Universidad de San Andrés

FROM FARMS TO BORDERS: AGRICULTURAL DISTORTIONS AND INTERNATIONAL MIGRATION

Braulio Britos (International Monetary Fund)
Manuel A. Hernández (International Food Policy Research Institute)
Danilo Trupkin (Universidad de San Andrés)

Primera versión: Diciembre 2025
Esta versión: Diciembre 2025

Documento de Trabajo Nº 174
Departamento de Economía
Universidad de San Andrés

Vito Dumas 284, B1644BID, Victoria,
San Fernando, Buenos Aires, Argentina
Teléfono +54 11 7078 0400
Email: economia@udesa.edu.ar

From farms to borders: Agricultural distortions and international migration*

Braulio Britos   Manuel A. Hernandez   Danilo Trupkin

December 17, 2025

Abstract

International migration has surged in recent years, especially from rural areas in developing countries. This paper examines how agricultural distortions contribute to these emigration patterns and affect welfare, using Guatemala as a case study. A structural model with agricultural and non-agricultural sectors, estimated with micro and aggregate data, shows that distortions drive emigration among more productive agents and cause factor misallocation, diminishing overall productivity and incomes. Reducing distortions to the most efficient departments lowers emigration by 2.3 points and raises agricultural productivity by 30.1% and median welfare by 4.5%. High-distortion areas are more isolated and lack institutional and financial access.

Keywords: Agricultural distortions, Emigration, Labor mobility, Productivity, Welfare

JEL Classifications: O15, O13, J24, J61, R23, Q1, O40, I31

*We thank the valuable comments of Tasso Adamopoulos, Lian Allub, Guillermo Alves, Julieta Caunedo, Evangelina Dardati, Dolores de la Mata, Javier Garcia-Cicco, Juan Odriozola, Diego Restuccia, and Mariano Tommasi, and seminar participants at the 2025 Meeting of the Latin American and Caribbean Economic Association (LACEA), 2025 RIDGE Forum, Development Bank of Latin America and the Caribbean (CAF), Universidad de San Andres, and International Monetary Fund (IMF). Britos: International Monetary Fund, bbritos@imf.org; Hernandez: International Food Policy Research Institute, m.a.hernandez@cgiar.org; Trupkin: Universidad de San Andres, dtrupkin@udesa.edu.ar. This work was supported by the donors who fund the CGIAR's Science Programs on Food Frontiers and Security and Scaling for Impact through their contributions to the CGIAR Trust Fund. We also gratefully acknowledge support from CAF, through the research call "The rural world in Latin America and the Caribbean."

1 Introduction

International migration is a recurrent phenomenon that has grown rapidly over the past two decades and at a faster pace than total population growth. According to the International Organization for Migration (IOM), there were 281 million international migrants in 2020 (3.6% of the world population), compared to 173 million in 2000 (2.8% of the population).[1] A substantial share of international migrants originates from rural areas, which also receive approximately 40% of global remittance inflows (FAO, 2016).[2] People generally emigrate from vulnerable areas in developing countries due to economic motives as well as extreme weather events and conflict (violence) such that migration can represent an important adaptation strategy to support livelihoods, build resilience, and protect against fragility (Hernandez et al., 2023). Clemens et al. (2019) and Clemens & Pritchett (2019) argue that barriers and restrictions to migration may result in important welfare losses.[3] Emigration similarly often involves irregular (undocumented) migration, particularly among food insecure households.[4]

This paper formally examines the role of agricultural distortions on explaining emigration patterns in developing countries and their effects on local aggregate productivity and welfare, using the case of Guatemala as an example. Agricultural distortions can arise from the existence of inefficient regulations and interventions, information asymmetries in favor of insiders, transaction costs, among other factors. Guatemala offers an interesting setting to evaluate the association between distortions in agricultural activities and migration. Agriculture employs around one third of the economically active population in the country (84% in rural areas) but most of the occupations (90%) are considered informal and over the past years the sector has been growing at a lower rate than the total economy, representing less than one tenth of the Gross Domestic Product (GDP) (Derlagen et al., 2020; INE, 2021). Land markets additionally present a high level of concentration and segmentation with limited market information, high informality, and cumbersome and costly

[1] `https://www.iom.sk/en/migration/migration-in-the-world.html` (accessed September 2025).

[2] Western Europe and the United States (US) represent the two main destinations for emigrants.

[3] Clemens et al. (2019) estimate that migration barriers of low-skilled men to the US represent over 13.7 thousand US dollars per worker per year.

[4] It is estimated that around 80% of the people forcibly displaced in the world have experienced acute food insecurity and high levels of malnutrition (Barchfield, 2021; USAID, 2021). Black & Sigman (2022) additionally point out that over 50,000 people have lost their lives during migratory movements since 2014.

transaction procedures (Britos et al., 2022), combined with poor governance conditions and weak institutions across the country (Gauster & Isakson, 2007; CRS, 2024).

Emigration, in turn, is recurrent and increasing, which situates the country among the top-15 recipient countries of international remittances (IOM, 2022), despite its relatively small population size.[5] It is estimated that over 850 thousand net people have migrated between 2002 and 2021,[6] while the most recent population census (INE, 2020) reveals a sustained increase in the number of both men and women in rural areas leaving the country between 2002 and 2018. Although natural disasters and insecurity are part of the longstanding emigration trigger factors in Guatemala and the Central American region, economic factors, including low wages, unemployment, and insufficient income, continue to be the main motivation, followed by family reunification (CAI, 2019; Soto et al., 2021). The US is by far the primary destination for Guatemalan migrants.[7]

We develop a theoretical framework where a household member has the option to stay in her country and work in the agricultural or non-agricultural sector, or emigrate. Agricultural activities are assumed to be performed at multiple (rural) locations while non-agricultural activities are performed at a specific (urban) location. In a given rural location, we allow for a distribution of individuals with heterogeneous agricultural and non-agricultural skills that face an idiosyncratic agricultural distortion. Multiple locations are considered with varying degrees of distortions to characterize land and labor (mis)allocations across agricultural and nonagricultural sectors and regions, and to quantify how these misallocations contribute to the rate of emigration, aggregate productivity, and welfare.

While the study pays particular attention to emigration decisions, the setup additionally permits to account for internal rural-urban migration choices, which are captured in the decision of individuals to work in the agricultural sector in their own location or to work in the non-agricultural sector in another given urban location. More important, the setting incorporates relevant features and regularities linked to emigration decisions in Guatemala and Central America such as the extent of household networks abroad, remittances, and

[5] International remittances represented close to 18% of the GDP in 2021 (Maldonado & Harris, 2022).

[6] `https://population.un.org/wpp/Download/Standard/MostUsed/` (accessed September 2025).

[7] According to the database of Chi et al. (2025), around 75% of Guatemalan migrants went to the US in 2019 and 85% in 2022.

migration costs. As noted above, transnational ties and connections (e.g., having a relative living abroad, receiving remittances) are important factors shaping international migration from the region.[8] Similarly, according to CID Gallup, one third of Central American migrants use traffickers ("coyotes") and pay between 5.5 and 12.5 thousand US dollars depending on the place of destination, route, and transportation method.[9] McKenzie & Rapoport (2007) show that migration networks can overall lower the costs of emigration and present evidence from Mexico.[10]

We estimate the model combining detailed micro and aggregate data available for Guatemala. These include population and agricultural census microdata, national household survey data, and price data. We also exploit rich subnational data—including socioeconomic, accessibility, institutional, cultural, climatic, and insecurity indicators— compiled from multiple sources to evaluate their association with agricultural distortions across areas. This analysis can help inform policies aimed at reducing agricultural market imperfections, enhancing local productivity and welfare, and mitigating emigration.

Theoretically, our model identifies two primary channels through which distortions operate. First, higher distortions influence the selection of emigrants by increasing the emigration probability of more productive individuals, while reducing it for less productive ones. This worsens the composition of the remaining workforce and lowers aggregate productivity. We refer to this as the migration channel contributing to reduced productivity. Second, higher distortions reduce productivity directly by misallocating labor and land across sectors and regions. The resulting decline in household incomes—both in agriculture and non-agriculture—increases the incentive to emigrate from both rural and urban areas. We refer to this as the productivity channel contributing to increased emigration.

Quantitatively, our model estimation reproduces the heterogeneity in agricultural distortions observed in the data, both across and within Guatemalan regions. Notably, the regions with higher distortions tend to be the less developed ones. Our counterfactual exercises of reducing the estimated distortions to benchmark scenarios within each region,

[8] Ceballos & Hernandez (2020) show that receiving remittances and being in communities with a higher rate of emigrants are highly correlated with the likelihood of emigration in Guatemala.

[9] `https://news.gallup.com/podcast/356501/exporting-people-central-america.aspx` (accessed September 2025).

[10] While the overall costs of migrating can vary between documented and undocumented migration, formally distinguishing between these two types of migration is beyond the scope of the study.

which result in a 30.3% aggregate reduction in efficiency loss, indicate an average decline of 2.3 percentage points (p.p.) in the share of Guatemalan emigrants.[11] Further, total agricultural productivity increases by 30.1% and median household welfare rises by 4.5%. Hence, reducing distortions and misallocation in agricultural markets can lead to substantial reductions in cross-border migration while increasing productivity and general well-being. An analysis at the sub-national level indicates that regions with greater distortions tend to be more isolated, with limited financial penetration and a weaker government presence.

The study contributes to several strands of the literature. First, the study relates to the literature on the root causes of emigration in developing countries, focusing on modeling and quantifying the role of agricultural distortions as an important push factor. While there have been important advances in the theoretical and empirical analysis of international migration (see, e.g., Borjas, 1999, for an early review of the economics of immigration), more work is needed to better understand the recent and increasing emigration flows. We develop a theoretical framework and exploit rich data to assess how the resulting misallocations from agricultural distortions contribute to internal employment decisions and emigration patterns in Guatemala. The modeling also permits to make some inference about positive or negative selection into migration by characterizing whether the most or least productive agricultural and non-agricultural workers emigrate. The evidence on selection and emigration in the developing world is still mixed (e.g., Chiquiar & Hanson, 2005; Orrenius & Zavodny, 2005; Borjas, 2008; Grogger & Hanson, 2011; Moraga, 2011; Clemens, 2020; Clemens, 2022).

In a related vein, the study adds to the broader literature on the factors that affect labor mobility. Recent work includes Bryan et al. (2014) that examine risks and monetary costs in the decision to migrate internally in Bangladesh; Brueckner & Lall (2015) that review the implications of tenure insecurity on rural-urban migration in developing countries; Munshi & Rosenzweig (2016) on the effect of rural insurance networks in male internal migration in India; and Morten & Oliveira (2023) and Asher & Novosad (2020) on the impact of transportation (road) infrastructure on migration in Brazil and India, respectively. In the agricultural sector, De Janvry et al. (2015) show that delinking land rights from land use

[11] By efficiency loss, we mean the difference between the aggregate agricultural production in a frictionless (distortion-free) environment and the actual total agricultural production under existing distortions.

(through the provision of ownership certificates) resulted in large-scale adjustments to labor and land allocations in Mexico and induced migration by reducing the opportunity cost of migrating. Ngai et al. (2019) and Adamopoulos et al. (2024) evaluate the role of insecure property rights as an important labor mobility friction in China for the reallocation of labor from agriculture to non-agriculture sectors as well as from rural to urban areas. While in our setup we allow for more general agricultural distortions (beyond land market frictions) and portray misallocations across sectors and regions within the country, we pay special attention to external migration given its importance for the context of Guatemala.

The study is also linked to the literature examining aggregate productivity and migration returns in developing countries. Recent studies focusing on internal migration include Bryan & Morten (2019), Lagakos et al. (2020), and Lagakos et al. (2023). Our paper is equally related to a growing literature on agricultural productivity and factor misallocation. Some examples include Gollin et al. (2014), Adamopoulos & Restuccia (2014), Chari et al. (2021), Adamopoulos et al. (2022), Britos et al. (2022), Chen et al. (2022), Chen et al. (2023), and Acampora et al. (2025). We illustrate land and factor misallocation in the agricultural (and non-agricultural) sector and its implications for external migration.

Lastly, the study contributes to the general discussion of migration drivers in Central America to the US, which has recently received special attention in the policy arena. Hanson et al. (2023) provides a historical perspective of US immigration from Latin America and shows that while emigration from Mexico to the US has slowed down since 2000 and decreased after 2010, emigration from Central America (and South America) has continued its upward trend mainly driven by people seeking better economic opportunities.[12] Besides quantifying the role of agricultural distortions on emigration, we assess the association of these distortions with other factors. Despite there are several studies characterizing emigrants from Central America to the US and the possible factors driving their decisions, these assessments are generally based on (limited) cross-sectional surveys and interviews that allow to approximate correlations rather than causality or provide anecdotal evidence (e.g., Cohn et al., 2017; CRS, 2019b, CRS, 2019a; CAI, 2019; NIF, 2019; Soto et al., 2021). Two quantitative studies that focus on specific migration trigger factors, and include Guatemala in their study sample, are Mahajan & Yang (2020) and Clemens (2021).

[12] See Abramitzky & Boustan (2017) for a broader overview of immigration in the US.

The former finds that hurricanes increase migration to the US and existing migrant networks in the US play an important function in hosting migrants after an extreme weather event in their home country. The latter shows that homicides in the Northern Triangle (El Salvador, Guatemala, and Honduras) increase undocumented migration to the US, measured through child-migrant apprehensions, while peer and family networks in the US can produce self-reinforcing migration waves after violence episodes.[13]

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework. Section 3 calibrates the base model for Guatemala following a two-stage approach. Section 4 discusses the estimation results, paying special attention to the role of agricultural distortions in the observed emigration patterns and local productivity and welfare, and examines the association of these distortions with key observable characteristics. Section 5 concludes and provides some policy recommendations.

2 Theoretical framework

Consider a two-sector economy with a unit measure of individuals distributed across J rural (agricultural) regions and one urban (non-agricultural) location. The agricultural good is produced using land in rural regions, where each region j faces a distinct level of agricultural distortions. Rural individuals are endowed with an idiosyncratic managerial ability to produce in the agricultural sector (i.e., farm the land) and an idiosyncratic ability to work in the non-agricultural sector. The non-agricultural good is produced using labor in the urban region, denoted by $j = u$, which operates without distortions. Thus, the subscript j identifies both rural and urban regions such that $j = 1, \ldots, J + 1$. Individuals from rural regions may internally migrate to the urban region and work in the non-agricultural sector. Similarly, individuals from either rural or urban regions can migrate to the rest of the world, receiving an exogenous income that is independent of their idiosyncratic abilities.

We assume that farmers face a production tax that summarizes the distortions in the agricultural sector across regions. These distortions may result from market regulations and interventions, information asymmetries, operational costs for implementing effective managerial control, among other transaction costs and factors discussed in the literature

[13]More recently, Barbosa-Alves & Britos (2025) find that in Guatemala, hotter-than-usual years reduce rural migration by lowering farm productivity and tightening liquidity constraints.

(see, e.g., Britos et al., 2022; Acampora et al., 2025) and are denoted by τ_{ij}, where i indicates the farmer and j the region. Based on their ability draw, an individual born in rural region j chooses among three options: (i) becoming a farmer in region j, in which case they use land to produce the agricultural good, (ii) migrating to the urban region to work in the non-agricultural sector, or (iii) emigrating, provided that the welfare from emigrating exceeds that of staying. The welfare of staying depends on the welfare received either as a farmer in the rural region or as a worker in the non-agricultural sector in the urban region.[14] An individual born in the urban region u chooses between: (i) working in the non-agricultural sector, or (ii) emigrating, provided that the welfare from emigrating exceeds that of staying.

2.1 Agricultural sector

Each farmer i in rural region j is endowed with ability z_{ij}^a and produces the agricultural good with the following technology:

$$y_{ij} = A^a z_{ij}^a l_{ij}^{\alpha}, \tag{1}$$

where A^a is an aggregate productivity parameter in the agricultural sector, l_{ij} is the farmer's land allocation, and $\alpha \in (0, 1)$.

The market is assumed to be competitive and the profit maximization problem for farmer i in region j is defined as:

$$\pi(z_{ij}^a, \tau_{ij}) = \max_{l_{ij}} \quad \tau_{ij} p A^a z_{ij}^a l_{ij}^{\alpha} - q_j l_{ij},$$

where p is the economy's relative price of the agricultural good and q_j is the rental price of a unit of land in region j – both in terms of the non-agricultural good.

[14] Assuming that an individual born in region j can only farm land in region j could be regarded as a restrictive assumption but it simplifies the solution of our model and, as commented below, this assumption is consistent with the fact that farmer mobility across rural regions in Guatemala is mostly temporary and relatively low. Below, we still relax this assumption by allowing individuals to work in the agricultural sector either in their own rural region or in another rural region.

2.2 Non-agricultural sector

The non-agricultural good is produced with labor and a linear technology. The aggregate non-agricultural output is given by:

$$Y^n = A^n \sum_j \int_0^1 \mathbb{W}_{ij} z_{ij}^n \, di, \tag{2}$$

where $\mathbb{W}_{ij}$ is an indicator variable that takes the value of 1 if individual i from region j becomes a worker in the non-agricultural sector and 0 otherwise, z_{ij}^n indicates the individual's ability to work in this sector, and A^n is an aggregate non-agricultural productivity parameter.

The market for the non-agricultural good is also assumed to be competitive. In equilibrium, the wage paid in this sector to each individual i from region j is equal to $w_{ij}^n = A^n z_{ij}^n$.

2.3 The household's domestic problem

Each household consumes two goods: an agricultural good and a non-agricultural good. Preferences are defined over both categories of consumption, with a subsistence requirement in the agricultural good. The non-agricultural good is taken as the numeraire, and all prices are expressed in units of this good.

Conditional on staying in the domestic economy, the utility maximization problem for individual i born in region j is defined as:

$$\begin{aligned} U_{ij}^d = \max_{c_{ij}^a, c_{ij}^n} \quad & \omega \log(c_{ij}^a - \overline{a}) + (1-\omega)\log(c_{ij}^n) \\ \text{s.t.} \quad & p c_{ij}^a + c_{ij}^n \leq I_{ij} + T, \end{aligned}$$

where $\overline{a} \geq 0$ is the subsistence level of consumption of the agricultural good, I_{ij} is the household income level, and $\omega \in (0,1)$ captures the preference weight on agricultural consumption. T refers to per capita transfers received by the household in terms of: (i) transfers from land rentals accrued by the government, who is assumed to be the land owner, which are distributed equally among individuals,[15] and (ii) lump-sum transfers of

[15] Farmers are assumed to rent the land they operate from the government such that we abstract from the implications of ownership distribution (see., e.g., Adamopoulos & Restuccia, 2014 and Britos et al., 2022).

tax revenues received by the government due to the (distorsionary) idiosyncratic tax on the agricultural output that are distributed equally among individuals. The per capita transfers from land rentals by the government, which are equal to $\sum_j q_j L_j/\mathbb{S}$, depend on the equilibrium rental price of land in each region (q_j), the aggregate land size at each region (L_j), and the number of individuals that stay in the domestic economy ($\mathbb{S}$).

