Land, Power and Population Density in the Global Population Shifts of 1500 – 1960

Bravo Working Paper # 2026-001

Mehmet Burak Akren*, Arhan S. Ertan† and Louis Putterman‡

July 22, 2026

*University of Duisburg-Essen, Department of Economics, 45141 Essen, Germany. e-mail: mehmet.akren@vwl.uni-due.de

†Boğaziçi University, Department of International Trade, 34342 Istanbul, Turkey. e-mail: arhan.ertan@bogazici.edu.tr

‡Corresponding Author. Brown University, Department of Economics, 02912 Providence RI, USA. e-mail: louis putterman@brown.edu

Land, Power and Population Density in the Global Population Shifts of 1500 – 1960

Mehmet Burak Akren*, Arhan S. Ertan† and Louis Putterman‡

July 22, 2026

*University of Duisburg-Essen, Department of Economics, 45141 Essen, Germany. e-mail: mehmet.akren@vwl.uni-due.de

†Boğaziçi University, Department of International Trade, 34342 Istanbul, Turkey. e-mail: arhan.ertan@bogazici.edu.tr

‡Corresponding Author. Brown University, Department of Economics, 02912 Providence RI, USA. e-mail: louis_putterman@brown.edu

Abstract

The regional distribution of world population underwent profound changes during the age of exploration and colonization, in which a handful of European powers colonized the better part of the remaining continents. One consequence is that many countries in the Americas, Australia and New Zealand, after sharp contractions of indigenous populations following exposure to Old World diseases, today have populations in which European or African descendants form majorities and speak predominantly European languages. Migration and in some cases colonization also altered population compositions in countries including South Africa, Cape Verde, Fiji and Singapore. And descendants of émigrés came to outnumber those of "stayers" even for non-colonizing countries, including, e.g., Ireland and Lebanon. We identify major determinants of who settled where, during this era of colonialism and migration, demonstrating that both "natural" factors like availability and fertility of agricultural land and climate and disease prevalence, on the one hand, and power inequalities and cultural variables, on the other, explain the broad patterns of migration.

JEL classification: F22, N30, N50

Keywords: International Migration, Columbian Exchange, Historical Demography, Power Asymmetries, Colonialism, Disease, Land Endowments

"Europe has definitively taken over the government of the globe. To ensure a better distribution of men over the habitable surface of the planet—some countries are overcrowded, while others are empty—nothing could be more advantageous to each of them, and to what one might call the good administration of our planet, than to bring to those who do not have enough of the population that the others possess in excess."

— Le Temps (1899), cited in Sven Beckert,
Capitalism: A Global History

1 Introduction

1.1 Overview

From the first European conquests in the Americas until the wave of decolonization that followed World War II, world population experienced significant contact shocks and redistribution. Two continents—Europe and Africa—saw substantial exoduses, the first being overwhelmingly voluntary and peaking in the early 20th Century, the other overwhelmingly involuntary, and peaking in the early 19th. At least two others–North and South America–endured initial shocks of massive population contraction. While flows to the Americas seem "efficiency enhancing" given differences in Old vs. New World population densities circa 1600, low indigenous population densities in destinations resulted mainly from an 80 - 85% contraction in Amerindian populations due to their lack of immunity to Old World diseases. Many smaller international population flows also occurred. Viewing a map of the world, the most dramatic change was the replacement of indigenous states, chiefdoms, and tribal territories in the Americas, Australia and New Zealand, by newly created countries with Spanish, Portuguese, English or French as official languages and Christianity as dominant religion. Thus, European cultures became politically and in some cases demographically dominant on four of the world's continents (North America, South America, Australia and Europe itself), or some 44.5% of the world's territory (including of the 7.5% accounted for by Europe itself). The Americas came to contain some countries whose majorities were still mainly of Amerindian ancestry (e.g., Peru, Bolivia), but in which dominant elites embodied and spread the influence of European language, culture, and religion. The Hemisphere's other countries came to have people of mainly European descent (United States, Argentina, Canada, Uruguay) or of mainly African descent (e.g. Haiti, Jamaica, Cuba). Much of the remainder of Old World territory beyond Europe and Africa also underwent population shifts, although most were relatively modest when compared to Europe's colonization of the Americas, Australia and New Zealand.

The goal of this paper is to investigate the patterns of migration during the era of colonization using the data on global migration between 1500 and 2000 assembled by Putterman and Weil (2010), and in particular the version of that data adjusted by Chanda et al. (2014) to cover 1500 to 1960 only.[1] Whereas both Putterman and Weil (2010) and Chanda et al. (2014) treat the population shifts between 1500 and 2000 as exogenous data with which to study the influence of early ancestral experience on modern outcomes, we treat who moved and where as outcomes in their own right.

Our main interest in this paper is to investigate what accounts for which populations moved, and where. Hence, we treat ancestry composition changes as outcomes in their own right. This structural shift marks a clean departure from the foundational work of Putterman and Weil (2010). While the existing deep-roots literature utilizes the global ancestry matrix primarily as a tool to compute ancestral adjustments for modern macroeconomic outcome regressions, our study is the first to treat the bilateral, dyadic flows within the matrix itself as the dependent variable. We thus turn the spotlight onto the underlying determinants of the matrix, mapping out the multi-variable gravitational frictions that generated the modern world's ancestral baseline. We model these global ancestral shifts as a long-run macroeconomic rebalancing, where old world populations track the massive exogenous shock of land-abundance—arising from shattered local indigenous density and modified by bilateral translation frictions such as geographic distance and the asymmetry of colonial power or enforcement. While the 'reversal of fortune' thesis of Acemoglu et al. (2001, 2002) treats 1500 population density as an exogenous driver of early institutional choice (extractive vs. settlement regimes) which in turn shaped modern GDP, we show that 1500 density functioned

[1] 1960 is used as a rounded end-point of the era of colonization, since 56 former colonies of Britain, France, Spain, Portugal, the Netherlands and Belgium achieved independence during the late 1940s, 1950s and 1960s, another 29 in the 1970s and 1980s, and the median independence year of these 85 countries (which exclude countries that achieved independence prior to 1940) was 1962. Note that most countries of the Americas were politically independent from their original colonizing powers for more than a century before the end of the global era of colonization and that a large share of the European migration to what became the majority-European countries of that hemisphere occurred under local rule by dominant groups descended from the original colonizers.

as a powerful, direct gravitational constraint on the multi-century global migration flows themselves. We explicitly analyze pre-existing local population density as an impediment to demographic reconfiguration.

We examine how well a modest set of key factors can explain what shares of the descendants of various territories' year 1500 populations continue to reside within their ancestral lands, what shares were instead found in other parts of the world by 1960, and which parts those are. Putterman and Weil (2010) illustrate the characteristics of their migration matrix by pointing out, for example, that today far more descendants of Ireland's year 1500 population live outside of Ireland than within it, whereas the vast majority of the descendants of those who lived in Japan and Korea in 1500 still live in those countries today. We seek to understand not only why Irish but not Japanese and Koreans emigrated in substantial numbers, but also why the descendants of all three countries' people became more numerous in what is now the United States, Canada and Australia rather than in India, Nigeria, Brazil or the Philippines. And we perform the same analysis for as many of the world's countries with populations exceeding 0.5 million as possible. Specifically, we consider the populations of 162 countries, which together accounted for an estimated 99.1% of global population in 1960. Like both Putterman and Weil (2010) and Chanda et al. (2014), our data are organized at the level of territories that constituted countries circa 2000, which puts former republics of the USSR and Yugoslavia among our country units. Note that the territory of a country of 2000 is treated as a territorial unit in 1500 for these exercises, even if borders then differed (e.g., France, China) or if no corresponding sovereign country existed at the time (i.e., most countries of the world, including Brazil, Germany, Kenya, Indonesia, etc.). Indeed, only a handful of national entities existed in 1500 with largely the same borders as those of late 20$^{\text{th}}$ Century counterparts, but for convenience we write, for example, of "17$^{\text{th}}$ Century migration from Germany to the U.S.," to avoid the more cumbersome phrasing "from what is now Germany to what is now the U.S." We likewise use "migration" when referring to the involuntary movements of the slave trade, rather than write "migration and transhipment" in the many instances where that phrasing would be more complete.

