Deforestation and Structural Transformation in Sub-Saharan Africa

Kenneth Houngbedji Clément Imbert Samuel Marshall
Julien Wolfersberger Liam Wren-Lewis *

Thursday 26th March, 2026

Abstract

This paper studies how structural transformation affects deforestation in Sub-Saharan Africa using data from 1429 sub-national units in 28 countries. The relationship between non-agricultural growth and deforestation is a priori ambiguous: while rising urban demand for agricultural goods may increase deforestation, migration of farmers to cities can reduce it, particularly in Africa where deforestation is driven by smallholders. We first provide evidence of these opposing migration and internal trade mechanisms using a shift-share approach, before building and estimating a three-sector spatial equilibrium model. Analysis of counterfactuals from the model reveal meaningful heterogeneity. While agricultural and non-agricultural productivity changes increased non-agricultural employment in almost all countries, their impact on deforestation changes sign depending on the spatial distribution of productivity changes, migration and trade flows.

JEL Codes: Q23, Q56, O13.

*Houngbedji: IRD Université Paris-Dauphine. Imbert: Sciences Po clement.imbert@sciencespo.fr. Marshall: University of Warwick. Julien Wolfersberger: AgroParisTech and Paris-Saclay University. Wren-Lewis: Paris School of Economics. We would like to thank numerous conference and seminar participants, as well as Kirill Borusyak, Bruno Conte, Nicolas Coeurdacier, Antoine Leblois, Thierry Mayer, Florian Oswald, and Anant Sudarshan for helpful comments. Edoardo Badii, Robin Benabid Jégaden, Iddo Glass, Flavia Grasso, and Alonso Venegas provided excellent research assistance. We gratefully acknowledge funding from the Agence Nationale de la Recherche (ANR-24-CE26-2254), INRAE EXPLOR'AE (ANR-24-RRII-000), and the STEG research initiative.

1 Introduction

Since the turn of the century, Africa has been the continent with the highest amount of deforestation, losing an average of 4 million hectares of forest every year (FAO, 2020), raising concerns about climate change, biodiversity losses and reduced ecosystem services (Pendrill et al., 2022). During this same period, many Sub-Saharan African countries underwent structural transformations, with steep declines in the share of labor in agriculture.[1] Did structural change cause deforestation in Africa? This seems plausible — rising incomes increase the demand for agricultural products, while increases in agricultural productivity lead to greater export competitiveness, both potentially driving up demand for agricultural land. On the other hand, unlike in other continents, deforestation in Africa is primarily linked with small-scale agriculture that produces food for local consumers (Curtis et al., 2018). This raises the possibility that structural change may decrease pressure on forests as a result of workers emigrating towards growing urban centers. Indeed, unlike in the rest of the world, African countries that saw higher rates of urban population growth had lower rates of deforestation (Figure 1).

In this paper, we investigate the effects of structural transformation on deforestation in Sub-Saharan Africa, using data on 1429 administrative units from 28 countries. We begin by providing motivational reduced form evidence that non-agricultural employment growth can reduce deforestation through migration out of rural areas (migration channel) but can also increase it through increased demand for agricultural products (internal trade channel). To do this, we construct two shift-share instruments, where shares are either based on historic migration patterns (migration shift-share) or based on market access (trade shift-share) between each origin and each destination, and shifts are the non-agricultural employment growth in the destination. These two shift-shares are far from perfectly correlated across origins (correlation of 0.3). We find evidence that deforestation was lower in origins exposed to structural change via migration and higher in those exposed via internal trade.

We provide three additional pieces of evidence to support our interpretation of these results as reflecting migration and internal trade. First, we show that the results are virtually unchanged when we control for local growth in non-agricultural employment. This suggests that our results are not driven by a correlation between economic shocks at destination and at origin, but by the transmission of the destination shock to the origin via migration and trade flows. Second, we split the sample equally between units where cash crops represent a higher vs lower share of harvested area. Since cash crops are not primarily for domestic consumption but tend to be exported, we do not expect them to respond as strongly to migration or internal trade. Indeed we show that our results are entirely driven by units with a lower share of cash crops. Finally, we split the sample between countries that have a higher vs lower share of agricultural imports in their GDP. If countries import their food, then structural change may move consumption across space as workers go from the agricultural to the non-agricultural

[1] For instance, the share of labor in agriculture fell from 55% to 30% in Ghana, and from 83% to 65% in Tanzania between 2000 and 2020 (ILO, 2021).

sector without affecting deforestation. Indeed, we find that our results are concentrated in countries that import fewer agricultural goods. Overall, our shift-share results lends credence to the idea that structural transformation can have positive or negative effect on agricultural land use and deforestation depending on the geography of migration and trade links.

To answer these questions and better understand how structural change impacts deforestation, we therefore build and estimate a spatial equilibrium model with internal migration and trade. Our model features costly movement of workers and goods as well as heterogeneity in land-productivity and in workers' location tastes. In each location, there are numerous plots of land which can be used for agriculture or left under forest. Opening the land to agriculture requires the payment of a fixed cost, which corresponds to a land conversion investment.

We then use this model to provide counterfactual analysis and estimate the impacts of historic productivity changes in both agriculture and non-agricultural sectors. We find that productivity growth accounts for most of the structural transformation observed between 2000 and 2010, with both agricultural and non-agricultural productivity playing important roles, though their relative importance varies substantially across countries. Overall, productivity growth increased deforestation, accounting for roughly 10 percent of observed forest loss, but this average effect conceals substantial cross-country heterogeneity, with negative net effects in many countries.

What explains this heterogeneity? One key factor is openness to international trade - agricultural productivity increases tend to increase deforestation in more open countries, but decrease deforestation in those that are closed, with non-agricultural productivity shocks having the opposite effect. This is, however, only part of the story, with another key factor being the location of the productivity shocks within the country relative to *internal* trade and migration flows. Indeed, following the intuition given by our reduced form analysis, we show that non-agricultural productivity shocks reduce deforestation if they take place in areas with relatively high migration flows from forested areas.

Our work is the first to study the effect of structural transformation on deforestation through its spatial equilibrium effects via internal migration and trade. It represents an important addition to the literature on deforestation in tropical countries. Early literature on the economics of tropical deforestation uses cross-country regressions to investigate economic and institutional factors behind deforestation (see Barbier and Burgess, 2001, for a review). The recent literature exploits advances in the measurement of deforestation using satellite data and new empirical strategies to get at the causal effect of determinants of deforestation (see Balboni et al., 2023, for a recent review).[2] The literature has highlighted the importance of international trade (Ferreira, 2004; Leblois et al., 2017; Abman and Lundberg, 2020; Hoang and Kanemoto, 2021) and market access through changes in

[2] A burgeoning literature investigates institutional and political factors, such as property rights (Fetzer and Marden, 2017; Wren-Lewis et al., 2020), political incentives (Burgess et al., 2012; Balboni et al., 2021), or anti-poverty programs (Ferraro and Simorangkir, 2020), and evaluates the effectiveness of payments for ecosystem services (Jack and Jayachandran, 2019; Souza-Rodrigues, 2019).

road infrastructures (Pfaff, 1999; Andersen, 2002; Pfaff et al., 2007; Asher et al., 2020; Kaczan, 2020; Gollin and Wolfersberger, 2023). Another strand of the literature has addressed the effects of population forces (Angelsen, 1999) and demand for fuelwood and wood products induced by rising incomes (Foster and Rosenzweig, 2003; Baland et al., 2013) on deforestation. Most of the empirical literature focuses on the local impacts of population and income growth, which tend to be negative (Geist, 2017; Seto et al., 2012; Simkin et al., 2022). Fewer papers look at the deforestation impacts of migration to urban centers on the sending rural locations. Angelsen et al. (2020) find mixed impacts of international migration on deforestation in sending communities in case studies in Guatemala. And Brewer et al. (2022), who find that labor loss and the inflow of remittances from rural urban migration translates into reduced tree cover loss in Uganda. Our reduced form results based on on a set of countries that includes Uganda confirm and generalize Brewer et al. (2022) findings.[3] Our contribution is to investigate migration and trade together as determinants of deforestation in the process of structural transformation, and to embed them in a spatial equilibrium model which allows us to quantify their net effect in equilibrium.

Our paper also contributes to the literature on internal trade and migration in developing countries, which increasingly employs new quantitative spatial models (Redding and Rossi-Hansberg, 2017). While several papers have developed spatial equilibrium frameworks to study agriculture (e.g., Sotelo, 2020; Pellegrina, 2022), or aggregate productivity (Tombe, 2015; Bryan and Morten, 2019), their focus is on economic rather than environmental outcomes.[4] Several recent papers use spatial models of trade and migration to study deforestation, but not focusing on structural change (Dominguez-Iino, 2021; Araujo et al., 2023; Gollin and Wolfersberger, 2024; Restrepo and Leite-Mariante, 2023). One exception is Farrokhi et al. (2023), who develop a spatial dynamic framework to study global deforestation. They study the effect of international trade on deforestation, including population growth and structural change. They show that decreasing agricultural trade costs can reduce deforestation if cheaper agricultural products increase demand for non-agricultural goods, which generates structural change. The focus of their analysis is on international trade and sectoral reallocation, rather than internal trade and migration, but their take away resonates with ours: structural change in developing countries can help curb deforestation.

The paper is structured as follows. Section 2 describes our data we use. Section 3 presents a shift-share strategy, whose results are presented in Section 4. Section 5 develops a model of deforestation and structural transformation. Finally, Section 6 quantifies the model and presents counterfactuals.

[3] Also related, Madhok et al. (2022) analyze the impact of rural-urban migration on agriculture in rural areas in India. They use a shift-share instrument which is similar to the one we use to identify trade effects, they do not use data on migration patterns and do not consider deforestation as an outcome.

[4] A related strand of the literature uses spatial models with trade and/or migration to quantify the effects of climate change on productivity (Desmet and Rossi-Hansberg, 2015; Costinot et al., 2016; Burzynski et al., 2022; Conte, 2022) Another estimates the effect of structural transformation on pollution (Grossman and Krueger, 1995; Herrendorf et al., 2014).

2 Data

Our study combines data on deforestation, agriculture, migration, and urbanization. For deforestation and agriculture, we rely on satellite sources which have global coverage. For migration however we need to use census data, which limits our geographical coverage. We first describe the definition of our sample, before presenting each data source in turn.

2.1 Sample

Our sample consists of the countries in sub-Saharan Africa for which we were able to collect migration information in the 1990s and 2000s. Of the 53 countries in Africa, 27 have census data on Integrated Public Use Microdata Series-International (IPUMS-I), and 23 have migration history (either birth location or previous location). For each of these countries, we use the most disaggregated administrative unit available in IPUMS-I for which there is previous location data. We also obtained census data for two additional countries (Angola and Madagascar) from the countries' statistical institutes, as well as aggregate tables of two census of Gabon. For two countries (Ethiopia and Nigeria), we use data from the nationally representative LSMS surveys.

Figure 2 shows the countries that we include in our sample, superimposed on a map of tree cover and deforestation in Sub-Saharan Africa based on Hansen et al. (2013) data (see below). Our sample includes West-African countries, East-African countries, and Southern African countries. Taken together, these countries make up for 43.8% of total forest cover in Sub-Saharan Africa in 2000, and 59.7% of deforestation between 2001 and 2020. Figure 2 also displays the administrative units we use in the analysis. We use admin 2 level data for 17 countries and admin 1 level data for the remaining eleven countries. Appendix Table A.1 shows the number of units and the administrative level for each country. Together this yields a total of 1429 administrative units in our main sample.

2.2 Satellite data on land use

Deforestation We construct our measure of forest loss using the Hansen et al. (2013) global forest cover change dataset. The data divides the world into $30m \times 30m$ pixels and reports, for each pixel, the share of area covered by tree canopy in 2000, as well as the year in which the tree cover loss occurred, defined as a clearing of at least half of the tree cover within a pixel. Once a pixel is classified as having experienced loss, it is masked from further updates in the loss year, so subsequent dynamics such as regrowth are not recorded. To help disentangle tree cover in forested regions from tree plantations for commercial purposes, we consider that pixels with tree canopy cover of 30% or more are likely to be forested. Then, we measure the area of tree cover lost across forested pixels located within each

administrative unit each year and define the share of administrative unit o that was deforested as

$$Deforestation_o = \sum_{t=T_1+1}^{T_2-1} \frac{forest\,loss_{ot}}{total\,area_o}$$

Where T_1 and T_2 are the years of the first and second census, respectively. $area_o$ is the total area of the spatially harmonized administrative unit. We exclude from our analysis any administrative unit that had no forest cover at the start of the period.

Crop Land To document the extent to which forest cover change is driven by changes in agricultural activities, we estimated the variation of cropland area for each administrative unit over time. The estimates were produced using a global consistent map of $30m$ spatial resolution that identifies the pixels with cropland over 4-year periods: 2000-2003, 2004-2007, 2008-2011, 2012-2015, and 2016-2019. Pixels are considered as cropland if a growing crop was detected during any of year of the 4 year epoch (see Potapov et al., 2022). Use of 4-year long interval increased the number of clear-sky satellite observations and the accuracy of cropland detection. Incidentally, area not used to grow crops for human consumption, forage or bio-fuel for more than 4 years are not considered as cropland. When the years do not align exactly with the census year, we use the year that is closest to each census and define the change in the log share of area that is cropped ($\Delta \log S^A$) as:

$$\Delta \log S_o^A = \log\left(\frac{area\,cropped_{o,t_2}}{total\,area_o}\right) - \log\left(\frac{area\,cropped_{o,t_1}}{total\,area_o}\right)$$

Area cropped is measured in hectares. To avoid losing units with no area cropped in one round, we replace zeros with 1 acre (0.01 hectares). This leaves areas with no cropped land in either period unchanged.

Deforestation in Tropical Moist Forests While the Global Forest Change dataset produced by Hansen et al. (2013) has the advantage of being global, this inevitably leads to inaccuracies when compared to datasets focused on particular forest types. We therefore also use data on deforestation of Tropical Moist Forests produced by Vancutsem et al. (2021). While some forested areas in our data do not fall into this categorization, tropical moist forests represent the most important kind of forest in our countries of interest. The advantage of using this dataset is that this captures deforestation often missed by Hansen et al. (2013), in particular deforestation which occurs through a gradual degradation process (Vancutsem et al., 2021). Such a gradual process is likely to be particularly important when areas are deforested due to the gradual spread of small-scale agriculture.

2.3 Census data on Employment and Migration

Migration We use census data to measure immigration in each location d as the number of individuals aged 15-65 who migrated into d during the period. There are several difficulties with creating a consistent measure of migration flows which we address. First, individual migration history is asked differently across countries.[5] Our aim is to estimate migration in the ten years preceding, but not including, the census year. We exclude migrants in the previous year, when possible, to abstract from seasonal and temporary migration flows. Second, international migrants are only observed in their destination country. That is, we do not observe the location within the sending country from which they came. Hence, our observed count of emigrants from any location is weakly less than the true number of emigrants. We adjust for this in the receiving location by including international migrants in our count of immigrants, but not in any origin-destination pair, *i.e.* $\sum_o m_{od} \leq m_d$. Third, the geo-political boundaries of administrative units are changing over time. We use spatially harmonized geographical units from IPUMS-I as our unit of analysis. While this reduces the number of observations, it prevents us from mislabeling people who change location because a border was redrawn as migrants.

We construct the annual emigration rate from origin administrative unit o as

$$e_o = \frac{1}{N_o^{pa}} \frac{\sum_{t=\bar{T}_1}^{\bar{T}_2} \sum_{d \in C \setminus o} m_{odt}^{pa}}{\bar{T}_2 - \bar{T}_1}$$

where C denotes the set of destinations within the country, N_o^{pa} is the prime-aged population in o in the first census and m_{odt}^{pa} is the total number of prime-aged migrants from o into d in year t. The years $\bar{T}_1$ and $\bar{T}_2$ do not correspond to the census years but the years of the window that we have for migration.

Non-agricultural employment We use data on employment from the census, where available, to compute the growth in non-agricultural employment, computed as the difference between the number of non-agricultural workers between the first and the second census divided by the number of years between the two censuses.

As an alternative measure of non-agricultural growth, we use measures of annual night time light intensity produced using satellite imagery from the Defense Meteorological Satellite Program (DMSP). The images acquired have been processed for the general public by the National Oceanic and Atmospheric Administration (NOAA) with records dating from 1992 to 2013. The data contains annual average nightlight intensity with normalized values ranging from 0, indicating no light, to 63 digital number (DN). They capture nocturnal lighting from human activity for grid pixels of approximately 1 kilometer squared area at the equator. For each unit of analysis, i.e. administrative

[5] See Appendix Section C for details on how migration is measured in each country.

unit, we measure the share (ntla) of the area of the unit that was lit $(DN > 0)$ in each given year.

3 Shift-Share Strategy

The aim of our empirical strategy is to capture the effect of urbanization on deforestation through two channels, migration and trade. For the migration channel, we use a shift-share based on previous migration patterns, similar to the "Card" instrument in the migration literature (Card, 2001; Jaeger et al., 2018; Imbert et al., 2022). For the trade channel, we use another shift-share based on inverse distance, similar to the "market access" measures in trade (Donaldson and Hornbeck, 2016).

3.1 Migration Shift-Share

Let o and $d \in C$ be the generic indices for origin and destination administrative units in country C. We define the migration shift-share (Bartik) instrument as

$$B_o^m = \sum_{d \in C \setminus o} \frac{m_{od,t-1}}{total\, area_o} \frac{\Delta N_{dt}^{nonag}}{N_{d,t-1}^{nonag}(T_2 - T_1)} \tag{1}$$

Here we can see that the 'shifts' are the annualized growth rates in non-agricultural employment, i.e. ΔN_d^{nonag} is the change in non-agricultural employment at destination between the first and second census, which is divided by the baseline employment $N_{d,t-1}^{nonag}$ and the time between the two censuses $T_2 - T_1$. The 'shares' composed of the total number of migrants out of origin o into destination d, $m_{od,t-1}$, divided by the origin land area $total\, area_o$. The shares are normalized by the land area to estimate the change in population density, which is the relevant variable when thinking about change in land use shares. Because the shares do not sum to one, we include as a control the sum of the migration shares W_o^m:

$$W_o^m = \sum_{d \in C \setminus o} \frac{m_{od,t-1}}{total\, area_o}$$

Since the units do not have a direct interpretation, we standardize both the migration shift-share and the sum of the shares.