Given the relative price p and income level I_{ij}, the solution to the household's maximization problem imply the following expressions for consumption allocations:

$$c_{ij}^a = \frac{\omega(I_{ij}+T)}{p} + (1-\omega)\bar{a}, \tag{3}$$

$$c_{ij}^n = (1-\omega)(I_{ij}+T) - (1-\omega)p\bar{a}. \tag{4}$$

For households born in the urban region, their income level at the domestic economy (I_{ij}) is equal to w_{ij}^n. For households born in a rural region, their income level at the domestic economy (I_{ij}) is a function of the occupational decision, which also encompasses an internal migration (rural-urban) choice. An individual born in rural region j chooses between being a farmer ($\mathbb{F}_{ij} = 1$) and receive profits $\pi(z_{ij}^a, \tau_{ij})$, or migrate internally to work in the non-agricultural sector ($\mathbb{W}_{ij} = 1$) and receive wage w_{ij}^n. The household income level is given by:

$$I_{ij} = \begin{cases} \pi(z_{ij}^a, \tau_{ij}) & \text{if } \mathbb{F}_{ij} = 1 \\ w_{ij}^n & \text{if } \mathbb{W}_{ij} = 1. \end{cases}$$

As a result, an individual born in rural region j decides whether to migrate to the urban region or stay, based on the following welfare maximization problem:

$$V_{ij}^d(p, I_{ij}, \epsilon_{ij}) = \max_{\mathbb{F}_{ij}, \mathbb{W}_{ij} \in \{0,1\}} \{U_{ij}^d(p, I_{ij}|\text{farmer}) + \epsilon_{ij}^f \ , \ U_{ij}^d(p, I_{ij}|\text{worker}) + \epsilon_{ij}^w\}.$$

Finally, the utility of an individual born in the urban region that stays in the domestic economy is equal to $V_{ij}^d(p, I_{ij}, \epsilon_{ij}^u) = U_{ij}^d(p, w_{ij}^n|\text{worker}) + \epsilon_{ij}^u$. We assume that ϵ_{ij}^f, ϵ_{ij}^w, and ϵ_{ij}^u are i.i.d. random variables across i and j that follow a Type-I extreme value (Gumbel) distribution, with mean zero and scale parameter $\sigma_{\epsilon j}$. Following the quantitative migration literature (e.g., Caliendo et al., 2019 and Lagakos et al., 2023), these additive

While addressing land ownership could be relevant in other contexts, it is not central to our analysis.

idiosyncratic taste shocks capture unobserved determinants of location and occupation choice. While they are often interpreted as reflecting non-economic motivations—such as family decisions, or aversion to relocation—they may also encompass household-specific economic considerations not explicitly modeled, such as access to local support systems and services within the domestic economy.

2.4 The migration problem

Let the utility of an individual who emigrates be defined as:

$$V_{ij}^{e}(U_{j}^{e}, \epsilon_{ij}^{e}) = U_{j}^{e} + \epsilon_{ij}^{e},$$

where U_j^e is the exogenous, deterministic component of the utility derived from emigrating from region j, interpretable as a net foreign wage after subtracting migration costs. We allow this component to be region-specific, as migration costs plausibly depend, for example, on the size of regional migrant networks (McKenzie & Rapoport, 2007). The idiosyncratic term ϵ_{ij}^e is mean-zero, i.i.d., and Gumbel distributed (Type I extreme value), as described above; it captures unobserved factors including non-pecuniary considerations (e.g., emotional support, destination familiarity) and economic benefits not modeled explicitly (e.g., likelihood of receiving remittances from relatives abroad).

An individual's decision on whether to emigrate or stay in the domestic economy is ultimately determined by the following maximization problem:

$$V_{ij}(p, I_{ij}, \epsilon_{ij}, U_{j}^{e}, \epsilon_{ij}^{e}) = \max\{V_{ij}^{d}(p, I_{ij}, \epsilon_{ij}) \ , \ V_{ij}^{e}(U_{j}^{e}, \epsilon_{ij}^{e})\}.$$

Lastly, we denote as $\mathbb{E}_{ij}$ the indicator variable that takes the value of one if the individual emigrates and zero if the individual stays in the country either as a farmer or as a non-agricultural worker. Consequently, the number of individuals, $\mathbb{S}$, that stay in the domestic economy is defined as $\mathbb{S} = 1 - \sum_j \mathbb{E}_j$, where $\mathbb{E}_j \equiv \int_0^1 \mathbb{E}_{ij} \, di$ is the share of emigrants from region j.

2.5 Equilibrium

A *competitive equilibrium* is defined as a set of prices: $\{p, q_j\}$ for $j = 1, \ldots, J$ and $w^n_{ij} = A^n z^n_{ij}$ for all pairs (i, j) with $j = 1, \ldots, J+1$; a set of occupational choices: $\{\mathbb{W}_{ij}, \mathbb{F}_{ij}\}$ for all (i, j) with $j = 1, \ldots, J+1$; a set of emigrate-stay choices: $\{\mathbb{E}_{ij}\}$ for all (i, j); a set of land allocations for farmers in each rural region: $\{l_{ij}\}$; and a set of consumption allocations for all households: $\{c^a_{ij}, c^n_{ij}\}$, given the set of agricultural distortions: $\{\tau_{ij}\}$, such that the markets clear.

The market clearing condition for land is given by:

$$\int_0^1 \mathbb{F}_{ij} l_{ij} \, di = L_j; \quad \text{for } j = 1, \ldots, J. \tag{5}$$

Given land rental price q_j, relative price p, and $\{z^a_{ij}, \tau_{ij}\}$ for $j = 1, \ldots, J$, the land allocation of each farmer i in region j is expressed by:

$$l_{ij} = \left(\frac{p A_{ij}}{q_j} \right)^{\frac{1}{1-\alpha}}, \tag{6}$$

where $A_{ij} \equiv \alpha A^a z^a_{ij} \tau_{ij}$.

From expressions (5) and (6), we have the following equilibrium relationship:

$$q_j = p \left(\frac{\int_0^1 \mathbb{F}_{ij} A_{ij}^{\frac{1}{1-\alpha}} \, di}{L_j} \right)^{1-\alpha}.$$

The market clearing condition for non-agricultural labor is given by:

$$\sum_j \int_0^1 \mathbb{W}_{ij} \, di = \mathbb{W}, \quad \text{for } j = 1, \ldots, J+1.$$

The market clearing condition for the agricultural good is defined as:

$$\sum_j \int_0^1 c^a_{ij} \, di = \sum_j \int_0^1 \mathbb{F}_{ij} y_{ij} \, di.$$

Lastly, the market clearing condition for the non-agricultural good is given by:

$$\sum_j \int_0^1 c_{ij}^n \, di = \sum_j \int_0^1 A^n \mathbb{W}_{ij} z_{ij}^n \, di.$$

From the market clearing conditions for the agricultural and non-agricultural good, we have the following expression that characterizes the equilibrium:

$$p[Y^a - \mathbb{S}\bar{a}] = \left(\frac{\omega}{1-\omega}\right) Y^n,$$

where $Y^a \equiv \sum_j \int_0^1 \mathbb{F}_{ij} y_{ij} di$ denotes the aggregate agricultural production across regions, and $Y^n \equiv \sum_j \int_0^1 A^n \mathbb{W}_{ij} z_{ij}^n \, di$ denotes the aggregate non-agricultural production.

2.6 Model mechanisms

We now turn to explore how agricultural market distortions affect emigration and production patterns across the distribution of abilities in each region. We assume that τ_{ij} is negatively correlated with agricultural ability z_{ij}^a—that is, more productive farmers tend to face lower values of τ_{ij}. This relationship reflects that policies or frictions distorting agricultural production disproportionately burden higher-ability farmers—for instance, by imposing greater barriers to scale or by exposing them more acutely to inefficient market institutions (Adamopoulos & Restuccia, 2014; Adamopoulos et al., 2024).[16]

Our model highlights two primary channels through which these distortions operate:

1. *Migration (or micro) channel* contributing to reduced productivity: Distortions disrupt the alignment between individual abilities and sectoral allocation, altering the selection of emigrants. In particular, higher distortions increase the likelihood that more productive individuals choose to emigrate, while reducing that probability for less productive individuals. Holding the total emigration rate constant, this results in a more positively selected emigrant pool compared to a scenario without distortions. The reallocation of higher-ability individuals abroad lowers aggregate domestic productivity, representing an additional channel beyond the usual productivity loss from

[16]See also Restuccia & Rogerson (2008), Bento & Restuccia (2017), and Restuccia & Rogerson (2017) for a broader discussion.

misallocation alone.

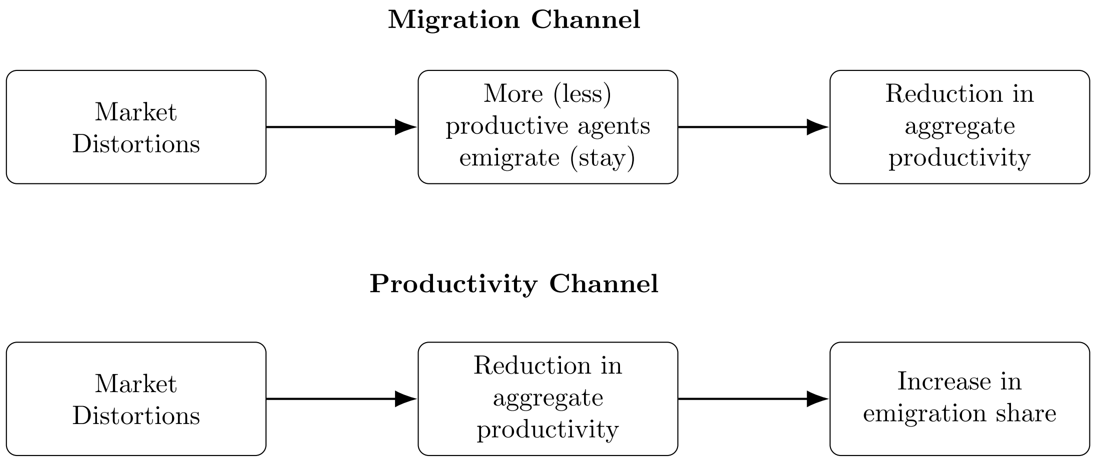
This channel illustrates how distortions increase the emigration probability of more able individuals while reducing that of less productive ones, thereby lowering aggregate productivity. It captures how distortions alter the composition of emigrants: in high-distortion regions, selection into emigration becomes more strongly positive, as the most productive individuals are more likely to leave the agricultural sector.

2. *Productivity (or macro) channel* contributing to increased emigration: Agricultural distortions lead to factor misallocation within the sector, which reduces output and lowers household incomes in both rural (agricultural) and urban (non-agricultural) areas. The resulting decline in domestic incomes increases the relative attractiveness of opportunities abroad, thereby raising the overall emigration rate.

   In contrast to the first channel, which concerns *who* migrates, this channel emphasizes how lower productivity and incomes increase the overall likelihood of migration. To isolate the contribution of this channel to emigration, consider the following exercise. First, we simulate a closed-economy scenario in which emigration is not allowed, and compute the resulting decline in aggregate productivity due solely to misallocation. Second, in a baseline model without distortions (i.e., assuming $\tau_{ij} = 1$ for all i and j), we replicate the same productivity loss by appropriately reducing the agricultural productivity parameter A^a. Comparing the share of emigrants across these two scenarios removes the influence of changes in the distribution of emigration probabilities induced by distortions. This procedure allows us to recover the component of emigration driven solely by the income effects of pure misallocation.

The two model mechanisms are schematically summarized in Figure 1.

Figure 1: Outline of model mechanisms



In formal terms, let p_i^k denote the probability that individual i chooses option k, where $k \in \{f, w, e\}$ corresponds to being a farmer, a non-agricultural worker, or an international migrant, respectively. Assuming idiosyncratic utility shocks that are independently and identically distributed with a Gumbel distribution, the choice probabilities follow a multinomial logit structure:[17]

$$p_i^k = \frac{\exp\left(\frac{U_i^k}{\sigma_\epsilon}\right)}{\sum\limits_{h \in \{f,w,e\}} \exp\left(\frac{U_i^h}{\sigma_\epsilon}\right)},$$

where U_i^k represents the utility that individual i derives from alternative k and $\sigma_\epsilon > 0$ is the scale parameter of the extreme value distribution. The exact expressions for U_i^k and the interpretation of σ_ϵ are provided in Subsections 2.3 and 2.4.

We argue that, as distortions increase, the migration channel implies that the probability of emigrating (p_i^e) increases among more productive agents and decreases among less productive ones. To illustrate this, consider two identical regions that differ only in their degree of distortions: one with high distortions ($j = H$) and the other with low distortions ($j = L$). We compare the probability of emigrating in each region under two scenarios: the distortionary case and the benchmark (no distortions) case.

The following theorems illustrate how distortions affect emigration probabilities and incomes. The proofs are detailed in Appendix A.

[17] See, for example, Iskhakov et al. (2017) and the references cited therein.

Theorem 1 (Distortions reduce (increase) emigration of less (more) productive agents)**.** *Let p^e_{ij} denote the probability that an agent with agricultural productivity z^a_{ij} in region j chooses to emigrate in the presence of distortions, and let $p^{e^*}_{ij}$ denote the corresponding probability in the benchmark case without distortions.*

Consider two otherwise identical regions, $j = H$ (high distortions) and $j = L$ (low distortions), and two productivity types: z^a_l (low) and z^a_h (high). Then, for low-productivity agents, the change in their emigration probability satisfies:

$$p^e_{lH} - p^{e^*}_{lH} < p^e_{lL} - p^{e^*}_{lL}.$$

Equivalently, since $p^{e^*}_{lH} = p^{e^*}_{lL}$, we have:

$$p^e_{lH} - p^e_{lL} < 0.$$

Similarly, for agents with high productivity z^a_h, the change in emigration probability satisfies:

$$p^e_{hH} - p^{e^*}_{hH} > p^e_{hL} - p^{e^*}_{hL}.$$

Or, equivalently:

$$p^e_{hH} - p^e_{hL} > 0,$$

since $p^{e^*}_{hH} = p^{e^*}_{hL}$.

Theorem 1 underscores how distortions differentially affect emigration patterns based on productivity levels, leading to varying impacts on emigration decisions between less and more productive individuals.

Theorem 2 (Distortions reduce incomes and raise emigration across all regions and sectors)**.** *Consider an economy with agricultural productivity level A^a. Let $p^e_{ij}(A^a)$ denote the probability of emigration for individual i in region j when agricultural total factor productivity is A^a.*

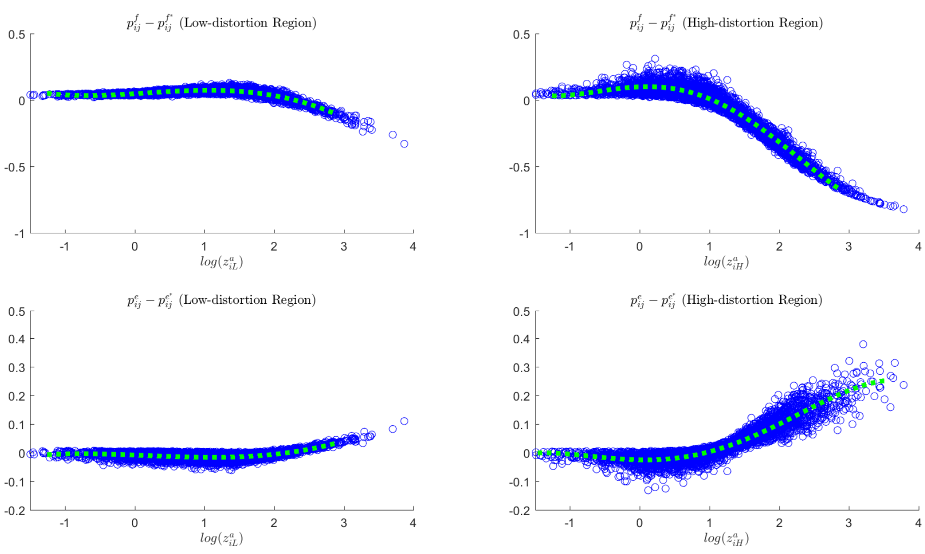
Suppose distortions reduce A^a to a lower level $A^a_\ell < A^a$, reflecting pure misallocation in the absence of selection effects. Then, for all individuals i and regions j, emigration *probabilities increase:*

$$p^e_{ij}(A^a_\ell) > p^e_{ij}(A^a), \quad \textit{for all } i \textit{ and } j.$$

Theorem 2 highlights how distortions affect incomes across regions and sectors, ultimately influencing the overall emigration rate.

Figure 2 illustrates the change in an agent's probability of being a farmer (upper panel) and an emigrant (lower panel) between the distorted and benchmark cases, across the distribution of agricultural abilities (z^a_{ij}) in regions with low versus high distortions (left and right panels, respectively). The analysis uses simulated data, with parameter values consistent with the macro-development literature (detailed in the next section) and assumes—following the arguments above—that the two regions are identical in all respects except for the level of agricultural distortions.

Figure 2: Change in the probability of being a farmer and emigrant in low- and high-distortion regions: Simulated data



Note: The vertical axis displays the change in the probability of being a farmer p^f_{ij} (upper panel) and an emigrant p^e_{ij} (lower panel) between the distorted and benchmark cases across the distribution of idiosyncratic agricultural abilities z^a_{ij} in regions with low versus high distortions (left and right panels). The horizontal axis represents the $\log(z^a_{ij})$. The green dashed line shows a polynomial fit to the observed data.

The upper left panel corresponds to a region with low distortions, showing minimal changes in the probability of being a farmer (p^f_{ij}) for most of the ability distribution but slight reductions in the higher ability range. The upper right panel corresponds to regions with high distortions, showing more pronounced reductions in the probability of being a farmer as abilities increase, indicating that high distortions discourage farming, particularly among more productive agents. The bottom left panel (low distortions) shows small increases in the probability of emigration (p_{ij}^{e}), particularly at higher ability levels, but overall changes remain modest. The bottom right panel (high distortions) shows, in turn, some decrease in emigration probabilities among agents with low to medium levels of ability, while displaying an important increase in emigration probabilities among individuals with higher abilities, indicating that high distortions strongly encourage emigration, particularly among more productive agents.

When we consider both channels together, we find that distortions affect overall emigration in a way that is not necessarily monotonic.

On the one hand, the migration channel implies that distortions lower the emigration probability of less productive agents, who are relatively more likely to emigrate, while raises the emigration probability of more productive agents, who are relatively less likely to emigrate. This channel tends to reduce the overall emigration rate by encouraging those more likely to leave to stay.

On the other hand, the productivity channel implies that distortions reduce aggregate productivity and, consequently, incomes across the entire distribution. This income reduction increases the likelihood of emigration across all groups, generally raising the overall emigration rate.

Interestingly, the first mechanism dominates at lower levels of distortion, while the second mechanism becomes dominant at higher levels. As distortions increase, the wealth effect reduces the income for everyone, raising the probability of emigration across the entire distribution.

2.7 Model extension

In Appendix B, we extend the model to incorporate the possibility that an individual can work in the agricultural sector as a wage earner in any rural region. An individual in this context can choose between being a farmer in their own region, an agricultural worker in any of the J regions, migrate to the urban region to become a non-agricultural worker, or emigrate. This setting permits a more comprehensive modeling and evaluation of the implications of agricultural distortions on observed external and internal (rural-rural) migration patterns. We do not estimate this extended setting but instead rely on the benchmark model introduced above due to data limitations, particularly regarding rural-to-rural migration. Data on these internal migration flows are only partially available and rural-to-rural migration is mostly a temporary rather than a permanent decision in Guatemala. We still show in the appendix that the mechanisms observed in the benchmark model—namely, the migration channel and the productivity channel—are preserved in the extended model.