While our analysis is conducted primarily at the level of territories that were countries in the late 20$^{\text{th}}$ Century, we sometimes adopt a more macro-regional outlook. Specifically, we invite the reader to think of the inhabited world of 1500 – 1960 as consisting of four 'super regions': the two that provided the largest shares of movers (Europe and sub-Saharan Africa), the one whose population origins were most dramatically changed ('greater New World' composed of North and South America and Australia and New Zealand),[2] and the remaining parts of Eurasia and North Africa. While this last "super-region"'s polities included no colonizing powers participating in the most lastingly transformative overseas colonization episodes of the era, the interaction between Europe and the Greater New World still had important economic and geo-strategic effects on it (e.g. Asia), and most North African and Asian territories also came under European colonial rule at least for a time. In addition to the "four super-regions" approach, we use a slightly finer regional breakdown at some points, treating Middle East and North Africa, Central Asia, South Asia, East Asia, Southeast Asia, the Americas, and Oceania (Australia, New Zealand, New Guinea, and Fiji–as we lack data for smaller countries of this region) as mid-sized regions additional to Europe and sub-Saharan Africa (the last two serving as both super-regions and regions, by our terminology). Appendix Table A.1 lists the countries included in each region, and Table A.2 lists the super-regions and the regions comprising those that are divisible.

We find the most important organizing framework for understanding global migration during the colonial era to be the idea that (1) long distance redistribution of human population selectively and partially reduced differences in population relative to land capable of supporting it following the post-contact decimation of indigenous populations, while transferring populations to warmer climates with higher agricultural potential and lower disease burdens, and that (2) while responsive to resource imbalances in those ostensibly efficiency-enhancing ways, population redistribution was also heavily influenced by political facts, especially which countries engaged in colonization (i.e. ones in Atlantic-facing Europe) and had the technological and military capability and economic incentive to engage in the slave trade, and which region lacked the social cohesion and political conditions to resist the selling of substantial numbers of its people into overseas slavery (sub-Saharan Africa).

Expanding on (1): because agriculture remained the most important economic activity and regions were mostly self-sufficient in food production during most of the period covered, a territory's ability to support population was largely a function of its area of arable land and that land's productive potential. Global population was unequally distributed relative

[2]Although the population of New Guinea, especially, was not as deeply transformed by colonization as those of Australia and New Zealand, we include it and Fiji in 'greater New World,' which thus includes all four Oceanian countries of our sample along with the countries of the Americas. This permits the simplification that "greater New World" equals what we treat as the regions of "the Americas" and "Oceania."

to fertile land, circa 1500, in part because independently developed suites of domesticated plants and animals had given Eurasia a global lead with respect to grains and domesticated animals and led to its earlier commencement of state building, as well as to a somewhat greater diffusion of sedentism and farming than in the Americas.[3] Some potentially cultivable regions such as the northern U.S. and Canadian plains, the Argentine Pampas and parts of eastern Australia and western South Africa, had previously seen little or no agricultural production. The American regions' "underpopulation" relative to more populous parts of Eurasia was increased by a full order of magnitude due to exposure of their populations, upon contact with European conquistadors and colonists, to Eurasian diseases to which they had not developed immunity (Denevan, 1992; Diamond, 1997; McNeill, 1976). HYDE estimates (revised 2023) already place New World population per ha. of arable slightly lower in 1500, at 0.174, than the corresponding levels for Europe (.275) and sub-Saharan Africa (.274), and much lower than those for North Africa and Asia (0.598)(HYDE, 2023). Owing mainly to the toll of disease, the Europe-Americas density difference (denominated by arable area) ballooned from a 1.6:1 in 1500 to a 12:1 ratio (0.354 vs. 0.029 people per hectare in Europe vs. the Americas) in 1600.

Since the adoptions of agriculture, movements of people and of their domesticated crops and animals had been limited to ones within the "Old World" of Eurasia and Africa, or within the Americas. Rice and millet moved from Asia westward, wheat and barley moved from the Near East eastward, cattle, sheep and goats moved from the Near East to India and Africa, and large human migrations including those of Indo-European, Bantu and Turkic language speakers also occurred within the "Old World." Within the Americas, Mesoamerican maize moved to South America and the present day U.S., and there were also large human migrations such as the Athabaskan, Uto-Aztecan and Tupi-Guarani migrations within North and South America and between the two. But a bridge between these worlds had been lacking, and thus land suited to growing wheat in Kansas or Saskatchewan, to grazing cattle in Ar-

[3] While Eurasia and the northern half of Africa developed numerous cereal crops including rice, numerous types of millet (independently domesticated in China, India and the Sahel), wheat, barley, oats, rye and sorghum, only one major cereal crop was domesticated in the Americas, maize. Knowledge of and experimentation with growing maize's wild ancestor may go back to almost as early a date as experimentation with wheat and millet, but its lengthy domestication process meant that it became the foundation for a mainly agricultural way of life two thousand years or more later than had wheat and barley (Larson et al., 2014). Domesticated animals in the size range of sheep, goats and pigs and in that of cattle, horses and water buffalo, were absent in the Americas. Only the llama and alpaca can be considered South American counterparts, in an intermediate size class, and these were largely restricted to the Andean region before European colonization. They were able to carry loads but not pull plows, so that soil preparation by plows was also absent in the Americas.

gentina and Texas, or supporting fruit, vegetable, and tree nut production in California, had been "underutilized" in terms of the set of domesticates available in global terms, while land suited to growing maize in Africa and China, potatoes in Ireland and Poland, and tomatoes in Italy, had likewise been "underutilized" (Chen and Kung, 2016; Nunn and Qian, 2010).

In addition to the important natural conditioning roles played in migration by availability of cultivable land and especially land with high productive potential, climate, disease, and proximity also played significant roles. Potential migrants, or where applicable their owners or labor recruiters, favored agriculturally advantageous climates, encouraging relocation to relatively warm locations. When able to do so, migrants avoided places with high disease burdens. And even moves overseas were more likely when entailing shorter distances (say, Ghana to Venezuela rather than Ghana to Chile, England to the U.S. eastern seaboard versus England to Australia), and moves to landlocked countries (say Paraguay, Niger) were less likely, all else equal.