3.2 Trade Shift-Share

We define a similar set of shift-share instruments using inverse distance weights. These instruments capture the effects of trade. We define the distance shift-share using non-agricultural employment as shifts:

$$B_o^d = \sum_{d \in C \setminus o} \frac{1}{d_{od}} \frac{N_{d,t-1}}{total\, area_o} \frac{\Delta N_{dt}^{nonag}}{N_{d,t-1}^{nonag}(T_2 - T_1)} \tag{2}$$

Where (d_{od}) is the travel distance in km between the centroids of o and d. The shares are made up of the inverse of this distance multiplied by the destination population (to approximate 'market access') and, as with the migration share, divided by the land area at origin. In this way, we estimate the intensity of demand from trade on a given area - this reflects the fact that a similar demand for agricultural goods will have a much larger effect on land use on a small region than on a large one. To control for the fact that the shares do not sum to one, we additionally define the sum of the distance weights W_o^d as

$$W_o^d = \sum_{d \in C \setminus o} \frac{1}{d_{od}} \frac{N_{d,t-1}}{total\ area_o}$$

As a robustness check, we also compute a trade shift-share where the distance d_{od} is computed as the travel duration (in minutes) between o and d. Since their units do not have a direct interpretation, we standardize all the trade shift-share and the sums of the trade shares.

While most of our analysis focuses on within country migration and trade, we investigate the robustness of our shift-share analysis to considering cross-country. For the pairs of countries which provide information on the origin of international migrants, we impute migration probabilities between pairs of units across countries to construct an international migration shift-share.[6] We also compute travel distance between all the units in our sample to construct an international trade shift-share.

3.3 Estimating Equation

Our main estimating equation is the following:

$$y_o = \beta_m B_o^m + \beta_d B_o^d + \mathbf{X_o} + \varepsilon_o \tag{3}$$

where y_o is the outcome in administrative unit o (deforestation or change in land use). We include in the estimation only administrative units with positive forest cover at baseline (764 out of 847). B_o^m is the migration shift-share, B_o^d the trade shift-share, $\mathbf{X_o}$ is a vector of controls which includes the sum of migration shares W_o^m, and the sum of trade shares W_o^d. These controls are important for identification, since origins which are better connected to the rest of the country in terms of migration (high W_o^m) or in terms of trade (high W_o^d), may be on different trends. We also include country fixed effects in $\mathbf{X_o}$ to account for national economic and policy shocks and use robust standard errors.

In order to interpret β_m as the effect migration and β_d as the effect of trade, it is crucial to include both instruments simultaneously. Most papers in the economics literature use either the migration shift-share or the distance shift-share but not both. Since migration patterns follow gravity, and hence are affected by distance, sometime the distance shift-share is even used to capture the effect of

[6] Since the origin of international migrants is not known at the level of the administrative unit, we allocate migrants from country A to country B following the allocation of migrants from country B within country A. This imputation can be justified in two ways: migration is often circular, and follow ethnic proximity across international borders.

migration, when reliable information on migration shares is not available (Madhok et al., 2022). Hence our claim of identification is that conditional on the migration shift-share, the distance shift-share will capture the effect of trade.

Also crucial for identification is the assumption that the destination shifts in the shift shares are exogenous to deforestation trends at origin (Borusyak et al., 2022). We believe that this assumption is reasonable, since the shift we use is the change in non-agricultural employment in another part of the country. We check the robustness of our results to controlling for the local shift, i.e. the change in non-agricultural employment at origin. Unfortunately, the forest cover data does not include the decade prior to the decade when we observe migration, which prevents us to test for pre-trends.

4 Shift-Share Results

This section presents empirical results on the effects of structural change on deforestation through migration and trade channels.

We first follow specification (3) and regress various variables on the migration and the distance shift shares at the origin level. Table 1 presents the results. In Column (1), we present something like a 'first stage' test of our migration shift-share instrument. The dependent variable here is the total emigration out of the origin region divided by the land area - in other words, the reduction in population density due to emigration. Reassuringly, the coefficient on the migration shift-share is significantly positive, suggesting we can interpret this coefficient as proxying for emigration. On the other hand, the coefficient on the trade shift-share is negative, which makes sense: if there is increased demand for agricultural goods from a place, wages should rise and emigration go down. Unfortunately, we do not have any comparable information on internal trade to test an equivalent first stage for this instrument.

We then move on to outcome variables. In column 2, the outcome is then the share of area deforested from Hansen et al. (2013). In column 3 the outcome is the deforestation of tropical moist forests, and in column 4 it is the change in crop land. In each of these columns, there is a negative coefficient on the migration shift-share term and a positive one on the trade shift-share term. The coefficients are all strongly significant (at the 1% level), except the effect of the migration shift share on the change in crop land, which is significant at the 10% level. These results suggest that agriculture is a main driver of deforestation, so that when agricultural land use decreases forest cover increases and vice versa. They also provide supportive evidence for our hypothesis, that growing non-agricultural employment in one location increases forest cover in areas with which it has migration links, but decreases it in areas that are geographically closer, i.e. from which it is likely to import food.

One potential issue with our shift-share specification is the spatial correlation of non-agricultural employment growth, so that we may confound the effect of the destination-level shocks with that of

correlated origin-level ones. To alleviate this concern, we control for local shocks to non-agricultural employment: the results shown in Appendix Table A.4 are very similar to our main estimates. We provide two additional robustness checks. We show that our results are similar if we consider structural change not only in the same country but also in other countries which could be migration or export destinations (Appendix Table A.5). We also show that are results are qualitatively similar if we use travel duration rather than distance as a share for the trade shift share (Appendix Table A.6).

A potential concern for this analysis is that our distance shift-share may not be capturing any trade effect if the agricultural goods are being exported. To test this, we divide the sample between administrative units that are above/below the median share of harvested area that is devoted to cash crops in 2000 according to the GAEZ/FAO data. Since cash crops are primarily exported, we expect our results to be strongest in the below-median-cash-crop units. The results in Table 3 confirm this prediction: in units with a low share of cash crops (Panel A) migration links to a booming non-agricultural center reduce deforestation and agricultural land, while geographic proximity increases them. The effects are much more muted, and mostly insignificant in units with a high share of cash crops (Panel B).

Finally, we divide the sample by countries that are above/below the median level of food imports as a share of gross domestic product between 2001 and 2020 according to the World Development Indicators (World Bank). Under our hypothesis, if a country primarily imports its food, then migration will shift the location of consumption across space, as workers migrate to urban centers and engage in non-agricultural work, but it will not affect deforestation. As shown in Appendix Table **??**, our results are entirely driven by countries that import fewer agricultural goods.

Overall, our shift-share results point to ambiguous effects of structural change, measured by non-agricultural employment growth on deforestation and agriculture. We find evidence that structural change reduces deforestation and agricultural land use in places that are likely to send migrants to growing non-agricultural centers, but might increase them in places that are likely to export food to these centers. These effects are suggestive, but can not deliver any quantitative conclusions, because they do not capture equilibrium effects, i.e. effects in other locations that are connected with origin and destination through migration and trade. For this, the next sections will develop and quantify a spatial equilibrium model.

5 The Model

In this section, we develop a model of structural transformation and deforestation. To fully capture the complex interactions described above, we develop a quantitative spatial model of migration and trade. Our model is inspired by frameworks developed by Sotelo (2020), for agricultural trade, and by Bryan and Morten (2019), for migration.

5.1 Model Set-up

Environment The economy is composed of a measure of individuals of size $\bar{N}$ and a discrete set of locations with generic index $i, j \in \{1, \ldots, D\}$. Each location is endowed with a fixed land area, $\bar{L}_i$, which captures the heterogeneity in size of administrative units. Some land is forested ($\bar{L}_i^F$) and some land is cleared already ($\bar{L}_i^C$). Land is composed of a continuum of plots $\omega \in \bar{L}_i$. Each plot is endowed with agricultural productivity $Z_i^A(\omega) \sim F(\theta, \Lambda_i)$ and clearings cost $Z_i^0(\omega) \sim F(\theta, \Lambda_i^0)$, where $F(\cdot, \cdot)$ are Frechet distributions. These assumptions imply that the distribution of both costs and agricultural productivity may vary across locations.

Technology There are three types of technology: agriculture (A), manufacturing (M), and services (S). Agricultural production combines land $L(\omega) \in [0, 1]$ and labor (N) via a CES technology. Output on plot ω is:

$$Q_i^A(\omega) = (N_i^A(\omega))^{\alpha_A} \left(Z_i^A(\omega) L(\omega)\right)^{1-\alpha_A} \tag{4}$$

where $\alpha_A \in (0, 1)$. The non-agricultural sectors take labor as the sole input to production to produce the manufacturing good (Q_i^M) and the services (Q_i^S) via a decreasing returns to scale technology:

$$Q_i^k = Z_i^k (N_i^k)^{\alpha_k}, \quad k \in \{M, S\} \tag{5}$$

where $\alpha_k \in (0, 1)$ and Z_i^k is TFP in sector k, location i. We allow the different sectors to have deferent labor demand elasticities.

Trade Shipping agricultural goods between locations is subject to iceberg costs: for one unit of agricultural good from i to arrive in j, $\tau_{ij} \geq 1$ units must be shipped. We assume that costs are symmetric, $\tau_{ij} = \tau_{ji}$ and that $\tau_{ii} = 1$. The manufacturing good is traded is not subject to trade costs, while services are consumed locally and are not traded.

Preferences Individuals are endowed with a birth location i, and a vector of idiosyncratic preferences for location-sector pairs $\{\zeta_{jk}\}$. Individuals choose a location j and a sector $k \in \{A, M, S\}$ in which they supply one unit of labor inelastically to maximize their utility of consumption c, given sector-location amenities (b_{jk}) and migration costs (δ_{ij}):

$$U_i = \max_{jk} \{\zeta_{jk} b_{jk} c_{jk} \delta_{ij}^{-1}\} \tag{6}$$

Where the cost of migrating from i to j $\delta_{ij} \geq 1$. As with trade costs, we assume that migration costs are symmetric, $(\delta_{ij} = \delta_{ji})$ and there is no cost to not migrating $(\delta_{ii} = 1)$. The idiosyncratic taste shocks $\{\zeta_{jk}\}$ are distributed via a nested generalized extreme value distribution. In the upper nest,

individuals choose between locations with elasticity κ, and in the lower nest they choose between sectors with elasticity γ. We assume that $b_{jA} = b_j$, and $b_{jk\neq A} = \beta_j^k b_j$, where β_j^k is the non-monetary penalty from working outside of agriculture in location j.

Individuals have non-homothetic preferences over consumption of the agricultural and manufacturing goods and services. Following Comin et al. (2021), we define the utility of consumption for an individual in location i, working in sector k, (c_{ik}) via the implicit function:

$$(\Omega_A)^{\frac{1}{\sigma}} \left(\frac{c_{ik}^A}{c_{ik}^{\varepsilon_A}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_M)^{\frac{1}{\sigma}} \left(\frac{c_{ik}^M}{c_{ik}^{\varepsilon_M}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_S)^{\frac{1}{\sigma}} \left(\frac{c_{ik}^S}{c_{ik}^{\varepsilon_S}} \right)^{\frac{\sigma-1}{\sigma}} = 1 \tag{7}$$

Where c_{ik}^A, c_{ik}^M, and c_{ik}^S are the composite agricultural good, the manufacturing good, and services, respectively, for an individual located in i and working in sector k. The parameter σ governs the elasticity of substitution across broad consumption categories. The parameters ε_k, $k \in \{A, M, S\}$, control the degree of non-homotheticity in preferences: when $\varepsilon_k = 1$, preferences are homothetic in good k, while deviations from unity imply that expenditure shares on that good vary with real income. Finally, Ω_k is a good-specific scale parameter.

As in Comin et al. (2021) we use the manufacturing good as base good, and set $\Omega_M = \varepsilon_M = 1$:

$$(\Omega_A)^{\frac{1}{\sigma}} \left(\frac{c_{ik}^A}{c_{ik}^{\varepsilon_A}} \right)^{\frac{\sigma-1}{\sigma}} + \left(\frac{c_{ik}^M}{c_{ik}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_S)^{\frac{1}{\sigma}} \left(\frac{c_{ik}^S}{c_{ik}^{\varepsilon_S}} \right)^{\frac{\sigma-1}{\sigma}} = 1 \tag{8}$$

We assume $\sigma, \varepsilon_A, \varepsilon_S, \Omega_A$, and Ω_S are all strictly positive for each i. If $\varepsilon_A \neq 1$ or $\varepsilon_S \neq 1$, preferences are non-homothetic. When $\varepsilon_A < 1$ and $\varepsilon_S > 1$, the consumption shares of agricultural goods and services decline and increase, respectively, as income increases relative to the manufacturing good. These preferences also nest the CES case when $\varepsilon_A = \varepsilon_S = 1$, thus we assume further that at least one of these parameters differs from unity. This preference structure allows that the preference weights on each good may vary across locations and as structural transformation occurs. The composite agricultural good is a CES aggregator of agricultural goods from different origins (c_{ijk}^A).

$$c_{jk}^A = \left(\sum_i (c_{ijk}^A)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

Where $\eta > 0$ governs the elasticity of substitution between origins.

Total income for an individual living in j and working in sector k is:

$$v_{jk} = w_{jk} + \frac{R_j + \Pi_j^M + \Pi_j^S}{N_j}$$

Where w_{jk} is the wage in sector k, R_j are land rents, and Π_j^M and Π_j^S are profits in the manufacturing and services sectors, respectively, which are both redistributed lump-sum to local residents.

5.2 Equilibrium

Production A plot is cleared whenever it is profitable to do so:

$$p_i^A N_i^A(\omega)^{\alpha_A} \left(Z_i^A(\omega)\right)^{1-\alpha_A} - w_{iA} N_i^A(\omega) > A_i^0(\omega)$$

Under the distributional assumptions on land productivity and clearing costs, integrating over ω yields the share of forest plots converted to agriculture in i:

$$S_i = \frac{\left(\Lambda_i \psi_i\right)^\theta}{\left(p^M \Lambda_i^0\right)^\theta + \left(\Lambda_i \psi_i\right)^\theta} \tag{9}$$

Where $\psi_i = \left(\tilde{\alpha}_A p_i^A w_{iA}^{-\alpha_A}\right)^{\frac{1}{1-\alpha_A}}$ and $\tilde{\alpha}_A = \alpha_A^{\alpha_A}(1-\alpha_A)^{(1-\alpha_A)}$. Equation (9) states that the share of forested land used for agriculture is increasing in the average local land productivity, and decreasing in the average local land clearing cost. The parameter θ measures the heterogeneity in costs and productivity across plots within a location. A higher value of θ implies greater dispersion in costs and productivity across space. While the cost of clearing a plot of land is fixed, whether it is profitable to do so depends on the productivity of that plot as well as the price of agricultural goods from i and the local wage. Higher values of θ mean that agricultural production will be more sensitive to changes in prices. Total labor demand in agriculture in region i is:

$$N_i^A = \frac{\alpha_A}{1-\alpha_A} \left(\frac{\psi_i \Lambda_i}{w_{iA}}\right) \left(S_i^{\frac{\theta-1}{\theta}} L_i^F + L_i^C\right). \tag{10}$$

Equation (10) states that the measure of workers engaged in agriculture is higher in locations with a comparative advantage for agriculture $(\psi_i \Lambda_i)$ and a larger amount of cropland $S_i^{\frac{\theta-1}{\theta}} L_i^F + L_i^C$. Total agricultural production in i is thus:

$$Q_i^A = \frac{1}{1-\alpha_A} \Lambda_i \frac{\psi_i}{p_i^A} \left(S_i^{\frac{\theta-1}{\theta}} L_i^F + L_i^C\right). \tag{11}$$

Then land rents are given by:

$$R_i = (1-\alpha_A) p_i^A Q_i^A - p^M \Lambda_i^0 \left(1 - (1-S_i)^{\frac{\theta-1}{\theta}}\right) L_i^F. \tag{12}$$

Consumption In equilibrium, the quantity of agricultural goods that an individual living in location j and working in sector k buys from location i is given by:

$$c_{ijk}^A = \tau_{ij}^{-\eta}(p_i^A)^{-\eta}P_j^{A^{\eta-1}}e_{jk}^A \tag{13}$$

Where e_{jk}^A is expenditure on agricultural goods in j and $P_j^A \equiv \left[\sum_i \left(\tau_{ij}p_i^A\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}$ is the agricultural price index in j. Iceberg costs imply that the price of a crop produced in i and consumed in j is $p_{ij}^A = \tau_{ij}p_i^A$. Since the manufacturing good is freely traded, we have $p_{ij}^M = p^M$ for all i, j. We use manufacturing as numeraire and set $p^M = 1$

Let e_{ik} denote individual expenditures on all goods of individual living in location i and working in sector k. Likewise, μ_{ik}^A, μ_{ik}^S denote the expenditure shares for agriculture and services, respectively. The expenditure share on manufacturing is defined residually as $\mu_{ik}^M = 1 - \mu_{ik}^A - \mu_{ik}^S$. Appendix B.2 expresses equilibrium expenditure shares as a function of prices of each good and price indices for each location i and sector of employment k denoted with P_{ik}. It also expresses price indices as a function of prices and expenditure shares. A fixed point problem determines equilibrium expenditure shares and price indices for given prices.

Sector Choice We first consider the choice of sector k by an individual born in i conditional on migrating to a given location j. The indirect utility of working in sector $k \in \{A, M, S\}$ is $V_{ijk} = b_{jk}/\delta_{ij} \times v_{jk}/P_{jk} \times \zeta_{ijk}$. Using the properties of the extreme value distribution, the fraction of workers in j that work in agriculture is:

$$n_{jA} = \frac{(v_{jA}/P_{jA})^\gamma}{(v_{jA}/P_{jA})^\gamma + (\beta_j^M v_{jM}/P_{jM})^\gamma + (\beta_j^S v_{jS}/P_{jS})^\gamma} \tag{14}$$

Migration We next consider the choice of migrating to j for an individual born in i. The migration rate from i to j is:

$$m_{ij} = \frac{\left(V_{ij}\right)^\kappa}{\sum_{j'}\left(V_{ij'}\right)^\kappa} \tag{15}$$

where $V_{ij} = b_j/\delta_{ij}\left((v_{jA}/P_{jA})^\gamma + (\beta_j^M v_{jM}/P_{jM})^\gamma + (\beta_j^S v_{jS}/P_{jS})^\gamma\right)^{1/\gamma}$ is the expected indirect utility from migrating to j .