3 Model estimation

In this section, we describe our estimation strategy. First, we briefly describe the different regions of Guatemala considered for the estimation exercise. Second, we set values for specific parameters based on (i) direct observations from the data, (ii) established standards in the literature, and (iii) external estimates. Third, we parameterize the distribution of idiosyncratic abilities (z_{ij}^a, z_{ij}^n) following a similar strategy as in Britos et al. (2022). Finally, we use indirect inference to estimate the remaining parameters using the Simulated Method of Moments (SMM). Appendix C provides additional details on the data sources and calibration.

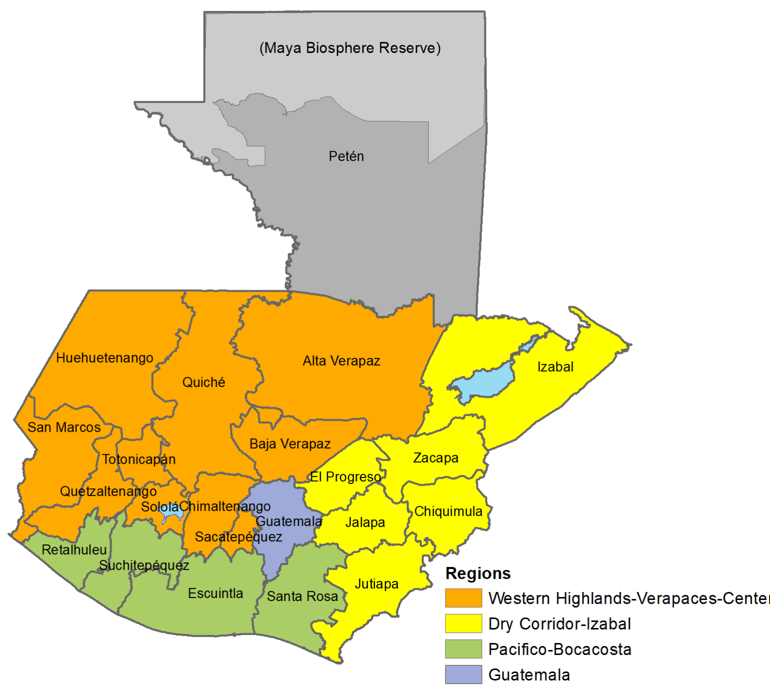
3.1 Regional division of Guatemala

Guatemala presents an important level of heterogeneity in terms of geography, weather, ethnic composition, and rural development. Agricultural activities are, in turn, mainly characterized by small-scale farming, where 65% of the landholdings have less than one hectare and only over 3% have more than ten hectares (INE, 2003; Durr, 2016). The main crops produced include maize, coffee, sugar cane, bananas, and beans, which together generate close to 80% of the total agricultural employment (MAGA, 2011; MAGA, 2013), while most of agricultural activities are informal as noted earlier.

For the analysis, we group the departments of Guatemala into three regions (see map in Figure 3): Western Highlands-Verapaces-Center comprising ten departments; Dry

Corridor-Izabal comprising six departments; and Pacifico-Bocacosta comprising four departments. We additionally consider the department of Guatemala as exclusively urban since the capital city is situated in this department and agricultural activities are significantly less prevalent compared to other sectors. We exclude the department of Peten from the study as natural grasslands and forest make up most of the department area (the Mayan Biosphere Reserve is also located in this department) and agricultural activity is very limited.

Figure 3: Map of Guatemala and regions



As shown by the ANOVA results presented in the last column of Appendix Table D.1, the departments within each region share more similar characteristics than across regions in terms of accessibility, climate vulnerability, government presence, cultural, and socioeconomic conditions. The indicators reported include travel time to the closest town of 20,000 habitants, risk of frosts, droughts, floodings, and geodynamic and geophysical disasters, density of public institutions, prevalence of indigenous population and Spanish as the main language reported by the household head, poverty rate, and chronic malnutrition rate, which are obtained from multiple data sources detailed in the appendix. Note that the

Western Highlands-Verapaces-Center region exhibits an important presence of indigenous population, a high risk of frosts, and is the poorest and with the highest rate of malnutrition, followed by the Dry Corridor-Izabal region that presents an important risk of droughts and floodings and is relatively less accessible. Pacifico-Bocacosta is more developed and better connected than the other two regions but presents an important risk of geodynamic and geophysical disasters as well as floodings.[18]

3.2 Externally calibrated parameters

We assign parameter values based on direct observations from the data or by referencing established values in the existing quantitative literature. Table 1 provides a summary of these externally calibrated parameters. We set the elasticity of output with respect to land (α) to 0.4, consistent with the macro-development literature that considers α values between 0.3 and 0.5 (see, e.g., Britos et al., 2022; Adamopoulos et al., 2022, 2024).[19] The parameter $\overline{a}$ represents the subsistence level of consumption for the agricultural good and is calibrated to 0.25, which is a value commonly used in other studies (see, e.g., Adamopoulos et al., 2024). Similarly, we calibrate the utility weight on agricultural consumption, ω, to 0.3 such that the agricultural employment share in the model matches Guatemala's, which is 30%.[20] Aggregate land size in each region (L_j) is obtained from the IV National Agricultural Census (INE, 2003) and normalized to be expressed as a percentage of the total land size in the country. Lastly, we normalize the agricultural and non-agricultural aggregate productivity to one.

3.3 Productivity and distortions

We assume that both agricultural and non-agricultural productivity follow a log-normal distribution. Specifically, for each region j, agricultural productivity is modeled as $\log(z_{ij}^a) \sim \mathcal{N}(\mu_j^a, \sigma_j^a)$, while non-agricultural productivity is modeled as $\log(z_{ij}^n) \sim \mathcal{N}(\mu^n, \sigma^n)$.

The estimation of the agricultural productivity measure follows the two-stage approach

[18] Pacifico-Bocacosta is characterized by a larger prevalence of commercial farming relative to the other two regions.

[19] We examine below the sensitivity of our results to alternative values of α (0.3 and 0.5).

[20] For 2013–2023, Guatemalan national accounts indicate that the consumption share of food and beverages (alcoholic and non-alcoholic) ranges from 27.0% to 29.5%, close to the model's 31% agricultural consumption share.

Table 1: Externally calibrated parameters

Parameter	Value	Explanation	Reference
A^n	1	Aggregate non-agricultural productivity	Normalized
A^a	1	Aggregate agricultural productivity	Normalized
α	0.4	Land share in agricultural production	Midrange value in the literature
$\overline{a}$	0.25	Subsistence consumption of agricultural good	Adamopoulos et al. (2024)
ω	0.3	Weight of agricultural consumption in utility	Employment share in agriculture
L_1	0.5	Normalized aggregate land size in Region 1	IV National Agricultural Census
L_2	0.3	Normalized aggregate land size in Region 2	IV National Agricultural Census
L_3	0.2	Normalized aggregate land size in Region 3	IV National Agricultural Census

Note: Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta.

of Britos et al. (2022). Full details are provided in Appendix C.3. We first derive a measure of total factor productivity (TFP) using agricultural census data (INE, 2003) on production and land and labor inputs. We then adjust this productivity measure by regressing the TFP value derived in the first step on a set of control variables not included in the theoretical framework that could influence our productivity measure. These include farmer's education, share of family labor force, use of machinery, equipment, enhanced seeds, fertilizer, pesticides, and irrigation system, if farmer has livestock, and number of different crops produced. We similarly include farmer's age and gender that proxy for the potential ascendancy and advantage position of older and male farmers, which can influence land allocation and input access beyond a farmer's inherent ability.[21] Agricultural productivity is finally calculated as the regression residual adjusted for productivity changes based on age and sex. We also compute the mean and standard deviation of agricultural productivity for each region j, yielding the parameters (μ_j^a, σ_j^a).

We employ a similar regression approach to estimate non-agricultural TFP. A detailed description of the calculation is available in Appendix C.3. We use in this case data from the 2019 National Survey of Labor and Income (INE, 2019) that include workers' non-agricultural income, education, age, and gender, and obtain the regression residual that serves as the non-agricultural productivity measure. We likewise obtain parameters (μ^n,σ^n).

[21]The regression additionally includes location (village) fixed effects to control for unobserved differences across locations that can influence productivity, such as biophysical and accessibility conditions.

To model agricultural distortions, we build on the framework proposed by Adamopoulos & Restuccia (2014), in which the distortions are represented as a function of agricultural productivity. Formally, we assume that agricultural distortions follow the process:

$$\log(\tau_{ij}) = \psi_j z_{ij}^a + \epsilon_{ij},$$

where τ_{ij} characterizes the idiosyncratic distortion faced by producer i in region j, z_{ij}^a represents the agricultural productivity of that producer, and ϵ_{ij} is a noise term following a distribution $\mathcal{N}(0, \sigma_j^\tau)$. The parameter ψ_j captures the correlation between distortions and agricultural productivity within region j. Market interventions or imperfections and other factors that drive factor misallocation are characterized by $\psi_j < 0$ (see, e.g., Adamopoulos et al., 2024), where a larger absolute value of ψ_j is indicative of greater distortions in region j and, consequently, higher levels of inefficiency in land use.

This specification allows us to quantitatively assess the degree of misallocation and its relationship with agricultural productivity across different regions.

3.4 Estimation of remaining parameters

We estimate the remaining parameters, $\{\psi_j, \sigma_j^\tau, U_j^e, \sigma_j^\epsilon\}_{j \in J}$, using the Simulated Method of Moments (SMM). This method identifies the parameter vector that minimizes the discrepancy between the empirical moments observed in the data and the corresponding moments generated by the model simulation. By aligning the simulated moments with the data, the model is calibrated to replicate key empirical regularities.

The choice of moments is informed by their relevance to the parameters being estimated. Parameter ψ_j is identified by matching the correlation, within each region j, between total factor productivity TFP_{ij} and revenue-based total factor productivity $TFPR_{ij}$, defined as in the misallocation literature (e.g., Hsieh & Klenow, 2009). This correlation captures the extent to which distortions are systematically related to productivity. Parameter σ_j^τ is estimated by targeting the standard deviation of $TFPR_{ij}$ for each j. For parameter U_j^e, which represents the utility of emigrating, we target the share of emigrants relative to the total employed population in each region j. These shares provide a direct empirical counterpart to the migration incentives embedded in the model. Finally, parameter σ_j^ϵ is identified by matching the share of household members employed in the non-agricultural sector from region j. See Appendix C for further details on the construction and calculation of these moments, including emigrant shares.

Table 2 presents the targeted moments from the data, the corresponding values generated by the model, and the estimated parameter values. The moments include both the share of emigrants across regions and the correlations between total factor productivity TFP and revenue-based total factor productivity $TFPR$. The discrepancy between the observed data moments and the model-generated moments is relatively small, indicating that the model performs well in capturing key empirical patterns.

Parameters $\{\psi_j, U_j^e\}_{j \in J}$ are intrinsically tied to the targeted moments in the model. Specifically, parameter ψ_j reflects the degree of correlation between total factor productivity and revenue-based total factor productivity within region j. A higher correlation indicates a stronger association between individual agricultural productivity, z_{ij}^a, and distortions, $(1 - \tau_{ij})$. This suggests that regions with larger (absolute) values of ψ_j exhibit a larger alignment between productivity levels and the distortions faced by agents, highlighting the extent of resource misallocation. As shown in the table, the Dry Corridor-Izabal region exhibits the highest level of distortions ($|\psi_2| = 0.09$), followed by the Western Highlands-Verapaces-Center region ($|\psi_1| = 0.07$) and the Pacifico-Bocacosta region ($|\psi_3| = 0.03$). The Dry Corridor-Izabal similarly exhibits the highest dispersion in revenue-based total factor productivity ($\sigma_2^\tau = 0.90$), indicating a significant degree of factor misallocation.

Given that parameter U_j^e captures the utility of emigrating from region j, a higher value of U_j^e implies a greater incentive for individuals to emigrate, which is expected to increase the share of emigrants ($\mathbb{E}_j$) from that region. This relationship directly links the observed migration patterns to the underlying economic drivers represented by U_j^e. As observed, the Western Highlands-Verapaces-Center region exhibits the strongest incentives to migrate, followed by Pacifico-Bocacosta and Dry Corridor-Izabal regions. Interestingly, the estimated utility of emigrating from the Pacifico-Bocacosta region (0.49) is comparable to that of the Dry-Corridor-Izabal region (0.42), despite the latter having a higher share of emigration (13.8 versus 8.7). This disparity may reflect the impact of greater agricultural distortions in the Dry Corridor-Izabal region, which appear to be driving more individuals from rural areas to emigrate.

Table 2: Targeted moments and parameter results

Moments	**Data**	**Model**
Share of Emigrants from Region 1, $\mathbb{E}_1$	16.50	16.50
Share of Emigrants from Region 2, $\mathbb{E}_2$	13.80	14.00
Share of Emigrants from Region 3, $\mathbb{E}_3$	8.70	8.80
Corr$(TFP_1, TFPR_1)$	0.48	0.48
Corr$(TFP_2, TFPR_2)$	0.57	0.57
Corr$(TFP_3, TFPR_3)$	0.50	0.50
S.D.$(TFPR_1)$	0.82	0.82
S.D.$(TFPR_2)$	1.09	1.07
S.D.$(TFPR_3)$	0.95	0.90
Share of Workers from Region 1, $\mathbb{W}_1$	52.20	54.69
Share of Workers from Region 2, $\mathbb{W}_2$	54.90	58.32
Share of Workers from Region 3, $\mathbb{W}_3$	62.50	62.76

Parameter	**Value**
Utility of emigrating from Region 1, U_1	0.61
Utility of emigrating from Region 2, U_2	0.42
Utility of emigrating from Region 3, U_3	0.49
Association between z_{ij}^a and τ_{ij} in Region 1, $\lvert\psi_1\rvert$	0.07
Association between z_{ij}^a and τ_{ij} in Region 2, $\lvert\psi_2\rvert$	0.09
Association between z_{ij}^a and τ_{ij} in Region 3, $\lvert\psi_3\rvert$	0.03
Dispersion of τ_{ij} in Region 1, σ_1^τ	0.70
Dispersion of τ_{ij} in Region 2, σ_2^τ	0.90
Dispersion of τ_{ij} in Region 3, σ_3^τ	0.85
Gumbel scale-parameter Region 1, σ_1^ϵ	1.21
Gumbel scale-parameter Region 2, σ_2^ϵ	1.42
Gumbel scale-parameter Region 3, σ_3^ϵ	0.81

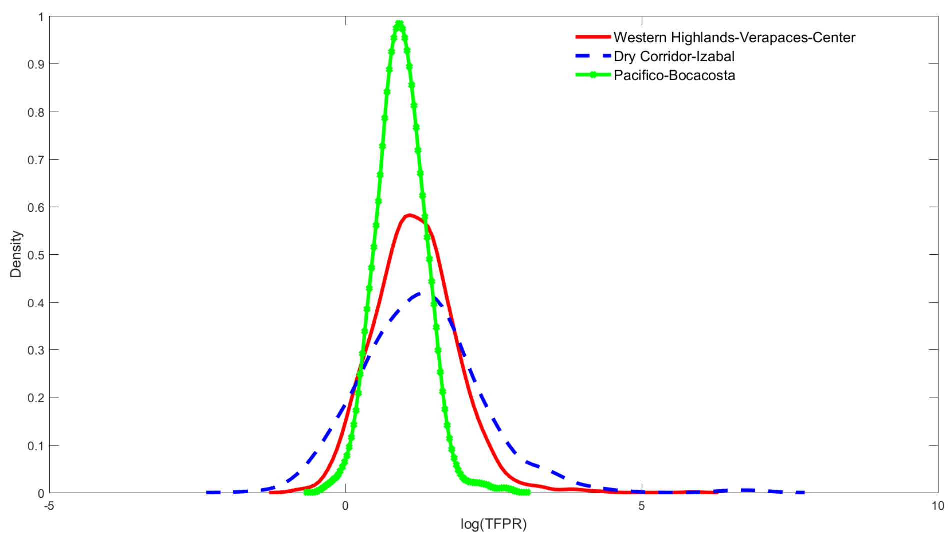
Note: This table displays the targeted moments from the data and their corresponding model-generated counterparts, as well as the values of the estimated parameters. Moments include the share of emigrants, correlations between total factor productivity TFP and revenue-based total factor productivity $TFPR$, standard deviations of $TFPR$, and *Gumbel* scale-parameters, for each region. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta.

Finally, the larger share of non-agricultural workers from the Pacifico-Bocacosta region (63%), compared to other regions (52%–55%), results in an estimated lower value for the

Gumbel scale parameter (0.81 versus 1.2 and 1.4). This lower scale parameter may indicate reduced variance in the (idiosyncratic) decisions influenced by non-economic factors, such as family considerations and specific household networks in the destination country.

Figure 4 illustrates the extent of misallocation across the three regions. The figure displays kernel density estimates of the distribution of $TFPR$ across Western Highlands-Verapaces-Center (red solid line), Dry Corridor-Izabal (blue dashed line), and Pacifico-Bocacosta (green line with markers). The Pacifico-Bocacosta region exhibits the most concentrated distribution, suggesting the lowest level of agricultural market distortions. In contrast, the Dry-Corridor-Izabal region displays the flattest distribution, indicating significant factor misallocation. The Western Highlands-Verapaces-Center region lies in between, with moderate dispersion. These patterns are consistent with the extent of distortions and misallocation highlighted in the estimation results discussed above.

Figure 4: Model-implied distortions: Revenue-based total factor productivity (TFPR) distribution by region



Note: The vertical axis displays the estimated kernel density, while the horizontal axis shows the log(TFPR) for each region.

Appendix Figure D.1 depicts, in turn, the relationship between the actual output ($\log(y_i)$) and the hypothetical no-distortion output ($\log(y_i)$) across the three regions. Each panel corresponds to one region, with red dots representing individual observations. The dashed 45-degree line indicates parity between actual and no-distortion output such that observations closely aligned with the 45-degree line suggest minimal distortions, while deviations from the line highlight the presence of distortions. The Pacifico-Bocacosta region shows the closest alignment with the the 45-degree line, pointing to lower levels of distortions, followed by the Western Highlands-Verapaces-Center, and the Dry Corridor-Izabal that exhibits the greatest scatter, indicating more substantial distortions.

The use of a complementary, although limited, dataset also permits to align these findings with observed differences in agricultural land markets across regions. Appendix Table D.2 reports the share of land rentals and perceived prices of the best land in the locality and their dispersion in each region, based on a three-year panel survey of households collected between 2012 and 2014 over half of the municipalities in the country.[22] We find a significantly larger share of land rentals and a considerably lower coefficient of price variation in Pacifico-Bocacosta compared to the other two regions. This relatively higher number of transactions and lower price dispersion is consistent with less market distortions in Pacifico-Bocacosta, as opposed to the Western Highlands-Verapaces-Center and Dry Corridor-Izabal.

