage, as consumption shares do not need to be calibrated. Related literature usually does so with household consumption data, which might be unfeasible to obtain for all the countries my study is covering.

index in i, respectively. Thus, the equilibrium per capita demand for j variety of sector k goods in i is $q_{ji}^{k*} = \lambda_{ji}^k \mu_i^k w_i$. By inserting it in eq. (4), one finds that

$$\left(\sum_{k \in \mathcal{K}} (C_i^k)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} = \frac{w_i}{P_i} \quad \forall i, \tag{9}$$

i.e. the per capita consumption in location i equals real wages. Moreover, the overall expenditure in i for goods produced in j, X_{ji}, is defined as

$$
\begin{aligned}
X_{ji} &= \sum_{k \in \mathcal{K}} \lambda_{ji}^k \mu_i^k w_i L_i \\
&= \sum_{k \in \mathcal{K}} (P_i^k / P_i)^{1-\eta} \left(\frac{w_j \tau_{ji}}{b_j^k A_j^k P_i^k} \right)^{1-\sigma} w_i L_i.
\end{aligned} \tag{10}
$$

Bilateral expenditures take a gravity-like form: for a given sector k, it is decreasing with respect to marginal cost of shipping from j to i ($w_j \tau_{ji} / b_j^k A_j^k$). Besides, the (partial) elasticity of trade with respect to trade frictions τ_{ij} is driven by the CES parameter in the format of $1 - \sigma < 0$.

Location choice. Workers choose where to live so to maximize welfare. The choice is subject to a location taste shock ε_j. Formally, a v worker chooses location j to solve

$$\max_j \quad W_j(v) = \frac{w_j}{P_j} \times \varepsilon_j(v). \tag{11}$$

Following Redding (2016), I assume that the taste shock is drawn independently (across workers and locations) from an extreme-value (Fréchet) distribution with shape parameter $\theta > 0$ and scale parameter $u_i L_i^{-\alpha}$. That is,

$$\varepsilon_i \sim G_i(z) = e^{-z^{-\theta} \times \left(u_i L_i^{-\alpha} \right)}. \tag{12}$$

The assumption above means that the workers' heterogeneity with respect to their location tastes (and the dispersion forces in the economy) is driven by the parameter θ. Higher values imply that agents are more homogeneous and that the economic components of welfare (real wages w_i / P_i) play a stronger role in the location decisions in Equation (11). In this case, there are weak dispersion forces in the economy. In contrast, lower values of θ imply more heterogeneous agents, who more likely draw higher values of taste shocks for every location. In this case, there are strong dispersion forces. The average of the preference draws are disciplined by $\left(u_i L_i^{-\alpha} \right)$; u_i stands for the fundamental amenity of location i and $\alpha > 0$ determines the extent to which population density diminishes the life quality in i.

The distributional assumption on the taste preference allows one to obtain closed-form solutions for the location choice of workers. As there is a continuum of workers in every location, the probability that a worker chooses to live in i is equivalent to the share of workers living in i in equilibrium. Following Eaton and Kortum (2002), one can show that the latter, defined as Π_i, is equivalent to

$$\Pi_i = \mathbb{P}\left(W_i(v) \geq \max\{W_j(v)\}_{s \neq i} \right) = \frac{(w_i/P_i)^\theta u_i L_i^{-\alpha}}{\sum\limits_{j \in S} (w_j/P_j)^\theta u_j L_i^{-\alpha}}. \tag{13}$$

Therefore, the number of workers that will choose to live in i is

$$L_i = \Pi_i \times \mathcal{L}. \tag{14}$$

The result above is quite intuitive: locations with higher real wages (w_i/P_i) and/or life quality $\left(u_i L_i^{-\alpha}\right)$ will have, in equilibrium, a higher share of workers. The magnitude of it is partially driven by θ, which is the elasticity of the location choice with respect to real wages.

4.2 Spatial Equilibrium

Given the geography $\mathcal{G}(S)$ and the exogenous parameters $\{\theta, \eta, \sigma\}$, a spatial equilibrium is a vector of factor prices and labor allocations $\{w_i, L_i\}_{i \in S}$ such that eqs. (2), (6), (8), (10) and (14) hold, and markets for goods clear. Market clearing, formally, requires that total GDP in i equals total exports to *and* total imports from all locations $j \in S$, including itself, i.e.

$$w_i L_i = \sum_{j \in \mathcal{S}} X_{ij} = \sum_{j \in \mathcal{S}} X_{ji}. \tag{15}$$

This condition is equivalent trade balancing in all locations. Note that factor markets clearing is determined by eq. (14), as $\Sigma_i \Pi_i = 1$ in by construction. Moreover, by using eq. (10) on (15), one can characterize the spatial equilibrium with the system of $4 \times N$ equations and $4 \times N$ unkowns below:

$$w_i L_i = \sum_{j \in S} \sum_{k \in \mathcal{K}} (P_j^k / P_j)^{1-\eta} \left(\frac{w_i \tau_{ji}}{b_i^k A_i^k P_j^k} \right)^{1-\sigma} w_j L_j \qquad P_i^k = \left(\sum_{j \in S} (w_j \tau_{ji} / b_i^k A_j^k)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

$$\text{(16)} \hspace{7cm} \text{(18)}$$

$$L_i = \frac{(w_i / P_i)^\theta u_i L_i^{-\alpha}}{\sum_{j \in S} (w_j / P_j)^\theta u_j L_j^{-\alpha}} \mathcal{L} \qquad \text{(17)} \qquad P_i = \left(\sum_{k \in \mathcal{K}} \left(P_i^k \right)^{1-\eta} \right)^{\frac{1}{1-\eta}} \qquad \text{(19)}$$

To solve this high-dimensional, non-linear system of equations, I apply an iterative algorithm whose intuition works as follows.[18] Given an initial guess for wages, I solve for prices and labor distribution. I then use the market clearing condition to solve for optimal wages conditional on its initial guess and the values calculated. I iterate this process until convergence.

Existence and Uniqueness. My model is not isomorphic to the general set up of Allen and Arkolakis (2014) and, as a consequence, the existence and uniqueness of the equilibrium cannot be guaranteed. The reason for that is the additional non-linearity introduced by the upper-level CES structure. I address that by solving my model for several parametric choices, starting from many different initial guesses. The equilibrium found is invariant across all cases.

4.3 Illustration of the spatial equilibrium

I illustrate how changes in the economy's fundamentals shape the distribution of the economic activity and population on the geography by representing it as a line with a discrete number of locations. I assume that locations are homogeneous with respect to amenity values ($u_i = u \; \forall i$) and sector productivities ($A_i^k = A^k \; \forall i, k$). Bilateral trade frictions are parametrized as

$$\tau_{ij} = e^{\tau \times |i-j|},$$

where $\tau = 0.05$. I set two sectors for the economy and assume that $A^2 > A^1$. I solve for equilibrium wages and labor allocations as described in Section 4.2; the distribution of L_i is plotted in Figure 4. In particular, Panel A shows that more central locations are those where most of the labor – thus economic activity – is allocated. The reason is that they are the most evenly distant from all other locations in the economy. Thus, shipping prices in these locations are lower due to the lower trade costs accrued in the trade with the other locations. There is more labor allocated to the second (more

---

[18]Appendix A.4 provides a detailed description of the algorithm.

Figure 4: Equilibrium values for $\{L_i, L_i^1, L_i^2\}_{i \in S}$ in the spatial model represented on a line.

Panel A: Homogeneous $\{A_i^k\}$ *Panel B: Heterogeneous $\{A_i^k\}$*
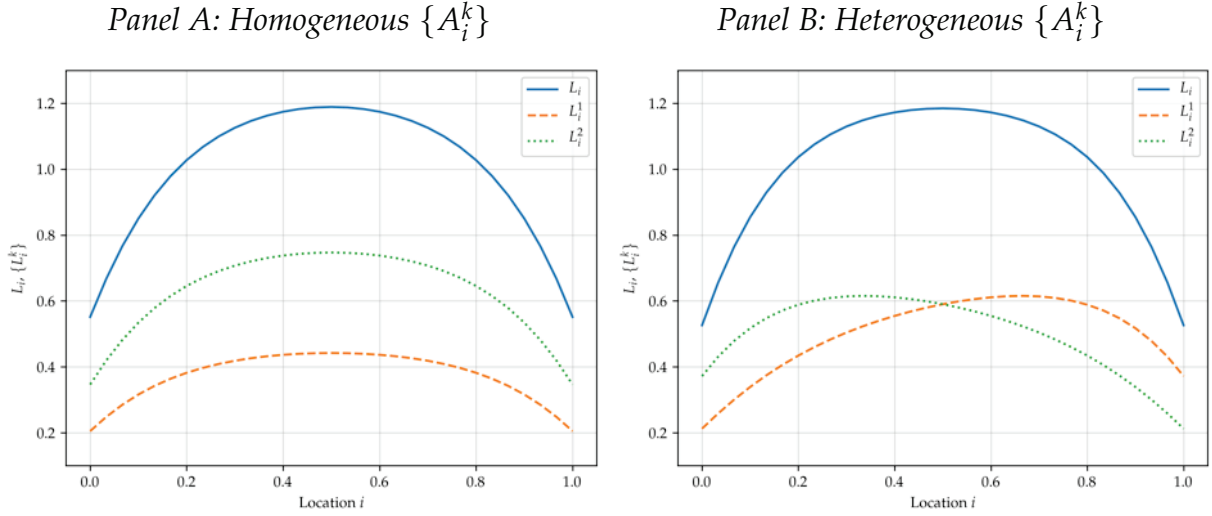
Notes: Equilibrium labor allocations for a simplified version of the model as described in Section 4.3. Panel A describes the allocation of workers (total and sector specific) if sector productivities do not change across locations ($A_i^1 = A^1 < A^2 = A_i^2 \ \forall i$). Panel B plots how the sector specific labor demands change if the right (left) most locations are the most productive ones in the first (second) sector.

productive) sector; however, the economy also produces goods for the first sector as the love for varieties' feature of the CES demand creates demand for it.

Panel B describes how the spatial equilibrium changes by altering the geography of the economy. In particular, it shows the equilibrium allocation of workers when the right (left)-most locations as the most productive in the first (second) sector. The equilibrium allocation becomes skewed accordingly, showing that the model implicitly determines that the most productive regions are those in which the economic activity is going to be agglomerated in each sector.

5   Calibration and goodness of fit

I calibrate my model to match SSA in the year of 2000. To do so, I use a mix of calibration and parametrization methods to map the model above to observable features of the SSA economy. The goal is to calibrate the exogenous parameters $\{\sigma, \eta, \theta, \{b_i^k\}_{i,k}\}$ and the geography fundamentals $\{\mathcal{L}, \mathcal{A}, \mathcal{U}, \mathcal{T}\}$. Table 1 summarises the methods and sources used and Appendix A.5 the numerical algorithms used.

5.1   CES and Frèchet dispersion

The exogenous parameters $\{\sigma, \eta, \theta\}$ are drawn from related literature. The lower-tier CES is set as $\sigma = 5.4$ following Costinot et al. (2016), who estimate it with similar

Table 1: Fundamentals, parameters, estimation methods and sources from the literature

Parameters	Description	Method	Reference
$\eta = 2.5$	Upper-tier CES	Literature	Sotelo (2020)
$\sigma = 5.4$	Lower-tier CES	Literature	Costinot et al. (2016)
$\theta = 3.4$	Workers (inv.) heterogeneity	Literature	Monte et al. (2018)

| Parameters	Subset	Description	Data source / Moment matched
$\mathcal{L}$	–	SSA's population endowment	Population data
$\{b_i^k\}_{i \in S}$	–	Sectoral shifters	Matched to location–sector production data in US$
$\mathcal{A}$	$\{A_i^k\}_{i \in S, k \neq K}$	Agricultural produc–tivities	GAEZ data
	$\{A_i^K\}_{i \in S}$	Non–agricultural pro–ductivities	Matched to GDP data in US$
$\mathcal{U}$	–	Amenities	Matched to population data
$\mathcal{T}$	dist(i,j)	Bilateral travel distance	Transportation data
	$\delta = 0.3$	Distance elasticity of τ	Moneke (2019)
	$\tau_{ij}^F = 1.15$	Trade friction to foreign markets	Baum-Snow et al. (2020)

data of mine at the same period, and for a geographical area comprising many SSA countries. The upper-tier CES is drawn from Sotelo (2020), i.e. $\eta = 2.5$. This value is estimated for the 1990's Peru, which is not a far-fetched approximation of the SSA economy in 2000. Finally, I set $\theta = 3.4$ following Monte et al. (2018) (the estimates for long term elasticities in development economies range between 2 and 4; see Morten and Oliveira (2018)).