Equilibrium Definition An equilibrium in this economy is given by prices $\{w_i, p_i^A, p_i^S\}\,\forall\, i$ and quantities $(c_{ij}^A, c_j^M, c_j^S, S_i, Q_i^A, Q_i^M, Q_i^S, N_i^A, N_i^M, N_i^S, N_i, m_{ij})\,\forall\, i, j$ such that

- Labor markets clear in each location $i \in \mathcal{L}$ and sector $k \in \{A, M, S\}$: $N_j^k = n_{jk}N_j$
- In each location j markets clear for

- Agricultural goods: $Q_i^A = \sum_j c_{ij}^A \tau_{ij} N_j$
- Services: $Q_i^S = c_i^S N_i$
- Manufacturing by Walras' Law

- Labor supply is consistent with migration decisions $\sum_i m_{ij} \bar{N}_i = N_j$
- Non-agricultural profits Π_j^M , Π_j^S and land rents (R_i) are consistent with profit maximization in each sector (4) and (5)
- Consumption is consistent with the budget constraint, $e_j N_j = \sum_k w_{jk} N_j^k + R_j + \Pi_j^M + \Pi_j^S$.

6 Quantification

We now turn to answer our main question of interest, i.e. what the net effect of structural transformation on deforestation is. For this, we quantify the model using data from 2000 and 2010.[7] We use the estimated model to conduct a set of counterfactual exercises that isolate the contribution of structural transformation to deforestation.

6.1 Calibration

To quantify the model, we use data from the year 2000 or 2010. When the data is taken directly from censuses, we choose the census closest to the respective year.

Data on agricultural output, Q_i^A, comes from GAEZ (Fischer et al., 2021). To calibrate our productivity parameter in agriculture, Λ_i, we use the data on potential yields from the same source, under low-inputs and rain-fed irrigation.[8] As in Porteous (2024), to account for potential measurement errors or multiple harvests, we consider a region-level scaling factor where we have $\Lambda_i = a_i \tilde{Z}_i^{\{\text{GAEZ}\}}$. The parameter $\tilde{Z}_i^{\{\text{GAEZ}\}}$ represents the average value of potential yields in region i, and to obtain Λ_i, we calibrate a_i by inverting the equation of agricultural output, given by (11).

Population data, N_i, is calculated from the gridded population of the world (CIESIN, 2018). The amount of under forest cover, L_i^F, is obtained from the data of Hansen (Hansen et al., 2013). To estimate the value of output in manufacturing and services, we exploit data on national GDP in those two sectors and divide it across admin units using nightlights, making an adjustment based on the number of people working outside of agriculture for places where there are minimal nightlights (Elvidge et al., 2022).

[7] We choose these years since they are those where we have estimates of agricultural production and yields at the grid cell level.

[8] We followed Adamopoulos and Restuccia (2022) and computed the potential agricultural yield across 22 major crops, valuing the yield with international prices from the year 2000 and weighting each crop based on its share of local land use in 2000.

We show in appendix that bilateral migration costs can be recovered from migration rates as in Bryan and Morten (2019):

$$\delta_{ij} = \left(\frac{m_{ij} m_{ji}}{m_{ii} m_{jj}} \right)^{-\frac{1}{2\kappa}}. \tag{16}$$

We construct bilateral trade costs τ_{ij} using travel-time between each pair of locations, mapped into iceberg trade costs using the transport-cost elasticity estimated by Sotelo (2020).

Finally, we calibrate a set of parameters that are held fixed across all specifications. Table 4 summarizes these parameters and their sources. We take the shape parameter governing land-use allocation, θ, from Sotelo (2020), and the elasticity of substitution across agricultural varieties, η, from Costinot et al. (2016). Sectoral labor shares, α_i, are taken from Gollin (2002) and Dinh and Clarke (2012). We borrow the elasticity of migration, κ, from Bryan and Morten (2019). From Comin et al. (2021), we take our parameter σ as well as ε_A and ε_S.

6.2 Estimation

The remaining model parameters are estimated through a sequence of nested fixed-point loops. Each loop targets a distinct set of equilibrium conditions and empirical moments covering goods markets, labor markets, land use, and migration. The procedure from the inner to the outer loop is as follows:

Inner loop The inner loop solves the fixed point problem and derives equilibrium expenditure shares μ_i^k and price indices P_{ik} for given agricultural and services prices P_i^A and P_i^S.

Price loop The price loop estimates prices P_i^A and P_i^S, productivity terms (Λ_i, Z_i^M, and Z_i^S), land clearing costs Λ_i^0, and sector-specific amenities β_i^k. It takes as given the wage w_i, quantity produced in each sector Q_i^k, employment in each sector N_i^k, and land use (L_i^C, L_i^F and S_i). More precisely, we iterate on P_i^A and P_i^S to find a solution that clears each market. Within each iteration, we set Λ_i to match Q_i^A ; Z_i^M, Z_i^S to match Q_i^M and Q_i^S ; Λ_i^0 to match S_i by inverting equation (9) ; β_i^k to match N_i^k ($k \in \{M, S\}$). Every iteration of the price loop solves the inner loop.

Wage loop To solve for the set of wages w_i, we then iterate on w_i in order to ensure total labor demand equals labor supply in each location for each sector, still taking employment in each sector N_i^k as given. Within each iteration, we run the price loop to ensure goods markets clear.

Ω loop We then iterate on Ω to match the total share of expenditure on agricultural goods at the national level. Within each iteration, we run the wage loop to ensure goods and labor markets clear.

τ_F loop We then iterate on τ_F to match total agricultural exports to the rest of the world. Within each iteration, we run the Ω loop so that goods and labor markets clear.

Migration loop Finally, this loop estimates amenities b_i that rationalize observed migration flows from the nearest census given migration costs calibrated from (16)

Three key parameters estimated by this model are productivity in the agricultural sector, Λ_i, productivity in manufacturing, Z_i^M, and productivity in services, Z_i^S. Estimated growth in productivity in these sectors for each country is presented in Figure 3. We can see that, for most countries, productivity increased in both sectors. However, there exists an important variation across countries: for some countries, productivity declined in at least one sector (Gabon, Ivory Coast, Madagascar, Zambia and Zimbabwe). In the next section, we will then explore how these estimated productivity changes impacted structural change and deforestation.

6.3 Validation with external data sources

In order to test how well our model performs, we compare the estimated average consumption in each location to that measured in households surveys, where we have them. Such a test is fairly demanding, because household surveys measure consumption noisily and are not generally designed to be representative at the level at which we undertake our analysis. Moreover, our model excludes various factors that are going to be important in consumption, such as government transfers (including public sector employment), remittances, etc. Nonetheless, we believe the exercise is valuable in testing whether to a first approximation we are predicting which areas are relaively better off than others.

In Figure 4, we plot relative real adult-equivalent consumption as estimated using the model and household surveys. We take the LSMS survey closest to 2010 is used for eight countries: Ethiopia, Ghana, Malawi, Mali, Nigeria, South Africa, Tanzania, and Uganda, and this is compared to our estimates for 2010. While the fit is not perfect, with the model estimating greater dispersion in consumption than in the survey data, there is a clear and significant positive correlation between the two measures.

Additionally, we also compare our estimated international trade cost τ_F to an estimate we obtain from international trade data.[9] This is plotted in Figure 5. Here we see that our estimation technique generally does a good job at predicting which countries have relatively higher trade costs. While our estimates are on average lower than we obtain from the trade value data, this likely stems from some extreme values that come from this data, with trade costs of over 5 likely an exaggeration of the true cost.

[9]To construct the benchmark data for τ_F, we use CEPII's data on Trade Unit Values, which report bilateral unit values (USD/ton) by year–exporter–importer–product, separately as declared by exporters (FOB) and by importers (CIF). We focus on agricultural products (HS chapters 01–24) and compute, for each exporting country, the ratio of total CIF to total FOB unit values in each year. Our benchmark τ_F is then defined as the average of this ratio over the period 2000–2005.

6.4 Effects of productivity growth on deforestation

Within our model, we consider structural transformation as mainly stemming from increases in productivity. Increases in productivity will increase consumption, which, due to non-homothetic preferences, will push relative demand towards non-agricultural goods and services, encouraging movement out of the agricultural sector. Of course, the overall effect will also depend heavily on the source of the productivity changes - if agricultural productivity increases, this may draw people into the sector, mitigating structural change. Alternatively, increases in agricultural productivity may free up labor and hence generate further structural change. In what follows, we therefore aim to isolate the impacts of changes in productivity changes in both the agricultural and the non-agricultural sectors, as well as all sectors together.

Rather than choose arbitrary values of productivity growth for our counterfactuals, we exploit the fact that we have estimated our model on two years worth of data, namely 2000 and 2010. In the following counterfactuals, we therefore compare outcomes in 2010 to model estimates of what would have happened if productivity hadn't changed over the period - i.e. if productivity had stayed at the 2000 levels. Note that this means that the counterfactual productivity growths considered are substantially different across country and across sector, as shown in Figure 3. The advantages of this strategy are that we consider meaningful counterfactuals and can answer the question how much of observed deforestation over the period can be explained by these productivity changes. In the figures that follow, to avoid clutter we remove countries with relatively small areas of forest cover, but we also show the results for the 28 countries as a whole.

In Figure 6 we first consider the role of productivity changes in explaining the structural change that has occurred over the period. In particular, we consider changes in the share of workers who work outside of agriculture, both in our counterfactuals and the actual change over the period. Consistent with our approach, we find that overall the productivity changes explain most of the structural transformation over the period. Both increases in agricultural and non-agricutlural productivity are important factors in driving this change, though the relative share varies importantly by country. In general, we see that increases in non-agricultural productivity growth lead to an increase in structural change (Nigeria is an exception), but the impact of agricultural productivity growth on structural transformation is more varied. For instance, while Ghana and Ethiopia saw substantial increases in agricultural productivity, this led to an important structural transformation in Ethiopia but almost no change in sectoral employment in Ghana.

With these varied effects of productivity changes on structural transformation in mind, we now turn to looking at the impact on deforestation. In Figure 7, we can note that the overall effect of the productivity changes on the 28 countries in our sample is positive - i.e. deforestation would have been about 0.3 percentage points lower had there been no structural change, which accounts for around 10 % of the deforestation that took place over the period. This overall small positive effect, however,

contains important heterogeneities. Countries that saw increases in agricultural productivity saw a very mixed picture in terms of impacts on deforestation, with an increase in deforestation of almost 2 percentage points in South Africa to a similarly substantial reduction in deforestation in Nigeria.[10] The impacts of non-agricultural productivity changes are also varied. Almost all countries increased their non-agricultural productivity over the period, and this most often led to an increase in deforestation, but this is not the case in South Africa, Kenya, or Ghana.

What explains this heterogeneity? Looking first at the impacts of agricultural productivity, an important factor is the extent to which the agricultural sector is oriented towards international trade. We see large increases in deforestation in South Africa, an important agricultural exporter, while almost no impact on deforestation in Nigeria, which is relatively closed, despite Nigeria experiencing a much larger increase in agricultural productivity. Figure 10a confirms this relationship for those countries that saw increases in agricultural productivity during the period. This is consistent with non-homothetic demand for agriculture leading to a limited domestic demand for agricultural production. In open economies, an increase in agricultural productivity pulls workers towards agriculture to exploit the comparative advantage, slowing structural change and increasing deforestation. In more closed economies, on the other hand, agricultural production doesn't increase so much, and the increase in agricultural productivity is partly used to free up workers to move to other sectors and land to be used as forest. When it comes to non-agricultural productivity increases, international trade plays a similar role, as shown in Figure 10b. Countries that export agriculture take advantage of gains in non-ag productivity to move resources out of agriculture and reduce deforestation, while those that are more closed have less ability to adapt, since importing food is very costly. Overall, this variation as a function of agricultural export orientation is consistent with the heterogeneity in effects in the reduced form by the share of cash crops.

Openness to international trade is clearly not the whole story, however, and doesn't explain, for instance, why Ghana saw a decrease in deforestation as a result of non-agricultural productivity increases, while Mozambique saw an increase. To understand this difference, we need to dig into the spatial distribution of productivity changes, and how this relates to forests. In Figure 9 we map the distributions of manufacturing and mining productivity gains and forest cover in 2000 for these two countries. In Ghana, we observe that manufacturing and mining productivity gains were highest in the north of the country, wheras the main forested area is in the South. In Mozambique, while manufacturing and mining productivity gains were also concentrated in the north, this is also an area of the country with a substantial amount of forest cover. This explains an important difference in impacts because the productivity growth leads to an increase in demand for local agricultural goods (since internal trade is costly).

To formally test this idea, we first construct measures of how much each region is 'exposed' to

[10]Note that three of the other countries that saw a reduction in deforestation as a result of agricultural productivity changes, notably Gabon, Ivory Coast, and Zambia, were those where agricultural productivity fell.

forests in two different ways - firstly through migration, and secondly through trade. To construct the migration measure for a region j, we consider at baseline the share of immigrants which come from region i - i.e. $M_{ij}/\sum_i M_{ij}$, where M_{ij} is the number of people who migrated from i to j - and multiply this by the share of forested land in region i, i.e. $L_i^F/(L_i^F + L_i^C)$. We then do the same for trade, where weights are instead the consumption shares - i.e. $c_{ij}/\sum_i c_{ij}$. We then construct the difference, which we label the 'relative forest exposure', FE_j, and is hence defined as follows:

$$FE_j = \sum_j \left(\frac{M_{ij}}{\sum_i M_{ij}} - \frac{c_{ij}}{\sum_i c_{ij}} \right) \frac{L_i^F}{(L_i^F + L_i^C)} \tag{17}$$

This variable then theoretically takes a value between -1 and 1, which higher values meaning that a region's exposure to forest mainly comes through migration, and a lower number meaning that it's exposure mainly comes through trade.

Armed with this geographical exposure measure, we then run two new counterfactuals where in each we consider what would happen if a single location in each country increased it's non-agricultural productivty by 50 %, and all other parameters stayed the same. In the first counterfactual, we pick the unit with the highest level of relative forest exposure, FE_j in each country, while in the second counterfactual we pick the region with the lowest. Since in each counterfactual we are shocking one unit in each country, variation in the impacts of the two shocks cannot stem from changes in openness to international trade. While both counterfactuals lead to a notable increase in non-agricultural employment, the impacts on deforestation are very different. When we shock those places that are most exposed to forests through migration, we see an overall decrease in deforestation, but, if we shock those places that are mainly exposed through trade, we see an increase in deforestation. This therefore demonstrates that the geographic location of non-agricultural productivity shocks is a crucial determinant of their impacts on deforestation.

Another way of considering the role of migration in explaining the impact of structural change on deforestation is to run a regression similar to our reduced-form estimation in Section 3. From Table 5, we can see that, consistent with our reduced form analysis, migration reduces deforestation while demand for agricultural products increases it. To explore the role of migration in general equilibrium, we undertake a counterfactual with productivity changes as before, except that we don't allow for migration to adapt to these productivity changes. In other words, we keep non-agricultural productivity at it's 2000 level as before, but we assume that people still move to wherever we find them in 2010. In this way, we can separate out the impacts of productivity changes that come through migration, and those that come through other channels.

The results of this analysis for changes in non-agricultural productivity are displayed in Figure 10. We can see that, in most countries, there is more structural change and less deforestation when we allow for productivity to have impacts on migration. This is consistent with our theoretical hypothesis

and reduced form results, with people moving away from forested areas towards the most productive cities. It should be noted, however, that the general equilibrium impacts of migration on deforestation are most substantial in those countries open to international trade such as South Africa and Ivory Coast. In relatively closed countries, migration has a limited effect since the food to feed these people must still be grown within the country, and hence forests will still be cut down even if people have moved relatively far away.

7 Conclusion

The overall impact of structural change on deforestation in sub-Saharan Africa is theoretically ambiguous. More employment in sectors with higher productivity than agriculture will increase domestic demand for agricultural products, thereby increasing the return to clearing land. At the same time, migration from rural areas to urban ones will decrease the return to land clearance. In this paper, we have found that both mechanisms play a significant role, with the dominant mechanism varying across countries.

Our results lead to important policy conclusions. In general, increases in agricultural productivity are likely to put forests least at risk when they occur in relatively closed economies, while increases in non-agricultural productivity are likely to be most beneficial for forests when they take place in parts of the country that are relatively far from forested areas. Hence, when pursuing the dual aims of protecting forests and increasing structural change, policy makers should consider the spatial distribution of where prodcutviity gains should be encouraged.

A number of important caveats should be kept in mind when interpreting our results. Crucially, we are only measuring one of the environmental impacts of structural transformation, and of course issues are likely to arise due to pollution and pressure on other natural resources. Our continent-wide approach also prevents us from looking at heterogeneity in the value of forests or the costs of tree-cover loss. As better data becomes available, an important avenue of future research will be to improve measures of trade-costs and environmental degradation.

References

Abman, Ryan and Clark Lundberg, "Does free trade increase deforestation? The effects of regional trade agreements," *Journal of the Association of Environmental and Resource Economists*, 2020, *7* (1), 35–72.

Adamopoulos, Tasso and Diego Restuccia, "Geography and Agricultural Productivity: Cross-Country Evidence from Micro Plot-Level Data," *The Review of Economic Studies*, July 2022, *89* (4), 1629–1653.

Andersen, Lykke E, *The dynamics of deforestation and economic growth in the Brazilian Amazon*, Cambridge University Press, 2002.

Angelsen, Arild, "Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights," *Journal of Development Economics*, 1999, *58* (1), 185–218.

_ , Mariel Aguilar-Støen, John Herbert Ainembabazi, Edwin Castellanos, and Matthew Taylor, "Migration, Remittances, and Forest Cover Change in Rural Guatemala and Chiapas, Mexico," *Land*, 2020, *9* (3).

Araujo, Rafael, Juliano Assunção, and Arthur Amorim Bragança, "The Effects of Transportation Infrastructure on Deforestation in the Amazon: A General Equilibrium Approach," resreport 10415, World Bank 2023.

Asher, Sam, Teevrat Garg, and Paul Novosad, "The ecological impact of transportation infrastructure," *Economic Journal*, 2020, *130* (629), 1173–1199.

Baland, Jean-Marie, Francois Libois, and Dilip Mookherjee, "Firewood Collections and Economic Growth in Rural Nepal 1995-2010: Evidence from a Household Panel," CEPR Discussion Papers 9394 March 2013.

Balboni, Clare, Aaron Berman, Robin Burgess, and Benjamin A. Olken, "The Economics of Tropical Deforestation," *Annual Review of Economics*, July 2023, *15* (3), 723–54.