4 Quantitative analysis

In this section, we approximate the magnitude of the estimated distortions in the agricultural sector along several dimensions of interest. We accordingly construct counterfactual scenarios for each region to compare the actual allocation of resources with a benchmark that represents a more efficient allocation. We then quantify changes in key outcomes related to emigration, local employment, agricultural productivity, and welfare. We also evaluate whether the quantified distortions are correlated with observable local characteristics across areas.

4.1 Counterfactual scenario

The procedure to construct the counterfactual scenario is outlined as follows:

[22] The survey was conducted as part of the evaluation of a large-scale program implemented by the Government of Guatemala to reduce food insecurity and malnutrition. It covered 176 of the country's 340 municipalities, focusing on the poorest areas with the highest stunting rates.

1. First, we re-estimate the distortions at the department level to obtain more granular estimates.

2. Second, we identify the most *'efficient'* department within each region. This is done by selecting the department with the lowest correlation between $TFPR_{ij}$ and TFP_{ij}, as a lower correlation indicates a closer alignment between productivity and resource allocation.

3. Third, we re-calibrate the two parameters that characterize the distribution of distortions in our model, ψ_j and σ_j^τ, for each region j. Specifically, given the remaining parameters estimated in Section 3, we calibrate ψ_j and σ_j^τ to match two key moments from our selected benchmarks: the correlation between $TFPR$ and TFP, and the standard deviation of $TFPR$.

4. Finally, using the model outcomes derived from the re-calibrated parameters, we construct the benchmark scenario and use it to perform the counterfactual exercises.

We consider this benchmark selection a more conservative and realistic approach than using the planner's allocation as the benchmark. Table 3 presents the estimated moments that characterize distortions across the selected *benchmark* departments in each region: Quetzaltenango (in Western Highlands-Verapaces-Central), Izabal (in Dry Corridor-Izabal), and Escuintla (in Pacifico-Bocacosta). These three departments show the lowest correlations between total factor productivity (TFP_{ij}) and revenue-based total factor productivity ($TFPR_{ij}$) within their respective regions.[23]

Table 3: Targeted moments from selected *benchmark* departments

Department, Region	Corr (TFP_{ij},$TFPR_{ij}$)	S.D. ($TFPR_{ij}$)
Quetzaltenango, Region 1	0.34	0.57
Izabal, Region 2	0.42	0.72
Escuintla, Region 3	0.41	0.86

Note: This table displays the correlation between TFP_{ij} and $TFPR_{ij}$ and the standard deviation of $TFPR_{ij}$, estimated from the data, for each resulting benchmark department. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta.

Quetzaltenango exhibits a relatively low correlation between TFP and $TFPR$ (0.34) and a standard deviation of $TFPR$ of 0.57, which suggest moderate distortions in factor

[23] Although not reported, these departments are among the most developed ones within each region.

allocation. Izabal has a higher correlation between total factor productivity and revenue-based total factor productivity (0.42) and a standard deviation of the latter of 0.72, indicating a greater degree of misallocation compared to Quetzaltenango. Finally, Escuintla shows a correlation of 0.41 and the highest standard deviation of $TFPR$ (0.86) among the benchmark departments. While the productivity-resource alignment in Escuintla is comparable to that in Izabal, the observed standard deviation of $TFPR$ in Escuintla is closer to the actual standard deviation for the entire Pacifico-Bocacosta region reported in Table 2 (0.95). This proximity suggests that the potential gains from eliminating distortions in the Pacifico-Bocacosta region may not be as large as those expected in other regions.

Table 4 reports the calibration results for the parameters characterizing the distributions of distortions (ψ_j and σ_j^τ) for each region j under the *benchmark* scenario. Compared to the parameters estimated under the actual scenario in Section 3, both ψ_j and σ_j^τ are significantly lower in the benchmark scenario for the Western Highlands-Verapaces-Center and Dry Corridor-Izabal regions; specifically, for the former region, $|\psi_1|$ decreases from 0.07 to 0.03 and σ_1^τ decreases from 0.70 to 0.52, while for the latter region, $|\psi_2|$ declines from 0.09 to 0.05 and σ_2^τ drops from 0.90 to 0.62. In contrast, in the Pacifico-Bocacosta there are only marginal differences between the two scenarios, with $|\psi_3|$ changing from 0.03 to 0.02 and σ_3^τ slightly decreasing from 0.85 to 0.76.

Figure 5 presents kernel density plots of the log of revenue-based total factor productivity ($\log(TFPR)$) for the three regions, comparing the distributions under the *actual scenario* (solid red lines) and the *benchmark scenario* (dashed green lines). For the Western Highlands-Verapaces-Center region, the benchmark scenario shows a narrower distribution (lower variance) and a higher peak compared to the actual scenario, indicating a more concentrated allocation of $TFPR$ values around the mean. Similarly, the benchmark scenario for the Dry Corridor-Izabal region results in a narrower distribution and a higher peak, reflecting significantly reduced distortions relative to the actual scenario. On the contrary, the Pacifico-Bocacosta region displays more closely aligned distributions for the actual and benchmark scenarios, with the benchmark scenario exhibiting only a slightly narrower spread and marginally higher peak, also pointing to less pronounced improvements in resource allocation under the benchmark scenario.

Overall, the plots illustrate that transitioning to the benchmark scenario leads to a

Table 4: Targeted moments and parameter results under the *benchmark scenario*

Moments	**Data**	**Model**
Corr($TFP_1, TFPR_1$)	0.34	0.34
Corr($TFP_2, TFPR_2$)	0.42	0.42
Corr($TFP_3, TFPR_3$)	0.41	0.40
S.D.($TFPR_1$)	0.57	0.58
S.D.($TFPR_2$)	0.72	0.73
S.D.($TFPR_3$)	0.86	0.84

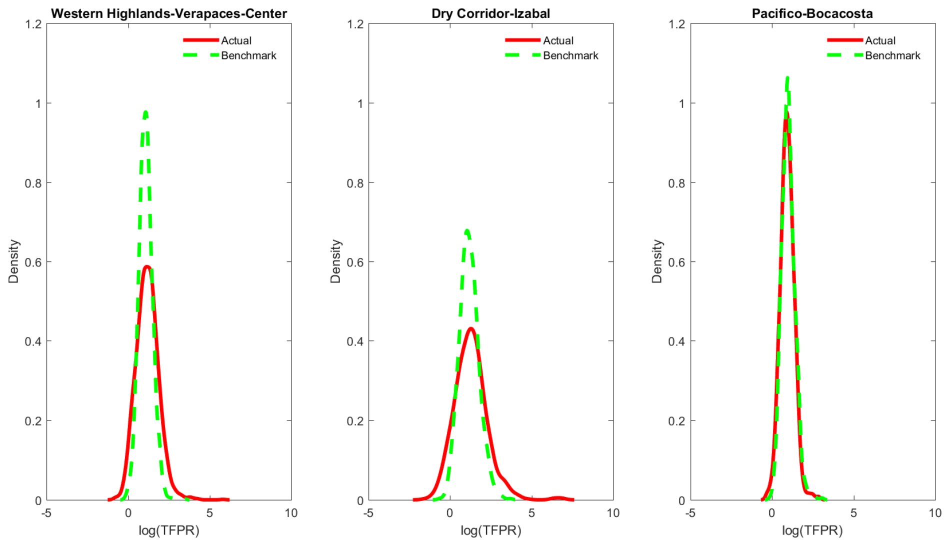
Parameter	**Value**
Association between z_{ij}^a and τ_{ij} in Region 1, $\lvert\psi_1\rvert$	0.033
Association between z_{ij}^a and τ_{ij} in Region 2, $\lvert\psi_2\rvert$	0.048
Association between z_{ij}^a and τ_{ij} in Region 3, $\lvert\psi_3\rvert$	0.023
Dispersion of τ_{ij} in Region 1, σ_1^τ	0.52
Dispersion of τ_{ij} in Region 2, σ_2^τ	0.62
Dispersion of τ_{ij} in Region 3, σ_3^τ	0.76

Note: This table displays the targeted moments from the data on the benchmark departments within their respective regions and their corresponding model-generated counterparts, as well as the values of the calibrated parameters. Moments include the correlations between total factor productivity TFP and revenue-based total factor productivity $TFPR$, and the standard deviations of $TFPR$. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta.

reduced dispersion of $\log(TFPR)$, particularly in the Western Highlands-Verapaces-Center and Dry Corridor-Izabal, indicating a more efficient allocation of resources in these regions, as opposed to Pacifico-Bocacosta with limited gains in resource reallocation efficiency.

Finally, Table 5 displays the reductions in the efficiency loss of aggregate agricultural production when transitioning from the distorted economy to the benchmark case. The first three values represent the decrease in the agricultural production gap for each of the three regions. Specifically, the gap (i.e., efficiency loss) narrows by 37.3%, 41%, and 14.5% in Western Highlands-Verapaces-Center, Dry Corridor-Izabal, and Pacífico-Bocacosta, respectively. The final value represents the overall reduction in the agricultural production gap across all regions, amounting to 30.3%. This indicates that the nationwide reduction in the gap is substantial—close to one-third of the total agricultural gap, which is measured as the difference between efficient (distortion-free) and actual production.

Figure 5: Model-implied distortions: Revenue-based total factor productivity (TFPR) distributions – Actual vs. Benchmark



Note: The vertical axis displays the estimated kernel density, while the horizontal axis shows the log(TFPR) for each region. *'Actual'* refers to the model-implied TFPR distributions derived in Section 3 using the actual regional data, whereas *'Benchmark'* refers to the model-implied TFPR distributions based on data from the selected benchmark departments.

Table 5: Reduction in the agricultural gap between actual and benchmark scenarios

Region 1	Region 2	Region 3	Total
37.3%	41.0%	14.5%	30.3%

Note: This table displays the reduction in efficiency loss (i.e., the difference between aggregate agricultural production in a frictionless (distortion-free) environment and actual production under existing distortions) between actual and benchmark scenarios. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacífico-Bocacosta.

4.2 Counterfactual estimates

Next, based on the counterfactual scenarios for each region, we quantitatively evaluate the effects of the estimated distortions on the emigration rate, local labor decisions, aggregate agricultural productivity, and welfare. We focus on addressing the following questions: How does the share of emigration change between the actual and benchmark scenarios across regions? What is the change in the share of non-agricultural workers? What are the gains in agricultural productivity and efficiency? What are the gains in median consumption under the benchmark economy, and who is most likely to experience increased welfare? How does the probability of emigration change across individuals with varying degrees of abilities? What is the contribution of the migration channel to the modeled changes in productivity?

4.2.1 Overall changes

Table 6 presents the estimated changes in emigration share, non-agricultural employment share, aggregate agricultural productivity, and median consumption when transitioning from the actual to the benchmark scenario in each region. We observe varying results across regions.

The Dry Corridor-Izabal region shows the largest improvement in agricultural productivity, with a 45.6% increase, reflecting the region's substantial potential for factor reallocation efficiency. The Western Highlands-Verapaces-Center region experiences the most substantial reduction in the share of emigrants (-3.0 p.p.), coupled with a modest increase in non-agricultural employment share (1.1 p.p.), implying improved local labor market opportunities. In contrast, the Pacifico-Bocacosta region shows the smallest increase in agricultural productivity (14.4%) but the largest increase in the share of non-agricultural workers (1.7 p.p.), suggesting a focus on labor reallocation rather than land-use efficiency.

The urban region, although not directly affected by agricultural distortions, also sees a notable rise in the share of non-agricultural employment (1.6 p.p.) and a decline in the share of emigrants (1.6 p.p.). This reflects the indirect general equilibrium effects of reducing rural distortions.

In terms of consumption, all regions experience gains, with the largest median consumption increase observed in the urban region (14.6%), followed by the Dry Corridor–Izabal (12.2%) and Western Highlands–Verapaces–Center (11.0%). The Pacifico–Bocacosta region shows the smallest gain, at 8.8%.[24]

On aggregate, the total emigration share decreases by 2.3 p.p., suggesting a decline in out-migration as distortions are reduced, while aggregate agricultural productivity rises by 30.1%. These are non-negligible changes in emigration and agricultural productivity, with the effects being more pronounced in the Western Highlands-Verapaces-Center (emigration) and Dry Corridor-Izabal (agricultural productivity), which are less developed than

[24] The change in consumption is calculated by comparing the median consumption basket in the benchmark scenario (at actual prices) relative to the actual consumption.

Table 6: Changes in emigration, non-agricultural employment, agricultural productivity, and median consumption between actual and benchmark scenarios

Region	Δ Share of Emigrants (p.p.)	Δ Share of Workers (p.p.)	Δ Agricultural Productivity (%)	Δ Median Consumption (%)
Region 1	-3.0	1.1	33.3	11.0
Region 2	-2.2	-0.1	45.6	12.2
Region 3	-1.8	1.7	14.4	8.8
Urban	-1.6	1.6	–	14.6
Total	-2.3	1.2	30.1	12.3

Note: This table presents changes in the emigration share ($\mathbb{E}_j$), the share of non-agricultural workers ($\mathbb{W}_j$), aggregate agricultural productivity, and median consumption when moving from the actual to the benchmark scenario in each region. Changes are expressed in percentage points (p.p.) for emigration and employment shares and in percentage terms (%) for aggregate productivity and median consumption. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta. Total refers to the aggregate, including the three agricultural regions and the urban region.

Pacifico-Bocacosta that shows the largest relative gains in non-agricultural employment. Considering that the total population of Guatemala is 16.3 million according to the last census (INE, 2020) and approximately 1.1 million Guatemalan migrants reside in the US (US Census Bureau, 2021), the overall reduction in emigration would be equivalent to a 35.5% decrease in the Guatemalan population currently living in the US.

In Appendix Table D.3, we assess the sensitivity of our main results to alternative values for the elasticity of output with respect to land (α). Panel A of the table presents outcomes for $\alpha = 0.3$ and Panel B for $\alpha = 0.5$. The direction of the effects across outcomes is consistent with our baseline results. The total change in the emigration share ranges from -2.7 p.p. (when $\alpha = 0.3$) to -1.3 p.p. (when $\alpha = 0.5$), compared to -2.3 p.p. under the baseline calibration. Changes in the share of non-agricultural workers are basically identical across α values (1.2 p.p.). Aggregate agricultural productivity, in turn, increases more markedly with a higher α, rising by 37.1% at $\alpha = 0.5$, compared to 21.7% at $\alpha = 0.3$. Median consumption follows a similar pattern, increasing by 8.4% under $\alpha = 0.3$ and by 11.7% under $\alpha = 0.5$.

4.2.2 Welfare

To assess the welfare gains under the benchmark scenario compared to the actual scenario, we employ an equivalent variation approach in consumption. The equivalent variation is defined as the change in consumption that leaves the individual indifferent between the two cases in terms of utility.[25]

Formally, for each individual i, the equivalent variation is the solution to the following equation:

$$V_{ij}^{act}(\hat{c}_{ij}^{a}, \hat{c}_{ij}^{n}) = V_{ij}^{ben}, \tag{7}$$

where $V_{ij}^{act}(\cdot)$ is the utility level in terms of the actual agricultural and non-agricultural consumption levels adjusted by the equivalent variation coefficient EV_{ij}, and $V_{ij}^{ben}(\cdot)$ represents the corresponding utility under the benchmark scenario. Following the optimal consumption allocations from (3) and (4), the adjusted consumption levels under the actual scenario are defined as $\hat{c}_{ij}^{a} = c_{ij}^{a} + EV_{ij}$ and $\hat{c}_{ij}^{n} = [(1-\omega)/\omega]p(c_{ij}^{a} + EV_{ij} - \bar{a})$.[26]

Using the functional form of the utility function, the equivalent variation EV_{ij} is, then, the solution to:

$$log(c_{ij}^{a} + EV_{ij} - \bar{a}) + (1-\omega)log([(1-\omega)/\omega]p) = V_{ij}^{ben}, \tag{8}$$

yielding:

$$EV_{ij} = e^{V_{ij}^{ben}} \left(\frac{\omega}{p(1-\omega)}\right)^{1-\omega} + \bar{a} - c_{ij}^{a}. \tag{9}$$

This expression captures the individual's equivalent variation in terms of the agricultural consumption good, representing the consumption adjustment needed to equalize utility across the two scenarios.

Using the expression defined above and the estimated model parameters, we find that the median household welfare increases by 4.6%, 4.5%, and 4.3% (in consumption-equivalent terms) in the Western Highlands-Verapaces-Center, Dry Corridor-Izabal, and Pacifico-Bocacosta regions, respectively. When considering all three regions together with the urban area, the welfare gain at the median amounts to 4.5%.

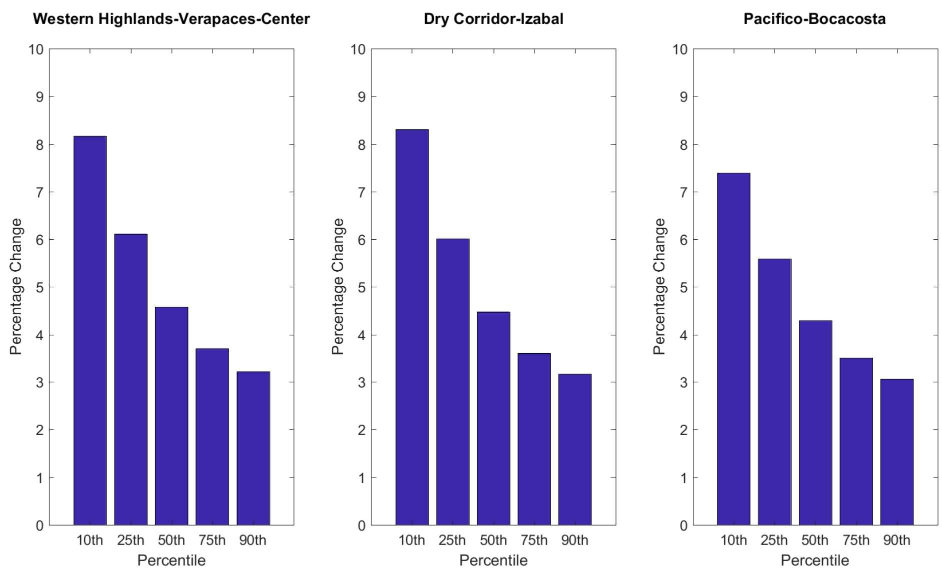
To examine the distributional effects of reducing agricultural distortions to their benchmark level, Figure 6 presents welfare changes across the income distribution within each

[25] The analysis that follows is carried out in partial equilibrium—that is, conditional on the equilibrium prices and allocations from the actual scenario.

[26] For individuals who migrate in the actual scenario, consumption levels are not directly observed. Instead, we impute the level of consumption—at actual scenario prices—that would yield the same utility as their migration outcome. The equivalent variation is then calculated based on this imputed consumption.

region. We observe that lower percentiles experience larger percentage increases in welfare across all regions, with the 10th percentile consistently exhibiting the greatest gains and the 90th percentile the smallest. In fact, the welfare gain at the 10th percentile is more than twice that at the 90th percentile across the three regions. The larger gains shown by individuals at the lower end of the income distribution are similarly more notable in the Dry Corridor-Izabal and Western Highlands-Verapaces-Center regions. These results suggest that reducing agricultural distortions has a relatively greater impact on lower-income groups in less developed regions.