5.2 Transportation network and trade costs

Trade frictions between locations are assumed to be proportional to the travel distance that separates them. In particular, I follow related research in the literature (e.g. Pellegrina and Sotelo, 2019; Donaldson, 2018, among others) by assuming that the trade costs of shipping goods from i to j take the following parametric format:

$$\tau_{ij} = \text{distance}(i, j)^\delta \times \tau_{ij}^F, \tag{20}$$

Figure 5: Estimated trade network for SSA – Western and Eastern Africa.
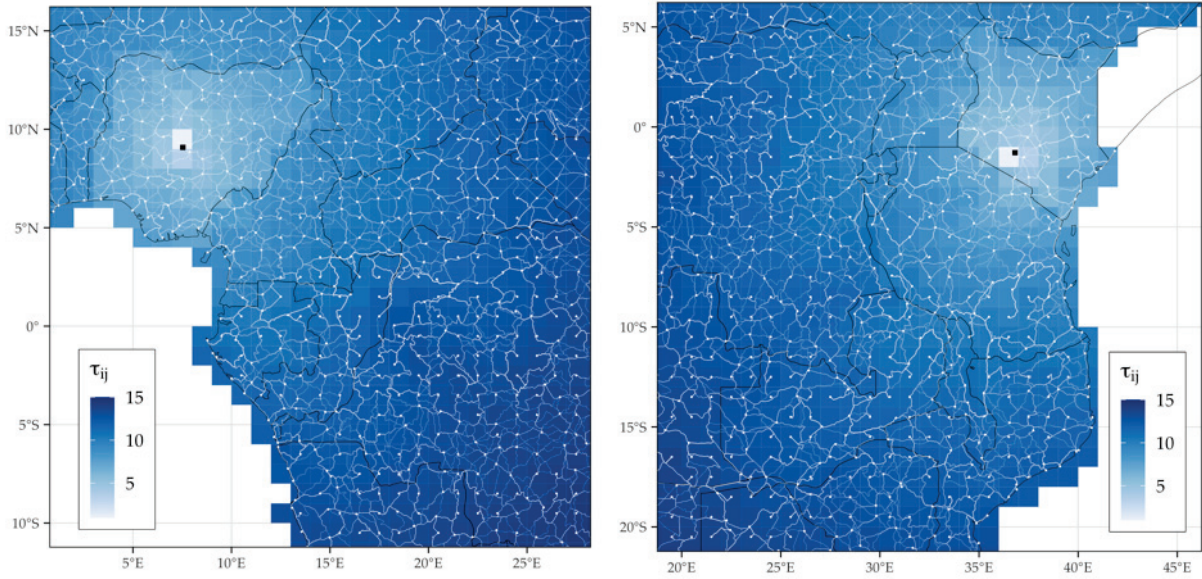
Notes: Estimated trade network for Western (left) and Eastern (right) Africa. The network is built by finding the shortest path between all neighboring cells over the road infrastructure with an optimal path algorithm. τ_{ij} stands for the estimated iceberg trade costs with respect to the capitals (black dots) of Nigeria (left) and Kenya (right). See Section 5.2 for details.

where distance(i, j) stands for the bilateral shortest distance between two locations and $\tau_{ij}^F > 1$ for an additional trade friction in case the location pair refers to places in different countries (i.e. $\tau_{ij}^F = 1$ if i, j belong to the same country). To calculate the bilateral distances between all location pairs, I proceed as follows. I overlay the roads' network data from gROADS onto the Accessibility to Cities' friction surface and set the pixels over the roads' data to be "cheapest" ones to be passed through.[19] I then use a pathfinding algorithm to calculate the shortest routes and respective distances between all neighboring cells.[20] With these distances in hand, I use the Dijkstra algorithm to calculate the shortest distance between all location pairs. The final step involves using these distances to build $\mathcal{T}$ following eq. (20): I set $\delta = 0.3$ following Moneke (2019), and $\tau_{ij}^F = 1.15$ following Baum-Snow et al. (2020). The result of it is a 2,032 x 2,032 matrix of trade costs; Figure 5 illustrates a subsample of this matrix. It can be seen that the trade network is very complex and reflects well the existing transportation infrastructure within and across countries. Moreover, it shows

---

[19]The advantage of my "two-input" strategy is that it provides additional information for my pathfinding algorithm when looking for the route between two coordinates that are not over a road. In such a case, the path would "go" to a road through an optimal route (i.e. considering the local geography) and then "move" over the road. This approach provides a more realistic outcome than if assuming a linear path to the closest road.

[20]The coordinates of each cell are obtained from the longitude and latitude of the most populated settlement/city in each cell. See Appendix B for more details and fig. C.4 for the results.

that further and/or foreign locations are those whose trade with is subject to a higher degree of frictions.

5.3 Fundamental productivities and sectoral shifters

I build the set of natural advantages and efficiency productivity parameters by partially drawing it from observable data; the remaining parameters are obtained with a calibration technique. In particular, I use the agro-climatic yields from GAEZ as the fundamental productivities of the agricultural sectors, i.e. $\{A_i^k\}_{i\in S, k\neq K}$.[21] The rationale for using the GAEZ estimates is that they measure the potential yield a certain location would obtain should its area be fully employed to grow a certain crop. Therefore, as the yield variation across location-crops are driven by each location's natural characteristics (such as soil and climate), it provides a reasonable measure for the parameters I am interested in.

The remaining elements to be quantified is the set of non–crop productivities $\{A_i^K\}_{i\in S}$, as well as the sectoral shifters $\{b_i^k\}_{i,k}$. I back them out with a standard inversion of my model conditional on observed endogenous variables in 2000: I solve for $\{\{b_i^k\}_k, A_i^K\}_{i\in S}$ that makes the model to simultaneously match GDP and sectoral production, in US\$, in all locations. This step is done by inverting the spatial equilibrium with numerical methods, carefully explained in Appendix A.5. Importantly, since I am not able to separately identify b_i^K from A_i^K, what I estimate is their product, which suffices for simulating the model.

5.4 Fundamental amenities

With these fundamentals and parameters in hand, I can solve for prices in the economy (eqs. (18) and (19)) and then solve for $\{u_i\}_{i\in S}$ by matching the model-implied population with the observed one from the data. In practice, I take advantage of the fact that my identification of the fundamentals holds up to scale and invert Equation (17) to obtain a closed–form solution for $\{u_i\}_{i\in S}$ as follows:

$$L_i = \frac{(w_i/P_i)^\theta u_i L_i^{-\alpha}}{\sum\limits_{j\in S}(w_j/P_j)^\theta u_j L_j^{-\alpha}}\mathcal{L} \propto (w_i/P_i)^\theta u_i L_i^{-\alpha} \quad \rightarrow \quad u_i \propto \frac{L_i^{1+\alpha}}{(w_i/P_i)^\theta} \tag{21}$$

---

[21]To be consistent with the SSA context, I draw the agro–climate potential yields calculated for rainfed agriculture with low usage of modern inputs; see Appendix B.1 for details.

5.5 Discussion of the inversion results

Figure 6 documents the results of the model inversion. First, wealthier locations are estimated to be particularly more productive in the non–agricultural sector. Moreover, densely populated locations with low income are estimated to have a higher value of amenities. In this case, however, the fundamental amenity parameter would be capturing not only intrinsic quality of life in locations, but also cultural and institutional characteristics that hinder the population in some locations to not be allocated therein.

Figure 6: Comparison between the results of the model inversion and observed endogenous variables.
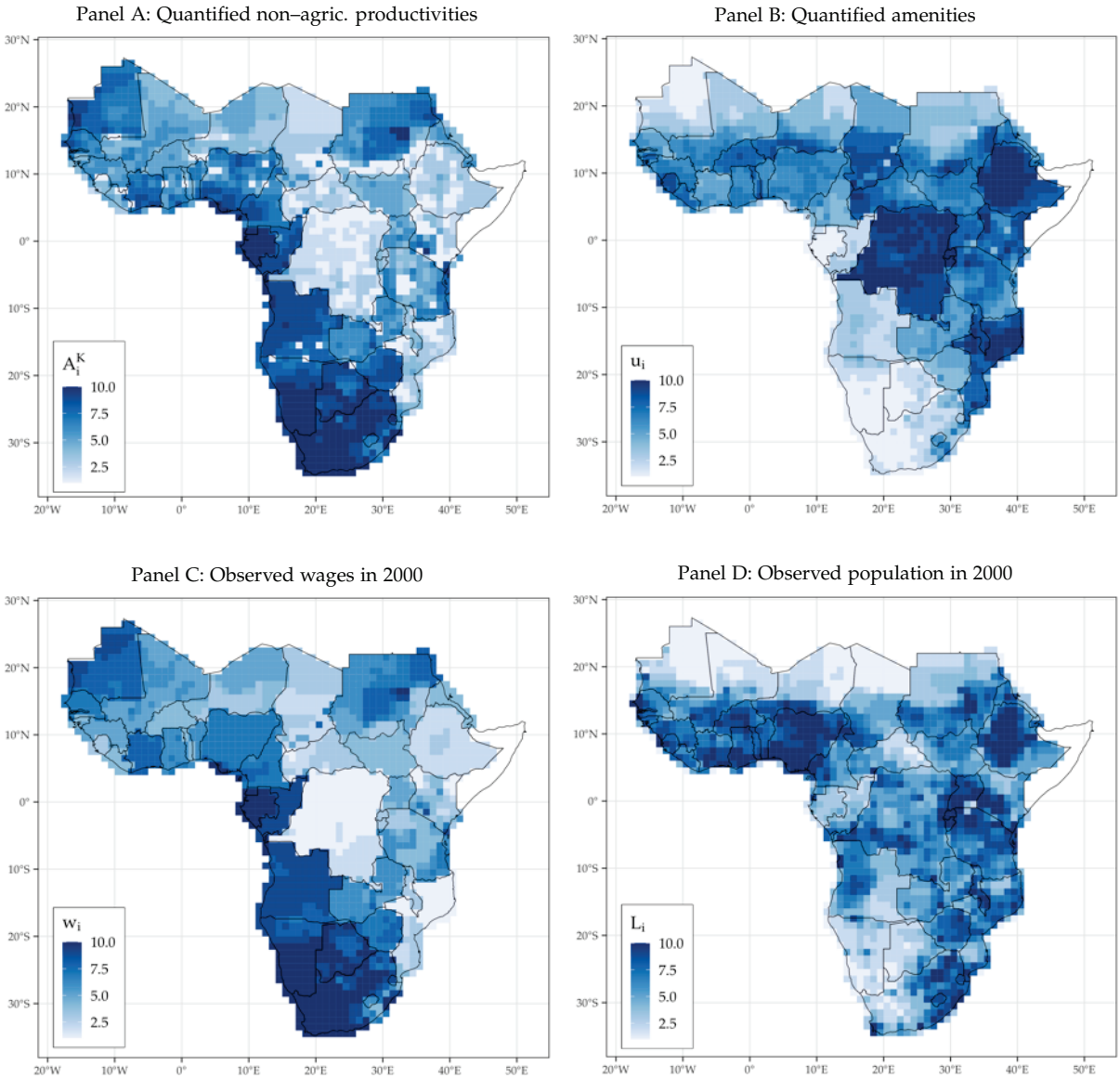
Notes: All results are shown in deciles, where 1 (10) stands for the bottom (top) decile of each sample. Each panel documents, respectively, the spatial distribution of the quantified non–agricultural productivities, quantified amenities, observed nominal wages and observed population.

DR Congo and Ethiopia illustrate this feature. Being among the poorest countries in SSA, but densely populated, is rationalized by the model as a consequence of high amenities. It is likely, however, that barriers to individuals to outmigrate from these countries also explain such an agglomeration of population. Therefore, the amenities I backout must be interpreted as a combination of the fundamental amenities and these other aspects, which will be held constant in my counterfactuals.

5.6 Model fit

Equipped with the calibrated model, I am able to test its capacity to replicate observed moments. I start with a backcasting exercise: I simulate the model after replacing the agricultural productivities and population endowments with their estimates for 1975. The result allows me to check whether my model replicates well the population distribution in SSA for that period. In addition to that, I check how well the model predicts the population displacement (differences) between 2000 and 1975 with respect to the observed values from the data, obtained from GHSP dataset.[22]

Figure 7 reports the results – the model is able to replicate closely the overall distribution of the population of 1975 within and across countries. Moreover, Panel A of Figure 8 documents the changes in the population counts between 1975 and 2000 are very well explained by the model (slope/R^2 of an OLS regression of 0.83/0.92).