_ , Robin Burgess, Anton Heil, Jonathan Old, and Benjamin A. Olken, "Cycles of Fire? Politics and Forest Burning in Indonesia," *AEA Papers and Proceedings*, May 2021, *111*, 415–19.

Barbier, E. B. and J. C. Burgess, "The Economics of Tropical Deforestation," *Journal of Economic Surveys*, July 2001, *15* (3), 413–433.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

Brewer, Julia, Ashley Larsen, and Frederik Noack, "The land use consequences of rural to urban migration," *American Journal of Agricultural Economics*, 2022, *n/a* (n/a), 1–29.

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber, "The Political Economy of Deforestation in the Tropics," *The Quarterly Journal of Economics*, 2012, *127* (4), 1707–1754.

Burzynski, Michal, Christoph Deuster, Frederic Docquier, and Jaime de Melo, "Climate Change, Inequality, and Human Migration," *Journal of the European Economic Association*, 2022, *20* (3), 1145–1197.

Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, January 2001, *19* (1), 22–64.

Central Statistics Agency (CSA) of Ethiopia and World Bank LSMS-ISA, "Socioeconomic Survey 2013-2014, Ethiopia [dataset]," World Bank Microdata Library 2013-2014.

CIESIN, "Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11 (Version 4.11)," 2018. [Data set].

Comin, Diego, Danial Lashkari, and Martí Mestieri, "Structural Change With Long-Run Income and Price Effects," *Econometrica*, 2021, *89* (1), 311–374.

Conte, Bruno, "Climate Change and Migration: The Case of Africa," CESifo Working Paper Series 9948 2022.

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.

Curtis, Philip G., Christy M. Slay, Nancy L. Harris, Alexandra Tyukavina, and Matthew C. Hansen, "Classifying drivers of global forest loss," *Science*, 2018, *361* (6407), 1108–1111.

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

Dinh, Hinh T and George RG Clarke, *Performance of manufacturing firms in Africa: An empirical analysis*, World Bank Publications, 2012.

Direction Générale de la Statistique (DGS), Gabon, "Recensement général de la population et de l'habitat, du 1er au 31 juillet 1993 : Principaux résultats [PDF]," Libreville, Gabon: Institut National de la Statistique (INSTAT Gabon) 1993.

_ , "Résultats globaux du Recensement Général de la Population et des Logements de 2013 du Gabon (RGPL-2013) [PDF]," Libreville, Gabon: Direction Générale de la Statistique (DGS) 2015.

Dominguez-Iino, Tomas, "Efficiency and redistribution in environmental policy: An equilibrium analysis of agricultural supply chains," *Cited on*, 2021, p. 6.

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

Elvidge, Christopher D., Kimberly Baugh, Tilottama Ghosh, Mikhail Zhizhin, Feng Chi Hsu, Tamara Sparks, Morgan Bazilian, Paul C. Sutton, Kenneth Houngbedji, and Ran Goldblatt, "Fifty years of nightly global low-light imaging satellite observations," *Frontiers in Remote Sensing*, 2022, *3*.

FAO, *Global Forest Resources Assessment 2020: Main report*, Rome: FAO, 2020.

Farrokhi, Farid, Elliot Kang, Heitor S. Pellegrina, and Sebastian Sotelo, "Deforestation: A Global and Dynamic Perspective," 2023. Manuscript.

Ferraro, Paul J. and Rhita Simorangkir, "Conditional cash transfers to alleviate poverty also reduced deforestation in Indonesia," *Science Advances*, 2020, *6* (24), eaaz1298.

Ferreira, Susana, "Deforestation, Property Rights, and International Trade," *Land Economics*, 2004, *80* (2), 174–193.

Fetzer, Thiemo and Samuel Marden, "Take What You Can: Property Rights, Contestability and Conflict," *Economic Journal*, 2017, *127* (601), 757–783.

Fischer, Günther, Freddy O. Nachtergaele, Harrij T. van Velthuizen, Federico Chiozza, Giuliano Franceschini, Michel Henry, David Muchoney, and Sylvia Tramberend, "Global Agro-Ecological Zones v4 – Model Documentation," techreport, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy 2021.

Foster, Andrew D. and Mark R. Rosenzweig, "Economic Growth and the Rise of Forests*," *The Quarterly Journal of Economics*, 05 2003, *118* (2), 601–637.

Geist, Helmut, *The causes and progression of desertification*, Routledge, 2017.

Gollin, Douglas, "Getting income shares right," *Journal of political Economy*, 2002, *110* (2), 458–474.

_ and Julien Wolfersberger, "Agricultural Trade and Deforestation: the Role of New Roads," 2023.

_ and _ , "Agricultural Trade and Deforestation: The Role of New Roads," CEPR Discussion Paper 19710, Centre for Economic Policy Research (CEPR), Paris & London 2024.

Grossman, Gene M and Alan B Krueger, "Economic Growth and the Environment," *The Quarterly Journal of Economics*, 1995, *110* (2), 353–377.

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

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi, "Chapter 6 - Growth and Structural Transformation," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 2 of *Handbook of Economic Growth*, Elsevier, 2014, pp. 855–941.

Hoang, Nguyen Tien and Keiichiro Kanemoto, "Mapping the deforestation footprint of nations reveals growing threat to tropical forests," *Nature Ecology and Evolution*, 2021, *5* (6), 845–853.

ILO, "ILO modelled estimates database (accessed via World Development Indicators)," Technical Report 2021.

Imbert, Clement, Marlon Seror, Yifan Zhang, and Yanos Zylberberg, "Migrants and Firms: Evidence from China," *American Economic Review*, June 2022, *112* (6), 1885–1914.

Institut National de la Statistique (INSTAT), Madagascar, "Recensement Général de la Population et de l'Habitation (RGPH-3) de 2018 [dataset]," 2018.

Jack, B. Kelsey and Seema Jayachandran, "Self-selection into payments for ecosystem services programs," *The National Academy of Sciences*, March 2019, *116* (12), 5326–5333.

Jaeger, David A., Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," resreport, NBER 2018.

Kaczan, David J., "Can roads contribute to forest transitions?," *World Development*, 2020, *129*, 104898.

Leblois, Antoine, Olivier Damette, and Julien Wolfersberger, "What has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data," *World Development*, 2017, *92*, 82–102.

Madhok, Raahil, Frederik Noack, Ahmed Mushfiq Mobarak, and Olivier Deschenes, "Internal Migration and the Re-organization of Agriculture," Mimeo 2022.

National Bureau of Statistics, Federal Republic of Nigeria, "General Household Survey, Panel 2010-2011, Wave 1, Nigeria [dataset]," World Bank Microdata Library 2010-2011.

Pellegrina, Heitor S., "Trade, productivity, and the spatial organization of agriculture: Evidence from Brazil," *Journal of Development Economics*, 2022, *156* (July 2020), 102816.

Pendrill, Florence, Toby A Gardner, Patrick Meyfroidt, U Martin Persson, Justin Adams, Tasso Azevedo, Mairon G Bastos Lima, Matthias Baumann, Philip G Curtis, Veronique De Sy et al., "Disentangling the numbers behind agriculture-driven tropical deforestation," *Science*, 2022, *377* (6611), eabm9267.

Pfaff, Alexander, Juan Robalino, Robert Walker, Steven Aldrich, Marcellus Caldas, Eustaquio Reis, Stephen Perz, Claudio Bohrer, Eugenio Arima, William Laurance et al., "Road investments, spatial spillovers, and deforestation in the Brazilian Amazon," *Journal of Regional Science*, 2007, *47* (1), 109–123.

Pfaff, Alexander S.P., "What drives deforestation in the Brazilian Amazon? Evidence from satellite and socioeconomic data," *Journal of Environmental Economics and Management*, 1999, *37* (1), 26–43.

Porteous, Obie, "Agricultural trade and adaptation to climate change in Sub-Saharan Africa," *Available at SSRN 4777962*, 2024.

Potapov, Peter, Svetlana Turubanova, Matthew C Hansen, Alexandra Tyukavina, Viviana Zalles, Ahmad Khan, Xiao-Peng Song, Amy Pickens, Quan Shen, and Jocelyn Cortez, "Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century," *Nature Food*, 2022, *3* (1), 19–28.

Redding, Stephen J. and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9* (1), 21–58.

Restrepo, Veronica Salazar and Gabriel Leite-Mariante, "Does Conservation Work in General Equilibrium," 2023. Manuscript.

Seto, Karen C., Burak Güneralp, and Lucy R. Hutyra, "Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools," *Proceedings of the National Academy of Sciences*, 2012, *109* (40), 16083–16088.

Simkin, Rohan D., Karen C. Seto, Robert I. McDonald, and Walter Jetz, "Biodiversity impacts and conservation implications of urban land expansion projected to 2050," *Proceedings of the National Academy of Sciences*, 2022, *119* (12), e2117297119.

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

Souza-Rodrigues, Eduardo, "Deforestation in the Amazon: A unified framework for estimation and policy analysis," *The Review of Economic Studies*, 2019, *86* (6), 2713–2744.

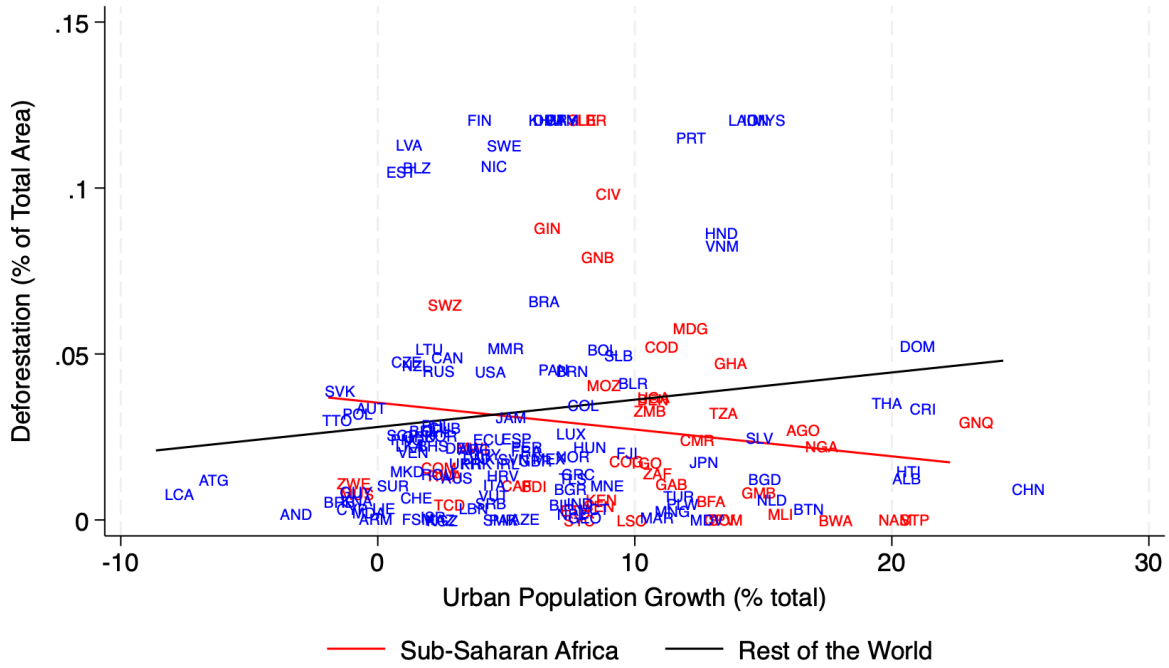
Tombe, Trevor, "The Missing Food Problem: Trade, Agriculture, and International Productivity Differences," *American Economic Journal: Macroeconomics*, July 2015, *7* (3), 226–58.

Vancutsem, C., F. Achard, J.-F. Pekel, G. Vieilledent, S. Carboni, D. Simonetti, J. Gallego, L. E. O. C. Aragão, and R. Nasi, "Long-term (1990–2019) monitoring of forest cover changes in the humid tropics," *Science Advances*, 2021, *7* (10), eabe1603.

Wren-Lewis, Liam, Luis Becerra-Valbuena, and Kenneth Houngbedji, "Formalizing land rights can reduce forest loss: Experimental evidence from Benin," *Science Advances*, 2020, *6* (26), 1–9.

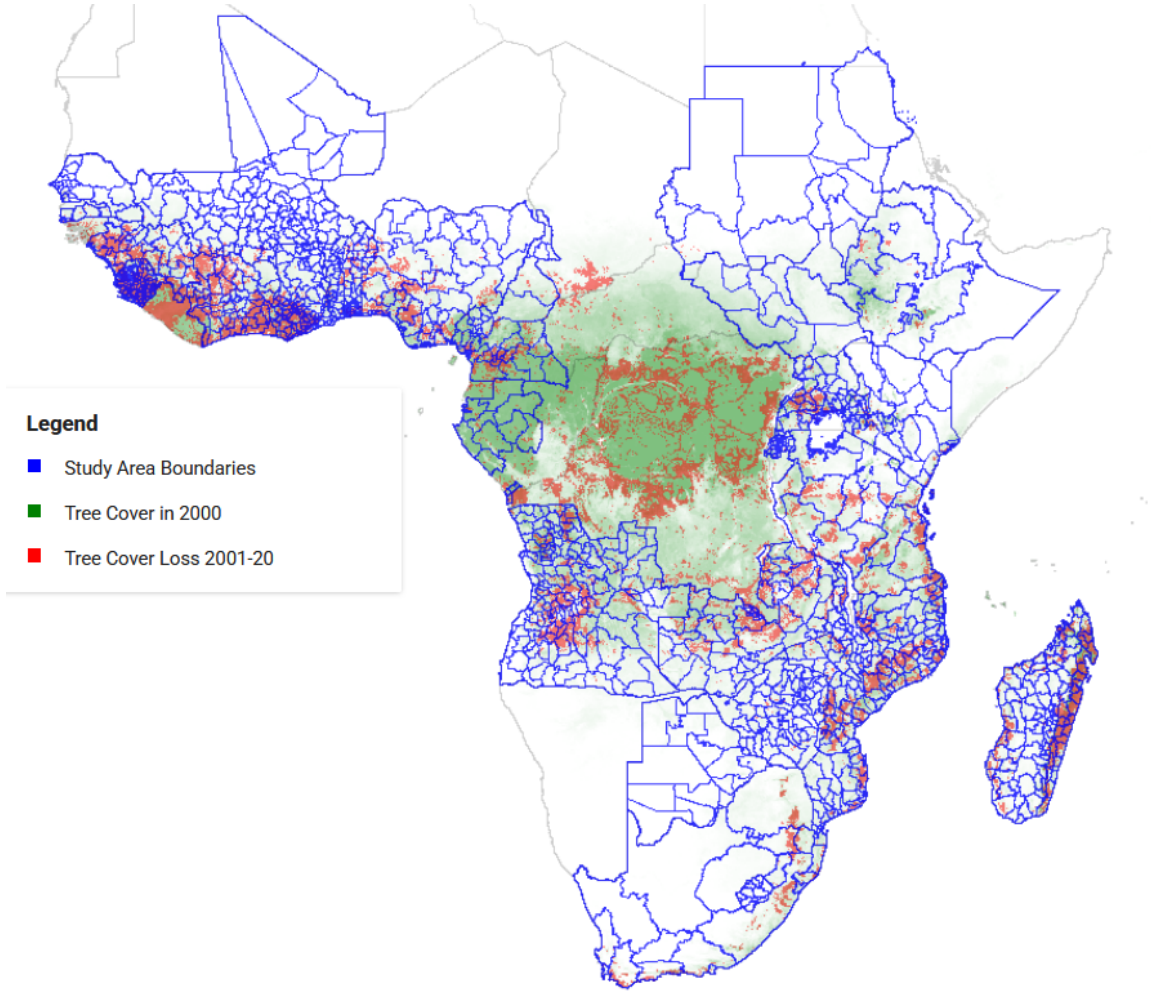
Figures and Tables

Figure 1: Deforestation and Urban Growth since 2000



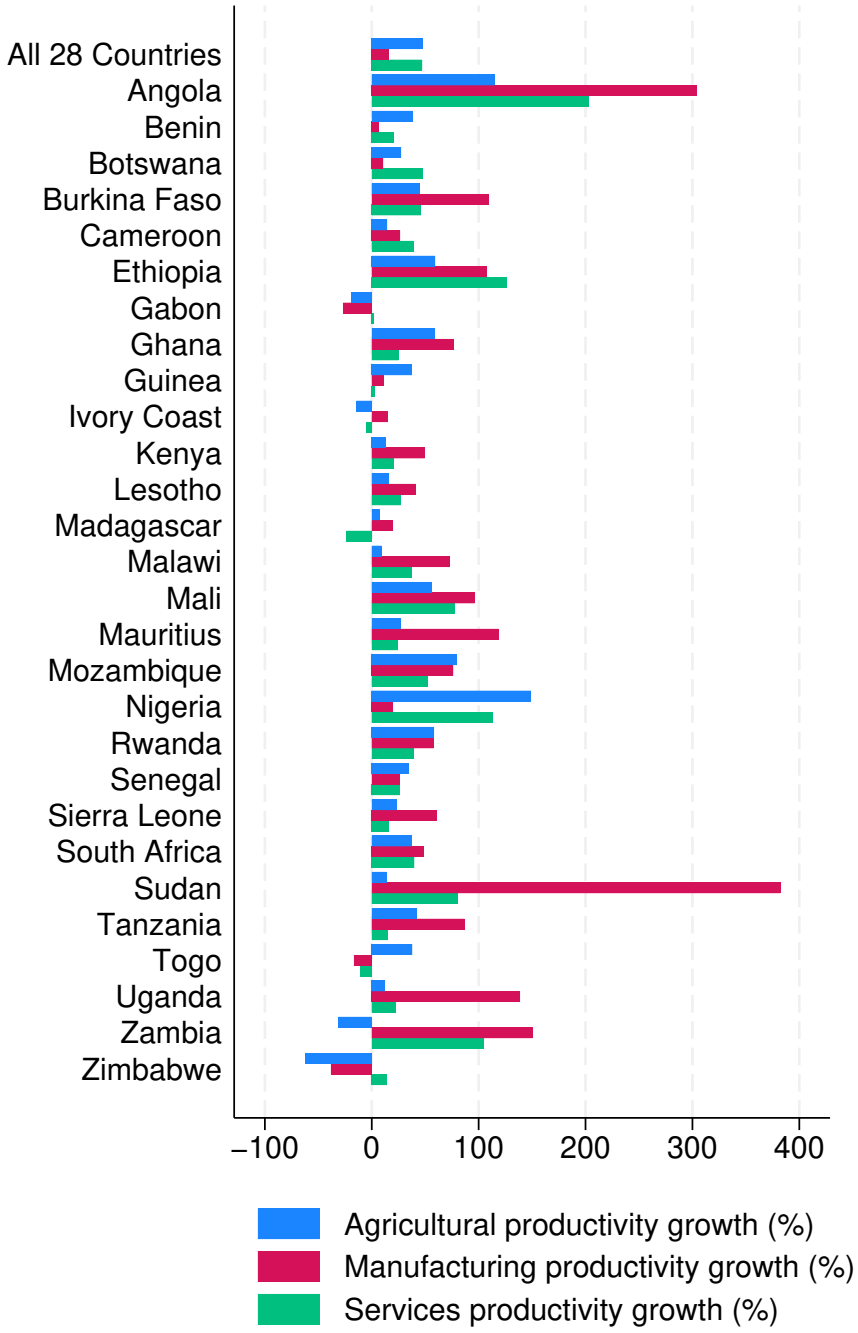
Notes: Along the x-axis is the change in the percentage of the population living in urban areas between 2001 and 2020 as defined by the World Development Indicators. The y-axis displays the percent of the total land area that was deforested during the same period using deforestation data form Hansen et al. (2013). The figure excludes countries with less than 2% forest cover in 2000 and winsorizes the deforestation share at the 5 and 95 percent levels.