Regarding the factors grouped in (2), geostrategic power differences tilted population movement away from what would have been predicted had the migration process been one of decentralized efficiency-seeking adjustments in a world without states and state boundaries. The most important examples are that (i) it was in the first instance countries in western Europe that established colonial rule over the continents previously unknown to Eurasia, rather than Eurasia's most populous and until shortly before more economically advanced parts, and that (ii) traders from the colonizing countries involuntarily moved a substantial share of the population of sub-Saharan Africa to serve as an enslaved labor force, thanks to the dearth of overarching African polities, the willingness of some African rulers to capture and sell non-co-ethnics in exchange for desired tradeables including guns, and perceived need for such labor due to the thinning of Amerindian populations by post-contact mortality. Other power-related rather than directly resource-based factors contributing to the skewing of long-distance migration away from what unequal population-to-arable land ratios, unequal land productivity, climate and disease alone would have predicted, include (iii) that it was from an ocean-facing subset of Old World states with long histories of state-building, military and metallurgical knowledge, a subset later to state-building than Eurasia's most ancient cores but still benefiting from technological cross-fertilization across the breadth of Eurasia, that colonizing powers arose, powers that proved able to vanquish still-younger New World empires lacking similar advantages (Diamond, 1997; Ertan et al., 2016), and (iv) that when slavery was outlawed as a source of low cost labor, migration of impoverished indentured and low-wage workers was arranged from densely populated colonies to less populous colonies of the given colonial power, e.g. “importing” laborers from India to British colonies such as British Honduras (today’s Belize) and British Guyana (today’s Guyana).[4]

Related to these power-based conditioners of migration, and likewise largely orthogonal to natural factors like land availability, are cultural determinants of greater colonizer preferences for accepting immigrants from some societies rather than others. Linguistic, cultural and religious proximity appear to be the main factors at work here. Familiar cases include the large numbers of Irish migrants who followed English and Scottish ones to North America and Australia, and Italian migrants who followed Spanish ones to Argentina. Although there are clearly exceptions, for example Morocco being geographically closer to Spain than Italy is, relative cultural, linguistic and religious proximity can be proxied adequately for our purposes by geographic distance, especially for countries within Europe. This would help explain why the numbers of Irish, Dutch and German migrants to the English colonies of the U.S. eastern seaboard was larger than the numbers of Italian, Polish or Russian migrants until the late 19th century, with dramatically higher diversity of migration sources picking up only as our period of study ends.

It is important to understand that while we have referred to conditioners or determinants of migration, our models properly provide estimates of statistical association only. We defer discussion of the issue until our modeling section, but acknowledge at the outset that our analytical work does not undertake to demonstrate causality, so words like ”determinant” apply in a loose sense, only.

We do not attempt, in this paper, to enter into the important debate about why it was western Europeans, among a more diverse set of technologically advanced and politically capable Eurasian powers on the eve of the early modern period, that ventured beyond the Old World and became world colonizers. It is important to our topic, however, to recognize

[4]An alternative way of breaking down the kinds of factors that influenced the population movements in question would treat the profit-seeking motives and financial leverage of financiers, plantation owners, etc., as roughly equal to and thoroughly intertwined with the territorial and military power motives of states, as described by Beckert (2025). Likewise, a considerable part of what is interpreted as efficient reallocation of labor in standard economic analysis and is not always explicitly distinguished from the latter in this paper would be labeled as either entirely (slavery) or largely coercive (penal exile, impressment into naval service) or only partially voluntary (indentured migration) in papers which place emphasis on economic power relations. We acknowledge our probable overuse of the term ”efficiency” due to the fact that our paper’s focus tends to leave political-economic concerns in the background

that Old World states including the Ottoman, Mughal, and Ming and Qing empires were as technologically advanced, in control of more territory, and had rulers commanding more wealth, than the early European colonizers, Spain and Portugal. It is also relevant that their neighbors including Safavid Persia, Russia, Korea, Japan, and parts of Southeast Asia, were also populous states with significant military capabilities, during the 16^{th} to 18^{th} centuries. Their large populations and their ability to rival the colonizing powers militarily within their home regions contribute importantly to both the geostrategic power and the human-to-natural resource balance factors explaining who migrated where, during 1500 – 1960. They did not absorb long-distance migration from distant lands on a large scale in part because several (e.g. Persia, China, and Japan) had the state capacity to stave off colonization (see Ertan et al. (2016)), and partly because, as long-lived agrarian societies, the existing populations in their arable regions meant that massive exercise of force would have been required to make room for European farmers or to permit a European colonizing elite to oversee landless or enslaved indigenous farm workers under institutions of the sort that formed in the Americas (Haciendas and Latifundia). Two of our explanatory variables play overlapping roles in representing these two factors. First, places that became major receivers of migration tended to have significantly shorter, if any, histories of rule by large political entities like states or empires. In contrast, places with long state histories like Anatolia, Persia, India, and China, were probably not significantly settled by colonists, even when political rule by European colonizing powers did occur, in part due to the anticipated resistance this would provoke (given that most of their suitable land was already being cultivated). In addition to state history, the ratio of population to arable land can be thought of as a barrier to colonization in its own right—or at least, the population to land ratio can be used alongside state history as another variable likely to capture a force working against large colonial settlement. Low or negative net migration to countries of high population density is part of the first explanatory factor discussed above. Two other natural factors—high disease burdens and high average temperature—may have played similar roles for some less densely populated places with shorter state histories, like Malaya and less populous islands of Indonesia. The high correlation between long state history and high population per unit of cultivable land poses a statistical challenge, discussed later.

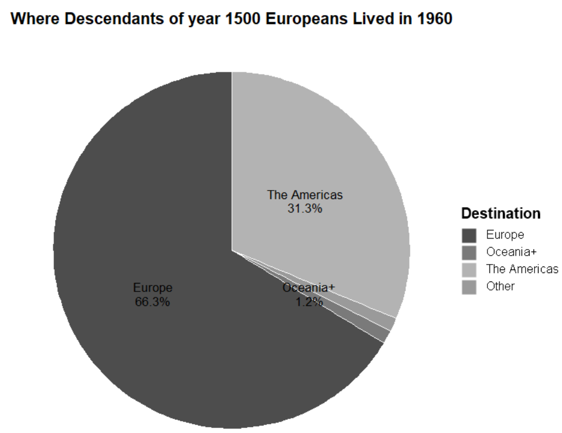
We acknowledge that countries are on average large, and also quite unequal in terms of territory and population, so using data at the level of provinces, states, etc., within the larger countries, would clearly be preferable, were such data available. In our study, the problem of unequal territorial and population sizes is partly mitigated by the fact that many world regions include both large and small countries–for example, South Asia is dominated in population and size by India, but also includes Nepal, Bhutan and Sri Lanka, along with Pakistan and Bangladesh. Europe includes the relatively small Baltic states and former Yugoslav republics along with large countries like Germany and France. Southeast Asia, North Africa and Middle East, and sub-Saharan Africa each contain multiple countries that vary in size. The multiple observations of varying size in most regions then yield thousands of pairings with the countries of other regions, allowing for a fairly rich picture of the direction of population shifts between them. Since each included country appears in sender and receiver pairs with well over 100 other countries (pairings with countries of its own region are omitted—see below), our estimations' sample sizes reach over twenty-thousand observations, consisting of slightly over ten thousand pairs of countries each yielding two sender-receiver observations.

In the rest of the paper, we conclude the introduction with some preliminary descriptive facts and figures. Section 2 then discusses the data employed and its translation into the variables used in our analysis. Section 3 presents our predictions, which involve further detailing the "nature related" and "power related" explanations introduced in groups (1) and (2) above, and spells out the reduced form models used to test them. Section 4 presents our results, beginning with estimates of factors explaining each country's descendants by location, then additional models to account for post-contact "die-off," then estimates of "stayer share" models. We find considerable support for the influences discussed above and detailed in Section 3, while also noting a few unexpected results. Section 5 summarizes and discusses conclusions, including simple counterfactual calculations of regional populations in the absence of migration and a perspective on the implications for more recent migration patterns.

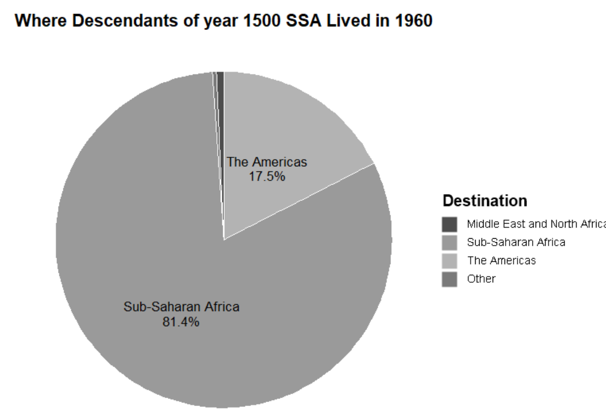
1.2 Preliminary Impressions

Figure 1 provides a visual sense of what shares of the year 1960 descendants of year 1500 populations were estimated to still be found in their home regions of Europe, sub-Saharan Africa, and North Africa plus Asia, versus in the Americas and other regions. As expected, the first two regions each had substantial diasporas, 33.5 and 18.5% of their descendants, respectively, living in other regions, primarily in the Americas, making Europe's diaspora (relative to its home population) a little short of twice as large as SSA's. In contrast, the North Africa and Asia super-region still had almost 99% of its descendants within their home regions, with the preponderance (78%) being in South and in East Asia.