Figure 7: Model goodness of fit with backcasting: population distribution in 1975.
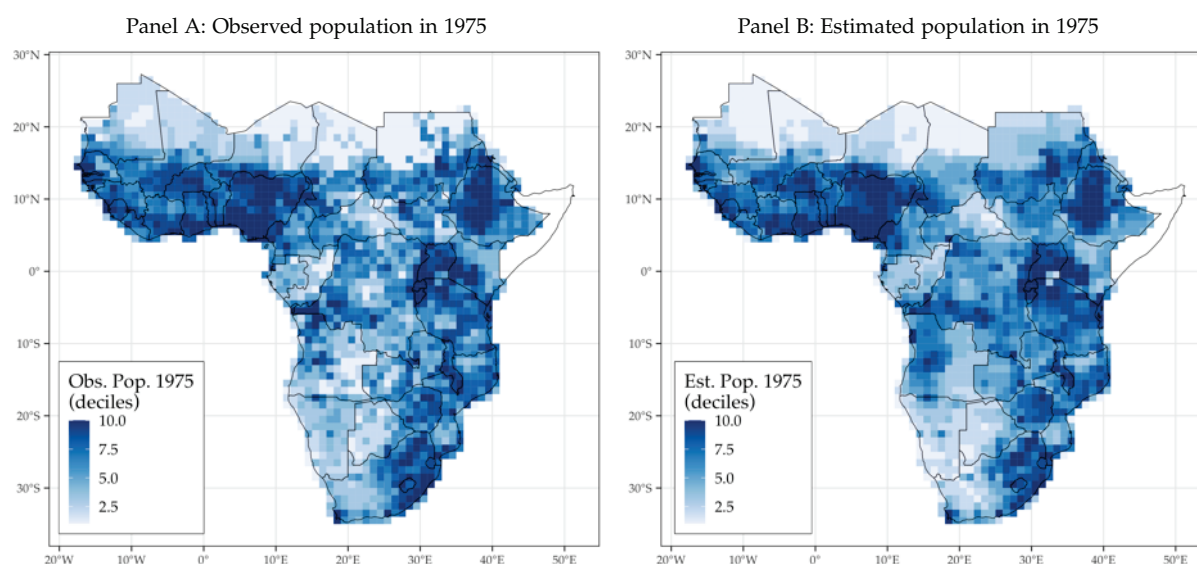
Notes: Panels A and B show the observed (from GHSP dataset) and model–implied population distribution in SSA for 1975, respectively. The values are shown in deciles; 1 (10) stands for the bottom (top) decile of each sample.

---

[22]As the data source for the population in 1975 (GHSP) comes from a different source than the data used in the calibration (G–Econ), I check their compatibility with the correlation of population in 2000 (available in both datasets) at the grid–cell and country level; see fig. C.1.

Figure 8: Model goodness of fit: backcasting results for differences in population and labor shares in agriculture for 2000.
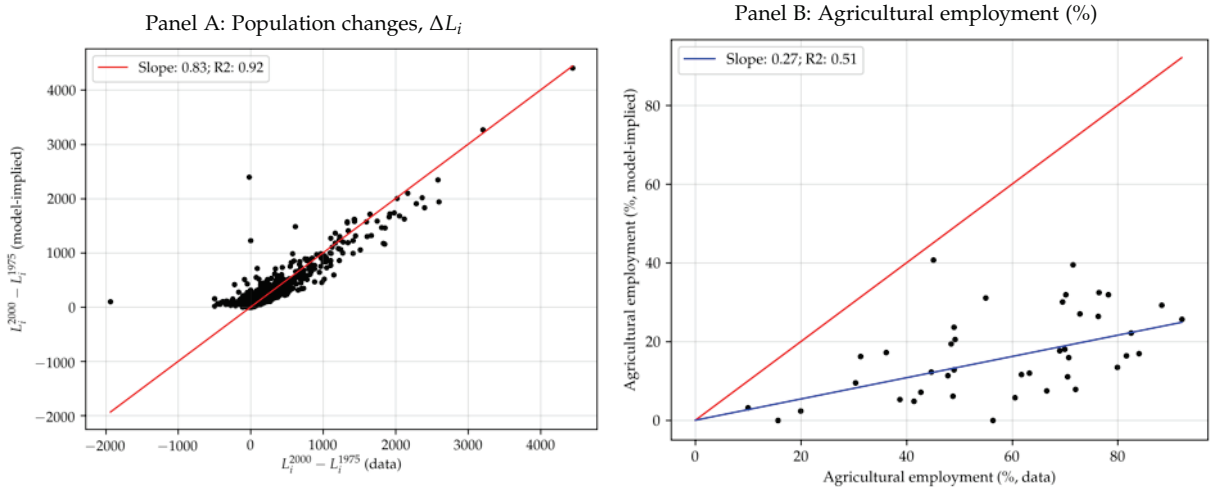
Notes: Panel A and B document the fit (observed versus model–implied) of the model for the changes in population (in thousands) between then and 2000, and 1975 and the country–level labor shares, respectively.

Importantly, the only change used in this exercise is on the agricultural suitabilities between the two periods. According to the GAEZ estimates, between 2000 and 1975 there have been already a substantial degree of changes in agricultural suitability – about 75% of the locations of my empirical setup experienced a decrease in average crop yields. Therefore, the fact that such changes can explain well the changes in population between periods in my model reassures its capacity of providing reliable numbers for the future.

As an additional overidentification test, I check the model's capacity to replicate sectoral employment shares. I focus on comparing the model implied agricultural shares (i.e. all crops) at the country level. I gather the agricultural share of employment in 2000 from the World Bank and compare with the shares generated by the model; Panel B of Figure 8 displays the results. The model does a good job when identifying the rank of countries with respect to agricultural employment shares, though underestimating their levels. On aggregate, the model predicts 20% of employment in agriculture compared to 58% in the data. This discrepency can be explained by the fact that I am only modeling a small fraction of the value generated by agricultural production.

6   Climate Change and Migration: The 2080 Forecast

I use my calibrated model to quantify potential climate migration with a series of counterfactual exercises. The benchmark counterfactual consists of solving for the

spatial distribution of the economic activity and population in 2080 after replacing the natural agricultural productivities with their estimates for 2080. The goal of this exercise is to quantify population reallocation with respect to a counterfactual SSA in 2080 in the absence of climate change.

Subsequently, I explore in detail the role that urbanization, crop–switching, and trade have in the resulting outcomes, and then perform a number of policy experiments to understand how technology adoption in agriculture may alter the estimated impact of climate change. I end by checking the robustness of my results to the degree of frictions in the economy, as well as to the climate model and climate change scenario underlying the data GAEZ data.

6.1 Benchmark counterfactual

I solve my model using the estimates of total population and natural agricultural productivities for 2080,[23] obtaining the spatial distribution of the economic activity and of the population. As forecasts point towards large population increases for 2080, a simple comparison with 2000's population in levels yields increases nearly everywhere. To address that, my metric for population displacement measures how the counterfactual population for 2080, L_i, compares with the model–implied distribution for 2080 without climate change, $\tilde{L}_i$. I define this metric ΔL_i, formally calculated in percentual changes as follows:

$$\Delta L_i = \left(L_i / \tilde{L}_i \right) - 1 \tag{22}$$

The results is shown graphically in Panel A of Figure 9. The first striking fact is the large number of locations – about 80% – whose population in 2080 are expected to be lower vis–à–vis a hypothetical scenario with no climate change. There is a large degree of heterogeneity, both within and across countries and locations. The median of the location–level population change is -9.3 percent, and the bottom and top deciles are -28 and 3 percent, respectively. Analogously, the median country experiences a -12 percent change in population, and the bottom and top deciles are of -22 and 8 percent change, respectively.

A comparison with Figure 3 (motivating fact 1) reveals a strong relationship between the intensity of climate change effects on productivities and the population movements. The fact that some locations, highly spatially clustered in Eastern Africa, are expected to become relatively more suitable to grow certain crops translates into

---

[23]When drawing the data for future periods, I choose the scenario that compares the closest to Representative Concentration Pathway (RCP) 8.5, a standard in the climatologist literature for a severe future (see Appendix B.1). I check the robustness of my results to less severe scenarios in Section 6.3.

Figure 9: Results of simulations of the SSA economy in a climate changed world in 2080.

| Panel A: Population changes | Panel B: Changes in non-crop employment |
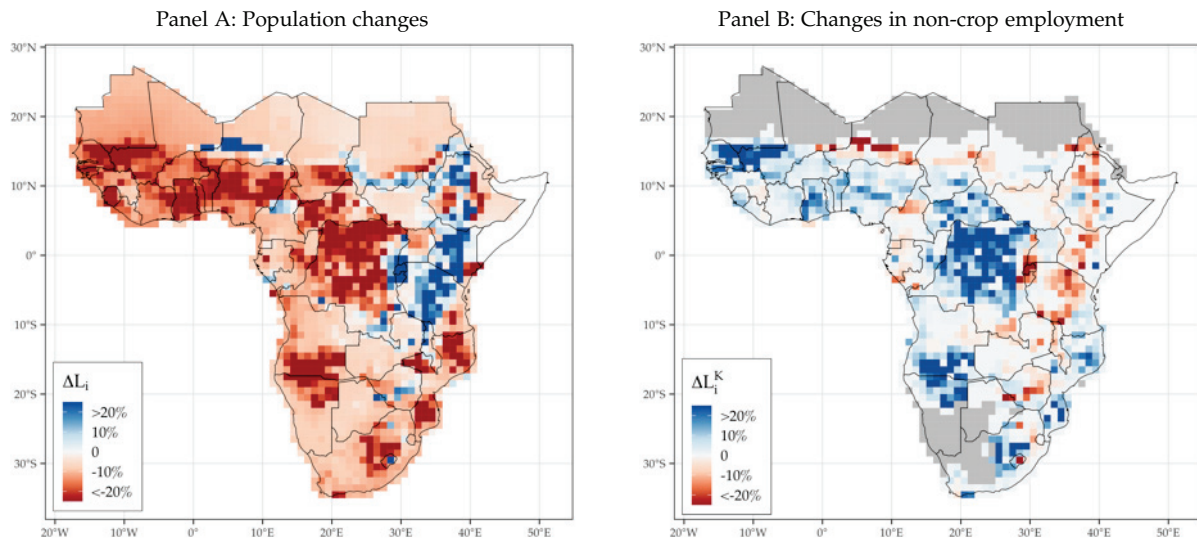
Notes: Results of counterfactual simulations for 2080 using agricultural productivity for 2080 estimated by GAEZ under a climate change scenario. The black dots stand for the main cities of countries. Panels A shows the changes in the cells' population compared to a scenario had the population distribution remained the same as in 2000. Panel B documents the difference between the within country population shares of the two scenarios.

higher specialization into agriculture inside my model. Such a pattern allows the SSA to adjust to climate change such that the overall supply of grain crops is not substantially affected, which would not be the case should climate change entail only productivity losses.[24]

Indeed, in terms of structural transformation, the aggregate employment share in the agricultural sector increases by 6 percentual points. This effect is very heterogeneous within and across countries, as shown in Panel B of Figure 9. Some countries (e.g. Ethiopia, Kenya, or Tanzania) go through an uneven process of sector–specialization, where some locations specialize more in agriculture and others in the non–agricultural sector. Other countries instead specialize more homogeneously. As a result, at the country level, the median change in non–agricultural employment increases by 1 percent, and the bottom and top deciles changes are of -2.8 and 5 percent, respectively.

Importantly, the results do not imply that all agricultural production would take place in Eastern Africa only. Indeed, it remains substantially spread out throughout SSA, and the most populated countries produce the highest bulk of grain production; see Figure C.2. Moreover, by comparing the least and most hit locations in terms of population loss, I find that they do not systematically differ in terms of fundamentals

---

[24]Indeed, if simulating a climate change scenario with suitability losses only, the magnitudes of the population displacements and agricultural production changes both decrease.

such as non–agricultural productivities, amenities, and market access. Such a fact, shown in Figure C.3, is very important to reject the hypothesis that the fundamentals that remain unchanged in my counterfactuals are the drivers of my results. Instead, a complex combination of several economic mechanisms and the climate change shock drives the population displacements, which I show next.

Heterogeneity of the results. I investigate how the economic mechanisms of my model interact with the estimated population displacements with a heterogeneity study. In particular, I focus on the locations expected to experience an outflow of people and investigate how, in the intensive margin, being able to urbanize, switch crops, and trade with other locations alters the observed climate change–migration relationship.

In order to do that, I estimate the relationship of several data moments and counterfactual outputs with a set of regressions. I begin with urbanization: I standardize the climate change–shock metric ΔA_i[25] and the estimated fundamental productivity in the non–agricultural sector, A_i^K, and use it to estimate the following regression:

$$\Delta L_i = \beta \times \Delta A_i \times A_i^K + \beta_1 \Delta A_i + \beta_2 A_i^K + \varepsilon_i, \tag{23}$$

where ΔL_i stands for the estimated population displacement in cell i in percentual terms. Standard errors are clustered at the country level. The coefficient of interest, β, estimates how much ΔL_i, in a location where the climate change shock is one standard deviation large, changes if being one standard deviation more productive in the non–agricultural sector. Therefore, it is informative of the role that the advantages in the non–agricultural sector have on the adjustment of the severely hit locations in terms of population changes.

The results are documented in Table 2, column 1. It shows that, on average, a large urban sector protects against the migration effects of climate change. This is shown by the point estimate of β: conditional on suffering a climate change shock one standard deviation high, being one standard deviation more productive on the urban sector decreases the outflow of people by 2.6 percentual points, one average. That stands for about one–fifth of the median estimated displacement of people in percentual points.