Figure 2: Tree Cover and Deforestation in Sub-Saharan Africa and in our Sample



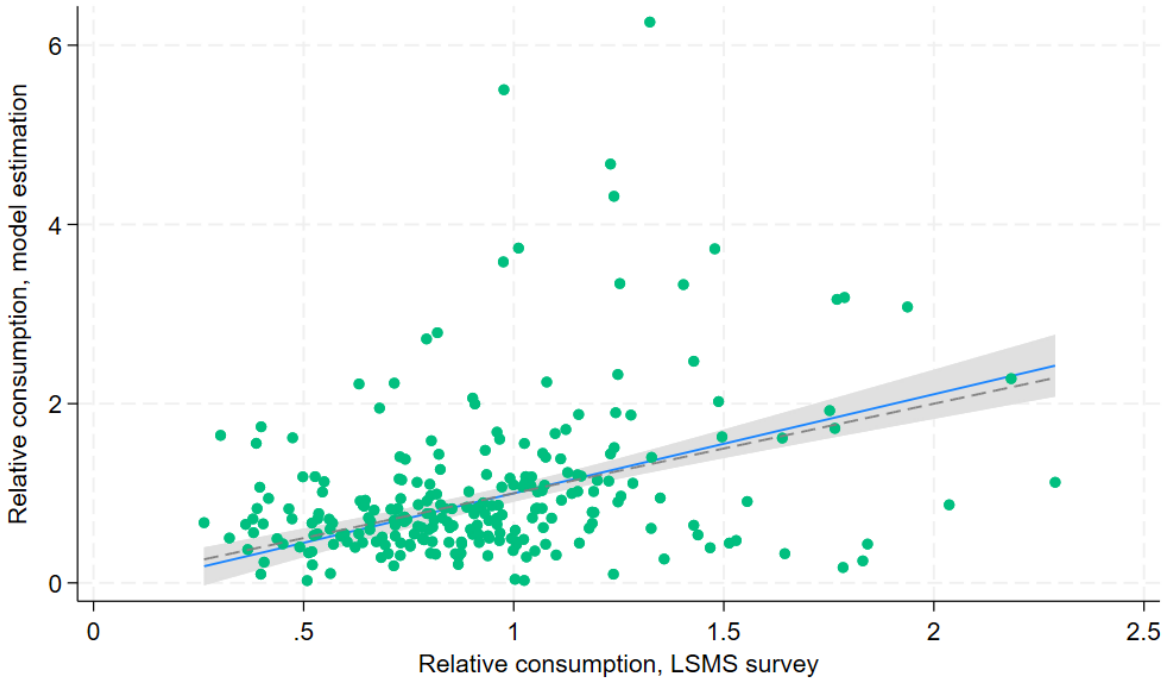
Notes: The map represents in green tree cover in 2000, in red tree cover loss between 2001 and 2020, and in blue the boundaries of the countries that meet the data requirement for the study.

Figure 3: Estimated Productivity Changes, 2000 to 2010



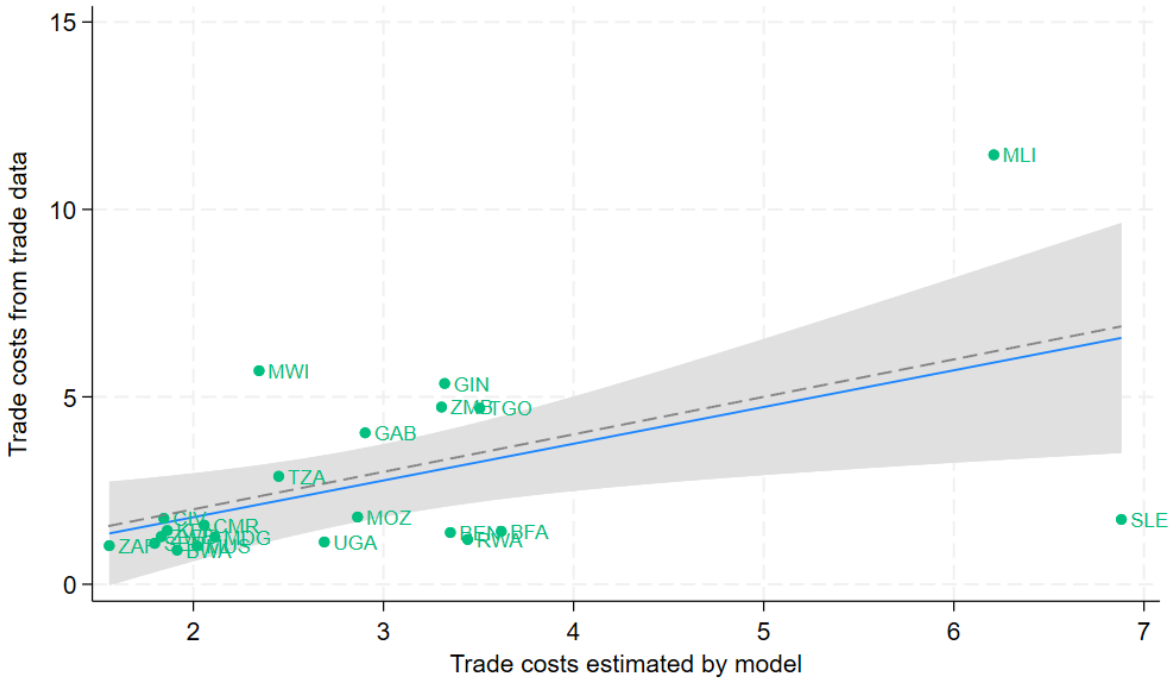
Notes: This figure plots the changes in productivity in agriculture (Λ_i), manufacturing (Z_i^M) and services (Z_i^S) in each country, with values below 1 representing a decrease and values above 1 an increase. The changes are calculated by estimating the structural model first with data from the year 2000 values and then with data from 2010. Averages are taken across districts weighted by baseline production in the respective sector.

Figure 4: Comparing estimated to surveyed real consumption



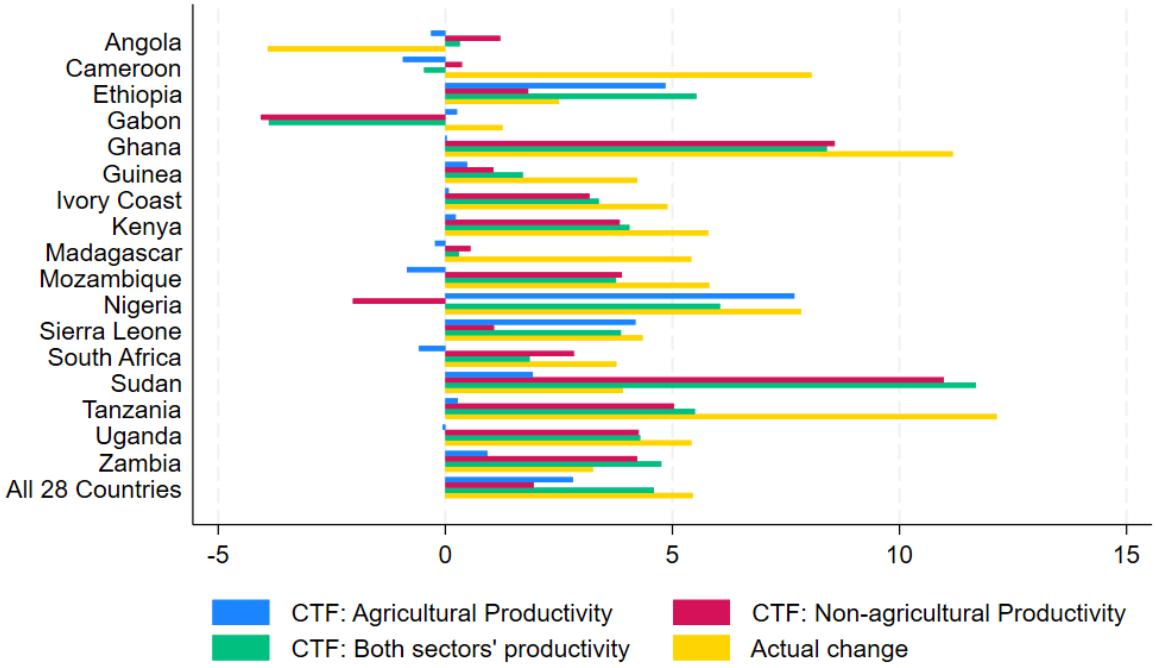
Notes: This figure plots real consumption as estimated in the model against real adult-equivalent consumption as measured using household surveys. We plot one point for each region, along with a line of best fit weighted by population and include a 95 % confidence interval. The LSMS survey closest to 2010 is used for eight countries: Ethiopia, Ghana, Malawi, Mali, Nigeria, South Africa, Tanzania, and Uganda, and this is compared to our estimates for 2010. We take the mean consumption in each geographical unit and divide by the mean across units.

Figure 5: Comparing estimated international trade costs from trade data to that estimated by the model.



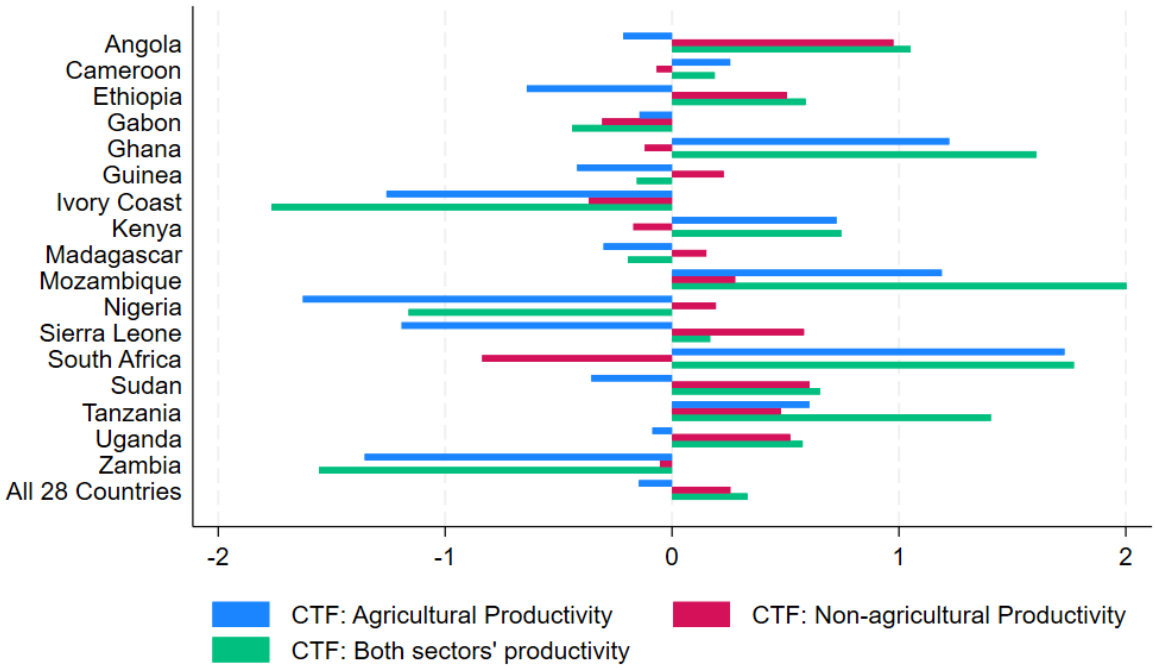
Notes: This figure plots international trade costs from trade data to that estimated by the model. We plot one point for each country, along with a line of best fit and include a 95 % confidence interval.

Figure 6: Counterfactual rates of structural transformation



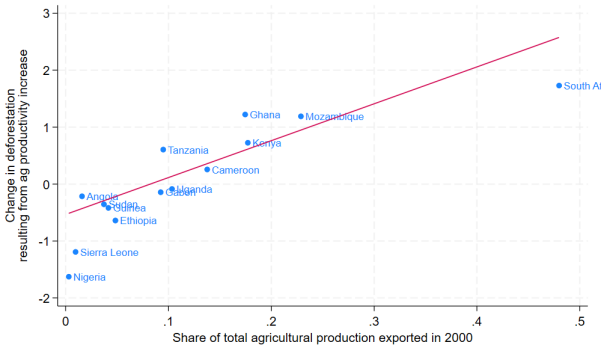
Notes:

Figure 7: Counterfactual rates of deforestation

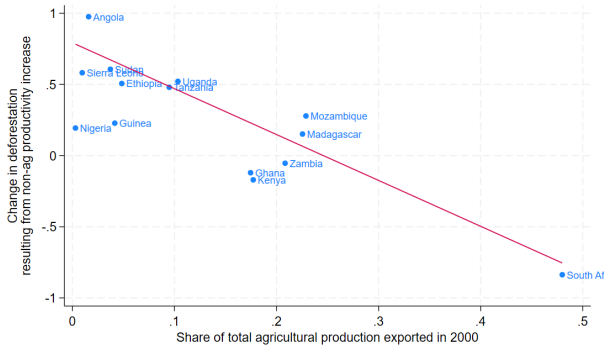


Notes:

Figure 8: Impacts of productivity changes on deforestation by trade openness



(a) Impacts of ag productivity increases by ag export share

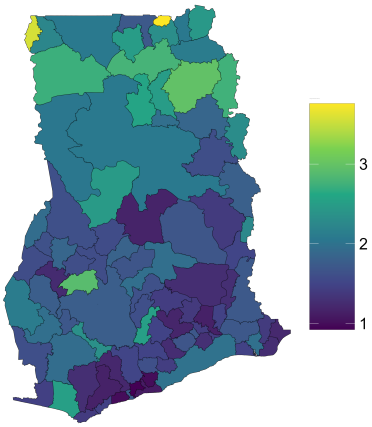


(b) Impacts of non-ag productivity increases by ag export share

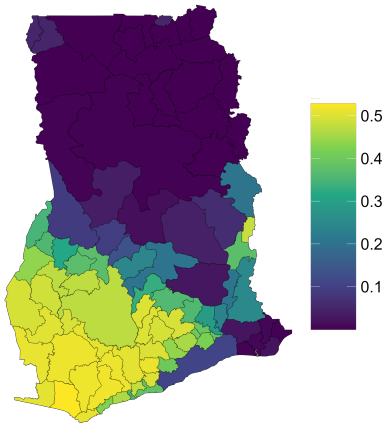
Notes:

Figure 9: Manufacturing growth and forest cover dynamics in Ghana and Mozambique

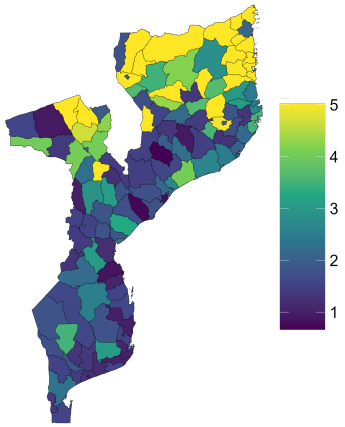
(a) Ghana: Manufacturing growth



(b) Ghana: Forest cover



(c) Mozambique: Manufacturing growth



(d) Mozambique: Forest cover

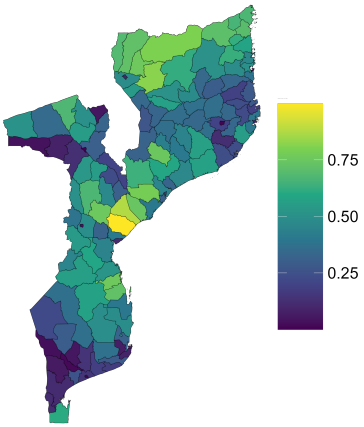
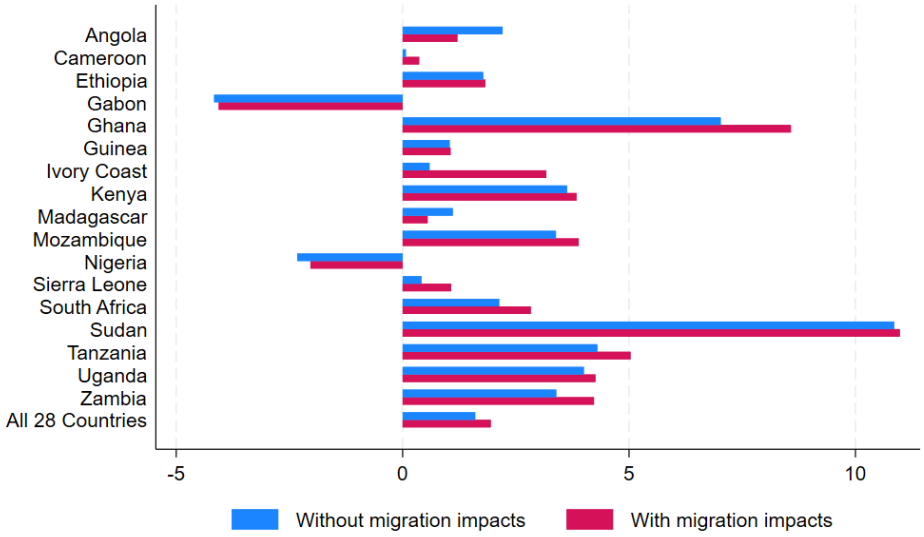
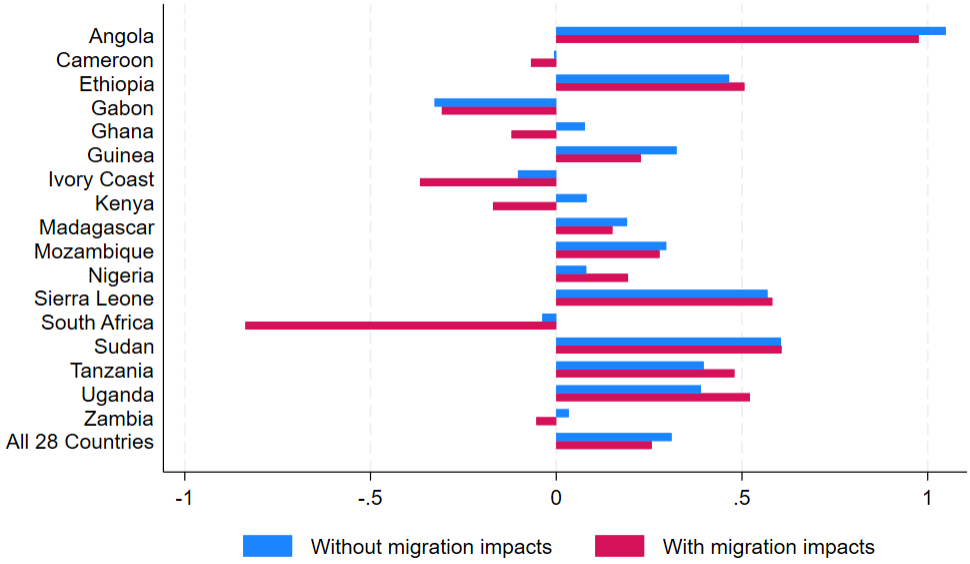