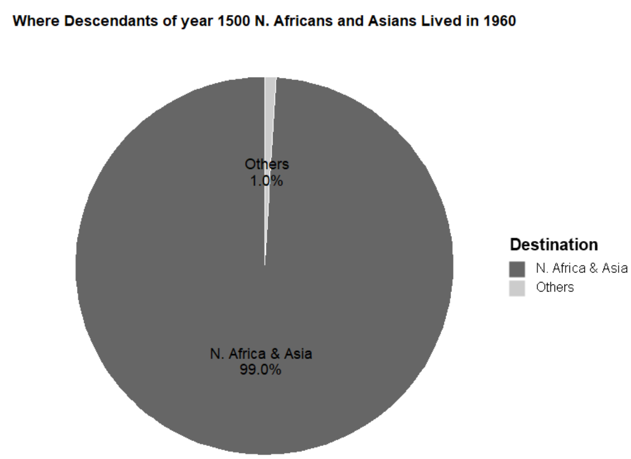
Figure 1: Descendant Structures for Europe, Sub-Saharan Africa, and North Africa and Asia



(a) Europe



(b) Sub-Saharan Africa

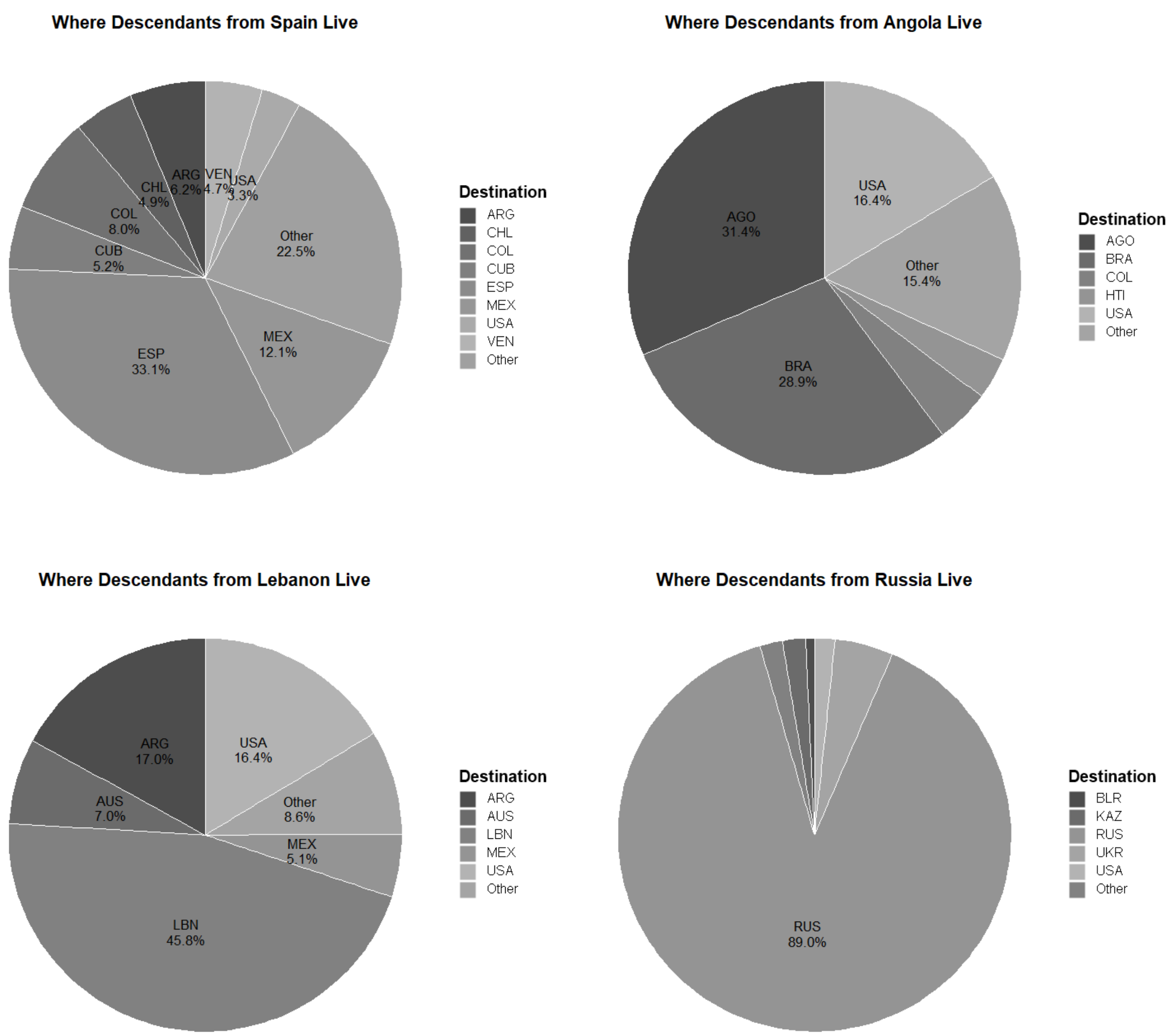


(c) North Africa and Asia

Notes: Figures display the shares of the descendants of the year-1500 population of the indicated region estimated to be living in 1960 in the same region versus other regions. These shares are calculated from the descendant share numbers in the migration matrix of Chanda et al. (2014) and population estimates for each country from the World Bank (2024).

Figure 2 provides similar charts for a few illustrative countries. The upper left pie chart shows that Spain, the first colonizer of the Americas, which thereby gave its language to 18 of that Hemisphere's 25 countries (omitting small Caribbean island nations) and Puerto Rico (treated in our paper, as in Putterman and Weil (2010) (hereafter PW), as a "country"), was estimated to have more descendants in the Americas (a total of 64.9%) than in Spain itself (33.1%). The upper right pie chart shows a corresponding break-out for Angola, which was among the large sources of slaves sent to the Americas. The share of descendants remaining at home (31.4 %) is only slightly larger than that of descendants residing in Brazil in 1960 (28.9 %), with another substantial share (16.4 %) estimated to be in the US, and the remainder in Colombia, Haiti and other countries.

Figure 2: Descendant Structures for Spain, Angola, Lebanon and Russia



Notes: Figures display the shares of the descendants of the year-1500 population of the indicated country estimated to be living in 1960 in the same country versus other countries. These shares are calculated from the descendant share numbers in the migration matrix of Chanda et al. (2014) and population estimates for each country from the World Bank (2024).

To bring home the fact that European and sub-Saharan countries were not the only ones sending large populations to distant destinations, we also display the pie chart of the atypical Near Eastern country of Lebanon, which borders the Mediterranean and had a history of projecting colonists overseas dating back (as Phoenecia) to the second millennium BCE. It also had more than 54% of its 1960 descendants estimated to be living elsewhere, including substantial numbers in Argentina and the U.S. These shares actually exceed those of some European countries distant from the western European colonizing powers, for example Bulgaria and Slovakia have estimated external shares of 5% and 20% respectively. Finally, we provide a breakdown for Russia (within the borders of the post-Soviet Russian Federation), of interest in part because its land empire in the 18th to 20th centuries, followed by its central role in the Soviet Union, led numerous Russians to take up residence in neighboring countries, with many still there in the post-Soviet era. Inspection of the Russian and Spanish pie charts, however, suggests that the long-term impact of Russian outmigration was quite small by comparison.[5] The relatively modest share of Russian descendants estimated to be in the U.S. (1.7%) and other countries of the Americas (0.7%) helps to illustrate, along with the Bulgarian and Slovakian examples, the relatively small impact of Europe's colonization on some countries near Europe's eastern edge.