Figure 6: Consumption-equivalent welfare gains by income percentile



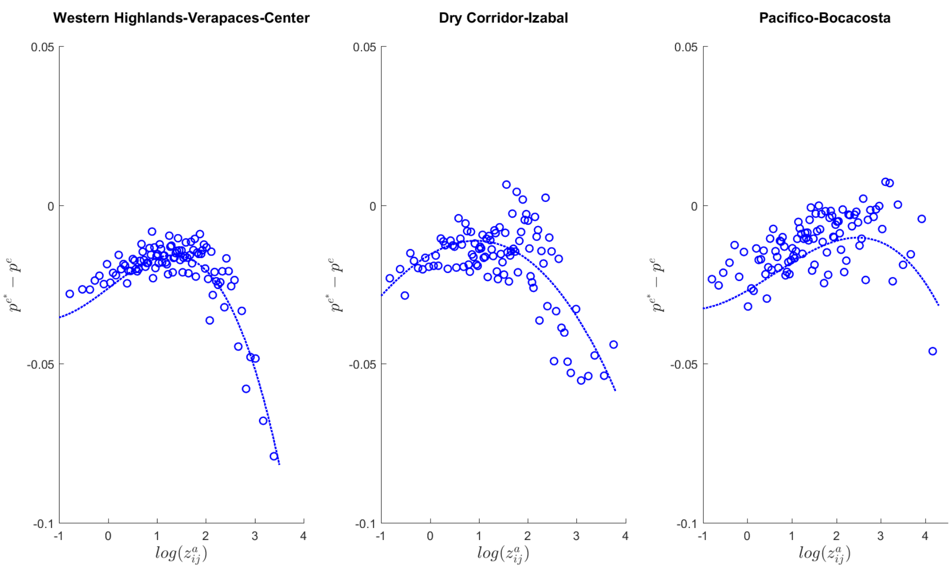
Note: This figure shows changes in welfare measured in consumption-equivalent terms across the income distribution for the 10th, 25th, 50th, 75th and 90th percentiles, comparing the benchmark and actual cases for each rural region.

4.2.3 Probability of emigration

We are also interested in examining the effects of distortions on individual migration incentives, which may vary by the level of ability. In Figure 7, we depict the change in an agent's probability of emigrating between the benchmark and actual cases $(p_{ij}^{e^*} - p_{ij}^{e})$ across percentiles of the distribution of agricultural abilities (z_{ij}^{a}) in each of the three regions. Several interesting patterns emerge from the figure. First, we observe a nonlinear (inverted-u-shaped) relationship between agricultural productivity and changes in migration probabilities, with variations across regions, consistent with our theoretical framework.

Second, among individuals with higher agricultural productivity, the decrease in the emigration probability under the benchmark scenario relative to the actual case is more pronounced in the Western Highlands-Verapaces-Center, particularly on the higher end of the distribution, followed by the Dry Corridor-Izabal and Pacifico-Bocacosta.[27] On average, emigrants' productivity under the benchmark scenario is 11.9%, 12.3%, and 13.9% lower than in the actual case for the Western Highlands-Verapaces-Center, Dry Corridor-Izabal, and Pacifico-Bocacosta regions, respectively.[28]

Figure 7: Change in the probability of emigration across the distribution of abilities



Note: This figure shows scatterplots of the change in the probability of being an emigrant between the benchmark and actual cases ($p^{e^*}_{ij} - p^e_{ij}$), plotted against the percentiles of the log of agricultural productivity ($\log(z^a_{ij})$), for each region: Western Highlands–Verapaces–Center (left), Dry Corridor–Izabal (center), and Pacific–Bocacosta (right), where $p^{e^*}_{ij}$ corresponds to the benchmark scenario and p^e_{ij} denotes the actual case. The dashed line shows a polynomial fit to the data.

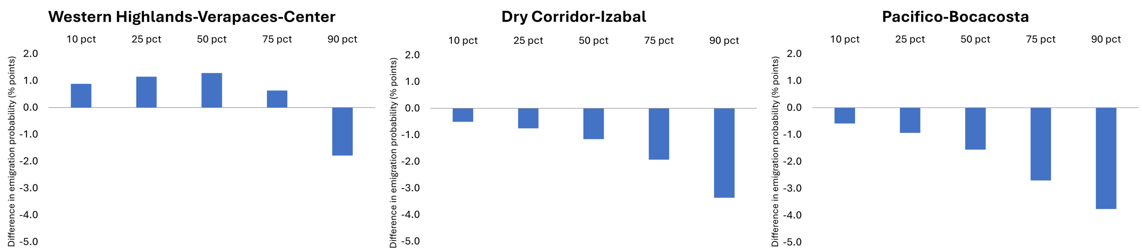
Using data from the Living Standards Measurement Survey in Guatemala (ENCOVI 2014), we can compare the probability of emigration between a household located in the

[27]Appendix Figure D.2 illustrates how changes in emigration probabilities across the ability distribution vary with α. When $\alpha = 0.5$, high-ability individuals experience a sharper decline in emigration probability compared to the case with $\alpha = 0.3$ (or $\alpha = 0.4$ in Figure 7), indicating stronger selection into domestic agriculture once most distortions are removed in the benchmark scenario. This pattern is aligned with the larger productivity and consumption gains reported in Table D.3 under $\alpha = 0.5$. In contrast, migration responses remain limited for individuals with low or intermediate abilities at this value of α.

[28]The same pattern holds when considering the median productivity of emigrants instead of the average, though the declines are relatively smaller. Specifically, the median agricultural productivity of emigrants decreases by 4.8%, 2.4%, and 8.4% in the Western Highlands-Verapaces-Center, Dry Corridor-Izabal, and Pacifico-Bocacosta regions, respectively.

benchmark department versus other departments in each region for specific percentiles of per capita household expenditure.[29] Consistent with the findings above, Figure 8 shows that households with higher expenditures (income), which could be regarded as more productive, typically display a notably lower likelihood of emigration in the benchmark department compared to other departments. In the case of Western Highlands-Verapaces-Center, this lower probability is only observed for the 90 percentile, while in the other two regions the probability differences grow steadily as income increases.[30]

Figure 8: Difference in the probability of emigration between benchmark and other departments by different levels of household per capita expenditure



Note: This figure shows the difference (in percentage points) in the probability of emigration between the benchmark and other departments in each region for different percentiles of household per capita expenditure, using data from the Living Standards Measurement Survey in Guatemala (ENCOVI 2014). The percentiles reported include the 10th, 25th, 50th, 75th, and 90th percentile (pct).

4.2.4 Contribution of *migration* channel

In this subsection, we evaluate the contribution of the migration channel to provide some guidance on the extent to which total productivity losses can be attributed to the emigration of more productive individuals induced by distortions. If emigration is eliminated (for instance, by sufficiently reducing the utility of emigrating across all regions) and the same counterfactual experiments are conducted, aggregate agricultural productivity increases by 29.3% in Western Highlands-Verapaces-Center, 40.9% in Dry Corridor-Izabal, and 14.2%

[29] The probability of emigration is obtained by regressing whether a household reports a recent emigrant (over the past five years) on their annual per capita expenditure and its squared term. Expenditures are adjusted by a geographic factor that accounts for price differences across locations, following the methodology established by the National Institute of Statistics of Guatemala (INE). We perform separate regressions for each region including as additional regressors the interactions of expenditures with an indicator variable to distinguish between the benchmark and other departments and derive the corresponding predicted probabilities for different per capita expenditure percentiles (10, 25, 50, 75, and 90 percentile).

[30] In Appendix Figure D.3 we report the results of a similar exercise comparing predicted probabilities of emigration in Pacifico-Bocacosta with the other two regions for different expenditure (income) percentiles. We likewise find a lower probability of emigration among households with higher income in Pacifico-Bocacosta, which exhibits a lower level of distortions than the other two regions.

in Pacifico-Bocacosta. Compared to the respective gains in aggregate productivity in the baseline case (see Table 6), this suggests that the migration channel accounts for 11.9%, 10.3%, and 1.3% of the total gains in each region, with the remainder driven by 'pure' misallocation.

At the national level, eliminating migration leads to a 27.3% increase in aggregate agricultural productivity, with the migration channel accounting for 9.2% of this gain. This highlights that while distortions affect productivity through both reallocation inefficiencies and migration, the latter plays a non-negligible role in shaping overall productivity outcomes. Table 7 presents a summary of these results.

Table 7: Contribution of the migration channel to productivity gains

Region	Change without Emigration (%)	Migration Contribution (%)
Region 1	29.3	11.9
Region 2	40.9	10.3
Region 3	14.2	1.3
Total	27.3	9.2

Note: This table presents the contribution of the migration channel to agricultural productivity gains. The second column shows the percentage increase in aggregate productivity between the actual and benchmark scenarios when emigration is eliminated. The third column represents the share of total productivity gains attributable to the migration channel, calculated as 100 minus the percentage change in agricultural productivity without emigration relative to the total productivity increase under the baseline. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta. Total refers to the aggregate, including the three agricultural regions.

4.3 Agricultural distortions and observable local characteristics

We now examine whether the quantified distortions in agricultural markets can be associated with observable characteristics across areas that can help provide specific policy recommendations. We work at the municipality level (instead of the department level) to perform a more detailed assessment using a regression framework. In Table 8, we regress the corresponding agricultural distortions (captured by ψ_j) derived at the municipality level on a set of indicators obtained from multiple data sources. We consider two model specifications. In column (1) of the table, we include as regressors contextual variables that may be correlated with agricultural distortions such as socioeconomic, accessibility, institutional, and cultural indicators. We particularly include an occupational precariousness index, per capita bank deposit accounts, mobile base stations (towers) per square kilometer, travel time to the closest town of 20,000 habitants, government (public institutions) density index, and share of indigenous population.[31] In column (2), we add as regressors climate vulnerability and insecurity indicators, specifically an index of natural hazards and rate of extortions, that may additionally be directly correlated with migration decisions.[32] All variables were standardized prior to the regression for comparability purposes. The regressions include regional fixed effects and the reported standard errors are robust.

We find that financial penetration, measured through the number of bank deposit accounts per person in a municipality, is negatively associated with agricultural distortions. Considering that a higher financial penetration and inclusion can contribute to a more formal and transparent economy and to the reduction of economic disparities (Cull et al., 2014; Klapper et al., 2016; Demirguc-Kunt et al., 2017), the limited financial access in certain locations could be playing some role in explaining distortions in agricultural markets.[33] We similarly observe that a lower road accessibility, measured through the average travel time to the nearest town of 20,000 or more habitants, is positively correlated with agricultural distortions, which suggests that reduced accessibility—leading to higher transaction costs—may partly account for the observed distortions. This finding is in line with Britos et al. (2022) who show that less accessible areas are less efficient. Government presence, in turn, measured though the density of public institutions, is negatively associated with agricultural distortions pointing to the importance of the public sector and institutional presence in an area to reduce market inefficiencies, especially in a context of fragile and severely limited government institutions (CRS, 2024). For the remaining indicators, in-

[31]The occupational precariousness index and government density index are obtained from the Food Insecurity and Malnutrition Vulnerability Index (IVISAN) calculated in 2012 by the Secretary of Food Security and Nutrition of Guatemala (SESAN); the per capita bank deposit accounts is obtained from the Superintendent of Banks of Guatemala (SIB) for 2018; the mobile base stations per square kilometer is obtained from the Superintendence of Telecommunications of Guatemala (SIT) for 2018; the travel time to the closest town of 20,000 habitants is obtained from the International Food Policy Research Institute (IFPRI) typology of micro-regions exercise performed in 2021; and the share of indigenous population is obtained from the 2018 Population and Housing Census implemented by the National Institute of Statistics of Guatemala (INE).

[32]The index of natural hazards, which comprises hydro-meteorological, geodynamic and geophysical disasters, is obtained from the National Institute for Seismology, Vulcanology, Meteorology, and Hydrology of Guatemala (INSIVUMEH) for the period 1530-2015 and the rate of extortions is obtained from INFOSEGURA-Guatemala database for 2018.

[33]More generally, Buera et al. (2011) show that financial development can account for significant cross-country differences in aggregate total factor productivity by inducing distortions across production units, which ultimately affect measured productivity.

Table 8: Relationship between agricultural distortions and observable characteristics at the municipality level

Coefficient	(1)	(2)
	Dependent variable: Agricultural distortions (ψ_j)	
Occupational precariousness index	-0.156	-0.156
	(0.095)	(0.100)
Per capita bank deposit accounts	-0.168**	-0.170**
	(0.074)	(0.074)
Mobile base stations per square kilometer	0.036	0.037
	(0.058)	(0.059)
Travel time to closest town of 20,000 habitants	0.178**	0.178**
	(0.070)	(0.071)
Government density index	-0.141**	-0.143**
	(0.062)	(0.069)
Share of indigenous population	0.008	0.009
	(0.073)	(0.073)
Index of natural hazards		0.003
		(0.029)
Rate of extortions		0.000
		(0.068)
Constant	-0.155**	-0.169
	(0.073)	(0.165)
Regional fixed effects	Yes	Yes
Observations	301	301
R-squared	0.062	0.062

Note: This table presents the results of regressing the estimated agricultural distortions (ψ_j) on observable characteristics at the municipality level (each observation is a municipality). See the main text for the definition of each regressor variable and source. All variables were standardized prior to the regression. Robust standard errors reported in parentheses. *, **, *** denotes statistical significance at 10%, 5%, and 1% level.

cluding the index of natural hazards and rate of extortions, we do not find statistically significant correlations with agricultural distortions. This also suggests that the estimated contribution of the modeled agricultural distortions on emigration, presented earlier, is not necessarily driven by climatic or insecurity events.

Overall, policies promoting more financial inclusion, road infrastructure, and government presence can be particularly helpful to reduce agricultural distortions and thereby increase local productivity and attenuate emigration.

5 Concluding remarks

This paper examines the role of agricultural distortions as a key driver of emigration and their effects on local aggregate productivity and welfare, using the case of Guatemala as an example. We develop a theoretical framework where household members can either stay in their country and work in the agricultural or non-agricultural sector, or emigrate, and estimate the model combining detailed micro and aggregate data. Our model identifies two key channels through which agricultural distortions affect migration and aggregate productivity: a migration channel, where distortions increase emigration among more productive agents, thereby reducing aggregate productivity, and a productivity channel, where distortions lead to factor misallocation, depressing incomes and encouraging cross-border migration. Quantitatively, reducing distortions in the agricultural sector to the most efficient department in each region can decrease the share of Guatemalan emigrants by 2.3 p.p., primarily among more productive workers—an amount equivalent to over 35% of the Guatemalan population currently residing in the US—, while increasing aggregate agricultural productivity by 30.1% and median household welfare by 4.5%.

An analysis at the sub-national level further reveals that regions with higher distortions tend to be more isolated and exhibit lower financial penetration and weaker government presence. This points to the importance of enhancing connectivity between regions, addressing financial barriers, and broadening the reach of public institutions as possible effective mechanisms to reduce agricultural distortions, avoid the need of people to relocate from their communities, and enhance local well-being. Certainly, addressing these issues (and others) require a combination of policies and interventions from the public and private sector that may not necessarily take effect in the short term. While building more roads and strengthening government institutions across the country require considerable resources and time, interventions promoting financial support and market transparency to farmers are probably more feasible and less costly to implement in the short run. Some examples include providing institutional and financial support to farmer cooperatives and associations to enhance their access to financial instruments and broader market information systems that can help reduce information asymmetries and other market inefficiencies in the agricultural sector. Similarly, reviewing current regulations and market interventions that may hinder the efficient allocation of resources is advisable.

Finally, while our analysis focuses on Guatemala, the study implications can be extended to other neighboring countries in the region that exhibit an important number of rural migration to the US, such as Mexico, Honduras, Nicaragua, and El Salvador.[34] Development programs focusing on addressing distortions in agricultural markets can play some role, whether directly or indirectly, in preventing people, especially productive workers in the sector, to leave their communities while improving agricultural productivity and local welfare. In the same vein, the study framework can be easily adapted to examine the role of agricultural distortions on migration decisions in other regions, such as Africa and South Asia, with a considerable outflow of migrants to Europe.

[34] https://www.pewresearch.org/short-reads/2024/09/27/key-findings-about-us-immigrants/ (accessed September 2025).

References

Abramitzky, R. & Boustan, L. (2017). Immigration in american economic history. *Journal of Economic Literature*, 55(4), 1311–1345.

Acampora, M., Casaburi, L., & Willis, J. (2025). Land rental markets: Experimental Evidence from Kenya. *American Economic Review*, 115(3), 727–771.

Adamopoulos, T., Brandt, L., Chen, C., Restuccia, D., & Wei, X. (2024). Land security and mobility frictions. *The Quarterly Journal of Economics*, 139(3), 1941–1987.

Adamopoulos, T., Brandt, L., Leight, J., & Restuccia, D. (2022). Misallocation, selection, and productivity: A quantitative analysis with panel data from China. *Econometrica*, 90(3), 1261–1282.

Adamopoulos, T. & Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6), 1667–1697.

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

Barbosa-Alves, M. & Britos, B. (2025). Climate change and international migration. Working Paper. Retrieved from `https://gbbritosh.github.io/resume/Climate_Change_and_International_Migration.pdf`.

Barchfield, J. (2021). Pandemic deepens hunger for displaced people the world over. The UN Refugee Agency (UNHCR). Retrieved from `https://www.unhcr.org/enus/news/stories/2021/3/6062fe334/pandemic-deepens-hunger-displaced-people-world.html`.

Bento, P. & Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3), 267–303.

Black, J. & Sigman, Z. (2022). 50,000 lives lost during migration: Analysis of missing migrants project data 2014–2022. International Organization for Migration (IOM GMDAC). Retrieved from `https://missingmigrants.iom.int/sites/g/files/tmzbdl601/files/publication/file/2022%2050k%20deaths.pdf`.

Borjas, G. J. (1999). *The Economic Analysis of Immigration.* Chapter 28 in Handbook of Labor Economics, Volume 3, Part A: Elsevier.

Borjas, G. J. (2008). Labor outflows and labor inflows in Puerto Rico. *Journal of Human Capital*, 2(1), 32–68.

Britos, B., Hernandez, M. A., Robles, M., & Trupkin, D. R. (2022). Land market distortions and aggregate agricultural productivity: Evidence from Guatemala. *Journal of Development Economics*, 155, 102787.

Brueckner, J. K. & Lall, S. V. (2015). Cities in developing countries: fueled by rural–urban migration, lacking in tenure security, and short of affordable housing. *Handbook of Regional and Urban Economics*, 5, 1399–1455.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, 82(5), 1671–1748.

Bryan, G. & Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from Indonesia. *Journal of Political Economy*, 127(5), 2229–2268.

Buera, F. J., Kaboski, J. P., & Shin, Y. (2011). Finance and development: A tale of two sectors. *American Economic Review*, 101(5), 1964–2002.

CAI (2019). Saliendo adelante: Why migrants risk it all. Creative Associates International, Washington. Retrieved from `https://www.saliendo-adelante.com/`.

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

Ceballos, F. & Hernandez, M. A. (2020). The Migration Propensity Index: An application to Guatemala. IFPRI Discussion Paper No. 1953. Retrieved from `https://www.ifpri.org/publication/migration-propensity-index-application-guatemala`.

Chari, A., Liu, E. M., Wang, S.-Y., & Wang, Y. (2021). Property rights, land misallocation, and agricultural efficiency in China. *The Review of Economic Studies*, 88(4), 1831–1862.