Figure 10: Impacts of non-agricultural productivity growth with and without migration



(a) Impacts on structural change



(b) Impacts on deforestation

Notes:

Table 1: Effect of Non-Agricultural Employment Growth on Deforestation via Migration and Trade

	(1) Emigration over land area	(2) Share of area deforested	(3) Share of area deforested (TMF)	(4) Δ share cropland
B_m	0.732***	-0.042***	-0.026***	-0.060**
	(0.235)	(0.012)	(0.010)	(0.030)
B_d	-0.702**	0.093***	0.069***	0.087**
	(0.279)	(0.023)	(0.020)	(0.037)
Observations	662	770	587	754
Outcome Mean in %	.49	.19	.06	.1

The unit of observation is an administrative unit. The sample in column 2 includes all units with non-zero forest coverage in 2000 within the 17 countries of our sample with two censuses. In column 1, the sample is reduced since we cannot compute the emigration rate in Gabon or Ghana, while in column 4 we do not have data on the change in cropland share for Cameroon or Mauritius. In column 3, we only include regions that have non-zero tropical moist forest cover at baseline. All regressions are weighted by baseline forest cover. Robust standard errors in parenthesis. Regressions include the sum of migration shares across all destinations and the sum of trade shares across all destinations for each origin as controls. All models include country fixed effects. $^{*}p < .1,^{**} p < .05,^{***} p < .01$

Table 2: Effect of Non-Agricultural Employment Growth on Deforestation via Migration and Trade depending on the importance of Cash Crops locally

	(1) Emigration over land area	(2) Share of area deforested	(3) Share of area deforested (TMF)	(4) Δ share cropland
Panel A: Low Cash Crop share				
B_m	1.176	-0.062***	-0.039*	-0.082**
	(0.724)	(0.018)	(0.020)	(0.037)
B_d	-0.930	0.081**	0.070*	0.101
	(0.921)	(0.038)	(0.042)	(0.064)
Observations	294	381	275	434
Panel B: High Cash Crop share				
B_m	0.720***	-0.025**	-0.018*	0.021
	(0.171)	(0.012)	(0.009)	(0.111)
B_d	-1.440	0.064**	0.052***	0.008
	(0.960)	(0.029)	(0.020)	(0.104)
Observations	368	380	303	315

The unit of observation is an administrative unit. The sample includes all 744 units with non-zero forest coverage in 2001 within the 17 countries of our sample. Panel A includes administrative units with below median share of cash crops per cultivated area; Panel B includes units with above median share of cash crops per cultivated area.In column 1, the sample is reduced since we cannot compute the emigration rate in Gabon or Ghana, while in column 4 we do not have data on the change in cropland share for Cameroon or Mauritius. In column 3, we only include regions that have non-zero tropical moist forest cover at baseline. Regressions weighted by baseline tree cover. Robust standard errors in parenthesis. Regressions include the sum of migration shares across all destinations and the sum of trade shares across all destinations for each origin as controls. Both shift shares and the sum of shares are standardized. All models include country fixed effects. $^{*}p < .1,^{**}p < .05,^{***}p < .01$

Table 3: Effect of Non-Agricultural Employment Growth on Deforestation via Migration and Trade depending on Agricultural Imports

	(1) Emigration over land area	(2) Share of area deforested	(3) Share of area deforested (TMF)	(4) Δ share cropland
Panel A: Low Agricultural Imports				
B_m	-0.108	-0.294	-0.533**	-0.012
	(2.137)	(0.314)	(0.245)	(0.980)
B_d	-5.399***	1.148***	0.799***	0.181
	(1.288)	(0.347)	(0.278)	(0.670)
Observations	348	357	287	277
Panel B: High Agricultural Imports				
B_m	9.108*	-0.422**	-0.004	-0.093
	(5.197)	(0.194)	(0.066)	(0.240)
B_d	-14.471	0.722**	-0.052	-0.340
	(13.163)	(0.357)	(0.178)	(0.478)
Observations	314	413	291	413

The unit of observation is an administrative unit. The sample includes all 764 units with non-zero forest coverage in 2001 within the 17 countries of our sample. Panel A includes countries with below median agricultural trade openness, measured as (agricultural exports + agricultural imports)/agricultural production ; Panel B includes units with above median share of agricultural trade openness. In column 1, the sample is reduced since we cannot compute the emigration rate in Gabon or Ghana, while in column 4 we do not have data on the change in cropland share for Cameroon or Mauritius. In column 3, we only include regions that have non-zero tropical moist forest cover at baseline. Regressions weighted by baseline tree cover. Robust standard errors in parenthesis. Regressions include the sum of migration shares across all destinations and the sum of trade shares across all destinations for each origin as controls. All models include country fixed effects. $^{*}p < .1,^{**}p < .05,^{***}p < .01$

Table 4: Parameters from the literature

Parameter	Value	Source
θ	1.7	Sotelo (2020)
η	5.4	Costinot et al. (2016)
α_A, α_S	0.7	Gollin (2002)
α_M	0.34	Dinh and Clarke (2012)
κ	3.2	Bryan and Morten (2019)
σ	0.48	Comin et al. (2021)
ε_A	0.2	Comin et al. (2021)
ε_S	1.37	Comin et al. (2021)

Table 5: Partial equilibrium impacts of migration and trade on deforestation from the model counterfactuals

	Share of area deforested	
	(1)	(2)
Δ Emigration	-0.017***	-0.013***
(N per Ha)	(0.003)	(0.003)
Δ Exports	0.016*	0.013*
(std)	(0.008)	(0.007)
Δ Local		0.000
manuf. productivity		(0.000)
Δ Local		0.000***
agricultural productivity		(0.000)
Observations	1324	1324

Results for all countries. Regressions weighted by tree cover . Robust standard errors in parenthesis. B_m is the migration shift-share, and B_d is the trade shift share. The shift is non-agricultural productivity, the migration shares and the trade shares are defined as in the reduced form. Regressions that include the migration shift-share also include the sum of migration shares for that origin across all destinations. Regressions that include the trade shift-share also include the sum of the trade shares. All models include country fixed effects.
$^{*}p < .1,^{**} p < .05,^{***} p < .01$

A Figures & Tables

Table A.1: Sample Administrative Unit level

Country	Administrative Unit Level	Number of Units	Census Year	
			First	Second
Angola	2	161	2014	
Benin	2	77	2002	2013
Botswana	1	21	2001	2011
Burkina Faso	2	45	1996	2006
Cameroon	2	39	1987	2005
Ethiopia	1	11	2014	
Gabon	1	9	1993	2013
Ghana	2	102	2000	2010
Guinea	1	33	1996	2014
Ivory Coast	2	50	1998	
Kenya	2	35	1999	2009
Lesotho	1	10	1996	2006
Madagascar	2	114	2018	
Malawi	1	26	2008	
Mali	2	47	1998	2009
Mauritius	1	10	2000	2011
Mozambique	2	137	1997	2007
Nigeria	1	37	2012	
Rwanda	2	30	2012	
Senegal	2	27	2002	2013
Sierra Leone	2	126	2004	2015
South Africa	2	18	2001	2011
Sudan	1	25	2008	
Tanzania	1	23	2002	2012
Togo	2	37	2010	
Uganda	1	36	2002	2014
Zambia	2	55	2000	2010
Zimbabwe	2	88	2012	

Notes: The table reports the administrative unit level used in our analysis and the total number of units in the country at that level. The first and second census years are those that we used in the analysis. Although data is available at the admin 2 level for Mauritius, we use admin level 1 because most of the migration at the admin 2 level is occurring between districts within Port Louis, so the more aggregated data on the island is more representative of the type of migration that we are interested in.

Table A.2: Descriptive statistics for the main outcome variables of the reduced form analysis (Sample of countries with two censuses)

Country	Emigration over land area	Share of area deforested	Share of area deforested (TMF)	Δ share cropland	N
BEN	3.73	0.64	0.21	-0.94	71
BWA	0.87	0.00	.	0.04	17
CMR	3.53	0.74	0.34	.	38
GAB	.	0.68	0.12	0.00	9
GHA	.	4.34	1.22	0.39	99
GIN	3.29	2.45	1.09	0.11	33
KEN	34.27	1.09	0.51	2.79	35
LSO	3.76	0.00	.	-0.52	10
MLI	1.03	0.02	0.00	0.96	20
MOZ	1.61	2.10	0.01	0.51	137
MUS	30.97	1.10	0.96	0.00	9
SEN	7.25	0.34	0.22	0.17	19
SLE	9.79	10.00	5.49	0.06	124
TZA	7.72	1.80	0.08	1.78	23
UGA	8.69	2.74	3.03	6.86	36
ZAF	2.59	1.66	.	0.14	18
ZMB	2.11	1.34	0.03	0.62	55
All Countries	4.88	1.85	0.54	1.01	753

Notes: The table reports the average of the outcomes considered in the reduced form analysis for the sample of countries with two censuses. Gabon only reports migration since birth, hence we cannot compute migration between the two censuses. Ghana has missing migration information in the second census. Botswana, Lesotho and South Africa are not in the tropical zone.

Table A.3: Annual Migration by Measure and Census

measure	outcome	census year	population (millions) (1)	migration (thousands) (2)	migration rate (3)
Benin					
Duration: 1992 - 2001	migration shares	2002	3.425	57.94	1.69
Duration: 2003 - 2012	emigration	2013	5.131	63.26	1.23
Birth	none	2002	3.214	110.86	3.45
Birth	none	2013	4.864	154.55	3.18
Botswana					
5 years: 1996 - 2001	migration shares	2001	0.986	51.33	5.21
5 years: 2006 - 2011	emigration	2011	1.267	66.84	5.28
Birth	none	2001	0.986	49.61	5.03
Birth	none	2011	1.267	64.82	5.12
Burkina Faso					
1 year: 1996 - 1997	none	1996	5.382	369.33	6.86
1 year: 2005 - 2006	emigration	2006	7.120	191.41	2.69
Birth	migration shares	1996	5.382	172.30	3.20
Birth	none	2006	7.120	220.11	3.09
Ghana					
5 years: 1996 - 2000	migration shares	2000	10.211	139.47	1.37
Guinea					
Duration: 1986 - 1995	migration shares	1996	3.698	103.16	2.79
Duration: 2004 - 2013	emigration	2014	5.445	83.35	1.53
Birth	none	1996	3.698	132.47	3.58
Birth	none	2014	5.445	163.79	3.01
Kenya					
Duration: 1989 - 1998	migration shares	1999	15.069	23.05	0.15
1 year: 2008 - 2009	emigration	2009	20.665	759.56	3.68
Birth	none	1999	15.069	484.80	3.22
Birth	none	2009	20.665	630.11	3.05
Lesotho					
10 years: 1987 - 1996	migration shares	1996	1.084	11.55	1.07
10 years: 1997 - 2006	emigration	2006	1.096	12.42	1.13
Birth	none	1996	1.084	20.08	1.85
Birth	none	2006	1.096	19.33	1.76

...Continued on next page

Annual Migration by Measure and Census – *continued from previous page*

measure	outcome	census year	population (millions) (1)	migration (thousands) (2)	migration rate (3)
Mali					
10 years: 1989 - 1998	migration shares	1998	5.006	121.42	2.43
10 years: 2000 - 2009	emigration	2009	7.074	247.53	3.50
Birth	none	1998	5.006	120.06	2.40
Birth	none	2009	6.993	297.93	4.26
Mauritius					
5 years: 1996 - 2000	migration shares	2000	0.826	12.65	1.53
5 years: 2007 - 2011	emigration	2011	0.878	9.21	1.05
Mozambique					
5 years: 1988 - 1997	none	1997	8.223	349.42	4.25
5 years: 1998 - 2007	emigration	2007	10.303	222.34	2.16
Birth	migration shares	1997	8.223	336.35	4.09
Birth	none	2007	10.303	391.33	3.80
Senegal					
5 years: 1998 - 2002	migration shares	2002	5.402	101.47	1.88
5 years: 2009 - 2013	emigration	2013	7.223	234.46	3.25
Birth	none	2002	5.402	217.85	4.03
Birth	none	2013	7.223	271.55	3.76
Sierra Leone					
14 years: 1990 - 2004	none	2004	2.237	45.32	2.03
5 years: 2011 - 2015	emigration	2015	4.012	133.88	3.34
Birth	migration shares	2004	2.691	145.57	5.41
Birth	none	2015	4.012	207.81	5.18
South Africa					
5 years: 2007 - 2011	migration shares	2011	33.544	736.46	2.20
5 years: 2012 - 2016	emigration	2016	36.188	201.08	0.56
Birth	none	2011	34.124	610.25	1.79
Birth	none	2016	36.197	367.28	1.01
Tanzania					
1 year: 2001 - 2002	none	2002	17.347	639.70	3.69
1 year: 2011 - 2012	emigration	2012	23.083	937.83	4.06
Birth	migration shares	2002	17.347	384.18	2.21
Birth	none	2012	23.083	569.86	2.47

...Continued on next page

Annual Migration by Measure and Census – *continued from previous page*

measure	outcome	census year	population (millions) (1)	migration (thousands) (2)	migration rate (3)
Uganda					
10 years: 1993 - 2002	migration shares	2002	12.207	117.19	0.96
10 years: 2005 - 2014	emigration	2014	17.640	175.58	1.00
Birth	none	2002	12.207	345.00	2.83
Zambia					
1 year: 1999 - 2000	none	2000	5.219	337.27	6.46
1 year: 2009 - 2010	emigration	2010	6.916	696.73	10.07
Birth	migration shares	2000	5.219	267.66	5.13
Birth	none	2010	6.916	375.12	5.42

Notes: Reporting the prime-aged population in millions, average annual number of migrants and the implied annual migration rate for each reported measure. Migration since birth is annualized by dividing each observed migration by the individual's age - 14. That is, for each person who has migrated since birth, we assume that the migration happened once the individual became prime-aged and that there was a single migration episode in that period. This transformation is for expositional purposes only. All migration since birth episodes are used as shares in the analysis.

Table A.4: Effect of Non-Agricultural Employment Growth on Deforestation via Migration and Trade controlling for Local Non-Agricultural Employment Growth

	(1) Emigration over land area	(2) Share of area deforested	(3) Share of area deforested (TMF)	(4) Δ share cropland
B_m	0.736***	-0.042***	-0.027***	-0.056*
	(0.237)	(0.012)	(0.010)	(0.030)
B_d	-0.703**	0.093***	0.069***	0.087**
	(0.279)	(0.023)	(0.020)	(0.037)
Local Shift	-0.002	-0.000	0.000	0.006
	(0.002)	(0.001)	(0.001)	(0.004)
Observations	662	770	587	754

The unit of observation is an administrative unit. The sample in column 2 includes all units with non-zero forest coverage in 2000 within the 17 countries of our sample with two censuses. In column 1, the sample is reduced since we cannot compute the emigration rate in Gabon or Ghana, while in column 4 we do not have data on the change in cropland share for Cameroon or Mauritius. In column 3, we only include regions that have non-zero tropical moist forest cover at baseline. All regressions are weighted by baseline forest cover. Robust standard errors in parenthesis. Regressions include the sum of migration shares across all destinations and the sum of trade shares across all destinations for each origin as controls. All models include country fixed effects. All models include country fixed effects. $^{*}p < .1,^{**}p < .05,^{***}p < .01$

Table A.5: Shift-share estimates using change in non-agricultural employment growth also in other countries.

	(1) Emigration over land area	(2) Share of area deforested	(3) Share of area deforested (TMF)	(4) Δ share cropland
iB_m	0.633***	-0.037***	-0.021**	-0.088***
	(0.241)	(0.011)	(0.009)	(0.033)
iB_d	-0.315	0.091***	0.061***	0.198***
	(0.202)	(0.024)	(0.019)	(0.066)
Observations	662	770	587	754

The unit of observation is an administrative unit. The sample in column 2 includes all units with non-zero forest coverage in 2000 within the 17 countries of our sample with two censuses. In column 1, the sample is reduced since we cannot compute the emigration rate in Gabon or Ghana, while in column 4 we do not have data on the change in cropland share for Cameroon or Mauritius. In column 3, we only include regions that have non-zero tropical moist forest cover at baseline. All regressions are weighted by baseline forest cover. Robust standard errors in parenthesis. Regressions include the sum of migration shares across all destinations and the sum of trade shares across all destinations for each origin as controls. All models include country fixed effects. All models include country fixed effects. $^{*}p < .1,^{**}p < .05,^{***}p < .01$

Table A.6: Shift-share estimates using travel duration instead of distance for the trade shift-share.