2 Data and Variable Construction

Our data on shifts of global population between 1500 and 1960 is taken from Chanda et al. (2014) (hereafter CCP)'s adaptation of the world migration matrix of PW. CCP converted PW's estimates of five centuries of population movements into estimates for 1500 – 1960—the period they and the present paper treat as the era of European overseas colonization—by subtracting and adding, as required, numbers of people reported as having migrated between the represented pairs of countries during the years 1960 – 2000. Migration data for 1960 – 2000 is based on estimates by Özden et al. (2011), relying on data from the U.N. Population Division's Global Migration Database supplemented by detailed census analyses for individual countries.[6] The basis for PW's estimates of population movement, which deter-

[5]The degree to which Russians dispersed from their original–say, year 1500–heartland in those centuries is understated, since many areas to the south and east of the earlier Russian core are today within the Russian Federation, so the relevant population shifts do not count as international migration.

[6]Özden et al. (2011) had access to detailed census analyses for 226 countries, of which the 176 countries covered by PW account for the overwhelming majority of population and of countries having at least 0.5 million people in year 2000.

mine the bulk of the population shifts in CCP's and our own estimates (since the Özden et al. (2011) adjustment accounts for only 8% of the years within the five centuries 1500 – 2000), is qualitatively different from that of Özden et al. (2011) Lacking global migration estimates for the longer period, PW relied primarily on late 20th Century efforts to account for the ethnic compositions of countries in encyclopedias and encyclopedia-like projects such as *CIA World Factbook*, *Encyclopedia Britannica*, *World Christian Encyclopedia*, *Library of Congress Country Studies*, and *JoshuaProject.net*. They averaged the estimates from the sources they deemed plausible and adopted the strong simplifying assumption that the ascribed ethnicity (for example, of Indians and Chinese in Singapore or of people of Spanish or Italian origin in Argentina) accurately reflects the actual ancestral origins of those covered. They used DNA-based estimates of the share of ancestors from the presumed main source groups (for example, Amerindian, Spanish, and African, for the "mestizo" population in Mexico) to assign shares of ancestry to members of ethnicities listed by their sources as having multiregional ancestry—e.g., groups called "mestizo," "mulatto," and "colored." To estimate origins within Africa according to boundaries of countries of today, PW relied on the detailed data on the trans-Atlantic slave trade by Eltis et al. (1999). Sources for other special cases, such as estimating where ancestors of people simply identified as Jews or as Gypsies had lived in 1500, are described in PW's Appendix. The differing methods were "interacted" where appropriate, for example the DNA analysis generated a share of African ancestors in Mexican mestizos, and the Eltis data were then used to assign sub-shares to specific countries-of-today within Africa.

As a caveat, we think it worthwhile to note that the compilation procedures employed by PW allow their data to provide a reasonable if rough approximation for the main ancestral origins of populations in countries having major changes in population origin during the period covered (especially countries of the greater New World), but it probably fails to pick up flows across borders within Old World regions. To give one example, the large majority of the year 2000 population of France is identified by the sources mentioned as being ethnically French, so the PW and CCP data treats all such individuals as having had all ancestors alive in 1500 residing within the territory of today's France in that year. This almost certainly overlooks a considerable number of cases in which genealogically curious families have grounds for believing themselves to have had ancestors who moved to today's French territory from the territories of neighboring or more distant European countries at various times between the 16th and early 20th centuries, with descendants becoming culturally assimilated over the generations. Neither PW nor CCP nor the present paper undertake to estimate such flows. Given that our main interest lies in "redistribution" of populations and cultures among the world's regions (from Asia to Latin America, and so on), and that the econometric analyses shown leave out observations for country pairs within the same region, the impact of this drawback is fairly unimportant, but finer-grained data on migrations built out of whatever detailed migration records can be constructed by better resourced scholars remains an interesting project for the future.[7] More generally, the migration matrix fails to pick up large numbers of small population flows that must have occurred over an almost five century period between every possible pair of countries, especially when a flow can entail a succession of movements in different generations. Zeros are more common than positive entries (see below) but are best understood as being approximations for small numbers not picked up in the matrix construction process. The conclusion that fully 35 countries in the sample had 100% of their descendants within the home territory in 1960 should be understood to be an artifact of this kind of imprecision in the development of the initial estimates.

The adapted version of the World Migration Matrix (WMM) is a 162-by-162 matrix in which country names serve as row and column labels.[8] The number of countries by region can be seen in Table 1.

[7]Researchers determined to do so could conceivably assemble large amounts of relevant material from family genealogies, church records, and other sources. Rough estimates could also be generated by examining surnames for signs of foreign ancestry and estimating the rate of surname transmission taking into account that during the period covered only male offspring typically retain their surname at birth throughout their lives.

[8]The WMM has 172 countries, however there are only 162 countries available for use because seven mostly small island countries in the original matrix that only have information as the destination country, which is not suitable for our bilateral structure (Channel Islands, Iceland, Niue, São Tomé and Príncipe, Tonga, Saint Vincent and the Grenadines, and Samoa), are excluded. Additionally, two countries (Taiwan, and East Timor) are missing information in other variable datasets that are required in our analyses, and there is a "Yugoslavia" entry that we omit in favor of data for the former Yugoslavia's constituent republics.

Table 1: Number of Countries by Region

Region	N
Central Asia	6
East Asia	6
Europe	38
Middle East and North Africa	21
Oceania+	4
South Asia	6
Southeast Asia	9
Sub-Saharan Africa	45
The Americas	27
Total	**162**

It is also worth reminding the reader that ”migration matrix” is to some degree a misnomer, because the numbers of descendants of each country at home and around the world is determined not only by migration but also by fertility, and along with both home and migrant families' fertility and mortality rates, the timing of migration would also be important in many contexts. The matrix attempts to estimate outcomes, but does not try to disentangle migrant numbers, timing, or population growth rates.

The situation is still more complex when mixed ancestries lead to our counting numbers of descendants of each source country based on the aggregation of fractional persons (for more details, see PW). For example, members of the mestizo population of Mexico or Peru, are on average each considered as being (simplifying, for illustration) about 2/3 of a descendant of one group (Amerindians) and 1/3 of a descendant of another (Spanish). In this case, following PW's approach, we would count 3 million mestizo residents of 1960 as if they were 2 million Amerindian descendants and 1 million Spanish descendants. Many of the notional Spanish descendants counted in Figure 2's illustration are indeed such fractional mestizos,or fractions of Argentines having both Italian and Spanish ancestors, etc.

Our analysis does not venture into efforts to disentangle these factors, as our purpose is solely to understand what accounts for where the descendants of each source country–whether they be “pure” or “admixed”–ended up living, as of the end of the colonial era. The CCP matrix does not attempt to estimate ancestral arrival time. We nevertheless suggest to keep in mind that our outcomes reflect not only numbers in the migrating generations but also their eventual fecundity in the places they moved to. In consequence of this, factors such as the caloric productivity of the destination country's farmland, its climate, the disease risks migrants encountered there, and the working conditions and socioeconomic constraints they faced as a result of institutions, power and inter-group relations, are likely to have influenced the number of émigré descendants in the latter country not only by affecting the level of attraction that word of mouth about the destination exerted on potential migrants back home, but also through the positive effects that, say, a propitious climate, fertile soil, and low risk of disease had on the number of 1960 descendants per migrant of 1600, 1700, 1800, etc.