Next, I show that the ability to switch crops very strongly protects against climate change. To do that, I identify in the counterfactual results the locations that switch the main crop produced if compared to the 2000's portfolio. I replace that for A_i^K, as a dummy, in Equation (23), so that β has an analogous interpretation but with respect

---

[25]The measure ΔA_i stands for the changes in the average potential yields due to climate change in cell i; see Section 3 for details. However, to make the econometric interpretations more intuitive in this application, I calculate it in losses, such that positive values stand for an estimated decrease in the average suitability between 2000 and 2080.

Table 2: Population displacement induced by climate change: sensitivity to advantages in the non–agricultural sector, crop switching and market access.

	(1)	(2)	(3)
CΔ Impact	−8.547***	−6.913***	−7.889***
	(0.948)	(1.233)	(1.260)
Non–agric. productivity	1.921**		
	(0.870)		
CΔ Impact × non–agric. productivity	2.638***		
	(0.929)		
Switch crops		6.412***	
		(1.414)	
CΔ Impact × switch crops		3.645*	
		(2.062)	
Market access			−0.574
			(0.453)
CΔ Impact × market access			1.136***
			(0.378)
Observations	1,784	1,784	1,784
R^2	0.315	0.265	0.244

 Notes: The dependent variable in all specification are the estimated population outflows, in percentual points, as of Equation (22). Urban productivity stands for the estimated fundamental productivity in the non–agricultural sector, in standard deviations. Switch crops stands for a dummy for switching the main crop of the production portfolio in 2080 vis–à–vis the production mix of 2000. Market access refers to a standard measure of market access in standard deviations. Standard errors are clustered at the country level; $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

to effectively changing the production mix as a consequence of climate change. The results, shown in column 2 of Table 2, provide qualitatively similar results: by switching crops, locations hit by climate change have population outflows 3.6 percentual points lower.

To conclude, I focus on the role of acces to trade, and find that it is an insufficient insurance against climate change. To show that, I calculate the degree of market access of every location in 2000, using a standard measure from the trade literature (Donaldson and Hornbeck, 2016; Pellegrina, 2019), and replace it in Equation (23) as before. Therefore, the coefficient of interest measures the average change in the population displacement in a location hit by climate change but with access to markets one standard deviation higher, which is estimated to be of about 1.2 percentual points (column 3).

Overall, these three exercises provide a sound message: the capacity of urbanizing, switching crops, or trading plays an important shock coping role among the locations severely hit by climate change. Therefore, they emphasize the importance of account-

ing for these mechanisms when understanding potential climate migration, and the limitation of quantitative frameworks that do not do so. Another novel feature of my framework is the feasibility of policy experiments, which I perform next.

6.2 Policy Experiment – Technology Adoption in Agriculture

An issue on which the policy debate and academic research center is the low level of modern input usage in agriculture in low–income countries.[26] I explore the most of the GAEZ database to investigate how my results, in terms of population displacement and aggregates, change should the degree of technological inputs used in agriculture increase during the next decades.

As the benchmark simulation resorts to the GAEZ data that assumes production at the lower technological frontier (subsistence level), I perform two additional experiments using the 2080' crop suitabilities that assume the usage of intermediate and high modern inputs in production.[27] Moreover, to isolate the effects of climate change under these two scenarios, I also simulate the SSA economy by 2080 in the absence of climate change but with these higher degrees of technology. Therefore, I can estimate the climate change effects on two hypothetical worlds – with and without modern inputs in agriculture.

The results are documented in Table 3. Compared to the climate change effects in a low–input SSA (benchmark results, column 1), the usage of intermediate or high inputs reverse considerably the negative baseline setbacks. In particular, the real GDP losses are fully compensated in the intermediate case, and even become gains in the extreme scenario in which the entire SSA adopts high technology in agriculture in all sectors. While perhaps counterintuitive at first glance, such a result is driven by specialization. As shown in Figure 2, one of the results of climate change is that the more suitable locations for agriculture are more spatially clustered (especially in Eastern Africa). Therefore, the economy benefits of agglomerating the production around these locations (as average marginal production costs decrease). Indeed, SSA specializes even more in agriculture if adopting higher degrees of technology, which is seen in the changes in the labor shares and the amount of population displaced.

---

[26]See Osborne (2005); Restuccia et al. (2008); Sheahan and Barrett (2017), or Dethier and Effenberger (2011); Brenton et al. (2014); FAO (2015) for related policy studies.

[27]The subsistence level assumes that production is based on the use of traditional cultivars, labor intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control. The intermediate level assumes that production is based on improved varieties, on manual labor with hand tools and/or animal traction and some mechanization, and uses some fertilizer application and chemical pest, disease and weed control. The high technological level assumes production at the technological frontier, i.e. a farming system that is mainly market oriented and commercial production is a management objective. Production is based on improved high yielding varieties, is fully mechanized with low labor intensity and uses optimum applications of nutrients and chemical pest, disease and

28

Table 3: Aggregate and disaggregated results of the policy experiments with respect to technology adoption.

	(1) Benchmark results	(2) Intermediate technology	(3) High technology
Panel A: Aggregate results			
Population displaced	11.84%	22.16%	28.15%
Real GDP Change	-4.24%	1.77%	13.32%
Change in non–agric. employment	-6.43%	-8.8%	-9.69%
Panel B: Country–level results			
Pop. displaced, bottom decile	-21.81%	-44.26%	-56.28%
Pop. displaced, median	-11.53%	-23.15%	-30.33%
Pop. displaced, top decile	7.93%	18.59%	16.19%
Real GDP change, bottom decile	-17.8%	-46.56%	-60.07%
Real GDP change, median	-9.47%	-20.8%	-26.33%
Real GDP change, top decile	11.39%	40.51%	40.32%
Panel C: Cell–level results			
Pop. displaced, bottom decile	-28.45%	-45.36%	-53.82%
Pop. displaced, median	-9.38%	-19.3%	-25.24%
Pop. displaced, top decile	2.99%	1.02%	-2.14%
Real GDP change, bottom decile	-34.23%	-52.23%	-60.72%
Real GDP change, median	-8.69%	-17.91%	-23.32%
Real GDP change, top decile	9.06%	12.12%	11.43%

Notes: Panel A documents the aggregate results, in terms of population displacement, real GDP change, and changes in the non–agricultural labor shares, of the benchmark counterfactual (column 1) and of the policy experiments related to technology adoption in agriculture (columns 2 to 3). Panel B provides moments of the country–level changes of each of these simulations, and Panel C of the grid cell–level changes.

6.3 Robustness checks

I conclude my study with a sequence of robustness checks. I first focus on the sensitivity of my results, in terms of climate migrant flows and other aggregates, to the trade frictions and the heterogeneity of workers' location preferences (driven by the δ and θ parameters, respectively). I check how my outcomes change if increasing/decreasing these by 10%. However, to properly isolate the climate change effects under each scenario, I proceed as in Section 6.2 by simulating also a scenario with no climate change but subject to the new degrees of frictions. The difference between these two exercises is the counterfactual to be compared with the benchmark results.

The results are reported in Table 4, Panel A (and in Figure C.5 at the grid–cell level). Lower trade frictions reduce the shipping costs in the economy and generate

---

weed control.

stronger agglomeration forces. Thus, the economic activity is more reshuffled if compared to the benchmark exercise. The mechanism behind it are the higher incentives for sectoral specialization between locations benefited/damaged by climate change and, as a consequence, more population displaced. With lower trade costs, fewer locations can produce and supply agricultural goods to a wider market (thus the decrease in the agricultural employment). The results in terms of real GDP changes also reflect that: the frictionless the economy becomes, the lower the climate–induced real GDP losses. The opposite holds for the case of higher trade frictions.

In terms of the degree of heterogeneity on workers' location preferences (dispersion forces), I find that the climate migration flows monotonically increase the more heterogeneous agents become. The reason is that agents choose where to live based increasingly more on idiosyncratic reasons (i.e. in locations with higher amenities rather than locations with higher real income). However, a large subset of the high–amenity locations in SSA are among the regions benefited by climate change (mostly Eastern Africa). As a consequence, higher dispersion forces pushes even more individuals into that region (see Figure C.5), which increases even more the employment shares in the agriculture sectors (crops). Nevertheless, GDP losses become larger because the economy does not organize as well, in terms of sectoral specialization based on comparative advantages, as in the benchmark scenario (where workers put a higher weight on the "economic component" of the location choice). The opposite holds should dispersion forces get weaker (workers become more homogeneous).

Finally, I verify the robustness of my results to changes in the climate models generating the agricultural suitability data. As extensively discussed by Costinot et al. (2016), the GAEZ forecasts are produced with climatic General Circulation Models (GCM), which simulate the evolution of the global climate under an assumption of the evolution of the world's stock of carbon (climate change scenario). As mentioned in Appendix B.1, the GCM from which my data is drawn is the Hadley CM3 model for the RCP 8.5 scenario (a severe scenario in which carbon emissions increase throughout the 21st and 22nd centuries). I test the sensitivity of my results to other GCM, under the same scenario for climate change. The results, reported in Panel B of Table 4, are qualitatively very similar. The range of the climate–led displacement of population goes from 8 to 12 percent, roughly, which provides a "confidence interval" for the possible effects given the uncertainty around the estimates for the future climate.[28]

Moreover, I check the sensitivity of my results to the severeness of the underlying climate change scenario, by using the forecasts for the RCP 4.5. scenario (which assumes that carbon emissions will peak by mid–century and decrease thereafter, due

---

[28]The estimates for the Hadley CM3 model/RCP 8.5 (benchmark counterfactual) are the most extreme (in terms of suitability losses) between all GCM models. That is the reason for the larger magnitudes of the results in terms of population displacement, real GDP losses, and sectoral specialization.

Table 4: Robustness of the benchmark results with respect to trade frictions, dispersion forces (workers' heterogeneity), GCM models, and climate change scenarios.

	(1) Population displaced	(2) Changes in real GDP	(3) Changes in non–agricultural empl.
Benchmark results	11.84%	-4.24%	-6.43%
Panel A: Robustness to frictions			
Lower trade frictions	12.33%	-3.88%	-6.28%
Higher trade frictions	11.46%	-4.41%	-6.46%
Lower worker's heterogeneity	11.15%	-2.13%	-6%
Higher worker's heterogeneity	12.01%	-5.66%	-6.55%
Panel B: Robustness to different GCM			
CCCma	8.29%	-3.61%	-4.52%
CSIRO Mk2	9.52%	-3.98%	-5.26%
MPI ECHAM	9.69%	-2.96%	-4.29%
Panel C: Robustness to RCP 4.5 scenario			
Hadley CM3	7.50%	-3.79%	-4.71%
CCCma	6.16%	-2.13%	-4.37%
CSIRO Mk2	8.02%	-3.72%	-5.20%
MPI ECHAM	7.32%	-2.84%	-5.43%

Notes: Panel A documents the aggregate effects of climate change to different degrees of trade frictions and workers' heterogeneity (dispersion forces), driven by the parameters δ and θ, respectively. Panel B provides the results of the benchmark simulation using climate change data from GAEZ generated by different GCM models. Panel C reports the sensibility of the benchmark results to a less severe climate change scenario.

to developments in the implementation of climate–friendly technologies). I simulate my model with the suitability data for this scenario produced by all GCM models available; the results are documented in Panel C of Table 4. As expected, all aggregate effects are attenuated under such a scenario.

7 Final remarks

The main message of this paper is that understanding (and quantifying) climate migration in agricultural economies is not straightforward. The forecasts by climatologists are spatially heterogeneous along many dimensions – e.g. along the geography

and types of crops. As a consequence, the reactions of economic agents when adapting to such a shock could be numerous and interconnected.

Using a multi-sector spatial trade quantitative framework, I model climate change as a spatially heterogeneous shock to agriculture crops. I draw on state–of–the–art climate data to measure how agricultural economies in SSA would be impacted. I find that the expected changes in agricultural productivity could lead to large population displacements, within and across countries, of about 12 percent of the SSA population. I also discover that the adoption of technology in agriculture could reverse dramatically the economic losses of climate change.

One of the main takeaways of my paper is the shock coping role of sectoral reallocations (structural transformation). In particular, the capacity of adjusting the production mix towards other sectors – within agriculture or to non–agricultural sectors – upon being hit by climate change is shown to weaken considerably the climate shock–outmigration link. Trade is also found to be key when allowing for such mechanism.

However, some questions remain to be addressed. First, would exporting (cash) crops be as affected as staple crops and, if so, which impact would that have in the adaptation of SSA to future changes in the global weather? Which role would international (i.e. out of Africa) trade play in such a scenario? Moreover, how would frictions to sector switching – a fact in developing economies – alter the steady–state equilibrium of SSA? Finally, how would other aspects of climate change – e.g. coastal flooding, extreme weather events – interact with the agricultural suitability losses that I account for in this project? Expanding my setup along these dimensions is left for future research.