Chen, C., Restuccia, D., & Santaeulàlia-Llopis, R. (2022). The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics*, 45, 41–54.

Chen, C., Restuccia, D., & Santaeulàlia-Llopis, R. (2023). Land misallocation and productivity. *American Economic Journal: Macroeconomics*, 15(2), 441–465.

Chi, G., Abel, G. J., Johnston, D., & Bailey, M. (2025). Measuring global migration flows using online data. *Proceedings of the National Academy of Sciences*, 122(18), e2409418122.

Chiquiar, D. & Hanson, G. H. (2005). International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2), 239–281.

Clemens, M. A. (2020). Migration from developing countries: Selection, income elasticity and Simpson's paradox. Centro Studi Luca d'Agliano Development Studies Working Paper No. 465.

Clemens, M. A. (2021). Violence, development, and migration waves: Evidence from Central American child migrant apprehensions. *Journal of Urban Economics*, 124, 103355.

Clemens, M. A. (2022). Migration on the rise, a paradigm in decline: The last half-century of global mobility. *AEA Papers and Proceedings*, 112, 257–261.

Clemens, M. A., Montenegro, C. E., & Pritchett, L. (2019). The place premium: Bounding the price equivalent of migration barriers. *Review of Economics and Statistics*, 101(2), 201–213.

Clemens, M. A. & Pritchett, L. (2019). The new economic case for migration restrictions: An assessment. *Journal of Development Economics*, 138, 153–164.

Cohn, D., Passel, J. S., & Gonzalez-Barrera, A. (2017). Rise in US immigrants from El Salvador, Guatemala and Honduras outpaces growth from elsewhere. Pew Research Center. Retrieved from `https://www.pewresearch.org/hispanic/2017/12/07`.

CRS (2019a). Central American migration: Root causes and U.S. policy. Congressional Research Service. Retrieved from `https://fas.org/sgp/crs/row/IF11151.pdf`.

CRS (2019b). Recent migration to the United States from Central America: Frequently asked questions. Congressional Research Service. Retrieved from `https://fas.org/sgp/crs/row/R45489.pdf`.

CRS (2024). Central American Migration: Root causes and U.S. Policy. Congressional Research Service Report. October 30.

Cull, R., Ehrbeck, T., & Holle, N. (2014). Financial Inclusion and Development: Recent Impact Evidence. Consultative Group to Assist the Poor (CGAP), Focus Note 92, April.

De Janvry, A., Emerick, K., Gonzalez-Navarro, M., & Sadoulet, E. (2015). Delinking land rights from land use: Certification and migration in Mexico. *American Economic Review*, 105(10), 3125–3149.

Demirguc-Kunt, A., Klapper, L., & Singer, D. (2017). Financial Inclusion and Economic Growth: A Review of Recent Empirical Evidence. Policy Research Working Paper 8040. World Bank.

Derlagen, C., Muñoz, G., & De Salvo, C. P. (2020). Análisis de políticas agropecuarias en Guatemala. Banco Interamericano de Desarrollo. Retrieved from `https://publications.iadb.org/es/analisis-de-politicas-agropecuarias-en-guatemala`.

Durr, J. (2016). The political economy of agriculture for development today: The "small versus large" scale debate revisited. *Agricultural Economics*, 47, 671–681.

FAO (2016). Migration, agriculture and rural development: Addressing the root causes of migration and harnessing its potential for development. Food and Agriculture Organization. Retrieved from `http://www.fao.org/3/a-i6064e.pdf`.

Gauster, S. & Isakson, R. (2007). Eliminating market distortions, perpetuating rural inequality: An evaluation of market-assisted land reform in Guatemala. *Third World Quarterly*, 28(8), 1519–1526.

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5), 165–170.

Grogger, J. & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1), 42–57.

Hanson, G., Orrenius, P., & Zavodny, M. (2023). US immigration from Latin America in historical perspective. *Journal of Economic Perspectives*, 37(1), 199–222.

Hernandez, M., Ecker, O., Läderach, P., & Maystadt, J. (2023). *Forced migration, fragility, and policy responses for developing countries*. Chapter 7 in 2023 IFPRI Global Food Policy Report. International Food Policy Research Institute (IFPRI): Washington, DC.

Hsieh, C.-T. & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4), 1403–1448.

INE (2003). IV Censo Nacional Agropecuario. Instituto Nacional de Estadistica. Guatemala: Ciudad de Guatemala [In Spanish].

INE (2019). Encuesta Nacional de Empleo e Ingresos 2019. Instituto Nacional de Estadistica. Guatemala: Ciudad de Guatemala [In Spanish].

INE (2020). 2018 Population and Housing Census. Instituto Nacional de Estadistica. Guatemala: Ciudad de Guatemala. Retrieved from `https://www.censopoblacion.gt/`.

INE (2021). Encuesta Nacional de Empleo e Ingresos 2021. Instituto Nacional de Estadistica. Guatemala: Ciudad de Guatemala. Retrieved from `https:///www.ine.gob.gt/tablero-enei/`.

IOM (2022). World Migration Report 2022. International Organization for Migration, Geneva. Retrieved from `https://publications.iom.int/books/world-migration-report-2022`.

Iskhakov, F., Jørgensen, T. H., Rust, J., & Schjerning, B. (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, 8(2), 317–365.

Klapper, L., El-Zoghbi, M., & Hess, J. (2016). Achieving the Sustainable Development Goals: The Role of Financial Inclusion. Consultative Group to Assist the Poor (CGAP), United Nations Secretary General's Special Advocate for Inclusive Finance for Development.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C., & Waugh, M. E. (2020). Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*, 113, 138–154.

Lagakos, D., Mobarak, A. M., & Waugh, M. E. (2023). The welfare effects of encouraging rural–urban migration. *Econometrica*, 91(3), 803–837.

MAGA (2011). El Agro en Cifras 2011. Ministerio de Agricultura, Ganaderia y Alimentacion. Guatemala: Ciudad de Guatemala [In Spanish].

MAGA (2013). El Agro en Cifras 2013. Ministerio de Agricultura, Ganaderia y Alimentacion. Guatemala: Ciudad de Guatemala [In Spanish].

MAGA (2025). Food price monitoring and analysis tool. Guatemalan Ministry of Agriculture, Livestock and Food. Retrieved from `https://precios.maga.gob.gt/tool/public/`.

Mahajan, P. & Yang, D. (2020). Taken by storm: Hurricanes, migrant networks, and US immigration. *American Economic Journal: Applied Economics*, 12(2), 250–77.

Maldonado, R. & Harris, J. (2022). Las remesas a Latinoamérica y el Caribe en 2021: El esfuerzo de los migrantes durante la crisis del COVID-19. Banco Interamericano de Desarrollo. Retrieved from `http://dx.doi.org/10.18235/0004343`.

McKenzie, D. & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics*, 84(1), 1–24.

Moraga, J. F.-H. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics*, 93(1), 72–96.

Morten, M. & Oliveira, J. (2023). The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, forthcoming.

Munshi, K. & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(01), 46–98.

Ngai, L. R., Pissarides, C. A., & Wang, J. (2019). China's mobility barriers and employment allocations. *Journal of the European Economic Association*, 17(5), 1617–1653.

NIF (2019). Push or pull factors: What drives central american migrants to the U.S.? National Immigration Forum. Retrieved from `https://immigrationforum.org/article/push-or-pull-factors-what-drives-central-american-migrants-to-the-u-s/`.

Orrenius, P. M. & Zavodny, M. (2005). Self-selection among undocumented immigrants from Mexico. *Journal of Development Economics*, 78(1), 215–240.

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

Restuccia, D. & Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3), 151–174.

Soto, A. R., Bottone, R., Waters, J., Williams, S., Louie, A., & Wang, Y. (2021). Charting a new regional course of action: The complex motivations and costs of Central American migration. Rome, Washington, DC, and Cambridge, MA: World Food Programme, Migration Policy Institute, and Civic Data Design Lab at Massachusetts Institute of Technology. Retrieved from `https://www.wfp.org/publications/complex-motivations-and-costs-central-american-migration`.

US Census Bureau (2021). 2021 American Community Survey. United States Census Bureau. Washington, DC. Retrieved from `https://www.census.gov/programs-surveys/acs/news/data-releases.html`.

USAID (2021). U.S. Government Global Food Security Strategy: Fiscal Year 2022–2026. United States Agency for International Development (USAID). Washington, DC. Retrieved from https://www.usaid.gov/what-we-do/agriculture-and-food-security/us-government-global-food-security-strategy.

APPENDIX

A Proofs of the migration and productivity channels

In this Appendix, we provide formal proofs for the migration and productivity channels discussed in Subsection 2.6.

A.1 Migration channel

We first show that, for individuals with low productivity in the agricultural sector (denoted by z_l^a), the change in the probability of emigration satisfies:

$$p_{lj}^{e} - p_{lj}^{e^*} < 0, \quad \text{for all } j,$$

where p_{ij}^{e} represents the probability of emigrating from region j under distortions, and $p_{ij}^{e^*}$ represents the probability under the benchmark case without distortions.

Proof. For Gumbel-distributed utility random terms, the probability of emigrating for individual i is given by:

$$p_{ij}^{e} = \frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^w}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^f}{\sigma_\epsilon}\right)},$$

where U^e is the deterministic component of the utility from emigration, assumed to be constant across regions, while U_{ij}^w and U_{ij}^f are the deterministic components of the utility from working in the non-agricultural and agricultural sectors, respectively.

For an individual with a low productivity z_l^a, we argue that the change in the probability of emigration satisfies:

$$\frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{lj}^w}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{lj}^f}{\sigma_\epsilon}\right)} < \frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{lj}^{w^*}}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{lj}^{f^*}}{\sigma_\epsilon}\right)}.$$

Rearranging terms, yields:

$$\exp\left(\frac{U_{lj}^{w^*}}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{lj}^{f^*}}{\sigma_\epsilon}\right) < \exp\left(\frac{U_{lj}^{w}}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{lj}^{f}}{\sigma_\epsilon}\right).$$

Since $w_{ij}^n = A^n z_{ij}^n$, the non-agricultural wage remains unaffected by distortions: $w_{lj}^{n^*} = w_{lj}^n$. That is, distortions do not alter A^n or z_{ij}^n, so the wage in the non-agricultural sector is the same in the distorted and benchmark economies. For sufficiently mild distortions in region j, this implies that $U_{lj}^{w^*} \approx U_{lj}^{w}$.[35]

Thus, we obtain:

$$\exp\left(\frac{U_{lj}^{f^*}}{\sigma_\epsilon}\right) < \exp\left(\frac{U_{lj}^{f}}{\sigma_\epsilon}\right),$$

which holds whenever:

$$U_{lj}^{f^*} < U_{lj}^{f}.$$

This inequality holds for individuals with low productivity in the agricultural sector. The key mechanism is that distortions are weakest for low-ability agents, as τ_{ij} decreases with productivity. Consequently, these individuals receive higher income from profits and lump-sum transfers—and attain higher utility—under distortions. □

Next, we show that, for individuals with high productivity in the agricultural sector (denoted by z_h^a), the change in emigration probabilities satisfies:

$$p_{hj}^e - p_{hj}^{e^*} > 0, \quad \text{for all } j.$$

Proof. For an individual with a high productivity z_h^a, we argue that the change in the probability of emigration satisfies:

$$\frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{hj}^w}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{hj}^f}{\sigma_\epsilon}\right)} > \frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{hj}^{w^*}}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{hj}^{f^*}}{\sigma_\epsilon}\right)}.$$

[35] Although $w_{lj}^{n^*} = w_{lj}^n$, distortions raise the relative price of agricultural goods ($p^* < p$), which reduces real income and welfare, particularly for agents in the non-agricultural sector. Hence, $U_{lj}^{w^*} \geq U_{lj}^{w}$. However, if the change in p is small—for instance, when distortions are concentrated in a single region and the number of regions is large—their impact on prices and welfare is limited, and we may approximate $U_{lj}^{w^*} \approx U_{lj}^{w}$.

Rearranging terms, yields:

$$\exp\left(\frac{U_{hj}^{w^*}}{\sigma_\epsilon}\right)+\exp\left(\frac{U_{hj}^{f^*}}{\sigma_\epsilon}\right)>\exp\left(\frac{U_{hj}^{w}}{\sigma_\epsilon}\right)+\exp\left(\frac{U_{hj}^{f}}{\sigma_\epsilon}\right).$$

Since $w_{lj}^{n^*}=w_{hj}^{n}$, but $p^*<p$, it follows that $U_{hj}^{w^*}\geq U_{hj}^{w}$. Therefore, we obtain:

$$\exp\left(\frac{U_{hj}^{f^*}}{\sigma_\epsilon}\right)>\exp\left(\frac{U_{hj}^{f}}{\sigma_\epsilon}\right),$$

which holds provided that:

$$U_{hj}^{f^*}>U_{hj}^{f}.$$

This inequality holds for individuals with high productivity in the agricultural sector. The key mechanism is that distortions are strongest for high-ability agents, as τ_{ij} decreases with productivity. Consequently, these individuals receive lower income from profits and lump-sum transfers under distortions, resulting in lower agricultural utility. As a result,

$$p_{hj}^{e}>p_{hj}^{e^*}.$$

In sum, for high-productivity individuals, distortions lead to a higher probability of emigration compared to the benchmark case, highlighting their differential impact on migration patterns discussed in Subsection 2.6. □

Having demonstrated the migration channel of distortions, it is straightforward to show that for two identical regions differing only in the degree of distortions in the agricultural sector—one with high distortions ($j=H$) and the other with low distortions ($j=L$)—the emigration probabilities satisfy:

$$p_{lH}^{e}<p_{lL}^{e},$$

$$p_{hH}^{e}>p_{hL}^{e},$$

indicating that the emigration probability for less productive agents (z_{lj}^{a}) is lower in highly distorted regions compared to their counterparts in less distorted regions, and vice versa for highly productive agents (z_{hj}^{a}).

Proof. For an individual with low productivity (z_l^a), we want to show that emigration probabilities satisfy:

$$\frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)+\exp\left(\frac{U^w_{lH}}{\sigma_\epsilon}\right)+\exp\left(\frac{U^f_{lH}}{\sigma_\epsilon}\right)} < \frac{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U^e}{\sigma_\epsilon}\right)+\exp\left(\frac{U^w_{lL}}{\sigma_\epsilon}\right)+\exp\left(\frac{U^f_{lL}}{\sigma_\epsilon}\right)}.$$

Rearranging terms yields:

$$\exp\left(\frac{U^w_{lL}}{\sigma_\epsilon}\right)+\exp\left(\frac{U^f_{lL}}{\sigma_\epsilon}\right) < \exp\left(\frac{U^w_{lH}}{\sigma_\epsilon}\right)+\exp\left(\frac{U^f_{lH}}{\sigma_\epsilon}\right).$$

Since the two regions differ only in terms of distortions, and wages in the non-agricultural sector are given by $w^n_{ij} = A^n z^n_{ij}$, it follows that $w^n_{lL} = w^n_{lH}$. Hence, non-agricultural utility is equal: $U^w_{lL} = U^w_{lH}$. Substituting this into the inequality gives:

$$\exp\left(\frac{U^f_{lL}}{\sigma_\epsilon}\right) < \exp\left(\frac{U^f_{lH}}{\sigma_\epsilon}\right),$$

which implies $U^f_{lL} < U^f_{lH}$.

This results from distortions being defined so that $\tau_{lH} > \tau_{lL}$ and $\tau_{hH} < \tau_{hL}$; that is, they are weaker for low-ability individuals in the high-distortion region, and stronger for high-ability individuals. A low-productivity farmer thus receives relatively higher income from profits and lump-sum transfers in the high-distortion region, while the opposite holds for a high-productivity farmer. Therefore:

$$U^f_{lL} < U^f_{lH} \quad \text{and} \quad U^f_{hL} > U^f_{hH}.$$

It follows that a low-productivity individual in the agricultural sector (z_l^a) has a lower probability of emigration in the high-distortion region: $p^e_{lH} < p^e_{lL}$. Conversely, a high-productivity individual exhibits a higher probability of emigration under high distortions: $p^e_{hH} > p^e_{hL}$. □

A.2 Productivity channel

Agricultural distortions reduce household incomes in both the agricultural and non-agricultural sectors by inducing factor misallocation. We now show that this decline in incomes enhances the relative attractiveness of foreign opportunities in equilibrium, thereby increasing emigration from both rural and urban areas. That is, even absent selection effects, distortions that depress aggregate productivity increase the overall emigration rate.

To isolate the productivity channel—capturing the effect of income losses on the share of emigrants—from the migration channel, which influences selection into migration, we propose the following two-step comparison:

1. First, compute a closed-economy equilibrium with agricultural distortions, where international emigration is not allowed. Let A^a denote the agricultural productivity parameter in this economy, and let $\bar{Y}^a$ be the resulting aggregate agricultural output. In this scenario, aggregate productivity is reduced due to misallocation.

2. Next, consider a counterfactual open-economy equilibrium with no agricultural distortions (i.e., $\tau_{ij} = 1$ for all i, j), but where agricultural productivity is reduced to $A_l^a < A^a$ such that the same aggregate agricultural output $\bar{Y}^a$ is obtained. That is, we replicate the same productivity loss purely via a decline in A^a, without inducing misallocation. This allows emigration while isolating the effect of income losses alone.

We aim to show that, under this change, the probability of emigration increases for all individuals:

$$p_{ij}^e(A_l^a) > p_{ij}^e(A^a), \quad \text{for all } i \text{ and } j.$$

Proof. Under the model's multinomial logit structure, the probability that individual i in region j chooses to emigrate is given by:

$$p_{ij}^e(A^a) = \frac{\exp\left(\frac{U_j^e}{\sigma_\epsilon}\right)}{\exp\left(\frac{U_j^e}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^w(A^a)}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^f(A^a)}{\sigma_\epsilon}\right)},$$

and similarly for $A_l^a < A^a$.

Since U_j^e is exogenous and unchanged, we need to verify that:

$$\exp\left(\frac{U_{ij}^w(A^a)}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^f(A^a)}{\sigma_\epsilon}\right) > \exp\left(\frac{U_{ij}^w(A_l^a)}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^f(A_l^a)}{\sigma_\epsilon}\right).$$

We examine each term on the left-hand side:

- *Non-agricultural utility:* A decrease in A^a reduces the supply of the agricultural good, raising its relative price p. This makes the subsistence agricultural consumption $\bar{a}$ more costly, tightening the household's budget constraint and lowering utility. Hence:

$$U_{ij}^w(A^a) > U_{ij}^w(A_l^a).$$

- *Agricultural utility:* Profits in farming are increasing in A^a, so a reduction in A^a reduces incomes for all farmers. Hence:

$$U_{ij}^f(A^a) > U_{ij}^f(A_l^a).$$

Combining these two results implies:

$$\exp\left(\frac{U_{ij}^w(A^a)}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^f(A^a)}{\sigma_\epsilon}\right) > \exp\left(\frac{U_{ij}^w(A_l^a)}{\sigma_\epsilon}\right) + \exp\left(\frac{U_{ij}^f(A_l^a)}{\sigma_\epsilon}\right),$$

which confirms that:

$$p_{ij}^e(A_l^a) > p_{ij}^e(A^a).$$

Therefore, reducing A^a to match the productivity loss induced by distortions—while holding emigration open and abstracting from selection effects—leads to a strictly higher probability of emigration for every individual. This completes the proof that misallocation, through the productivity channel, increases the emigration rate. □

B Model extension with agricultural labor and rural-rural migration

The model setup is similar to the base model described in the main text, with one key difference: based on the productivity draw, an individual born in region j can choose among the following options: (i) farming in region j, where they use land and labor to produce the agricultural good; (ii) working in the agricultural sector as a wage earner in any region; (iii) working in the non-agricultural sector; or (iv) emigrating, provided that the welfare from emigrating exceeds that of staying. The welfare of staying depends on the utility derived from being a farmer or a worker in either the agricultural or non-agricultural sector.[36]

B.1 Agricultural sector

Each farmer i in region j is endowed with productivity z^a_{ij} and produces the agricultural good with the following technology:

$$y_{ij} = A^a z^a_{ij} l^\alpha_{ij} n^\beta_{ij},$$

where A^a is the aggregate productivity in the agricultural sector, l_{ij} and n_{ij} are the allocations of land and labor for the farmer, respectively, $\alpha, \beta \in (0,1)$, and $\alpha + \beta < 1$.