Table 2: Data Coverage

Dataset Structure	
source countries	153
destination countries	162
total dyadic observations	24,786
Estimation Samples	
Global sample	20,194
Europe (Source) sample	4,712
Sub-Saharan Africa (Source) sample	4,914
North Africa & Asia (Source) sample	5,130

Notes: "Total dyadic observations" reports the number of origin–destination pairs in the dataset. Sample sizes report the number of maximum observations omitting same-region pairings used in the corresponding estimations.

We transformed the matrix form dataset into a long-format dataset organized by origin-destination pairs. In this format, each country appears 161 times as an origin and the same number of times as a destination. For example, population movements from Greece to Afghanistan and from Afghanistan to Greece are recorded as separate observations. However, 9 countries record zero outward migration to all potential destinations, leaving us with 153 source pairings for each of 162 destination countries.[9] Our main estimations exclude country

[9]The countries that account for 0% of ancestors in other included countries are Cape Verde, Haiti, Hong Kong, Jamaica, Mauritius, Qatar, Singapore, Swaziland, and Trinidad and Tobago. Absence of recorded descendants of year 1500 population living abroad in 1960 follows for Cape Verde and Mauritius because those islands are assumed to have been unpopulated in 1500, while for Haiti, Jamaica, and Trinidad and Tobago, the indigenous Taino-speaking people are assumed by most sources to have died out due to disease and overwork shortly after European arrival. The remaining countries had fairly small populations in 1500. Apart from the two uninhabited islands, the 0 values for descendants abroad resemble those of the large

pairs located within the same region, as our primary interest lies in long-distance, inter-regional migration; regional definitions and the rationale for this restriction are discussed in Section 3. Table 2 summarizes the resulting sample sizes.

Explanatory Variables

Below, we provide an overview of the key drivers and control variables used in the analysis. Detailed descriptive statistics are presented in the Appendix.

1. **State History, 1500:** The variable *"State History, 1500"*, introduced by Borcan et al. (2018, 2021), builds upon the earlier index developed by Bockstette et al. (2002). It measures the historical existence and persistence of centralized political structures within the geographic boundaries of present-day countries, reaching back to the earliest recorded states in Mesopotamia. The index divides history into 50-year segments and assigns scores based on three criteria: the existence of governance beyond tribal organization, whether the ruling authority was domestic or foreign, and the extent of territorial control.[10]

2. **Population per Arable Hectare, 1500:** Historical population in 1500 CE is constructed from the 2023 update of the HYDE database (HYDE, 2023; Klein Goldewijk et al., 2017), which we access via Our World in Data (2026), which provides historical population estimates from 10,000 BCE onward.[11] To account for agricultural potential, we divide the 1500 population by the arable land area in hectares, as reported by the World Bank. Because dramatic shrinkage of the pre-Colombian population of the Americas caused the destination country population densities observable to most migrants of later centuries to be far lower by 1600 (for example, HYDE's 2023 estimate puts Mexico's 1600 population at only about 4% of its population in 1500 and Koch et al. (2019) document a hemispheric depopulation rate of approximately 90%, resulting in the abandonment and secondary succession of 56 million hectares of anthropogenic

majority of country pairs in the WMM; there is no reason to assume that literally no descendant exists in each destination country, but flows were probably very small, leading to an absence of mention in the consulted sources.

[10]In the original dataset, the following 12 countries are missing: Bahrain, Belize, Bhutan, North Korea, Kuwait, Luxembourg, Malta, Oman, Puerto Rico, Qatar, the United Arab Emirates, and Guinea-Bissau. For these countries, new State History indices were calculated by us using the same methodology as in the original paper. Detailed notes on the countries and calculations are provided in the Appendix.

[11]For two missing countries (Mauritius, Ireland) estimates are supplemented using data from the Maddison Project Database (Bolt and Van Zanden, 2025). The details are explained in the Appendix.

land by the early 1600s) we also use *population per arable hectare, 1600*, from the same source, for some exercises.

3. **Historical Disease Prevalence:** Developed by Murray and Schaller (2010) and complementing the work of Fincher et al. (2008), this variable measures the historical prevalence of several major infectious diseases based on early epidemiological studies (Rodenwaldt and Jusatz, 1961; Simmons et al., 1944).[12] The index is constructed by averaging the z-scores of each disease, with higher values indicating greater historical exposure.

4. **Pre-1500 Average Maximum Calories per Hectare:** It is one of the caloric suitability indices created by Galor and Özak (2016) which measures the average maximum potential caloric yield across a country's cultivable grid-cells, assuming the optimal pre-1500 crop for each location. The crops included are those available in the relevant region before 1500 CE and these averages are calculated for the contemporary country borders.[13]

5. **Colonial Relationships:** Several indicator variables to capture colonial status and colonial ties between source and destination countries are created, using colonization data from Head et al. (2010) and Conte et al. (2022).[14] *Source is Colonizer* equals 1 if country i was a colonizing power. *Source is Colony* equals 1 if country i was colonized at any point during 1500–1960. *Destination is Colony* equals 1 if country j was colonized during this period. *Destination is Colonizer* equals 1 if destination country j was the colonizer of source country i (e.g., i = Algeria, j = France). *Colonizer of Destination* equals 1 if source country i was the colonizer of destination country j (e.g., i = United Kingdom, j = India). *Same Colonizer* equals 1 if both source and destination were colonized by the same power (e.g., i = Angola, j = Brazil, both colonized by Portugal). Each indicator takes value 0 if the condition described does not hold.

[12]The seven diseases are leishmanias, trypanosomes, malaria, schistosomes, filariae, leprosy, dengue, typhus, and tuberculosis. Due to large number of missing values in the full version, we use the 7-disease version (excluding leprosy and tuberculosis.), which offers full coverage for our sample. Sensitivity checks with the 9-disease version is provided in the Appendix.

[13]There are several versions of this measure. In our study, we use the one that excludes Japonica and Indica rice in sub-Saharan Africa, due to uncertainty about its time of arrival prior to 1500. More details about the variable versions can be found in Galor and Özak (2016) and the author's website: https://ozak.github.io/Caloric-Suitability-Index/

[14]Details on the construction of these variables and the complete list of colonizer countries are provided in the Appendix.

6. **Temperature:** We use country-level average temperature data from Ashraf and Galor (2013). Obolensky et al. (2024) demonstrate that historical migrants exhibited strong "climate matching" behavior—seeking out destinations with familiar temperature profiles to maximize the transferability of their specific agricultural skills and human capital.

7. **Arable Land Area (Hectare):** In addition to *Population per Arable Hectare, 1500*, we include source and destination *Arable Land Area* from World Bank (2026).

Control Variables

1. **Bilateral Distance:** Our baseline distance measure from Bertoli et al. (2016) combines *maritime navigation distance* with *land distance* (for details, see Appendix.).

2. **Landlocked Dummy:** This indicator equals one if the destination country has no coastline (i.e., no direct access to the sea), and zero otherwise.

In order to conserve space, we show pairwise correlations in Table A.3 and Table A.4, and present descriptive statistics of all variables in Table A.5, both in the Appendix. Importantly, ln(*statehist1500*) and ln(*pop1500/arable ha.*) have positive correlation 0.386, while ln(*(pop1500/arable*) and ln(*arable hectare*) have negative correlation -0.280, both significant at the 1% level. The first correlation makes it difficult to gauge the relationship of each to dependent share, leading us to show variant regression specifications with only one or the other, in addition to "horse race" versions that include both.[15]

3 Models and Predictions

We focus on two main kinds of models, of which the first has four sub-divisions. We discuss also some additional variants that are reported in the Appendix. The main models attempt to explain (1) the proportion of each country's year 1500 population's descendants who are estimated to live in 1960 in each other country outside of its home region, and (2) the degree to which each country's population out-migrated between 1500 and 1960 as indicated by its inverse, the share of 1960 descendants still estimated to be in the source country—called

[15]Other statistically significant correlations include a negative correlation between *colony* (ever colonized status) and *statehist1500* (consistent with Ertan et al. (2016)) and positive correlations between *colony* and both *temperature* and *historical disease prevalence*.