References

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

__ **and Dave Donaldson**, "The Geography of Path Dependence," *mimeo*, 2018.

__ **, Cauê de Castro Dobbin, and Melanie Morten**, "Border walls," Technical Report, National Bureau of Economic Research 2019.

Baez, Javier, German Caruso, Valerie Mueller, and Chiyu Niu, "Heat Exposure and Youth Migration in Central America and the Caribbean," *American Economic Review*, 2017, *107* (5), 446–50.

Balboni, Clare Alexandra, "In harm's way? infrastructure investments and the persistence of coastal cities." PhD dissertation, The London School of Economics and Political Science (LSE) 2019.

Barrios, Salvador, Luisito Bertinelli, and Eric Strobl, "Climatic change and rural-urban migration: The case of sub-Saharan Africa," 2006.

Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt, "Does investment in national highways help or hurt hinterland city growth?," *Journal of Urban Economics*, 2020, *115*, 103124.

Benveniste, Hélène, Michael Oppenheimer, and Marc Fleurbaey, "Effect of border policy on exposure and vulnerability to climate change," *Proceedings of the National Academy of Sciences*, 2020.

Berlemann, Michael and Max Friedrich Steinhardt, "Climate change, natural disasters, and migration—a survey of the empirical evidence," *CESifo Economic Studies*, 2017, *63* (4), 353–385.

Brenton, Paul, Alberto Portugal-Perez, and Julie Régolo, *Food prices, road infrastructure, and market integration in Central and Eastern Africa*, The World Bank, 2014.

Burke, Marshall and Kyle Emerick, "Adaptation to climate change: Evidence from US agriculture," *American Economic Journal: Economic Policy*, 2016, *8* (3), 106–40.

__ **, Solomon M Hsiang, and Edward Miguel**, "Global non-linear effect of temperature on economic production," *Nature*, 2015, *527* (7577), 235.

Burzyński, Michał, Christoph Deuster, Frédéric Docquier, and Jaime De Melo, "Climate Change, Inequality, and Human Migration," 2019.

__ , **Frédéric Docquier, and Hendrik Scheewel**, "The Geography of Climate Migration," *(mimeo)*, 2019.

Cai, Ruohong, Shuaizhang Feng, Michael Oppenheimer, and Mariola Pytlikova, "Climate variability and international migration: The importance of the agricultural linkage," *Journal of Environmental Economics and Management*, 2016, *79*, 135–151.

Cattaneo, Cristina, Michel Beine, Christiane J Fröhlich, Dominic Kniveton, Inmaculada Martinez-Zarzoso, Marina Mastrorillo, Katrin Millock, Etienne Piguet, and Benjamin Schraven, "Human migration in the era of climate change," *Review of Environmental Economics and Policy*, 2019, *13* (2), 189–206.

CIESIN, "Global roads open access data set, version 1 (gROADSv1)," *Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)*, 2013.

COMTRADE, UN, "United Nations commodity trade statistics database," *URL: http://comtrade. un. org*, 2010.

Conte, Bruno, Klaus Desmet, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "Local sectoral specialization in a warming world," Technical Report 2020.

Costinot, Arnaud, Dave Donaldson, and Cory Smith, "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*, 2016, *124* (1), 205–248.

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

__ , __ , **and** __ , "What do we learn from the weather? The new climate-economy literature," *Journal of Economic Literature*, 2014, *52* (3), 740–98.

Desmet, Klaus and Esteban Rossi-Hansberg, "On the spatial economic impact of global warming," *Journal of Urban Economics*, 2015, *88*, 16–37.

__ , **Dávid Krisztián Nagy, and Esteban Rossi-Hansberg**, "The geography of development," *Journal of Political Economy*, 2018, *126* (3), 903–983.

__ , **Robert E Kopp, Scott A Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H Strauss**, "Evaluating the Economic Cost of Coastal Flooding," *forthcoming: American Economic Journal: Macroeconomics*, 2020.

Dethier, Jean-Jacques and Alexandra Effenberger, *Agriculture and development: A brief review of the literature*, The World Bank, 2011.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ **and Richard Hornbeck**, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Ducruet, César, Réka Juhász, Dávid Krisztián Nagy, and Claudia Steinwender, "All aboard: The aggregate effects of port development," Technical Report, Technical report, manuscript 2019.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

FAO, "The Impact of Natural Hazards and Disasters on Agriculture and Food Security and Nutrition–A call for Action to Build Resilient Livelihoods," 2015.

Florczyk, AJ, C Corbane, D Ehrlich, S Freire, T Kemper, L Maffenini, M Melchiorri, M Pesaresi, P Politis, M Schiavina et al., "GHSL Data Package 2019," *Luxembourg. EUR*, 2019, *29788.*

Gröger, André and Yanos Zylberberg, "Internal labor migration as a shock coping strategy: Evidence from a typhoon," *American Economic Journal: Applied Economics*, 2016, *8* (2), 123–53.

Henderson, J Vernon, Adam Storeygard, and Uwe Deichmann, "Has climate change driven urbanization in Africa?," *Journal of development economics*, 2017, *124*, 60–82.

IIASA and FAO, "Global Agro-Ecological Zones (GAEZ v3. 0)," 2012.

IPCC, "IPCC Special Report," 2000.

_ , "Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change," *International Panel on Climate Change*, 2007.

_ , *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*, Cambridge University Press, 2012.

_ , "Global warming of 1.5 C," *An IPCC Special Report on the impacts of global warming of 1.5 degrees*, 2018, *1.*

Jones, Bryan and Brian C O'Neill, "Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways," *Environmental Research Letters*, 2016, *11* (8), 084003.

Lustgarten, Abrahm, "The Great Climate Migration Has Begun," *The New York Times*, Jun 2020.

Moneke, Niclas, "Can Big Push Infrastructure Unlock Development? Evidence from Ethiopia," Technical Report, Mimeo 2019.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–90.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *mimeo*, 2018.

Nagy, Dávid Krisztián, "Hinterlands, city formation and growth: evidence from the US westward expansion," Technical Report 2020.

Nath, Ishan B, "The Food Problem and the Aggregate Productivity Consequences of Climate Change," Technical Report, National Bureau of Economic Research 2020.

Nordhaus, William, "Integrated economic and climate modeling," in "Handbook of computable general equilibrium modeling," Vol. 1, Elsevier, 2013, pp. 1069–1131.

— , "Projections and uncertainties about climate change in an era of minimal climate policies," *American Economic Journal: Economic Policy*, 2018, *10* (3), 333–60.

— , "Climate change: The ultimate challenge for economics," *American Economic Review*, 2019, *109* (6), 1991–2014.

Nordhaus, William D, "An optimal transition path for controlling greenhouse gases," *Science*, 1992, *258* (5086), 1315–1319.

Nordhaus, William, Qazi Azam, David Corderi, Kyle Hood, Nadejda Makarova Victor, Mukhtar Mohammed, Alexandra Miltner, and Jyldyz Weiss, "The G-Econ database on gridded output: methods and data," *Yale University, New Haven*, 2006, *6*.

of Economic United Nations, Department and Population Division Social Affairs, "World Population Prospects 2019: Highlights," 2019.

Osborne, Theresa, "Imperfect competition in agricultural markets: evidence from Ethiopia," *Journal of Development Economics*, 2005, *76* (2), 405–428.

Pellegrina, Heitor S., "The Causes and Consequences of the Spatial Organization of Agriculture: Evidence from Brazil," *mimeo*, 2019.

Pellegrina, Heitor S and Sebastian Sotelo, "Migration, Specialization, and Trade: Evidence from the Brazilian March to the West," *mimeo*, 2019.

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

＿ , "Trade and Agricultural Technology Adoption: Evidence from Africa," *mimeo*, 2019.

Redding, Stephen J, "Goods trade, factor mobility and welfare," *Journal of International Economics*, 2016, *101*, 148–167.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu, "Agriculture and aggregate productivity: A quantitative cross-country analysis," *Journal of monetary economics*, 2008, *55* (2), 234–250.

Rigaud, KK, B Jones, J Bergmann, V Clement, K Ober, J Schewe, S Adamo, B McCusker, S Heuser, and A Midgley, "Groundswell: Preparing for Internal Climate Migration (Washington, DC: World Bank)," 2018.

Schlenker, Wolfram and David B Lobell, "Robust negative impacts of climate change on African agriculture," *Environmental Research Letters*, 2010, *5* (1), 014010.

＿ , **W Michael Hanemann, and Anthony C Fisher**, "Will US agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach," *American Economic Review*, 2005, *95* (1), 395–406.

Shayegh, Soheil, "Outward migration may alter population dynamics and income inequality," *Nature Climate Change*, 2017, *7* (11), 828–832.

Sheahan, Megan and Christopher B Barrett, "Ten striking facts about agricultural input use in Sub-Saharan Africa," *Food Policy*, 2017, *67*, 12–25.

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

Veronese, N and H Tyrman, "MEDSTATII: Asymmetry in Foreign Trade Statistics in Mediterranean Partner Countries," Technical Report, Eurostat Methodologies Working Papers 2009.

Weiss, D, A Nelson, HS Gibson, W Temperley, S Peedell, A Lieber, M Hancher, E Poyart, S Belchior, N Fullman et al., "A global map of travel time to cities to assess inequalities in accessibility in 2015," *Nature*, 2018, *553* (7688), 333.

Appendix

This appendix contains additional material related to the main text. In particular, appendix A documents theoretical derivations that support the main results of Section 4. Appendix B provides more details about the data sources mentioned in Section 2 and other data sources not mentioned therein. Appendix C contains additional figures and tables.

A   Theory Appendix

A.1   Derivation of shipping prices

The representative firm in location i uses labor as the unique input of a linear production technology. Locations trade with one another; following the iceberg-like formulation of trade costs, the quantity of a good from sector k produced by the representative firm from i shipped to location j is

$$q_{ij}^k = \frac{b_i^k A_i^k L_i^k}{\tau_{ij}}.$$

Thus, the representative firm solves

$$\max_{L_i^k} p_{ij}^k \frac{b_i^k A_i^k L_i^k}{\tau_{ij}} - w_i L_i^k \quad \forall k.$$

As a constant returns to scale problem, the solution is straight-forward: at an interior optimum, shipping prices will equal marginal shipping costs, i.e.

$$p_{ij}^k = \left(\frac{w_i}{b_i^k A_i^k} \right) \tau_{ij} \quad \forall i, j, k. \tag{A.1}$$

A.2   Derivation of Bilateral Trade Shares

When maximizing welfare with respect to consumption of varieties, worker v solves

$$\max_{\{q_{ji}^k\}_{j,k}} \left(\sum_{k \in \mathcal{K}} \left(C_i^k \right)^{(\eta-1/\eta)} \right)^{\eta/\eta-1} \varepsilon_i(v) \quad \text{s. to} \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{K}} p_{ji}^k q_{ji}^k \leq w_i \quad \forall i,$$

where $C_i^k = \left(\sum\limits_{j \in \mathcal{S}} (q_{ji}^k)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$. Suppose first workers choose sector composites; i.e. taking first order conditions with respect to C_i^k (μ stands for the Lagrange multiplier):

$$\frac{\eta}{\eta-1} w_i^{1/\eta-1} \frac{\eta-1}{\eta} (C_i^k)^{-1/\eta} - \mu P_i^k \leq 0 \quad \forall i,j, \quad = 0 \text{ for interior solution. Assume so, then}$$

$$(C_i^k)^{\frac{-1}{\eta}} = \mu P_i^k \times w_i^{1-\eta} \quad \forall i,j, \tag{A.2}$$

Then, one can write ratio of two sector consumptions as

$$\frac{C_i^k}{C_i^s} = \left(\frac{P_i^k}{P_i^s} \right)^{-\eta} \quad \rightarrow \quad C_i^k = \left(\frac{P_i^k}{P_i^s} \right)^{-\eta} \times C_i^s \quad \forall i,j,s. \tag{A.3}$$

Then, by defining μ_i^k as the share of i's spending in k-sector goods and making use of eq. (A.3),

$$\mu_i^k = \frac{P_i^k C_i^k}{\sum\limits_{k \in \mathcal{K}} P_i^s C_i^s} = \frac{P_i^k (P_i^k/P_i^s)^{-\eta} C_i^s}{\sum\limits_{k \in \mathcal{K}} P_i^j (P_i^j/P_i^s)^{-\eta} C_i^s} = \left(\frac{P_i^k}{P_i} \right)^{1-\eta} \quad \forall i,j,$$

where the last equation takes advatage of the definition of the price index from eq. (8). By proceeding analogously for the choice of crop varieties (q_{ji}^k), one finds that the share of spending on each location variety is defined as eq. (5).