The market is assumed to be competitive and the profit maximization problem for farmer i in region j is defined as:

$$\pi(z^a_{ij}, \tau_{ij}) = \max_{l_{ij}, n_{ij}} \quad \tau_{ij} p A^a z^a_{ij} l^\alpha_{ij} n^\beta_{ij} - q_j l_{ij} - w^a_j n_{ij},$$

where p is the economy's relative price of the agricultural good, q_j is the rental price of a unit of land in region j, and w^a_j is region j's wage in the agricultural sector –all in terms of non-agricultural goods.

Given factor prices q_j and w^a_j, relative price p, and $\{z^a_{ij}, \tau_{ij}\}$, the allocations of land and

[36] By assuming that an individual born in region j can only engage in farming within their region but can work as a wage earner in other regions, we implicitly account for sufficiently high (prohibitive) costs associated with directly farming in a different region. Although somewhat restrictive, this assumption aligns with existing evidence.

labor for each farmer i in region j are determined by the following expressions:

$$l_{ij} = \left[(\alpha/\beta)(w_j^a/q_j)\right]^{(1-\beta)/(1-\alpha-\beta)} \left[\beta(p/w_j^a)A_{ij}\right]^{1/(1-\alpha-\beta)},$$

$$n_{ij} = \left[(\alpha/\beta)(w_j^a/q_j)\right]^{\alpha/(1-\alpha-\beta)} \left[\beta(p/w_j^a)A_{ij}\right]^{1/(1-\alpha-\beta)},$$

where $A_{ij} \equiv A^a z_{ij}^a \tau_{ij}$.

B.2 Non-agricultural sector

We assume a linear production function in labor, which is the only factor used to produce the non-agricultural good. The aggregate non-agricultural output is given by:

$$Y^n = A^n \sum_j \int_0^1 \mathbb{W}_{ij} z_{ij}^n \, di,$$

where $\mathbb{W}_{ij} = 1$ is a dummy variable equal to one if individual i from region j becomes a worker in the non-agricultural sector and zero otherwise, z_{ij}^n denotes the individual's productivity to work in this sector, and A^n is the aggregate non-agricultural productivity.

The market for the non-agricultural good is also assumed to be competitive. In equilibrium, the wage paid to each individual indexed by the pair (i, j) working in this sector is equal to $w_{ij}^n = A^n z_{ij}^n$.

B.3 The migration problem

The decision of an individual to emigrate is determined by maximizing welfare between emigrating and staying. Specifically, each individual's migration decision is defined as the following problem:

$$\max_{\mathbb{S}_{ij}, \mathbb{E}_{ij} \in \{0,1\}} \{V_{ij}^d, V_{ij}^e\},$$

where $V_{ij}^d = U_{ij}^d + \epsilon_{ij}^d$ represents the utility of individual i, born in region j, for staying in the country. This utility depends on U_{ij}^d (defined below) and a random variable ϵ_{ij}^d, which follows a Gumbel distribution as described in the main text. Similarly, $V_{ij}^e = U_j^e + \epsilon_{ij}^e$ denotes the utility of emigrating, determined by the foreign wage (net of region-specific moving costs) U_j^e, and an idiosyncratic random variable ϵ_{ij}^e, also following a Gumbel distribution.

Finally, $\mathbb{S}_{ij} = 1$ indicates that the individual stays in the country, while $\mathbb{E}_{ij} = 1$ indicates that the individual emigrates.

B.4 The household's domestic problem

Conditional on staying, the utility maximization problem for individual i born in region j is defined as:

$$
\begin{aligned}
U_{ij}^{d} = \max_{c_{ij}^{a}, c_{ij}^{n}} \quad & \omega \log(c_{ij}^{a} - \overline{a}) + (1-\omega)\log(c_{ij}^{n}) \\
\text{s.t.} \quad & pc_{ij}^{a} + c_{ij}^{n} \leq I_{ij} + T \\
& I_{ij} \equiv \max_{\mathbb{F}_{ij}, \mathbb{W}_{ij}, \mathbb{N}_{ij}^{j}, \mathbb{N}_{ij}^{k} \in \{0,1\}} \{\pi(z_{ij}^{a}, \tau_{ij}), \quad w_{ij}^{n} - b_{ij}, w_{j}^{a}, w_{k\neq j}^{a} - \kappa_{ik}\},
\end{aligned}
$$

where $\overline{a} \geq 0$ is the subsistence level of consumption of the agricultural good, I_{ij} is the income level of the household as a function of the occupational decision, and $\omega \in (0,1)$. The decision problem for the individual is: (i) being a farmer ($\mathbb{F}_{ij} = 1$), (ii) work in the non-agricultural sector ($\mathbb{W}_{ij} = 1$), (iii) work in region j's agricultural sector ($\mathbb{N}_{ij}^{j} = 1$), or (iv) work in region k's agricultural sector ($\mathbb{N}_{ij}^{k} = 1$), in which case she pays an idiosyncratic commuting cost κ_{ik}, expressed in terms of non-agricultural goods. T represents the per capita transfers received by households, including land rental payments from the government (assumed to own the land) and lump-sum transfers from the redistribution of revenues collected through the (distortionary) idiosyncratic tax on agricultural output, all distributed equally among individuals and expressed in terms of non-agricultural goods.

Given the relative price p and income I_{ij}, the solution to the household's maximization problem imply the following expressions:

$$c_{ij}^{a} = \frac{\omega(I_{ij} + T)}{p} + (1-\omega)\bar{a},$$

$$c_{ij}^{n} = (1-\omega)(I_{ij} + T) - (1-\omega)p\bar{a}.$$

B.5 Equilibrium

A *competitive equilibrium* is defined as a set of prices: $\{p, q_j, w_j^a\}$ for all j and $w_{ij}^{n} = A^{n} z_{ij}^{n}$ for all pairs (i,j); a set of occupational choices: $\{\mathbb{W}_{ij}, \mathbb{F}_{ij}, \mathbb{N}_{ij}^{j}, \mathbb{N}_{ij}^{k}\}$ for all $(i, j, k \neq j)$; a set of emigrate-stay choices: $\{\mathbb{S}_{ij}, \mathbb{E}_{ij}\}$ for all (i, j); a set of factor allocations for the farmers in each region: $\{l_{ij}, n_{ij}\}$; and a set of consumption allocations for all households: $\{c^a_{ij}, c^n_{ij}\}$, such that the markets clear.

The market clearing condition for land is defined as:

$$\int_0^1 \mathbb{F}_{ij} l_{ij}\, di = L_j; \quad \text{for all } j.$$

The market clearing condition for agricultural labor is:

$$\underbrace{\int_0^1 n_{ij}\, di}_{\text{Labor demand from farmers in } j} \quad = \quad \underbrace{\textstyle\sum_k \displaystyle\int_0^1 \mathbb{N}^j_{ik}\, di}_{\text{Total labor supply to region } j} \quad \equiv \mathbb{N}^j; \quad \text{for all } j.$$

The market clearing condition for non-agricultural labor is:

$$\sum_j \int_0^1 \mathbb{W}_{ij}\, di = \mathbb{W}.$$

The market clearing condition for the agricultural good is:

$$\sum_j \int_0^1 c^a_{ij}\, di = \sum_j \int_0^1 \tau_{ij} \mathbb{F}_{ij} y_{ij}\, di.$$

Lastly, the market clearing condition for the non-agricultural good is:

$$\sum_j \int_0^1 \left[c^n_{ij} + \kappa_{ik} \mathbb{N}^k_{ij} \right] di = A^n \sum_j \int_0^1 \mathbb{W}_{ij} z^n_{ij}\, di,$$

where $\sum_j \int_0^1 \kappa_{ik} \mathbb{N}^k_{ij} di$, for $k \neq j$, represents the total commuting costs expressed in terms of the non-agricultural good.

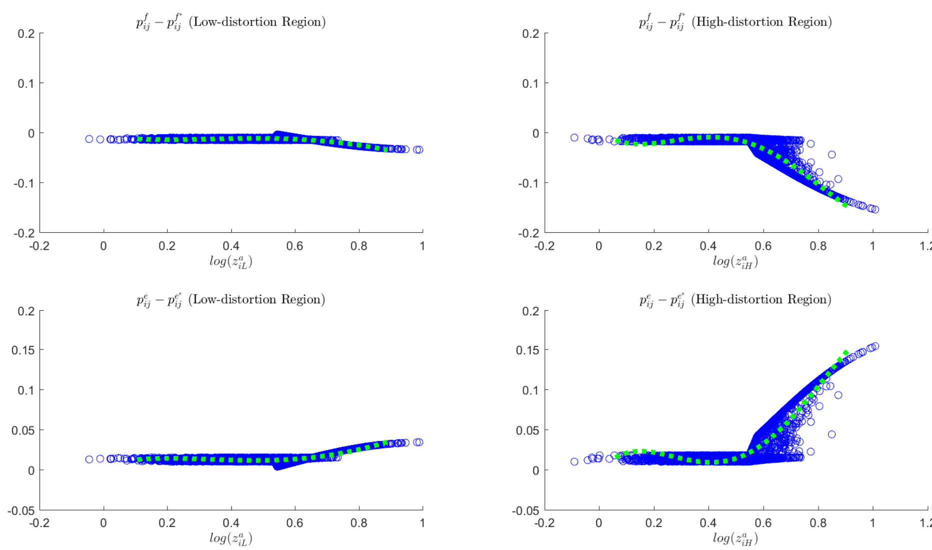
B.6 Model mechanisms

As discussed in Section 2.6, distortions in this extended model affect aggregate productivity and the emigration rate via two channels–the migration and productivity channels. Higher agricultural distortions increase the emigration probability of more productive agents while decreasing that of less productive agents, leading to a decline in aggregate productivity.

In addition, higher agricultural distortions drive factor misallocation, reducing household incomes across both sectors, which, in turn, increases emigration from both rural and urban areas.

Figure B.1 illustrates the change in an agent's probability of being a farmer (upper panel) and an emigrant (lower panel) between the distorted and benchmark cases, across the distribution of agricultural abilities z^a_{ij} in regions with low versus high distortions (left and right panels, respectively). The analysis is based on simulated data, assuming regions are identical in all aspects except for the level of agricultural distortions.

Figure B.1: Change in the probability of being a farmer and emigrant in low- and high-distortion regions: Simulated data



Note: The vertical axis displays the change in the probability of being a farmer p^f_{ij} (upper panel) and an emigrant p^e_{ij} (lower panel) between the distorted and benchmark cases across the distribution of idiosyncratic agricultural abilities z^a_{ij} in regions with low versus high distortions (left and right panels). The horizontal axis represents the $\log(z^a_{ij})$. The green dashed line shows a polynomial fit to the observed data.

In regions with low distortions (left panels), changes in the probability of being a farmer as well as that of emigrating are minimal across the ability distribution, with slight reductions (increases) in the probability of being a farmer (emigrating) observed for individuals with higher agricultural abilities. In regions with high distortions (right panels), the probability of being a farmer (emigrating) decreases (increases) substantially for individuals with medium-to-high agricultural abilities, suggesting a stronger incentive for these agents to transition out of farming. In line with the baseline model discussed in Section 2, these patterns highlight how higher distortions strongly encourage emigration among more productive agents.

The main difference between the benchmark and extended models, as illustrated in Figures 2 and B.1, lies in the introduction of new choices in the extended model—-specifically, the options to become an agricultural worker (rather than a farmer) and to internally migrate to another rural region while incurring a commuting cost. In this extended framework, distortions mainly impact the probabilities of being a farmer or emigrating for medium- to high-ability agents. This is because agents now have the additional possibility of remaining in a rural area as an agricultural worker, either in their current location or in another rural location. Importantly, despite these added choices, the overall effects on the emigration rate and aggregate productivity, which are central to our study, remain consistent with those observed in the benchmark model.

C Data and calibration

C.1 Employment and migration shares

$\mathbb{F}_j$, $\mathbb{W}j$, and $\mathbb{E}j$ denote the shares for farmers, workers, and migrants, respectively, at location j. To calculate these shares, it is necessary to define the corresponding data categories by region. Table C.1 below summarizes the construction and data sources for each of these categories.

Number of farmers. Using data from the 2018 Population and Housing Census (INE, 2020), we define farmers as individuals identified as either entrepreneurs or workers within occupational categories related to market-oriented farming and skilled agricultural labor; subsistence agricultural workers, fishers, hunters, and gatherers; and laborers in agriculture, fishing, and forestry. This is the numerator used to calculate the share of farmers, $\mathbb{F}_j$.

Number of Non-agricultural workers. Non-agricultural (NA) workers are defined as all entrepreneurs and workers employed outside the agricultural sector, encompassing all occupations not classified under farming, fishing, forestry, or subsistence activities. This serves as the numerator in the calculation of the share of workers, $\mathbb{W}_j$.

Number of international migrants. This corresponds to the stock of migrants from each region of Guatemala working in the US. We focus on the US as it is by far the largest destination country for Guatemalan emigrants. The total number of Guatemalan migrants currently working in the US is obtained from the 2021 American Community Survey (US Census Bureau, 2021). Since this dataset lacks information on migrants' regional origins, we calculate the regional share of individuals who migrated to the US between 2002 and 2018 using Guatemalan Census data (INE, 2020) and multiply these shares by the total number of Guatemalan migrants currently employed in the US, to obtain an estimate of the number of migrants from region j working in the US. This represents the numerator for calculating the share of international migrants, $\mathbb{E}_j$.

Number of employed people. The total number of employed individuals is computed as the sum of the number of farmers, number of NA-workers, and number of international migrants. This aggregate measure serves as the denominator to calculate the shares $\mathbb{F}_j$, $\mathbb{W}_j$, and $\mathbb{E}_j$.

Table C.1: Definition of variables for employment and migration shares

Variable	**Definition and Source**
Number of farmers	Individuals classified as entrepreneurs or workers in agriculture, forestry, or fishing, including market-oriented farmers, skilled agricultural workers, subsistence agricultural workers, fishers, hunters, gatherers, and agricultural laborers. *Source: 2018 Population and Housing Census (INE, 2020)*
Number of non-agricultural workers	All employed individuals (entrepreneurs and workers) in occupations outside of agriculture, forestry, and fishing. *Source: 2018 Population and Housing Census (INE, 2020)*
Number of international migrants	Stock of Guatemalan migrants working in the U.S., allocated across regions using the 2002–2018 regional migration shares from the Guatemalan Census. *Source: 2021 American Community Survey (US Census Bureau, 2021); 2018 Population and Housing Census (INE, 2020)*
Number of employed people	Sum of the number of farmers, non-agricultural workers, and international migrants.

C.2 Agricultural production and income data

Our agricultural production dataset consists of microdata from the IV National Agricultural Census (INE, 2003), the most recent agricultural census conducted in Guatemala, corresponding to the 2002–03 crop year. The dataset includes information on crop output, land allocation, production inputs, labor use, machinery and equipment ownership, and producers' socioeconomic characteristics by geographic location. Since the census does not record agricultural income generated by crops, we use 2002 price data from the Ministry of Agriculture, Livestock and Food of Guatemala (MAGA) to calculate crop-level revenues.[37]

Table C.2 summarizes the construction and data sources for the key variables used to estimate agricultural revenue and input use as well as non-agricultural income. We use this information to estimate agricultural and non-agricultural productivity.

Total revenue. From the agricultural census, we obtain crop-level physical output, which we multiply by the corresponding average monthly prices for 2002 from MAGA to compute revenue per crop. We then aggregate these revenues at the producer level to derive total revenue.

[37] Price data was retrieved from the Food Price Monitoring and Analysis Tool (MAGA, 2025).

Land input. We use land allocated to permanent and seasonal crops as our measure of land size.

Labor input. We consider labor used in the farm as our measure labor input. Unfortunately, the dataset does not provide information on labor allocation by crop. Labor is measured as the number of people involved in farm work, including both family members and hired workers.

Non-agricultural income. We obtain non-agricultural income data from the 2019 National Survey of Labor and Income (INE, 2019). Non-agricultural income is defined as wages and earnings from activities outside the agricultural sector in which the worker is engaged.

Table C.2: Definition of variables for the estimation of productivity

Variable	**Definition and Source**
Total revenue	Physical output by crop is obtained from the IV National Agricultural Census. Each crop's output is multiplied by its average monthly price in 2002, sourced from the Ministry of Agriculture (MAGA). The total revenue per producer is the sum of crop-level revenues. *Source: IV National Agricultural Census (INE, 2003); Food Price Monitoring and Analysis Tool (MAGA, 2025)*
Land input	Total land used for both permanent and seasonal crops. *Source: IV National Agricultural Census (INE, 2003)*
Labor input	Total number of individuals working on the farm, including family members and hired workers. Labor is measured at the farm level, not by crop. *Source: IV National Agricultural Census (INE, 2003)*
Non-agricultural income	Income earned from non-agricultural employment or activities, including wages and earnings from self-employment outside agriculture. *Source: 2019 National Survey of Labor and Income (INE, 2019)*

C.3 Estimation of agricultural and non-agricultural productivity

Agricultural TFP and productivity. Based on our framework, we define total factor productivity (TFP) as the following:

$$TFP_{ij} = \frac{y_{ij}}{l_{ij}^{\alpha}},$$

where y_{ij} is farm's i total revenue by unit of labor at region j, and l_{ij} is the farm's total land used by unit of labor.

In our setting, TFP_{ij} is equivalent to the farmer's agricultural ability z_{ij}^a. We estimate this productivity measure following the two-stage approach of Britos et al. (2022). First, we derive a measure of z_{ij}^a using the equation above and the previously described data from the IV National Agricultural Census (INE, 2003) and Food Price Monitoring and Analysis Tool (MAGA, 2025). Second, to account for observable factors not included in the theoretical framework that could influence our productivity measure, we estimate z_{ij}^a using the following regression by region j:

$$\log\left(\frac{y_{ipj}}{l_{ipj}^{\alpha}}\right) = \beta_j^1 X_{ipj}^1 + \beta_j^2 X_{ipj}^2 + \eta_p + \varepsilon_{ipj},$$

where X_{ipj}^1 is a vector of controls that includes age and sex of farmer i in village p at region j; X_{ipj}^2 is a vector that includes farmer's years of education, the ratio of farm labor from household members over total farm labor, if the farmer has machinery and equipment, use of high-performance seeds, use of fertilizer and pesticides, if farm has an irrigation system, if farmer has livestock, and number of different crops produced; η_p is a fixed effect term at the village level, which accounts for unobserved differences across locations that can contribute to productivity differences such as biophysical and accessibility factors; and ε_{ipj} corresponds to the residual.