"stayer share." The explanatory variables of the two types of models largely overlap, although which variables have greater predictive powers differ.

We denote the dependent variable for the first type of model as DS_{ij}, where "DS" stands for "descendant share," i is the country of year 1500 population ancestors, and j is the country of residence in 1960. The calculation can be expressed formally as follows:

$$DS_{ij} = \frac{\text{Pop1960}_j \times \text{WMM Value}_{ij}}{\sum_{k \in \mathcal{J}} (\text{Pop1960}_k \times \text{WMM Value}_{ik})} \tag{1}$$

In this equation:

- DS_{ij} represents the DS value for a pair of countries i and j, where i is the source country and j is the destination country.
- Pop1960_j denotes the population of the destination country j in the year 1960.[16]
- WMM Value_{ij} corresponds to the ancestry-matrix share from WMM for the country pair i and j.
- The denominator is the total number of descendants of source country i living anywhere in 1960. It is constructed by converting each destination share into a headcount using destination population and then summing these implied headcounts across all countries k in the set of all countries ($\mathcal{J}$) which includes i itself.

We estimate two main and two auxiliary types of regression model to test our hypotheses about the determinants of what countries outside their home region the descendants of year 1500 populations were found in as of 1960. We use all country dyads for which the required data are available, except those for pairs of countries located within the same sub-region. Each dyad (say country i of region x and country j of region y) is represented by two observations: one in which i is the source (sender) and j is the destination (receiver), and the other in which j is the source and i is the destination.

All models are estimated by Poisson Pseudo Maximum Likelihood (PPML), a preferred approach for dyadic gravity-style models in international trade and migration research (Beine et al., 2016). A first practical advantage is its ability to handle data sets in which the dependent variable includes a large number of zeros (Silva and Tenreyro, 2006, 2011, 2022).

[16] From World Bank (2026).

This property makes it particularly effective for sparse datasets such as ours.[17] A common workaround in earlier studies has been to replace zero values by adding a small arbitrary constant. However, as Head and Mayer (2014) emphasize, this approach is problematic because it makes the results dependent on the units of measurement and undermines the interpretation of gravity coefficients as elasticities.[18] Another strategy has been to completely drop zero-valued observations, yet recent literature increasingly regards the preferred approach as applying PPML while retaining the complete dataset, including zeros (Larch et al., 2025; Yotov et al., 2016). A second major advantage is that PPML delivers consistent estimates under general forms of heteroskedasticity that are common in bilateral data (Silva and Tenreyro, 2006, 2022).

The term "Poisson" often causes confusion, but PPML is not restricted to count data, and it is frequently applied to continuous dependent variables, such as trade or migration ratios (Alesina et al., 2016; Beine and Parsons, 2015; Beine et al., 2011; Foubert and Ruyssen, 2024; Larch et al., 2025; Silva and Tenreyro, 2022; Yotov et al., 2016).[19] Beyond point estimation, inference in gravity models crucially depends on how standard errors are computed. Given the lack of consensus on the optimal approach– robust standard errors, one-way clustering, and multi-way clustering are all common in the literature–it is recommended to report results under alternative clustering schemes. Larch et al. (2025) recommend experimenting with clustering by (i) country-pair, or (ii) two-way clustering by exporter and importer.

The first main model type takes as independent variables the *difference* between the values for destination and source, for factors like population density that are defined in the same way for both source and destination countries. For example, we hypothesize that all else equal, there is a larger share of country i descendants in country j the smaller is the population density of j relative to i in 1500 (computing density by amount of arable land rather than overall territory), so we subtract pop/arable of source i from that of destination j to form our independent variable controlling for this factor. In addition to density (pop/arable), we use the *difference* form for arable land area, historical disease prevalence,

[17]Zero entries constitute more than 85% of the DS_{ij} data, due to failure of the sources used to pick up small flows not indicated by identifiable ethnic minorities, as mentioned above. Correia et al. (2020) specifically designed PPML to identify and drop "separated" observations (where fixed effects perfectly predict zero flows).

[18]PPML estimates the aggregate elasticity, while ordinary least squares (OLS) and Gamma PML estimate the average elasticity.

[19]We used the *ppmlhdfe* command developed for STATA by Correia et al. (2020), which is the most recent and widely used one in recent studies.

avg. max calories per hectare (a measure of land productivity), state history, and average temperature.

For those variables of which we take natural logs, we use the difference of the logs (not log of the difference). In addition to the six variables captured by the destination minus source difference, other variables enter our estimating models for which the difference form is infeasible or inappropriate. These include landlocked dummy variables (entered separately for source and for destination), the distance between the two countries, and the set of indicators relating to colonizer/colony status, namely: *destination colony*, *colonizer of destination*, *same colonizer*, and *destination is colonizer*.

$$DS_{ij} = \exp\Big(\alpha_0 + \boldsymbol{\beta}'\big(\mathbf{X}_i - \mathbf{X}_j\big) + \boldsymbol{\gamma}'\mathbf{Z}_{ij}\Big)\,\varepsilon_{ij}. \tag{2}$$

The second main model type replaces the difference terms of the first type with their source and destination components as distinct independent variables for which separate best-fitting coefficients are estimated. Thus, in place of the difference between, say, arable area of destination and that of source entering the model as one independent variable, arable area of source and arable area of destination enter as separate additive terms to each of which the model fits a coefficient, and likewise for population per arable hectare and each other variable entered in difference form in the previous model type. While the difference variables of the first model also include the destination and source values of each variable, that model finds a best fit under a strong set of constraints on the coefficients of each pairable destination and source variable. The unconstrained or non-difference-form models also includes landlocked, distance and the colony indicators, as above.

$$DS_{ij} = \exp\Big(\alpha_0 + (\boldsymbol{\beta}^D)'\mathbf{X}_i + (\boldsymbol{\beta}^S)'\mathbf{X}_j + \boldsymbol{\gamma}'\mathbf{Z}_{ij}\Big)\,\varepsilon_{ij}. \tag{3}$$

The key distinction between the model represented by Eq. (3) and that of Eq. (2) lies in the constraints imposed. Equation 2 tests the strict assumption of "symmetry in push and pull frictions". Relaxing this in Equation 3 is theoretically vital because structural gravity literature consistently shows that origin characteristics (push factors) and destination characteristics (pull factors) almost never operate with perfectly symmetric elasticities. Equation 2 requires that the coefficient on each destination characteristic equal the negative of the coefficient on the corresponding source characteristic ($\beta^D = -\beta^S$), embedding the assumption that push and pull forces are mirror images. Equation 3 relaxes this restriction, allowing β^D and β^S to differ freely. The separate-coefficients specification can therefore reveal whether a factor operates primarily on the sending side, the receiving side, or both—and whether the magnitudes are comparable. Because Equation 2 is nested within Equation 3, the two yield similar results when symmetry holds empirically. Wheresource and destination effects differ in magnitude or sign, however, the constrained model may produce attenuated, biased orinsignificant coefficients. We present both sets of estimates and note where they diverge.

Along with these models, we present a model in which the variables that enter in difference form or as separate destination and source pairs in the first and second model types, respectively, are included only as destination country values, with source country characteristics controlled by source country fixed effects. Finally, we present a complementary model in which only source country values are included and destination country characteristics are captured by destination country fixed effects.