A.3 Derivation of Population Shares

Take the definition of the welfare attained by a worker v living in i and moving to j as $W_i(v) = (w_i/P_i)\varepsilon_i(v)$, $\varepsilon_i \sim G_i(v) = e^{-v^{-\theta}u_i}$.[29] Following Eaton and Kortum (2002), one can obtain the distribution of the welfare from one specific location i as

$$A_i(w) \equiv \mathbb{P}(W_i \leq w) = G_j(w \times P_i/w_i) = e^{-(w \times P_i/w_i)^{-\theta}u_i}.$$

Thus, the joint distribution of welfare of all destinations from i can be derived as

$$A(w) = \prod_{i \in S} e^{-(w \times P_i/w_i)^{-\theta}u_i} = e^{-\Phi \times w^{-\theta}}, \quad \text{where } \Phi = \sum_{i \in S} (P_i/w_i)^{-\theta}u_i.$$

---

[29]For the sake of neatness, I omit the congestion forces present in the main model; i.e. I assume that $\alpha = 0$. The results are analogous if otherwise.

Now, recalling the share of workers living in i is equivalent to the probability that the welfare attained at i, w, is the highest among all other locations, one writes

$$\Pi_i(w) = \mathbb{P}\left(W_i(v) \equiv w \geq \max\{W_j(v)\}_{j\neq i}\right) = \prod_{j\neq i} \mathbb{P}(W_j \leq w) = e^{-\Phi^{-i} \times w^{-\theta}}.$$

With that, it is possible to obtain the unconditional probability Π_i by integrating over all possible values of $w \in \mathbb{R}_+$, i.e.

$$
\begin{aligned}
\Pi_i &= \int_0^\infty \Pi_i(w) \, d\mathbb{P}(W_i \leq w) \, dw \\
&= \int_0^\infty e^{-\Phi^{-i} \times w^{-\theta}} \times \left(e^{(w \times P_i/w_i)^{-\theta} u_i}(-\theta) w^{-\theta-1} (P_i/w_i)^{-\theta} u_i \right) dw \\
&= u_i \left(\frac{P_i}{w_i}\right)^{-\theta} \times \int_0^\infty e^{-\Phi \times w^{-\theta}}(-\theta) w^{-\theta-1} dw; \quad \text{multiply/divide by } \Phi \\
&= u_i \left(\frac{P_i}{w_i}\right)^{-\theta} \frac{1}{\Phi} \times \underbrace{\int_0^\infty \overbrace{e^{-\Phi \times w^{-\theta}} \Phi(-\theta) w^{-\theta-1}}^{dA(w)} dw}_{=1} \\
&= \frac{(w_i/P_i)^\theta u_i}{\sum\limits_{j \in S} (w_j/P_j)^\theta u_j},
\end{aligned}
$$

which is the equivalent of eq. (13) without congestion forces.

A.4 Numerical Algorithm for Solving the Model

To solve for the model's spatial equilibrium, I use a fixed-point approach starting from eq. (16). In particular, one can re-write it as

$$w_i L_i = \sum_{j \in S} \sum_{k \in \mathcal{K}} (P_j^k/P_j)^{1-\eta} \left(\frac{w_i \tau_{ji}}{b_i^k A_i^k P_j^k}\right)^{1-\sigma} w_j L_j$$

$$w_i^\sigma = L_i^{-1} \sum_{j \in S} \sum_{k \in \mathcal{K}} (P_j^k/P_j)^{1-\eta} \left(\frac{\tau_{ji}}{b_i^k A_i^k P_j^k}\right)^{1-\sigma} w_j L_j \rightarrow$$

$$w_i \equiv g_i(\mathbf{w}) = \left[L_i^{-1} \sum_{j \in S} \sum_{k \in \mathcal{K}} (P_j^k/P_j)^{1-\eta} \left(\frac{\tau_{ji}}{b_i^k A_i^k P_j^k}\right)^{1-\sigma} w_j L_j\right]^{1/\sigma}. \tag{A.4}$$

Noting that prices and labor are all explicit function of wages (eqs. (17) to (19)), starting from an initial guess of $\mathbf{w}$, eq. (A.4) provides an updated, model implied value of it. By iterating this procedure until the differences between steps are sufficient small, I solve for optimal wages up to a normalization. I then use the values found to solve

for optimal prices and labor distributions in Equations (17) to (19).

A.5 Model Inversion

I invert the spatial equilibrium of my model to back out the unobserved non-agricultural productivities, $\mathbf{A^K} \equiv \{A_i^K\}_i$, and productivity shifters for all sectors, $\mathbf{b^k} \equiv \{b_i^k\}_{i,k}$. First, to quantify $\mathbf{A^K}$, one can use eq. (16) to write

$$
w_i L_i = \sum_{j \in S} \sum_{k \in \mathcal{K}} \left(\frac{P_j^k}{P_j} \right)^{1-\eta} \left(\frac{w_i \tau_{ij}}{b_i^k A_i^k P_j^k} \right)^{1-\sigma} w_j L_j
$$

$$
w_i L_i = \sum_{j \in S} \sum_{k \neq K} \left(\frac{P_j^k}{P_j} \right)^{1-\eta} \left(\frac{w_i \tau_{ij}}{b_i^k A_i^k P_j^k} \right)^{1-\sigma} w_j L_j + \sum_{j \in S} \left(\frac{P_j^K}{P_j} \right)^{1-\eta} \left(\frac{w_i \tau_{ij}}{b_i^K A_i^K P_j^K} \right)^{1-\sigma} w_j L_j \rightarrow
$$

$$
A_i^K \equiv g_i(\mathbf{A^K}) = \left[\frac{w_i L_i - \sum_{j \in S} \sum_{k \neq K} \left(P_j^k / P_j \right)^{1-\eta} \left(w_i \tau_{ij} / b_i^k A_i^k P_j^k \right)^{1-\sigma} w_j L_j}{\sum_{j \in S} \left(P_j^K / P_j \right)^{1-\eta} \left(w_i \tau_{ij} / b_i^K P_j^K \right)^{1-\sigma} w_j L_j} \right]^{1/\sigma - 1},
$$

(A.5)

i.e. an expression for A_i^K as a function of all exogenous parameters, endogenous variables $\{w_i, L_i\}_{i \in S}$, fundamental productivities, and producitivity shifters. Amenities are not accounted conditional on observing labor distribution. Therefore, for an initial guess for $\mathbf{A^K}$, eq. (A.5) provides updated, model implied optimal values for $\mathbf{A^K}$ itself.

Subsequently, one can analogously solve Equation (10) for $\mathbf{b^k}$ as follows:

$$
X_i^k = \sum_{j \in S} \left(\frac{P_j^k}{P_j} \right)^{1-\eta} \left(\frac{w_i \tau_{ij}}{b_i^k A_i^k P_j^k} \right)^{1-\sigma} w_j L_j, \rightarrow
$$

$$
b_i^k \equiv f_i(\mathbf{b^k}) = \left[X_i^k / \sum_{j \in S} \left(\frac{P_j^k}{P_j} \right)^{1-\eta} \left(\frac{w_i \tau_{ij}}{A_i^k P_j^k} \right)^{1-\sigma} w_j L_j \right]^{1/\sigma - 1}
$$

(A.6)

Equations (A.5) and (A.6) provide to fixed–point solutions for $\{\mathbf{A^K}, \mathbf{b^k}\}$, conditional on observing endogenous variables $\{w_i, L_i, X_i^k\}_{i,k}$. As they both depend on one another, my algorithm consists of solving them sequentially, starting from an initial guess, until they both hold (i.e. until the difference between left and right hand sides are sufficiently small in both equations).

Importantly, given my model structure, I cannot separately identify in my estimation b_i^K from A_i^K; i.e. the fundamental producitvity and shifter of the non–agricultural sector. Therefore, it requires me to normalize $b_i^K = 1 \; \forall i$, so that the parameter that

I identify in Equation (A.5) stands for the product of them. Moreover, the shifters I quantify with Equation (A.6) have a relative interpretation, that is, with respect to the shifter of the non–agricultural sector whose level I cannot pin down.

B   Data Appendix

Several sources of data were put together for the purpose of this project. Here I describe in some more detail some of the sources detailes in section 2, and provide information of auxiliar data not mention thereof. A summary of all data sources used and their temporal coverage is described in Table C.1.

B.1   GAEZ agro-climatic yields.

The GAEZ's database provides estiamtes of agricultural potential yields for several crops, in different time periods, and for different degrees of technology usage in agriculture. As my interest in subsistence agriculture setup of SSA, I aim at building a time varying dataset of potential yields over the entire subcontinent, for several crops, at low usage of modern inputs: with rainfed water access, labor intensive techniques, and no application of of nutrients, no use of chemicals for pest and disease control and minimum conservation measures.

A challenge, however, is that the time varying potential yields from GAEZ are available only for **high usage of modern inputs** (based on improved high yielding varieties, fully mechanized with low labor intensity techniques, and usage of optimum applications of nutrients and chemical pest, disease and weed control). The estimates for different input levels are only available for the long–run estimates (averages between 1960–1990).

Therefore, to obtain a time varying dataset of the agro–climatic yields at low input usage, I first use the long–run values to calculate the GAEZ–implied ratio between high inputs ($A_i^{k,h}$) / low inputs ($A_i^{k,l}$) yields for each crop. This procedures reveals how the gains from adopting higher input levels differ across locations and crops – Figure B.1 illustrates the results for two selected crops in deciles. I use the calculated ratios to scale down the time varying estimates for high inputs that I collect.

Armed with the location–crop technology scales, I collect the time varying estimates of agro–climatic yields for high input usage. For the estimates in the past, retrieve those for 1971–1975 and 1996–2000. I average out the 5 years' blocks so to avoid year-specific outliers. The reason is to capture long term changes, which could be contaminated if a certain year faces unusual climate conditions.

The yield estimates for future periods require another parametrical selection: the

Figure B.1: Yield gains from adoption of high inputs in agriculture vis–à–vis low inputs for selected crops.

Panel A: Gains for adoption of high inputs for rice

Panel B: Gains for adoption of high inputs for sorghum
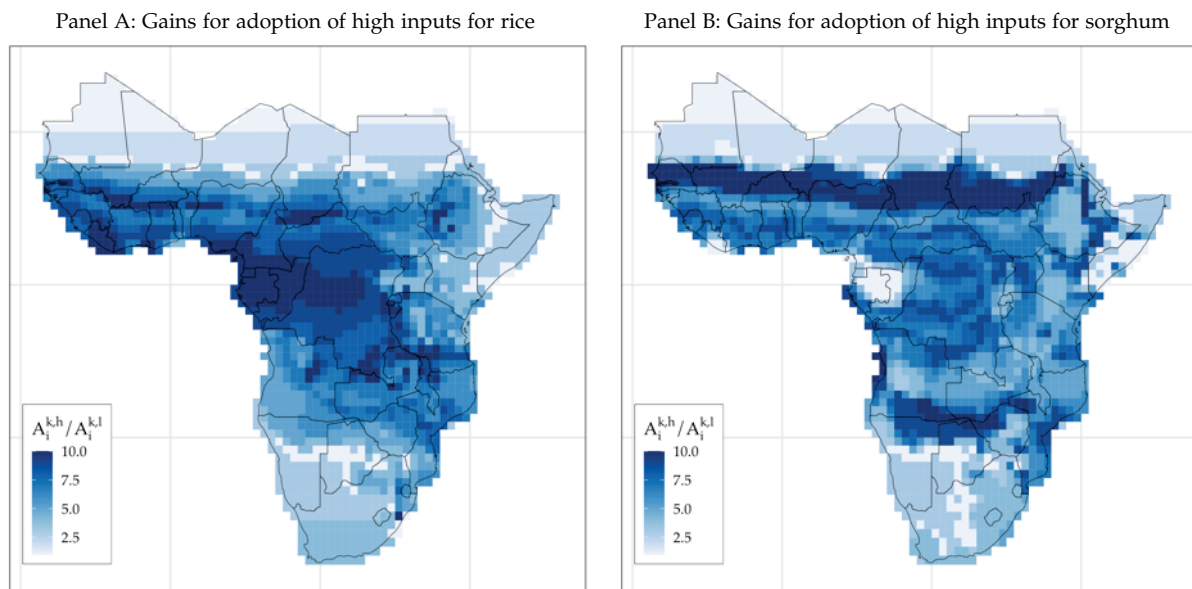
Notes: Panels A and B show the ratio of high/low input usage yields for growing two selected crops according to GAEZ long–run estimates. The values are shown in deciles; 1 (10) stands for the bottom (top) decile of each sample.

underlying scenario for which the data is produced and with which climatic (general circulation) model (GCM) the data is produced. As carefully discussed by Costinot et al. (2016), the GAEZ v3.0 database provides such estimates produced with four main GCM, and for several future scenarios. The latter is of key importance: it contains the underlying assumption on how the global carbon emissions are going to evolve in the future so to produce the changes in the climate.

I choose the scenario A2 from the GAEZ database, which matches closely the current standard of severe evolution of the global climate for the future: the Representative Concentration Pathway (RCP) 8.5.[30] This scenario assumes a steady increase in carbon stocks in the atmosphere througout the 21st and 22nd centuries, becoming stable by mid–23rd century. A milder scenario that I use for my robustness checks is the B1, which is similar to the nowadays–standard RCP 4.5. It assumes that the global stock of carbon will peak by late 21st century, becoming stable thereafter. Figure B.2 illustrates the equivalence between the SRES and RCP scenarios.