Based on the regression estimations, we obtain the residual $\hat{\varepsilon}_{ij}$ and the coefficients corresponding to age and sex, $\hat{\beta}_j^1$, to estimate the farmer's agricultural productivity z_{ij}.[38]

[38] Age and sex are included in the regression above to control for the potential ascendancy and advantage position of older and male farmers that is typical in rural Guatemala (especially among Mayan cultures), which could be correlated with asymmetric land allocation (and input access) patterns beyond a farmer's inherent skill; these two exogenous variables are then put back into our productivity measure.

This is equal to:

$$\log\left(\hat{z}_{ij}^{a}\right) = \hat{\beta}_j^1 X_{ij}^1 + \hat{\varepsilon}_{ij}.$$

where $\hat{\beta}_j^1 X_{ij}^1$ reintroduces the influence of age and sex and $\hat{\varepsilon}_{ij}$ captures unobserved productivity. We remove the top and bottom percentiles to account for outliers.

We then compute the region-specific mean of $\log(\hat{z}_{ij}^a)$, denoted by $\hat{\mu}_j^a$, and the standard deviation of the regression residuals, denoted by $\hat{\sigma}_j^a$. These parameters are reported in Table C.3.

Table C.3: Means and Standard Deviations of Agricultural Productivity by Region

Region	**Mean** ($\hat{\mu}_j^a$)	**Std. Dev.** ($\hat{\sigma}_j^a$)
Region 1	1.31	0.90
Region 2	1.39	1.05
Region 3	1.47	1.12

Note: This table reports the region-specific means and standard deviations of the estimated log-normal agricultural productivity, $\log(\hat{z}_{ij}^a)$. The mean $\hat{\mu}_j^a$ is calculated as the average of the model-predicted productivity measure, $\hat{\beta}_j^1 X_{ij}^1 + \hat{\varepsilon}_{ij}$, and the standard deviation $\hat{\sigma}_j^a$ is the dispersion of the regression residuals. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta.

Agricultural TFPR. We define revenue-based total factor productivity (TFPR) as the following:

$$TFPR_{ij} = \frac{y_{ij}}{l_{ij}}.$$

To calculate the $TFPR_{ij}$ we plug $\hat{z}_{ij}^a$ into y_{ij} and obtain:

$$TFPR_{ij} = \frac{\hat{z}_{ij}^a l_{ij}^{\alpha}}{l_{ij}}.$$

From the measures of TFP and TFPR across i, we calculate each region's j mean, standard deviation, and correlation. These statistics are used in the estimation of the regional distortions, τ_{ij}.

Non-agricultural productivity. We estimate non-agricultural productivity in a similar fashion as the agricultural productivity. Using data from the 2019 National Survey of Labor and Income (INE, 2019), we first run the following estimation:

$$\log\left(NA\ Income_{ip}\right) = \lambda^1 W_{ip}^1 + \lambda^2 W_{ip}^2 + \phi_p + \nu_{ip},$$

where $NA\ Income_{ip}$ is the non-agricultural (NA) income reported by worker i in sampling unit p; W^1_{ip} is a vector of controls that includes the age and sex of the worker; W^2_{ip} corresponds to the worker's education; ϕ_p is a fixed effect term at the sampling unit level controlling for unobserved heterogeneity between locations influencing income; and ν_{ip} is the residual.

As in the estimation of $\hat{z}^a_{ij}$, the non-agricultural productivity $\hat{z}^{na}_i$ is equal to:

$$\log\left(\hat{z}^{na}_i\right) = \hat{\lambda}^1 W^1_i + \hat{\nu}_i.$$

From this measure of non-agricultural productivity, we keep individuals living in urban areas and remove the top and bottom percentile. The resulting mean and standard deviation for the non-agricultural productivity are $\hat{\mu}^n = 0.8$ and $\hat{\sigma}^n = 0.59$.

D Supplementary Tables and Figures

Table D.1: Region characteristics

Variable	Western Highlands-Verapaces-Center	Dry Corridor-Izabal	Pacifico-Bocacosta	Within- vs Between-region variation ANOVA (*p*-value)
Travel time to closest town of 20,000 habitants (hours)	1.317	1.503	0.761	0.091
	(0.514)	(0.603)	(0.122)	
Risk of frosts (index 0-1)	0.240	0.013	0.002	0.008
	(0.197)	(0.016)	(0.001)	
Risk of droughts (index 0-1)	0.647	0.760	0.497	0.014
	(0.118)	(0.112)	(0.147)	
Risk of floodings (index 0-1)	0.061	0.246	0.247	0.087
	(0.085)	(0.259)	(0.193)	
Geodynamic and geophysical disasters (index 0-10)	3.555	1.611	4.981	0.006
	(1.876)	(0.360)	(0.853)	
Government density (index 0-1)	0.206	0.258	0.224	0.001
	(0.021)	(0.020)	(0.021)	
Share of indigenous population (0-100)	70.088	18.046	20.490	0.000
	(24.290)	(14.772)	(14.888)	
Share of households Spanish main language (0-100)	42.515	93.461	93.243	0.000
	(27.546)	(11.188)	(4.895)	
Poverty rate (0-100)	69.172	60.443	59.990	0.185
	(12.696)	(7.735)	(4.384)	
Chronic malnutrition rate (0-100)	43.105	25.942	23.897	0.002
	(10.781)	(8.394)	(5.233)	
Number of departments	10	6	4	20

Note: This table presents average characteristics of the regions considered for the analysis (standard deviations reported in parentheses). The Western Highlands-Verapaces-Center region includes 10 departments; Dry Corridor-Izabal region includes 6 departments; and Pacifico-Bocacosta region includes 4 departments. The travel time to the closest town of 20,000 habitants is obtained from the International Food Policy Research Institute (IFPRI) typology of micro-regions exercise performed in 2021; the risk of frosts, droughts, and floodings and the government (public institutions) density index are obtained from the Food Insecurity and Malnutrition Vulnerability Index (IVISAN) calculated in 2012 by the Secretary of Food Security and Nutrition of Guatemala (SESAN); the index of geodynamic and geophysical disasters is obtained from the National Institute for Seismology, Vulcanology, Meteorology, and Hydrology of Guatemala (INSIVUMEH) for the period 1530-2015; the share of indigenous population and households with Spanish as the main language reported by the head are obtained from the 2018 Population and Housing Census implemented by the National Institute of Statistics of Guatemala (INE); the poverty rate is obtained from INE for 2014; and the chronic malnutrition rate for school children in first grade is obtained from the 2015 Height Census in Schoolchildren implemented by INE. The last columns reports the results (*p*-value) of a one-way ANOVA comparing within- versus between-variation across regions in the reported variables (H_0: all region means are equal).

Table D.2: Land rentals and perceived prices of best land

Region	**% Renting Land**	**Mean Price**	**Std. Dev.**	**CV**
Region 1	24.6%	26.47	40.25	1.52
Region 2	45.4%	18.13	30.71	1.69
Region 3	61.9%	32.88	38.98	1.19

Note: This table presents the percentage of rural individuals who report renting land from others over the past 12 months (column 2), the average perceived price (in quetzales per vara^2) of the best land in the municipality where the individual is located (column 3), the standard deviation of the perceived price (column 4), and the resulting coefficient of variation (column 5), disaggregated by region. The data is obtained from a three-year panel survey conducted between 2012 and 2014 by IFPRI, covering 176 of Guatemala's 340 municipalities. The survey was conducted as part of the evaluation of a large-scale program implemented by the Government of Guatemala to reduce food insecurity and malnutrition, focusing on the poorest areas with the highest stunting rates. Region 1 corresponds to Western Highlands-Verapaces-Center (4,788 observations), Region 2 to Dry Corridor-Izabal (625 observations), and Region 3 to Pacifico-Bocacosta (139 observations).

Table D.3: Changes in emigration, non-agricultural employment, agricultural productivity, and median consumption between actual and benchmark scenarios under alternative values for the land elasticity of output (α)

Panel A: $\alpha = 0.3$

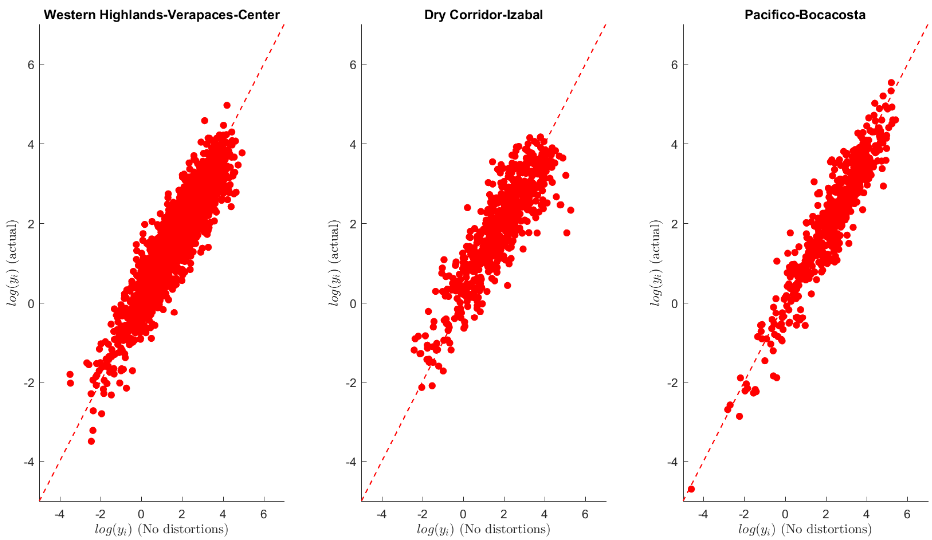
Region	Δ Share of Emigrants (p.p.)	Δ Share of Workers (p.p.)	Δ Agricultural Productivity (%)	Δ Median Consumption (%)
Region 1	-3.0	0.6	19.2	4.5
Region 2	-2.4	-0.6	32.8	8.5
Region 3	-2.2	2.4	19.6	7.8
Urban	-2.4	2.4	–	13.3
Total	-2.7	1.2	21.7	8.4

Panel B: $\alpha = 0.5$

Region	Δ Share of Emigrants (p.p.)	Δ Share of Workers (p.p.)	Δ Agricultural Productivity (%)	Δ Median Consumption (%)
Region 1	-1.5	1.0	37.6	13.1
Region 2	-1.2	0.8	49.4	12.9
Region 3	-0.9	2.7	24.8	10.0
Urban	-1.0	1.0	–	11.2
Total	-1.3	1.2	37.1	11.7

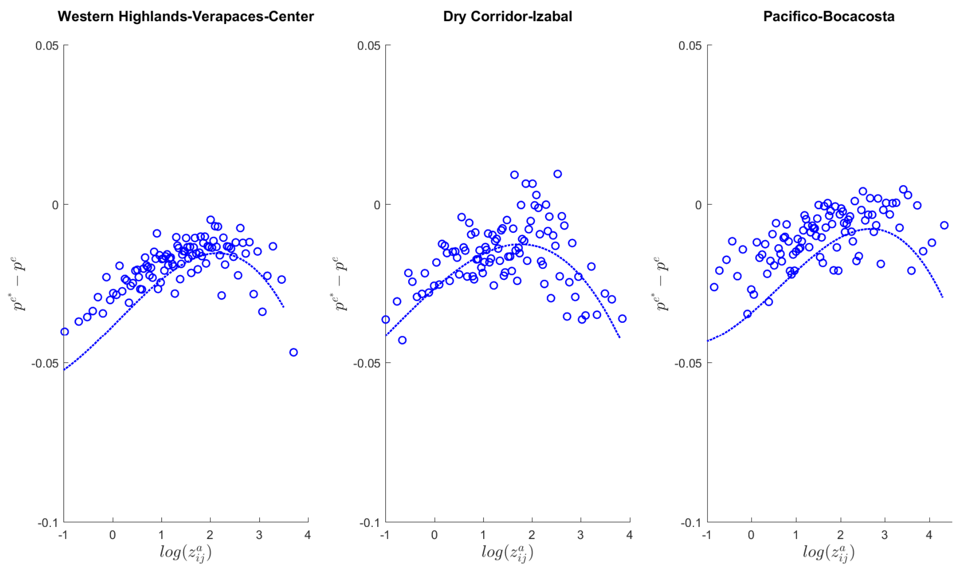
Note: This table presents changes in the emigration share ($\mathbb{E}_j$), the share of non-agricultural workers ($\mathbb{W}_j$), aggregate agricultural productivity, and median consumption when moving from the actual to the benchmark scenario in each region, for $\alpha = 0.3$ (Panel A) and $\alpha = 0.5$ (Panel B). Changes are expressed in percentage points (p.p.) for emigration and employment shares and in percentage terms (%) for aggregate productivity and median consumption. Region 1 corresponds to Western Highlands-Verapaces-Center, Region 2 to Dry Corridor-Izabal, and Region 3 to Pacifico-Bocacosta. Total refers to the aggregate, including the three agricultural regions and the urban region.

Figure D.1: Comparison of actual vs. no-distortion output

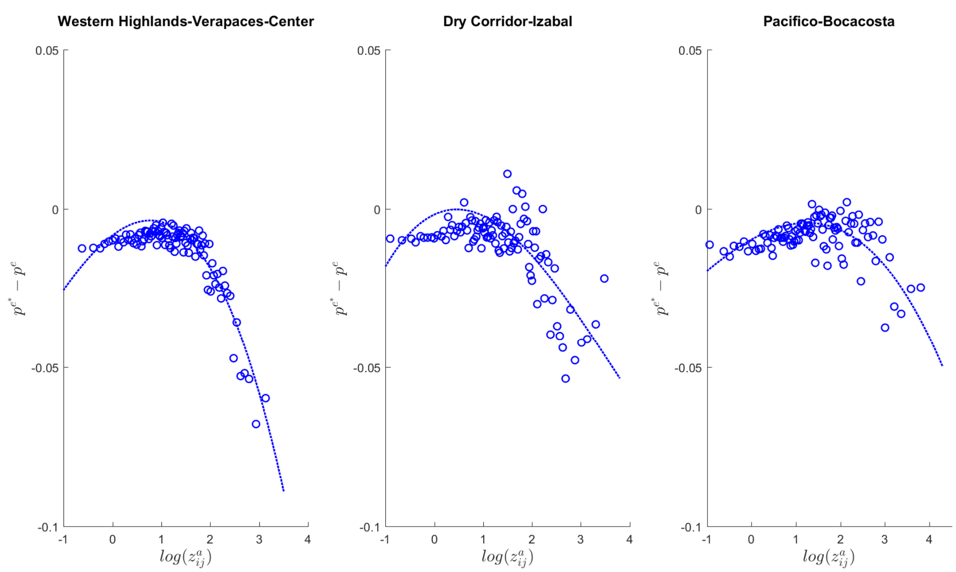


Note: The figure shows scatterplots comparing the actual output ($\log(y_i)$, vertical axis) with the hypothetical no-distortion output ($\log(y_i)$, horizontal axis) for the three regions: Western Highlands-Verapaces-Center, Dry Corridor-Izabal, and Pacifico-Bocacosta. Each red dot represents an individual observation, while the dashed red line represents the 45-degree line where actual output equals no-distortion output. Deviations from the line indicate the extent of distortions in output.

Figure D.2: Change in the probability of emigration across the distribution of abilities under alternative values for the land elasticity of output (α)



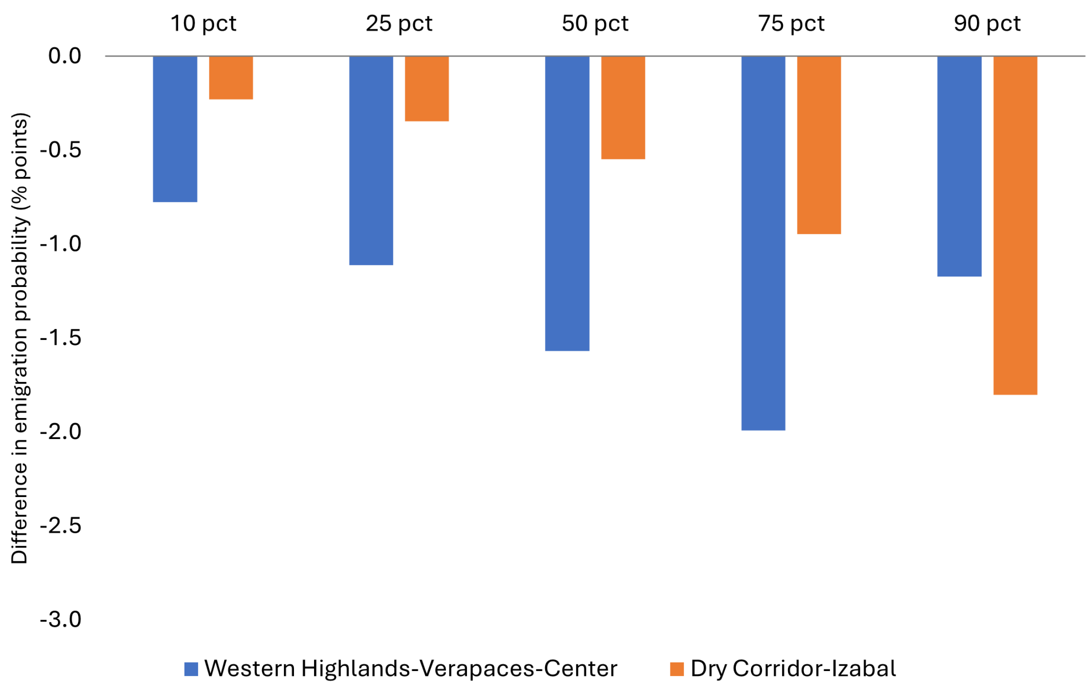
$\alpha = 0.3$



$\alpha = 0.5$

Note: This figure shows scatterplots of the change in the probability of being an emigrant between the benchmark and actual cases ($p^{e^*}_{ij} - p^e_{ij}$), plotted against the percentiles of the log of agricultural productivity ($\log(z^a_{ij})$), for each region: Western Highlands–Verapaces–Center (left), Dry Corridor–Izabal (center), and Pacific–Bocacosta (right), for $\alpha = 0.3$ (top panel) and $\alpha = 0.5$ (bottom panel), where $p^{e^*}_{ij}$ corresponds to the benchmark scenario and p^e_{ij} denotes the actual case. The dashed line shows a polynomial fit to the data.

Figure D.3: Difference in the probability of emigration between Pacifico-Bocacosta and other regions by different levels of household per capita expenditure



Note: The figure shows the difference (in percentage points) in the probability of emigration between Pacifico-Bocacosta versus Western Highlands-Verapaces-Center and Dry Corridor-Izabal for different percentiles of household per capita expenditure, using data from the Living Standards Measurement Survey in Guatemala (ENCOVI 2014). See the main text for details on the derivation of the probabilities. The percentiles reported include the 10th, 25th, 50th, 75th, and 90th percentile (pct).