	(1) Emigration over land area	(2) Share of area deforested	(3) Share of area deforested (TMF)	(4) Δ share cropland
B_m	0.705***	-0.040***	-0.024***	-0.054*
	(0.230)	(0.011)	(0.009)	(0.029)
B_d	-39.160***	6.298***	4.480***	5.800**
	(13.905)	(1.498)	(1.256)	(2.354)
Observations	662	770	587	754

The unit of observation is an administrative unit. The sample in column 2 includes all units with non-zero forest coverage in 2000 within the 17 countries of our sample with two censuses. In column 1, the sample is reduced since we cannot compute the emigration rate in Gabon or Ghana, while in column 4 we do not have data on the change in cropland share for Cameroon or Mauritius. In column 3, we only include regions that have non-zero tropical moist forest cover at baseline. All regressions are weighted by baseline forest cover. Robust standard errors in parenthesis. Regressions include the sum of migration shares across all destinations and the sum of trade shares across all destinations for each origin as controls. All models include country fixed effects. All models include country fixed effects. $^{*}p < .1,^{**}p < .05,^{***}p < .01$

B Mathematical Appendix

B.1 Bartik instruments

Jaeger et al. (2018) define the immigration shift-share instrument as

$$\tilde{m}_{dt} = \sum_o \frac{m_{od,t^0}}{m_{o,t^0}} \frac{\Delta m_{ot}}{N_{d,t-1}}$$

Where $m_{od,t^0}/m_{o,t^0}$ is the share of immigrants from o in d at reference date t^0 which predates t. Δm_{ot} is the total number of immigrants across all locations at time t and $N_{d,t-1}$ is the local population in d in the previous period. In this context, we do not necessarily have the stock of immigrants in each location, so we replace with the flow from the previous period and estimate the Bartik instrument for immigration as

$$B_{dt} = \sum_o \frac{m_{od,t-1}}{m_{o,t-1}} \frac{m_{ot}}{N_{d,t-1}}$$

Where $m_{od,t-1}/m_{o,t-1}$ is the share of migration out of o into d in the previous period, m_{ot} is total migration out of o in the current period, and $N_{d,t-1}$ is the total population in d at the start of the period. We use the reverse of this to define the emigration shift-share instrument.

B.2 Model Equilibrium

Agricultural Production Labor demand on a plot that is cleared equates marginal cost of that plot with marginal revenue

$$w_i = \alpha p_i^A (N_i^A(\omega))^{\alpha-1} \left(Z_i^A(\omega)\right)^{1-\alpha} \tag{18}$$

$$N_i^A(\omega) = \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} Z_i^A(\omega) \tag{19}$$

Then profit on a plot abstracting from the cost of opening the plot is given by

$$\begin{aligned}
p_i^A \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{\alpha}{1-\alpha}} Z_i^A(\omega) - w_i \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} Z_i^A(\omega) &= Z_i^A(\omega)\left((p_i^A)^{\frac{1}{1-\alpha}} \alpha^{\frac{\alpha}{1-\alpha}} w_i^{\frac{-\alpha}{1-\alpha}} - (p_i^A)^{\frac{1}{1-\alpha}} \alpha^{\frac{1}{1-\alpha}} w_i^{\frac{-\alpha}{1-\alpha}}\right) \\
&= Z_i^A(\omega)(p_i^A)^{\frac{1}{1-\alpha}} w_i^{\frac{-\alpha}{1-\alpha}} \left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right) \\
&= Z_i^A(\omega)\left(p_i^A w_i^{-\alpha} \tilde{\alpha}\right)^{\frac{1}{1-\alpha}} \\
&= \psi_i Z_i^A(\omega)
\end{aligned}$$

Forested plot ω is cleared whenever it is profitable to do so

$$\psi_i Z_i^A(\omega) > Z_i^0(\omega)$$

Where $\tilde{\alpha} = \alpha^\alpha (1-\alpha)^{1-\alpha}$. Integrating over ω the share of forested land for which it is profitable

to clear is then

$$S_i^A = Pr\Big(\psi_i Z_i^A(\omega) > Z_i^0(\omega)\Big) = \frac{\big(\Lambda_i\psi_i\big)^\theta}{\big(p^M\Lambda_i^0\big)^\theta + \big(\Lambda_i\psi_i\big)^\theta}$$

Total labor demand in j is found by summing over demand on each plot, forested and deforested:

$$\begin{aligned}
N_i^A =& \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} \Big[\bar{L}_i^F S_i E[Z_i^A | \psi_i Z_i^A > Z_i^0] + \bar{L}_i^C E[Z_i^A]\Big] \\
=& \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} \Big[\frac{1}{\psi_i}\bar{L}_i S_i E\Big[Z_i^A\psi_i | Z_i^A\psi_i > Z_i^0\Big] + \bar{L}_i^C E[Z_i^A]\Big] \\
=& \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} \Big[\frac{1}{\psi_i}\bar{L}_i S_i \Big(\big(p^M\Lambda_i^0\big)^\theta + \big(\Lambda_i\psi_i\big)^\theta\Big)^{1/\theta} + \bar{L}_i^C\Lambda_i\Big]\Gamma\Big(1-\frac{1}{\theta}\Big) \\
=& \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} \Big[\frac{1}{\psi_i}\bar{L}_i S_i \frac{\Lambda_i\psi_i}{(S_i)^{1/\theta}} + \bar{L}_i^C\Lambda_i\Big]\Gamma\Big(1-\frac{1}{\theta}\Big) \\
=& \left(\frac{\alpha p_i^A}{w_i}\right)^{\frac{1}{1-\alpha}} \Lambda_i\Big[\bar{L}_i S_i (S_i)^{-1/\theta} + \bar{L}_i^C\Big]\Gamma\Big(1-\frac{1}{\theta}\Big) \\
=& \frac{\alpha}{1-\alpha}\Big(\frac{\psi_i}{p_i^A}\Big)\Big[\bar{L}_i (S_i)^{\frac{\theta-1}{\theta}} + \bar{L}_i^C\Big]\Gamma\Big(1-\frac{1}{\theta}\Big)
\end{aligned}$$

The final expression is found by normalizing the Frechet distributions by $\bar{\Gamma} = \Gamma\Big(1-\frac{1}{\theta}\Big)^{-1/\theta}$ so that $E[Z_i^A] = \Lambda_i$ and $E[A_i^0] = \Lambda_i^0$.

Note that on each plot $Q_i^A(\omega) = (N_i^A(\omega))^\alpha (Z_i^A(\omega))^{(1-\alpha)}$. Substituting out for $Z_i^A(\omega)$ using (19) then gives $Q_i^A(\omega) = N_i^A(\omega)\frac{w_i}{\alpha p_i^A}$. Integrating over ω then gives

$$Q_i^A = N_i^A \frac{w_i}{\alpha p_i^A} \tag{20}$$

$$= \frac{1}{1-\alpha}\Big(\frac{\psi_i}{p_i^A}\Big)\Lambda_i\Big[\bar{L}_i (S_i)^{\frac{\theta-1}{\theta}} + \bar{L}_i^C\Big] \tag{21}$$

Land rents, which are returned to local workers are equal to agricultural profits minus the fixed cost of opening the land. The expected value of Z_i^0 conditional on the land being used for agriculture is given by

$$\begin{aligned}
E[Z_i^0 | Z_i^A\psi_i > Z_i^0] * Pr(Z_i^A\psi_i > Z_i^0) =& E[Z_i^0] - E[Z_i^0 | Z_i^0 > Z_i^A\psi_i] * Pr(Z_i^0 > Z_i^A\psi_i) \\
E[Z_i^0 | Z_i^A\psi_i > Z_i^0] * Pr(Z_i^A\psi_i > Z_i^0) =& p^M\Lambda_i^0 - \Big(\big(p^M\Lambda_i^0\big)^\theta + \big(\Lambda_i\psi_i\big)^\theta\Big)^{1/\theta} * (1-S_i)
\end{aligned}$$

We can then substitute using the expression for the share of land not used for agriculture

$$1 - S_i = \frac{\left(p^M \Lambda_i^0\right)^\theta}{\left(p^M \Lambda_i^0\right)^\theta + \left(\Lambda_i \psi_i\right)^\theta}$$

So that

$$\begin{aligned} E[Z_i^0 | Z_i^A \psi_i > Z_i^0] * Pr(Z_i^A \psi_i > Z_i^0) =& p^M \Lambda_i^0 - p^M \Lambda_i^0 (1 - S_i)^{1/\theta} * (1 - S_i) \\ E[Z_i^0 | Z_i^A \psi_i > Z_i^0] * Pr(Z_i^A \psi_i > Z_i^0) =& p^M \Lambda_i^0 \left(1 - (1 - S_i)^{\frac{\theta-1}{\theta}}\right) \end{aligned}$$

Total land rents are then

$$\begin{aligned} R_i =& p_i^A Q_i^A - w_i N_i^A - \bar{L}_i E[Z_i^0 | Z_i^A \psi_i > Z_i^0] * Pr(Z_i^A \psi_i > Z_i^0) \\ R_i =& (1 - \alpha_A) p_i^A Q_i^A - \bar{L}_i^F p^M \Lambda_i^0 \left(1 - (1 - S_i)^{\frac{\theta-1}{\theta}}\right) \end{aligned}$$

Manufacturing Profits Profit maximization in the manufacturing sector implies that labor demand in manufacturing is given by

$$N_d^M = \left(\frac{\alpha Z_d}{w_d}\right)^{\frac{1}{1-\alpha}}$$

Then profit in the manufacturing sector is given by

$$\begin{aligned} \Pi_d^M =& Z_d \left(\frac{\alpha Z_d}{w_d}\right)^{\frac{\alpha}{1-\alpha}} - w_d \left(\frac{\alpha Z_d}{w_d}\right)^{\frac{1}{1-\alpha}} \\ \Pi_d^M =& Z_d^{\frac{1}{1-\alpha}} w_d^{\frac{-\alpha}{1-\alpha}} \left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right) \\ \Pi_d^M =& Z_d^{\frac{1}{1-\alpha}} w_d^{\frac{-\alpha}{1-\alpha}} \alpha^{\frac{\alpha}{1-\alpha}} \left(1 - \alpha\right) \\ \Pi_d^M =& \left(\frac{Z_d \tilde{\alpha}}{w_d^\alpha}\right)^{\frac{1}{1-\alpha}} \end{aligned}$$

Consumption Decisions Fixing d, let $U(C^M, C^A)$ be the utility value of C_d when consuming quantities C^M and C^A. The individuals problem can be expressed as choosing quantities to maximize the constrained optimization problem

$$\begin{aligned} & \max_{C_d^M, \{C_{od}^A\}} U(C_d^M, C_d^A) \\ \text{subject to} \qquad & C_d^A = \left(\sum_o (\beta_{od})^{\frac{1}{\eta}} (C_{od}^A)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} \\ & p_M C_d^M + \sum_o p_o^A \tau_{od} C_{od}^A = w_d + R_d / N_d \end{aligned}$$

The first order condition for C_{od}^A is

$$U_2 \beta_{od}^{\frac{1}{\eta}} {C_{od}^A}^{\frac{-1}{\eta}} {C_d^A}^{\frac{1}{\eta}} - \lambda p_o^A \tau_{od} = 0$$

Rearranging and summing over o yields an expression for λ / U_2

$$\begin{aligned}
\beta_{od}^{\frac{1}{\eta}} (C_{od}^A)^{\frac{\eta-1}{\eta}} =& \frac{\lambda p_o^A \tau_{od} C_{od}^A}{U_2 (C_d^A)^{\frac{1}{\eta}}} \\
\sum_o \beta_{od}^{\frac{1}{\eta}} (C_{od}^A)^{\frac{\eta-1}{\eta}} =& \frac{\lambda}{U_2 (C_d^A)^{\frac{1}{\eta}}} \Big(\sum_o p_o^A \tau_{od} C_{od}^A \Big) \\
\left(\sum_o \beta_{od}^{\frac{1}{\eta}} (C_{od}^A)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} =& C_d^A = \left(\frac{\lambda}{U_2 (C_d^A)^{\frac{1}{\eta}}} \right)^{\frac{\eta}{\eta-1}} \Big(\sum_o p_o^A \tau_{od} C_{od}^A \Big)^{\frac{\eta}{\eta-1}} \\
C_d^A =& \left(\frac{\lambda}{U_2 (C_d^A)^{\frac{1}{\eta}}} \right)^{\frac{\eta}{\eta-1}} (X_d^A)^{\frac{\eta}{\eta-1}} \\
\frac{C_d^A}{X_d^A} =& \frac{\lambda}{U_2}
\end{aligned}$$

Where X_d^A is total expenditure on agricultural goods in d. Returning to the FOC, we can express demand for C_{od}^A as

$$C_{od}^A = \beta_{od} \left(\frac{X_d^A}{p_o^A \tau_{od}} \right)^{\eta} (C_d^A)^{1-\eta}$$

This expression still depends upon C_d^A. We can solve for it by substituting this expression into the definition

$$\begin{aligned}
C_d^A =& \Big(\sum_o (\beta_{od})^{\frac{1}{\eta}} (C_{od}^A)^{\frac{\eta-1}{\eta}} \Big)^{\frac{\eta}{\eta-1}} \\
C_d^A =& \Big(\sum_o (\beta_{od})^{\frac{1}{\eta}} \left(\beta_{od} \left(\frac{X_d^A}{p_o^A \tau_{od}} \right)^{\eta} (C_d^A)^{1-\eta} \right)^{\frac{\eta-1}{\eta}} \Big)^{\frac{\eta}{\eta-1}} \\
(C_d^A)^{\eta} =& \left(\sum_o \beta_{od} \Big(p_o^A \tau_{od} \Big)^{1-\eta} \right)^{\frac{\eta}{\eta-1}} (X_d^A)^{\eta} \\
P_d^A C_d^A =& X_d^A
\end{aligned}$$

Where $P_d^A = \left(\sum_o \beta_{od} \Big(p_o^A \tau_{od} \Big)^{1-\eta} \right)^{\frac{1}{1-\eta}}$ is the price index that equates consumption of the composite good with total expenditure on agricultural goods. Demand for C_{od}^A can then be expressed as

$$C_{od}^A = \beta_{od} \big(p_o^A \tau_{od} \big)^{-\eta} (P_d^A)^{\eta-1} X_d^A$$

Non-Homothetic Demand Functions To make the exposition simpler, we will consider the general case in which there are non-homotheticities in both goods which I will denote by ε_M and ε_A. Recall that C_d is the utility function that the individual seeks to maximize. By rearranging (7), we can express this quantity as:

$$C_d = \left[(\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_d^A)^{\frac{1}{\sigma}} \left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

Then we can express the maximization problem in the standard way

$$C_d = \max_{C_d^M, C_d^A} \left[(\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_d^A)^{\frac{1}{\sigma}} \left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} + \lambda(E - p_M C_d^M - p_A C_d^A)$$

Implicit differentiation with respect to C_d^M yields

$$\begin{aligned}
\frac{\partial C_d}{\partial C_d^M} =& \frac{\sigma}{\sigma-1} \left[(\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_d^A)^{\frac{1}{\sigma}} \left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}} \times \\
& \left(\frac{\sigma-1}{\sigma} (\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{-1}{\sigma}} \left(\frac{(C_d)^{\varepsilon_M - 1} - C_d^M(\varepsilon_M - 1)(C_d)^{\varepsilon_M - 2} \frac{\partial C_d}{\partial C_d^M}}{((C_d)^{\varepsilon_M - 1})^2} \right) + \right. \\
& \left. \frac{\sigma-1}{\sigma} (\Omega_d^A)^{\frac{1}{\sigma}} \left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}} \right)^{\frac{-1}{\sigma}} \left(\frac{-C_d^A(\varepsilon_A - 1)(C_d)^{\varepsilon_A - 2} \frac{\partial C_d}{\partial C_d^M}}{((C_d)^{\varepsilon_A - 1})^2} \right) \right) - \lambda p_M
\end{aligned}$$

Under the assumption that we are at an interior solution, at an optimum, $\frac{\partial C_d}{\partial C_d^M} = 0$. Thus the equation simplifies to

$$\begin{aligned}
\lambda p_M =& \frac{\sigma}{\sigma-1} \left[(\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_d^A)^{\frac{1}{\sigma}} \left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}} \times \\
& \left(\frac{\sigma-1}{\sigma} (\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{-1}{\sigma}} \left(\frac{(C_d)^{\varepsilon_M - 1}}{((C_d)^{\varepsilon_M - 1})^2} \right) \right) \\
\lambda p_M =& \left[(\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{\sigma-1}{\sigma}} + (\Omega_d^A)^{\frac{1}{\sigma}} \left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}} \left((\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{-1}{\sigma}} \left(\frac{1}{(C_d)^{\varepsilon_M - 1}} \right) \right) \\
\lambda p_M =& (C_d)^{\frac{1}{\sigma}} \left((\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{-1}{\sigma}} \left(\frac{1}{(C_d)^{\varepsilon_M - 1}} \right) \right) \\
\lambda p_M C_d^M =& (C_d)^{\frac{1}{\sigma}} \left((\Omega_d^M)^{\frac{1}{\sigma}} \left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}} \right)^{\frac{\sigma-1}{\sigma}} \right)
\end{aligned}$$

Using the equivalent logic for C_d^A we solve for λ by

$$\begin{aligned}
\lambda\left(p_M C_d^M + p_A C_d^A\right) =& (C_d)^{\frac{1}{\sigma}}\left((\Omega_d^M)^{\frac{1}{\sigma}}\left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}}\right)^{\frac{\sigma-1}{\sigma}}\right) + (C_d)^{\frac{1}{\sigma}}\left((\Omega_d^A)^{\frac{1}{\sigma}}\left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}}\right)^{\frac{\sigma-1}{\sigma}}\right) \\
\lambda X_d =& (C_d)^{\frac{1}{\sigma}}\left((\Omega_d^M)^{\frac{1}{\sigma}}\left(\frac{C_d^M}{(C_d)^{\varepsilon_M - 1}}\right)^{\frac{\sigma-1}{\sigma}} + (\Omega_d^A)^{\frac{1}{\sigma}}\left(\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}}\right)^{\frac{\sigma-1}{\sigma}}\right) \\
\lambda X_d =& (C_d)^{\frac{1}{\sigma}}(C_d)^{\frac{\sigma-1}{\sigma}} \\
\lambda =& \frac{C_d}{X_d}
\end{aligned}$$