---

[30]Unfortunatelly, the GAEZ v3.0 database contains data produced under old standards for future climate scenarios – those produced in the Special Report on Emission Scenarios (SRES; see IPCC, 2000). The SRES scenarios were later updated by IPCC as the RCP scenarios, which are now the standards in the climate community (IPCC, 2012).

Figure B.2: Equivalence between long and longer–run estimates of radiative forcing (proportional to carbon emissions) between SRES and RCP scenarios.
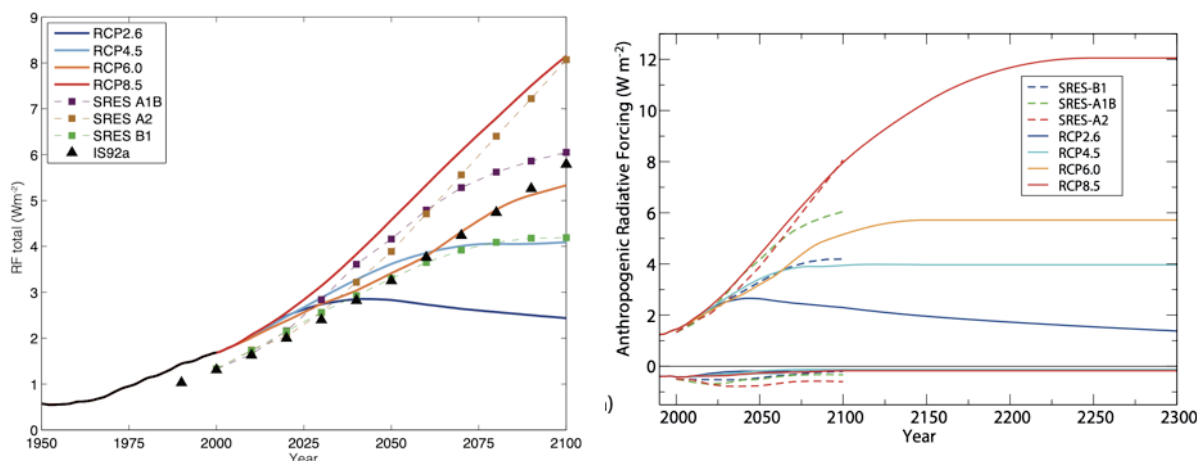
B.2   COMTRADE data.

The trade data used in this paper is obtained from the COMTRADE database (COM-TRADE, 2010). I collect bilateral trade flows, in current US$, for all crops of my study, between all country–pairs of my empirical setup. To do that, I rely on the COM-TRADE API system, which allows me to retrieve imports and exports trade flows between country at standard HS product codes.

Consistent with good practice with trade data, I collect import flows rather than exports. The reson for that is the usual discrepancy between total import and exports at the country–pair–product level. While import flows are registered between country of production and final country of shipment, export data usually register intermediate countries on the trade chain as final destination, biasing the trade flows (Veronese and Tyrman, 2009).

Finally, to transform the trade data to monetary unit of my study (US$ PPP from G–Econ), I proceed as follows. First, I calculate the share of trade flows, at the importer–exporter–crop–year levels, over the GDP of of the importing country in each year, in current values. Subsequently, I average out the shares over the 2000–2010 period, so to avoid outliers in the year of 2000. Finally, I multiply the shares at the importer–exporter–crop level by the importer GDP of G–Econ for the year of 2000.

B.3   Building the agricultural production data.

To build a dataset for agricultural production at the location–crop level for 2000, I combine the GAEZ data of production (in tonnes) with the FAOSTAT agricultural production data (country–crop level) and World Bank country GDP data (both in

current US$). First, I use the GAEZ data at the cell–crop level to calculate the share that each cell is observed to produce, of each crop, over its country's total production. Second, I obtain with the FAOSTAT and WB data the share of each country crop production for the years of 2000 to 2010. I average out such shares and multiply them by the country GDP implied by the G–Econ data, so that the unit is consistent with the monetary unit of the model (US$ PPP). Finally, I multiply the country–crop PPP values by the location–crop shares. For very liitle locations, however, the outcome can exceed the their total GDP. In these cases, I simply trim the value by 99.99% of its GDP.

B.4   Additional data sources

Main populated places. I collect the coordinaes of the main populated places of SSA from the Populated Places data set from Natural Earth. It consists of a geo-referenced dataset with the coordinates of about 90% of all cities, towns and settlements in the World. I use it to set coordinates for each of the cells of SSA. If a certain cell contains more than one location, I pick the one with the highest population. If another does not have any location to obtain the coordinates, I set them to be the cell's centroid. Finally, if any of the centroids are not located in the mainland (i.e. ocean, lakes), I set it to be the closest coordinate to the centroid that is on the mainland. See fig. C.4 for the result.

C   Additional figures and tables

Table C.1: Main data sources

Type of data	Coverage	Source
GDP and Population	2000	G-Econ Project v4.0 (Nordhaus et al., 2006)
Population	1975, 2000	Global Human Settlements Project (Florczyk et al., 2019)
Population projections	2021 – 2100	United Nations and Social Affairs (2019)
Area Harvested/Crop	2000	GAEZ v3.0 (IIASA and FAO, 2012)
Agric. Productivities	1960–2000	GAEZ v3.0 (IIASA and FAO, 2012)
Climate Δ projections	2020, 2050, 2080	GAEZ v3.0 (IIASA and FAO, 2012)
Transportation data	2000	gROADS project (CIESIN, 2013)
Friction transportation surface	2000	Accessibility to Cities' project (Weiss et al., 2018)
Bilateral crop trade data	2000–2010	COMTRADE (COMTRADE, 2010)

Table C.2: Share of grain crop production (in tonnes) over total production of the main staple and cash crops in SSA.

| Crop	Share of production
Grain crops:	
Cassava	56.65%
Maize	11.75%
Millet	4.59%
Rice	2.18%
Sorghum	6.15%
Wheat	1.13%
Total:	82.45%
Cash crops:	
Coffee	1.13%
Cotton	1.14%
Groundnut	2.72%
Palm oil	4.93%
Soybean	0.33%
Sugarcane	7.31%
Total:	17,55%

Source: GAEZ production data for 2000 aggregated in over the 42 countries of my empirical setup. SSA includes all sub–Saharan African countries but Somalia.

Figure C.1: Correlations between populations from G–Econ and GHSP datasets for the year of 2000.
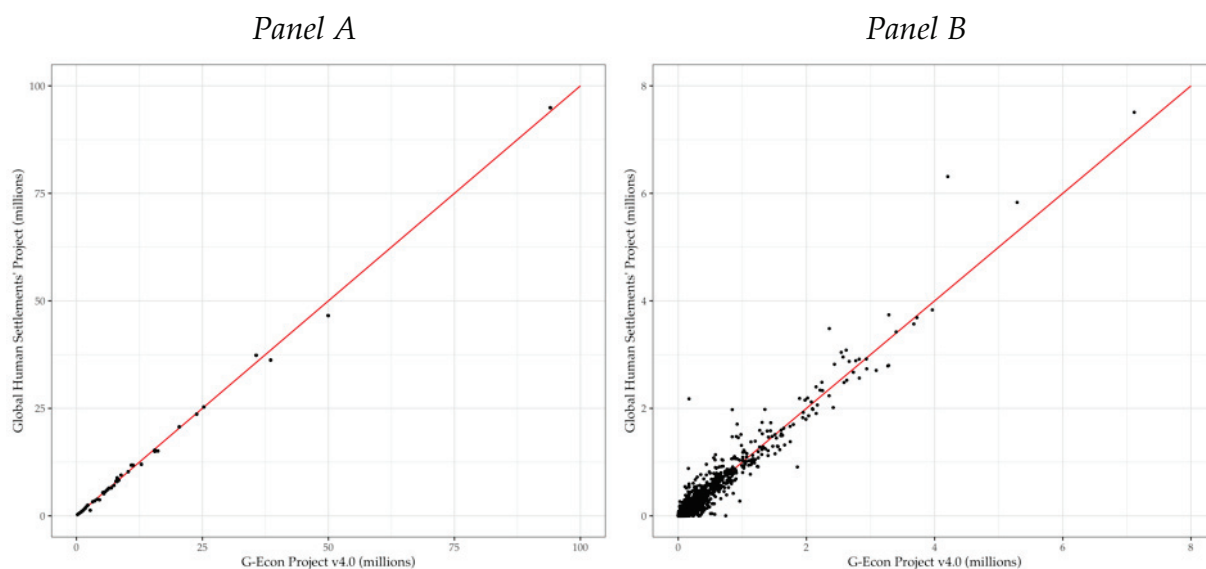

Notes: Panel A: Population counts in SSA from G–Econ (x axis) and GHSP (y axis) aggregate at country level. Panel B: Population counts in SSA from G–Econ (x axis) and GHSP (y axis) aggregate at 1 degree grid cells.

Figure C.2: Model implied agricultural output in 2080 under climate change for two selected crops.
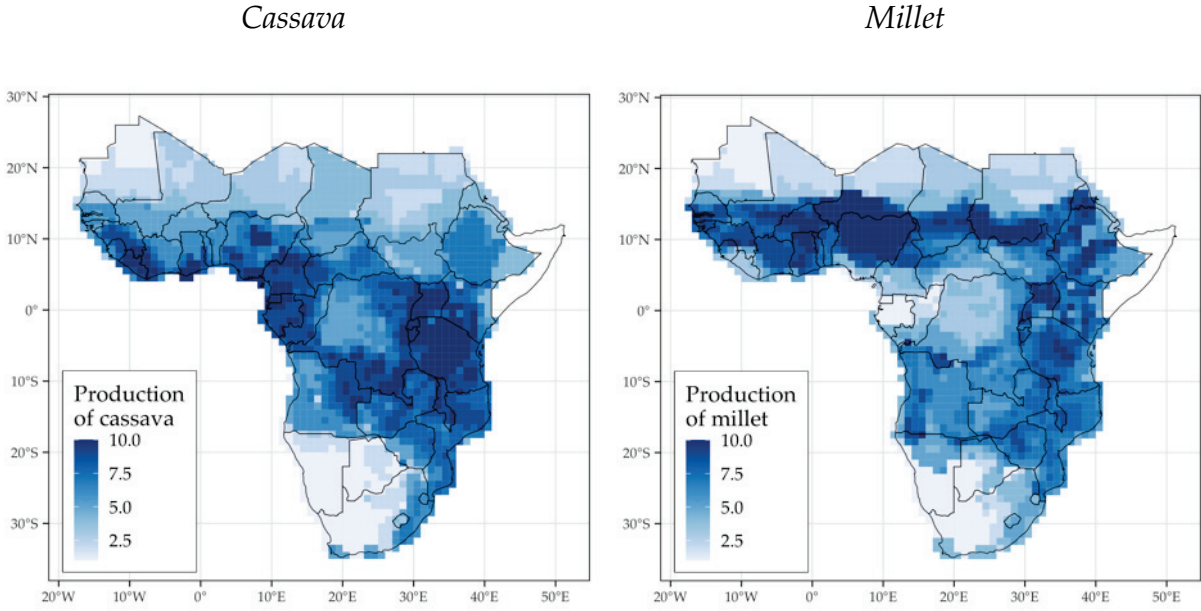
*Cassava*   *Millet*

Notes: Total crop output from the benchmark counterfactual simulation for cassava and millet, shown in deciles of each sample. 1 (10) stands for the bottom (top) deciles.

Figure C.3: Kernel density estimates of most/least hit locations (in terms of population loss) with respect to fundamental amenities, non–agricultural productivities, and market access in 2000.
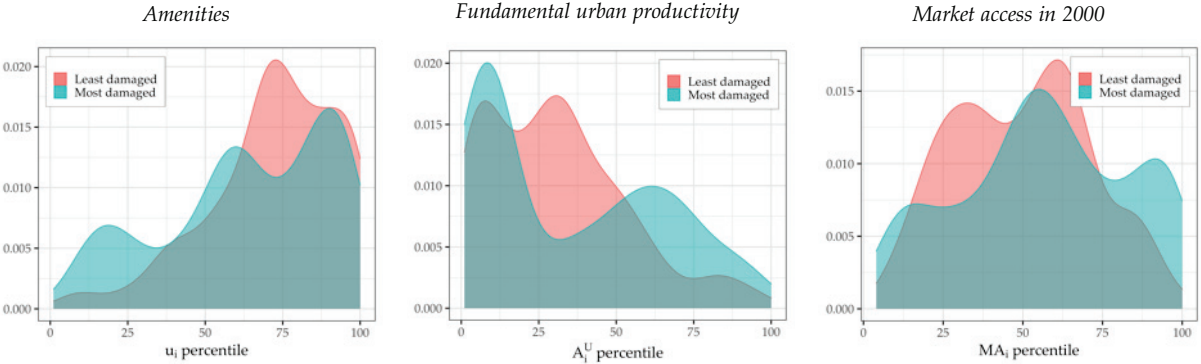

Notes: Densities with respect to the quantile of amenities, urban productivities and market access. The most hit locations stand for the bottom quintile of the subset of locations that are estimated to experience outflows of people, and account for about 17% of the total sample. The least hit locations stand for all of those expected to receive inflows of population, and represent about 20% of the sample.

Figure C.4: Coordinates for SSA grid cells (localities) for Western (left) and Eastern (right) Africa.
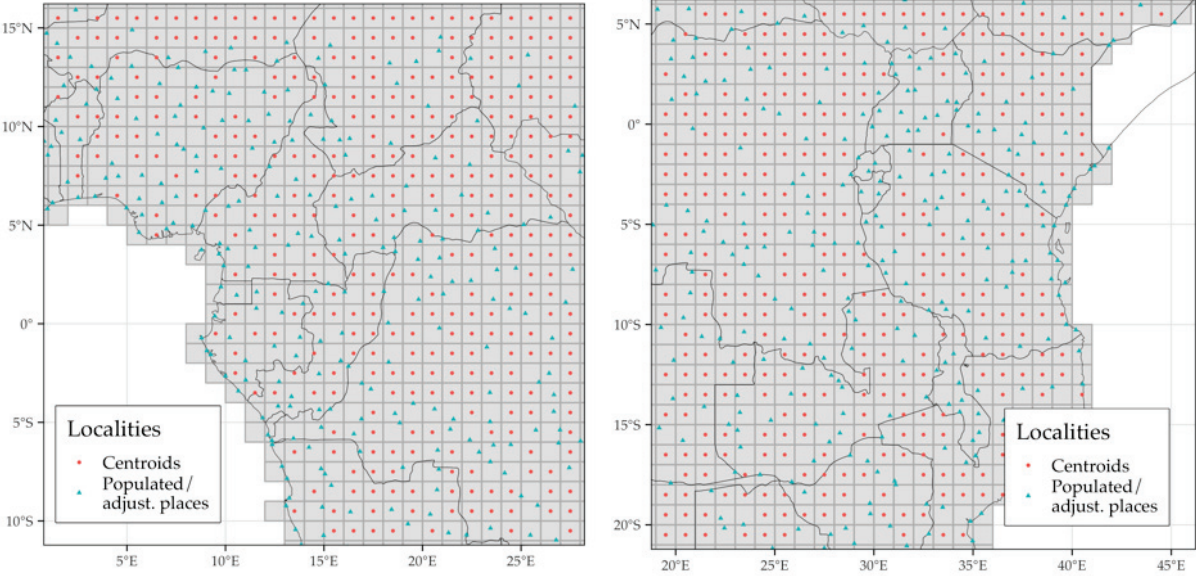
Figure C.5: Differences in differences results in terms of climate change–induced population changes between scenarios with different degrees of frictions and the benchmark results at the grid–cell level.

Lower trade frictions   *Higher dispersion forces*
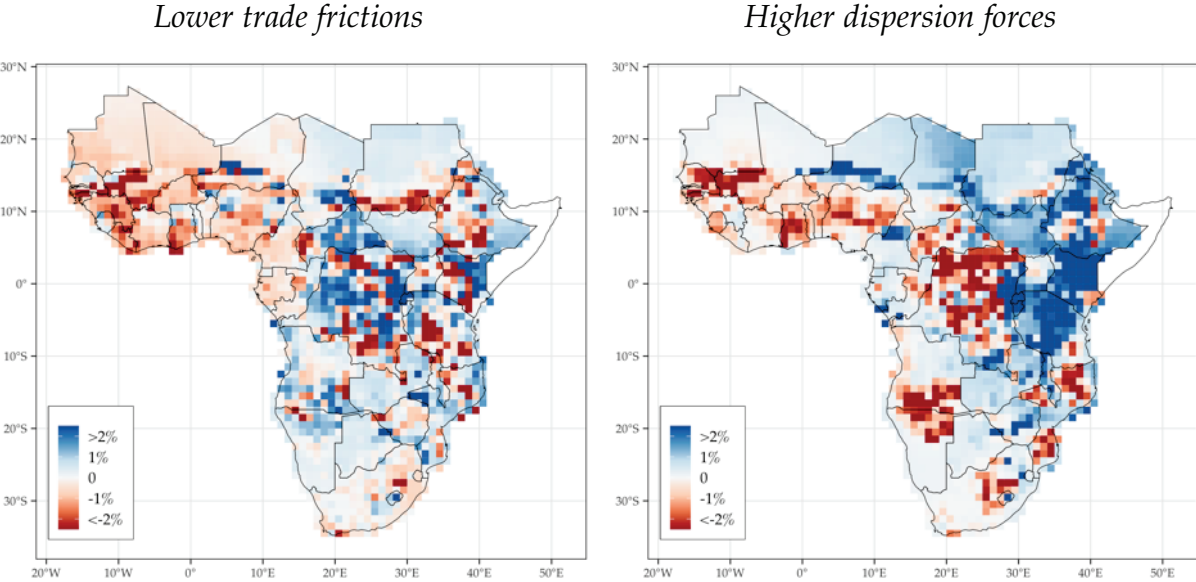
Notes: Both plots document the disaggregated effect of different degree of frictions (trade led by δ and dispersion forces by θ) on the estimated population changes induced by climate change. These are obtained as a differences in differences; i.e. the differences between the population changes in a scenario with lower trade frictions and those of the benchmark results (the same holds for the dispersion forces' plot).

C.1 Regression results for production and trade explained by comparative advantage

I explain here how I use the GAEZ data to investigate the relation between advantages to growing crops and the spatial patterns of agricultural production across and within countries of SSA and trade.

Comparative advantages and production. I combine three sources of GAEZ data: the agricultural production, in thousands of tonnes, the total harvested land, in thousands of hectares, and the average agro-climatic potential yields, for each crop. Formally, I define X_i^k as the dependent variable, standing for either total production/harvested land of crop k in grid cell i. Moreover, A_i^k is defined as the degree of suitability to grow k in cell i, in tonnes/hectares. I use it to estimate the following regression:[31]

$$\log\left(X_i^k\right) = \delta \log\left(A_i^k\right) + a_k + \varepsilon_i^k. \tag{C.1}$$

The parameter of interest, δ, estimates the average change in crop production (or harvested land), in percentual points, associated with a one percent increase in the crop suitability. Importantly, a_k stands for a set of crop fixed–effects (dummies); thus, the variation that identifies δ is within crops across geographical locations. Intuitively, δ shows the degree of association, throughout the entire SSA, of being more suitable to grow a certain crop and effectively producing it. That is, it provides evidence of an association between natural comparative advantage and specialization in agriculture at the crop level across SSA.

In order to understand whether such a relation holds at the country level, I replace a_k with a set of country–crop fixed effects. In that case, δ is identified with variation within country–crop , and its interpretation analogous. It identifies the intensity of the geographical clustering of agricultural activity on more suitable locations, according to GAEZ, within countries.

The results are documented in Table C.3. It provides strong evidence for the hypothesis in question. In particular, Panel A locations with potential yields one percent higher are found to produce, on average, 0.75% percent more if compared with all locations in SSA (column 1), and 0.62% if compared with locations from the same country (column 2). The results for harvested land (Panel B) are qualitatively equivalent. Overall, it conveys a sound message: crop–specialization happens both across and within countries. To generate this pattern, my general equilibrium model will take the perspective of subnational units that specialize in crops based on comparative advantage.

---

[31]If not otherwise specified, the regression models throughout the paper omit the constant as of neatness.

Importantly, the resuls shown in Figure 3 stand for the residuals of X_i^k and A_i^k from a regression on location and crop–country fixed effects (the most demanding possible). The estimated line stands for the results of a semi–parametric (polynomial) regression of these two controlling for the same fixed effects.

Table C.3: Suggestive evidence for the relation between natural comparative advantage (relative potential yields) and crop-specialization (effective production) in sub-Saharan Africa.

	(1)	(2)	(3)	(4)
	Panel A: Crop production (in logs)		*Panel B: Harvest land (in logs)*	
Potential yields (logs)	0.753***	0.623***	0.694***	0.612***
	(0.057)	(0.071)	(0.053)	(0.068)
Observations	12,192	12,192	12,192	12,192
R²	0.504	0.732	0.469	0.689
Crop FE	Yes	No	Yes	No
Country–crop FE	No	Yes	No	Yes

Notes: Estimation using GAEZ data (of year 2000) for agricultural production (thousands of tonnes), harvested land (hectares), and potential yields of agriculture (agro-climatic potential yields, in tonnes/hectare). Panel A uses crop production as the dependent variable; Panel B uses harvested land instead. *$p<0.1$; **$p<0.05$; ***$p<0.01$

Comparative advantage and trade. I provide further evidence of the importance of comparative advantage on economic outcomes in SSA by focusing on trade. In order to do that, I collect the average bilateral crop trade, in US\$, between all country pairs of my empirical setup, from 2000 to 2010, from the UN Commodity Trade Statistics Database (COMTRADE, 2010).[32] By combining it with the GAEZ potential yields, I learn whether trade flows are somehow determined by the degree of comparative advantage between countries at the crop level.

I start by looking at aggregate flows. I sum up total exports at the country–crop level, defining X_c^k as the total exports of crop k from country c. Next, I average out the GAEZ yields at the country level, analogously defined as A_c^k, and use it to estimate the following regression:

$$\log\left(X_c^k\right) = \delta \log\left(A_c^k\right) + a_k + b_c + \varepsilon_c^k, \tag{C.2}$$

where a_k and b_c stand for crop and country set of fixed effects, respectively. Therefore, the parameter of interest δ is identified with variation at the country–crop level, net out of invariant country and crop characteristics.

Subsequently, I use the trade flows at the bilateral level to investigate further the relation of interest. I first define bilateral exports at the country pair–crop level as $X_{cc'}^k$, where c and c' stand for the exporting and importing countries, respectively. I

---

[32]Appendix B.2 describes carefully how I collect and aggregate the raw COMTRADE data for the following empirical exercise.

proceed by calculating a bilateral measure of comparative advantage across countries, formally defined as $A^k_{cc'} = A^k_c / A^k_{c'}$,[33] and use it to estimate the following regression:

$$\log \left(X^k_{cc'} \right) = \delta \log \left(A^k_{cc'} \right) + a_k + b_c + d_{c'} + \varepsilon^k_c. \tag{C.3}$$

Equation (C.3) contains a set of crop, exporter and importer fixed effects, respectively a_k, b_c, and $d_{c'}$. Therefore, the correlation of interested, δ, is estimated with variation at the country pair–crop level, net out of the fixed effects. As of robustness, I add to Equation (C.3) a set of controls at the country pair level, such as distance between capitals, and dummies for geographical contiguity, common language, and ethnicity. Finally, I exploit the most of the variation in the data by adding country pair fixed effects, which net out all the variation in trade flows and comparative advantage at the bilateral level.

The results are documented in Table C.4. First, Panel A shows that the country's natural suitabilities are strong correlates of their exports: countries 1% more productive to grow a certain crop are observed to export about 0.5% more of that crop, on average. The same pattern holds at the bilateral level, shown in Panel B. The volume of crop exports from countries 1% more suitable, relative to the importing country, is about 0.4% higher, on average (column 2). The estimates become more precise with the inclusion of controls, as seen in column 3. Moreover, the magnitude of the estimates remains somehow stable even with the inclusion of more demanding (country pair) fixed effects, which absorb a substantial degree of the variation on relative suitabilities due to the high degree of spatial correlation between crop yields.

The results shown in Figure 3 stand for a semi–parametric regression (polynomial) of the specification of Table C.4, column 3. Moreover, the residuals plot are those of a regression of the dependent and independent variables on all fixed effects and controls.

---

[33]Defined as such, the data allows me to exploit the most of the trade flows, which contain exports between the same countries pairs in both directions.

Table C.4: Suggestive evidence for the relation between comparative advantages (relative potential yields) and crop exports in sub-Saharan Africa.

	(1)	(2)	(3)	(4)
	Panel A: Total crop exports	*Panel B: Total bilateral crop exports*		
Country potential yields (A_c^k, in logs)	0.473** (0.190)			
Bilateral relative yields ($A_{cc'}^k$, in logs)		0.360** (0.157)	0.490*** (0.148)	0.257 (0.203)
Observations	198	924	924	924
R^2	0.645	0.350	0.415	0.677
Crop FE	Yes	Yes	Yes	Yes
Exporter FE	Yes	Yes	Yes	No
Importer FE	No	Yes	Yes	No
Bilateral controls	No	No	Yes	No
Exporter–Importer FE	No	No	No	Yes

Notes: The dependent variable in Panel A is average country crop exports, in US$, between 2000 and 2010, and bilateral crop exports (average flows between 2000 and 2010, in US$) in Panel B. The trade data is collected from COMTRADE, and the potential yields from GAEZ (as of year 2000). Refer to Appendix B for details on the construction of the data. *p<0.1; **p<0.05; ***p<0.01