Where X_d is total expenditure in d. Returning to the expression for C_d^M

$$\begin{aligned}
(C_d^M)^{\frac{1}{\sigma}} =& (C_d)^{\frac{1}{\sigma}}\left((\Omega_d^M)^{\frac{1}{\sigma}}\left(\frac{1}{(C_d)^{\varepsilon_M - 1}}\right)^{\frac{\sigma-1}{\sigma}}\right)\frac{X_d}{P_d^M C_d} \\
C_d^M =& C_d \Omega_d^M \left(\frac{1}{(C_d)^{\varepsilon_M - 1}}\right)^{\sigma-1}\left(\frac{X_d}{P_d^M C_d}\right)^{\sigma} \\
C_d^M =& \Omega_d^M \left(\frac{1}{(C_d)^{\varepsilon_M}}\right)^{\sigma-1}\left(\frac{X_d}{P_d^M}\right)^{\sigma} = \Omega_d^M \left(\frac{X_d}{P_d^M}\right)^{\sigma}(C_d)^{\varepsilon_M(1-\sigma)} \\
C_d^A =& \Omega_d^A \left(\frac{1}{(C_d)^{\varepsilon_A}}\right)^{\sigma-1}\left(\frac{X_d}{P_d^A}\right)^{\sigma} = \Omega_d^A \left(\frac{X_d}{P_d^A}\right)^{\sigma}(C_d)^{\varepsilon_A(1-\sigma)}
\end{aligned}$$

The expression for total consumption can be expressed as

$$\begin{aligned}
C_d =& \left[(\Omega_d^M)^{\frac{1}{\sigma}}\left(\Omega_d^M\left(\frac{1}{(C_d)^{\varepsilon_M}}\right)^{\sigma-1}\left(\frac{X_d}{P_d^M}\right)^{\sigma}\frac{1}{(C_d)^{\varepsilon_M - 1}}\right)^{\frac{\sigma-1}{\sigma}}\right. \\
&\left. + (\Omega_d^A)^{\frac{1}{\sigma}}\left(\Omega_d^A\left(\frac{1}{(C_d)^{\varepsilon_A}}\right)^{\sigma-1}\left(\frac{X_d}{P_d^A}\right)^{\sigma}\frac{C_d^A}{(C_d)^{\varepsilon_A - 1}}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
C_d =& \left[\Omega_d^M\left(\left(\frac{C_d}{(C_d^{\varepsilon_M})^{\sigma}}\right)\left(\frac{X_d}{P_d^M}\right)^{\sigma}\right)^{\frac{\sigma-1}{\sigma}} + \Omega_d^A\left(\left(\frac{C_d}{(C_d^{\varepsilon_A})^{\sigma}}\right)\left(\frac{X_d}{P_d^A}\right)^{\sigma}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\
C_d =& \left[\Omega_d^M\left((C_d)^{\varepsilon_M - 1}P_d^M\right)^{1-\sigma} + \Omega_d^A\left((C_d)^{\varepsilon_A - 1}P_d^A\right)^{1-\sigma}\right]^{\frac{\sigma}{\sigma-1}} C_d^{1-\sigma}X_d^{\sigma} \\
C_d P_d =& X_d \\
P_d =& \left[\Omega_d^M\left((C_d)^{\varepsilon_M - 1}P_d^M\right)^{1-\sigma} + \Omega_d^A\left((C_d)^{\varepsilon_A - 1}P_d^A\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}
\end{aligned}$$

In the context of our model, this yields the two Hicksian demand functions

$$
\begin{aligned}
C_d^M =& \Omega_d^M \left(\frac{1}{X_d} \right)^{-\sigma} (C_d)^{(1-\sigma)} \\
C_d^A =& \Omega_d^A \left(\frac{P_d^A}{X_d} \right)^{-\sigma} (C_d)^{\varepsilon(1-\sigma)} \\
C_d =& \left[\Omega_d^M + \Omega_d^A \left((C_d)^{\varepsilon-1} P_d^A \right)^{1-\sigma} \right]^{\frac{1}{\sigma-1}} X_d \\
P_d =& \left[\Omega_d^M + \Omega_d^A \left((C_d)^{\varepsilon-1} P_d^A \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}
\end{aligned}
$$

Expenditure shares We now derive an expression of expenditure share of each good as a function of its price and of price indices for each location i and sector of employment k. The ratios of the expenditure shares of agricultural and service goods to the manufacturing good are given by:

$$
\frac{\mu_{ik}^A}{\mu_{ik}^M} = \Omega_A \left(\frac{P_i^A}{p^M} \right)^{(1-\sigma)} c_{ik}^{(\epsilon_A-1)(1-\sigma)}, \tag{22}
$$

$$
\frac{\mu_{ik}^S}{\mu_{ik}^M} = \Omega_S \left(\frac{p_i^S}{p^M} \right)^{(1-\sigma)} c_{ik}^{(\epsilon_S-1)(1-\sigma)}. \tag{23}
$$

Since $\mu_{ik}^M = 1 - \mu_{ik}^A - \mu_{ik}^S$, this implies:

$$
\mu_{ik}^A = \frac{\Omega_A \left(\frac{P_i^A}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_A-1)(1-\sigma)}}{1 + \Omega_A \left(\frac{P_i^A}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(\frac{p_i^S}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_S-1)(1-\sigma)}}, \tag{24}
$$

$$
\mu_{ik}^M = \frac{1}{1 + \Omega_A \left(\frac{P_i^A}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(\frac{p_i^S}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_S-1)(1-\sigma)}}, \tag{25}
$$

$$
\mu_{ik}^S = \frac{\Omega_S \left(\frac{p_i^S}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_S-1)(1-\sigma)}}{1 + \Omega_A \left(\frac{P_i^A}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(\frac{p_i^S}{p^M} \right)^{1-\sigma} c_{ik}^{(\epsilon_S-1)(1-\sigma)}}. \tag{26}
$$

Defining $P_{ik} = \frac{e_{ik}}{c_{ik}}$ and multiplying numerators and denominators by $(p^M)^{1-\sigma}$, then we can write this as

$$\mu_{ik}^A = \frac{\Omega_A \left(P_i^A\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_A-1)(1-\sigma)}}{(p^M)^{(1-\sigma)} + \Omega_A \left(P_i^A\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(p_i^S\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_S-1)(1-\sigma)}}, \quad (27)$$

$$\mu_{ik}^M = \frac{(p^M)^{(1-\sigma)}}{(p^M)^{(1-\sigma)} + \Omega_A \left(P_i^A\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(p_i^S\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_S-1)(1-\sigma)}}, \quad (28)$$

$$\mu_{ik}^S = \frac{\Omega_S \left(p_i^S\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_S-1)(1-\sigma)}}{(p^M)^{(1-\sigma)} + \Omega_A \left(P_i^A\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(p_i^S\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_S-1)(1-\sigma)}}. \quad (29)$$

Price indices Conversely we can derive an expression of price indices for each location i and sector of employment k as a function of expenditure shares and prices of agricultural goods, manufacturing, and services. To derive an expression for P_{jk} in terms of e_{jk} but not c_{jk}, we use equation (13) of Comin et al. (2021) which gives:

$$\frac{e_{jk}}{P_{jk}} = c_{jk} = \frac{e_{jk}}{p^M}(\mu_{jk}^M)^{\frac{1}{1-\sigma}} \quad (30)$$

Rearranging therefore gives the price indices for each location i and sector of employment k

$$\begin{aligned} P_{ik} &= \frac{p^M}{(\mu_{ik}^M)^{\frac{1}{1-\sigma}}} \\ &= \left((p^M)^{(1-\sigma)} + \Omega_A \left(P_i^A\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_A-1)(1-\sigma)} + \Omega_S \left(p_i^S\right)^{(1-\sigma)} \left(e_{ik}/P_{ik}\right)^{(\epsilon_S-1)(1-\sigma)}\right)^{\frac{1}{1-\sigma}} \\ &= ((p^M)^{(1-\sigma)} + \Omega_A \left(P_i^A\right)^{(1-\sigma)} \left(e_{ik}/p_M\right)^{(\epsilon_A-1)(1-\sigma)} (1-\mu_{ik}^A-\mu_{ik}^S)^{(\epsilon_A-1)} \\ &+ \Omega_S \left(p_i^S\right)^{(1-\sigma)} \left(e_{ik}/p_M\right)^{(\epsilon_S-1)(1-\sigma)} (1-\mu_{ik}^A-\mu_{ik}^S)^{(\epsilon_S-1)})^{\frac{1}{1-\sigma}} \end{aligned} \quad (31)$$

Migration Costs Equation 15 gives the migration rate from i to j :

$$m_{ij} = \frac{\left(V_{ij}\right)^\kappa}{\sum_{j'} \left(V_{ij'}\right)^\kappa}$$

where $V_{ij} = 1/\delta_{ij} \times \left((\beta_{jA} v_{jA}/P_{jA})^\gamma + (v_{jM}/P_{jM})^\gamma + (\beta_{jS} v_{jS}/P_{jS})^\gamma\right)^{1/\gamma}$ is the expected indirect utility from migrating to j.
Hence

$$\frac{m_{ij} m_{ji}}{m_{ii} m_{jj}} = \frac{\left(V_{ij} V_{ji}\right)^\kappa}{\left(V_{ii} V_{jj}\right)^\kappa} = \delta_{ij}^{-2\kappa} \quad (32)$$

Which yields the desired expression:

$$\delta_{ij} = \left(\frac{m_{ij} m_{ji}}{m_{ii} m_{jj}} \right)^{-\frac{1}{2\kappa}} \tag{33}$$

C Census Data Details

All IPUMS-I censuses are 10% samples unless otherwise indicated. Migration flows are constructed using the most disaggregated level of administrative geography available, except for Mauritius. Our preferred question is, "in which year did you move here and where did you live previously?" When we know the year of migration, we exclude migrants in the previous year to exclude temporary and transitional migrants from our measure. When migration history is asked in more than one way and we do not have the year of migration, we give preference to (in this order), "where did you live 10 years ago?", "where did you live five years ago?", "where were you born?", and "where did you live one year ago?".

Angola: 2014 census. Respondents were asked in which municipality (level 2) and province (level 1) they lived in 5 years ago. Information is also available for the municipality and province of residence 1 year before the census, and for the municipality and province of birth.

Benin: 2002 and 2013 censuses. Migration is asked as, "length of time in current residence," (translated from French to English). For 2013, the values range from 1948-2013, and for 2002, from 1937-2002. Our measures of migration uses flows between 1992-2001 and 2003-2012 for the first and second censuses, respectively.

Botswana: 2001 and 2011 censuses. Respondents were asked where they lived at this time last year and in August of 1996/2006, respectively. This information, as well as the birth location, are collected at the district (level 1 geography) level.

Burkina Faso: 1996 and 2006 censuses. Respondents were asked which province (level 1) and department (level 2) they lived one year ago. The province and department of birth are also available.

Cameroon: 1987 and 2005 censuses. In 1987, respondents were asked in which province (level 1) and department (level 2) they lived in 11 years ago. In 2005, respondents were again asked in which province and department they lived in 11 years ago.

Côte d'Ivoire: 1998 census. Respondents were asked about their region (level 1) and department (level 2) of residence one year ago. Information is also available about the location of birth.

Ethiopia: socioeconomic survey 2013-2014 (Central Statistics Agency , CSA). The survey began in 2011/12 as the Ethiopia Rural Socioeconomic Survey and was expanded in wave 2 to include all urban areas, creating a nationally representative sample of 5,262 households. We use data from wave 2 of the ESS. Respondents were asked whether they had migrated to a different region (level 1) since the timing of the previous wave. As the median time between waves is 2 years, we construct migration variables as moving in the last 2 years.

Gabon: 1993 and 2013 censuses. In the absence of micro data about these censuses, we used aggregated tables that appeared in reports about the results of the censuses. For the 1993 census, the report includes migration matrices showing the number of people living in each province (level 1) and their province of birth (Direction Générale de la Statistique , DGS). Similarly, for the 2013 census, the report includes migration matrices of current province and province of birth (Direction Générale de la Statistique , DGS).

Ghana: 2000 and 2010 censuses. In the 2000 census, respondents were asked what was the district of usual residence five years ago? This question is not repeated in the 2010 census. Instead respondents were asked for how many years have they lived in the current town/village if they haven't lived there since birth. In both censuses, respondents were asked for the region of birth. We use the district of residence five years ago as our measure of migration.

Guinea: 1996 and 2014 censuses. Respondents were asked which prefecture (level 2) they lived

in before moving to the current prefecture. In 1996, respondents were asked if they have lived in the current location for less than one year, between 1-4 years, 5-9 years, 10 or more years, or since birth. In 2014, respondents were asked an open-ended question of how many years they have lived in the current prefecture. Information is also available for the location of birth. We construct migration flows using the the periods 1986-1995 and 2004-2013 to construct migration flows in the first and second censuses respectively. The period 1986-1995 will include all migrants from the previous ten years and exclude migrants from the previous year.

Kenya: 1999 and 2009 censuses. Migration history is obtained in two stages. First respondents were asked, "where was [the respondent] living in August of [the previous year]?" Next, respondents who are not living in their district of birth were asked for the year and month that they moved there. We construct flows using the location of previous residence rather than the birth location.

Lesotho: 1996 and 2006 censuses. Respondents were asked where they were living at the time of the previous census (April 1986 and 1996, respectively). If the respondent was not living in the same village they were then asked in what year they moved to their current location.

Malawi: 2008 census. Respondents were asked about their district (level 1) of previous residence, and how long they lived in their current place of residence. We use migrants who lived in a different district within the previous 10 years to construct migration.

Mali: 1998 and 2009 censuses. Data on the year of migration, previous location, and birth location are all available. The exact wording in the migration question is slightly different in the two censuses, but the premise is the same. Respondents were asked if they have lived anywhere else for at least six months. If yes, respondents were asked where they lived immediately preceding their current address. Respondents are then asked for the duration of their current residence in years. We use previous location and duration to construct migration.

Mauritius: 2000 and 2011 censuses. No data is available on the location of birth. We construct migration flows using the location of residence five years ago. Respondents were asked, "what was your usual address five years ago?" This data is available at both the district and region level. However, given the large amount of movement between level 2 administrative units (MVCA/wards) within Port Louis, we use the level 1 administrative units.

Madagascar: 2018 census. This census is not harmonized in IPUMS, and we obtain it from the statistical institute of Madagascar (Institut National de la Statistique , INSTAT). Respondents were asked about their previous region (level 1) and district (level 2) of residence and about the duration of living in the current location. We construct the migration variables based on those.

Mozambique: 1997 and 2007 censuses. In 1997, respondents were asked where they lived in October 1992, when the war ended. In 2007, this text was changed to where was the respondent living five years ago. In both rounds, respondents were also asked where they were living one year ago and where they were born. Migration in the five years after the war will likely contain a lot of moves back home, which is not what we want to measure. For this reason, we choose to estimate migration using the district of birth.

Nigeria: 2010/11 general household survey (National Bureau of Statistics, Federal Republic of Nigeria, 2010-2011). This is the first wave of a nationally representative panel survey, with 5,000 households. Migration variables are constructed using data from the second wave (2012/13), when respondents were asked whether they had migrated to a different state (level 1) since the previous wave. As the median time between waves is 2 years, we define migration as moving within the last 2 years.

Rwanda: 2012 census. Respondents were asked about the district (level 2) of previous residence

and about the duration of residence in their current district. We take migration in the last 10 years to construct the migration values.

Senegal: 2002 and 2013 censuses. Although information on the location of residence is available at one, five, and ten year intervals in the 2013 census, it is only available in a five year interval in 2002. Respondents who are residents are asked to write the name of the commune or rural community of residence five years ago, and visitors are asked to record the place where they came from. Birth location is also available, although we use the location of residence five years ago.

Sierra Leone: 2004 and 2015 censuses. In 2004, respondents were asked where they were living in December 1990 (14 years ago), and in 2015 they were asked where they were living in December 2010 (5 years ago). Due to the lack of consistency of these measures, we use the location of birth.

South Africa: The post-apartheid government of South Africa conducted its first census in 1996 and was followed by a second census in 2001. A third census was scheduled for 2006, but Statistics South Africa did not have the capacity to conduct a full census that time and it was replaced with a community survey. The third census took place in 2011 and a second community survey took place in 2016. As a result, the questions in each year varied. In 1996, respondents were asked in which year they moved to their current dwelling and responses were open ended. In 2001, the responses were recorded only for people who moved in the past five years (since the first census). In the 2007 community survey, moves were again only recorded if they happened after 2001. In 2011, the next census year, the question was asked for moves to the current place since 2001. This is different in two dimensions. First, it covers the previous ten years rather than five. Second, the question is less broad. The original question asked about when the respondent moved to their current dwelling (*i.e.* housing unit), while the latter asks when they moved to the current place (*i.e.* municipality). Finally, in 2016, respondents were asked for the year that they moved to the current place again which was top-coded at five-years. Information on the previous location is not available in the 1996 census. The province (level 1 geography) of previous residence and birth is available in all years beginning in 2001. The district of previous residence is only available in 2011 and 2016, and is never available for the birth location. Hence, we ultimately use these latter two surveys (2011 and 2016) in our analysis and use migration flows over the previous five years at the district level.

Sudan: 2008 census. Respondents were asked about their previous region (level 1) of residence, and about the duration of residence in current region. We keep migration that occurred in the 10 years prior to the census. Information about the location of birth is also available, as well as the region of residence 1 year ago. As in 2008, South Sudan has not yet become independent, we treat units in the two modern-day countries, that appear separately in the IPUMS dataset, as part of a single country.

Tanzania: 2002 and 2012 censuses. Respondents were asked where they lived in 2001/2011, respectively, *i.e.* one year prior. The text did not specify in which month. Previous residence information was collected at the region and district level in 2012, but only at the region (level 1 administrative unit) in 2002. We use the location of birth to construct migration flows which is only available at the region level.

Togo: 2010 census. Respondents were asked about their region (level 1) and prefecture (level 2) of previous residence, and about the duration of their residence in the current canton (level 3). We construct migration variables based on migration in the previous 10 years. Information is also available about the location of birth.

Uganda: 2002 and 2014 censuses. Information on the birth location is only available for the 2002 census. Respondents were asked in which district they were living prior to the current district and for how many years they have been living in the current district. We use this information to construct

migration flows during the ten-year windows preceding each census, 1992-2001 and 2004-2013. This excludes all migration that happened in the year just preceding the census.

Zambia: 2000 and 2010 censuses. Respondents were asked where were they living before August of the previous year. This information is available at the first and second level of geographic administrative unit. However, given the short duration, we instead use the location of birth.

Zimbabwe: 2012 census. Respondents were asked about their district (level 2) of resident 10 years ago.