The role of the state history variable and its high correlation with population density *(pop/arable)* requires attention. As Ertan et al. (2016) show, places that had experienced long histories as large, locally-based polities, were less likely than others to have been colonized, and if colonized, it occurred later in the colonial era. They suggest that the reason for this is that state history is associated with proximity to the global technological frontier, and that more experienced and technologically advanced states could resist military conquest by the western European colonial powers.[20] A problem, however, is that the earliest states came into being following millennia of gradual domestication of plants and animals culminating in agricultural intensification and increasing population density on and around land suited to crop production. In consequence, there is a high correlation across countries between population density, denominated by arable area, and state history ($\rho = 0.42$). That correlation produces a problem of multicollinearity when the variables appear together on the right hand side of regression models, causing the coefficients on both variables to suffer from large variance. While there are also grounds for worrying about correlation between *(pop/arable)* and *(arable area)* arable area, since the latter is the denominator of the former,

[20]Ertan *et al.* (2016) show that a longer state history is associated with both a significantly higher likelihood of not becoming colonized by the European colonizing powers, and a later year of colonization, if it did occur. A similar finding is reported in Hariri (2012).

their correlation (in the log format employed) is actually smaller, $\rho = 0.07$. Fortunately, the problems are not so acute as to render these coefficients statistically insignificant, in most specifications, but it is nevertheless helpful to compare specifications that include only one or the other of *(pop/arable)* and *(statehist1500)* with specifications including both, to obtain a fuller picture of the relationship each one has to migration, in the face of the other controls.

In addition to the two kinds of estimates of the above models covering all inter-regional country pairs in our global sample, we also report estimates for three restricted sub-samples. This is motivated in part by our interest in whether the way in which the various factors studied affect migration between Europe and the rest of the world is qualitatively different from that for migration between sub-Saharan Africa and the rest of the world. Since most of the European migration went to parts of the greater New World voluntarily while most African migration to the greater New World was involuntary, there might be differences in the patterns seen. For example, we would expect Europeans to have avoided the hottest climates and places with higher disease risk more than those transporting Africans to be sold as slaves avoided such ports of disembarkation, in the first case because they perceived their African captives to be acclimated to hot climates and in the second because slave owners may have placed a lower value on their slaves' lives.[21] We also report restricted sample estimates for the country dyads in which one member is located in North Africa and Asia, a form of "rest of the Old World" super-region. With this, our region-by-region estimates cover all three mainly-migrant-sending super-regions. We report estimates of models for the three sending super-regions using both the difference approach and the approach separating source and destination values.

Finally, we compare the models discussed, in which the population densities of receiving countries in the Americas are represented by their pre-Colombian estimates, to ones reflecting the much smaller indigenous populations present after what demographers term "the Great Dying." We do so in a set of regressions that either add 16$^{\text{th}}$ Century population change, in percentage terms, or substitute population density (in terms of arable land) of 1600 for that of 1500. Global and region by source region regressions are presented and discussed.

[21] The degree to which the economic logic of chattel slavery led to owner investment in the health of slaves is a much-studied topic in economic history. See, for example, Lander and Pritchett (2009) The fact that British settlement of Australia and New Zealand took off on a large scale only after Britain abolished the slave trade played a part in restricting African emigration almost entirely to the Americas and to previously uninhabited islands near Africa, such as Cape Verde.

While all of the above models use observations pairing source and destination countries, we also report models that investigate the determinants of the share of descendants that each country retains on its own territory versus having "sent" abroad. For these models we have only one observation per country i, with dependent variable being the "stayer (within region) share," or 1 minus the sum of the shares of all year 1960 descendants of the country's year 1500 population estimated to be living in each country $j \neq i$ in 1960, with j still excluding countries in i's own region.[22] The explanatory variables are the same as those in the initial, non-difference models, but the attributes such as amount of arable land, temperature, etc., are now those of country i itself. We estimate "stayer share" models both for all countries in our global sample, pooled, and for each of the three "sender" super-regions (Europe, SSA, North Africa & Asia). Note that we could, alternatively, have simply included descendants remaining in the ancestral country in the previous model as a special subset of dyadic DS_{ij}'s in which "destination" country j is the same as "source" country i. We choose to examine the determinants of our "DS_{ij}"'s ($j \neq i$) separately, however, in part because these observations would present problems in a difference analysis like Table 4, and omitting them from both the non-difference and difference analyses discussed above preserves comparability between it and Table 5 (the non-difference version). The relative importance of the various factors in determining the overall likelihood of "propelling" descendants to other regions versus retaining descendants at home might also be different from the determinants of shares abroad, as the analysis will in fact end up suggesting.

In all models reported, we focus on long-distance migration which can be defined as migration beyond the origin region. Thus, country pairs within a given region or super-region, such as Cambodia and Malaysia, France and Poland, or Egypt and Morocco, are not among the included observations, but pairs in different sub-regions, e.g. France and Morocco or Turkey and Greece, are included even when relatively proximate.

As noted in the Introduction, we acknowledge that our estimates indicate the values and significance levels of partial correlations which establish association but cannot be claimed to establish causality. Finding a significant positive coefficient on our indicator of land fertility (*maximum average calories per ha.*), for example, suggests to us that more migrants were attracted to what is now a country and produced more surviving descendants there due to the high productivity of its land, but we do not attempt to prove that that productivity

[22]A variant using the *Stayer Share w/in Country* as the dependent variable is reported in the Supplementary Tables.

caused more migrants to make the trip or is a cause of their higher fertility there. One reason is that several of our independent variables are related to each other, so genuine causal relations might be masked and significant coefficients could emerge spuriously. Given our roughly equal interest in a somewhat large number of potential explanatory variables, identifying good instruments for any large number of them would be challenging. If we use causal-sounding language, such as "highly productive land seems to have attracted more migrants," we are referring to what should be understood to be a correlation or association without proof of causality.

We make the following predictions, for the accompanying briefly explained reasons.

All else equal, more descendants of migrants will be found in countries with greater amounts of arable land, because individual migrants sought advantageous farming opportunities, because labor-intensive plantations were created on some previously less intensely exploited land, and because markets supplied food to non-farm workers mainly from domestic rather than imported sources, so countries with more arable land could also support more migration to cities and towns. In difference models, a larger difference between destination country j's arable land and that of source country i will be associated with larger DS_{ij}.

All else equal, less descendants of migrants will be found in countries with more year 1500 population per unit of arable land, since there were limits to the displacement of existing agriculturalist inhabitants in densely populated countries. This tendency would help to explain why there was less migration to colonized countries like India and Sri Lanka than to ones like the U.S. and Argentina, although factors like climate (*temperature*) and disease prevalence may have also been at work. Although dispossession of indigenous people from their land was common in the Americas, in those places more densely settled before colonization, existing population density might still be said to have served as a partial check on in-migration, because surviving descendants of the indigenous people, including ones of mixed Spanish and indigenous ancestry, formed the main agricultural labor force, with limited "importation" of either slave or free European labor. In difference models, a larger difference between destination country *(j)*'s *(pop/arable)* and that of source country i will be associated with smaller DS_{ij} (on average, people will have migrated to places of lower rather than higher population density than that of their country of origin). Our predictions about population density are expected to hold most strongly when we substitute population of 1600 for that of 1500 to capture the effect of post-contact disease mortality in the Americas, whereas they may fail to hold with estimated year 1500 population since so much of the 1500 - 1960 migration took place after the "Great Dying" in response to the radically transformed comparative population densities of Old vs. New World that it led to.

All else equal, more descendants of migrants will be found in countries in which the arable land is more productive, measured by maximum calories per ha. This may help to explain, for example, the larger influx of Old World people per hectare of cultivable land to the fertile islands of Cuba, Jamaica and Hispañola than to the vast plains of Canada and Argentina, although the better part of the migration to those islands was involuntary.[23] In difference models, a larger difference between destination country j's *avg. max-cal./ha. (Pre-1500 Maximum Calories cal./ha.)* and that of source country i will be associated with larger DS_{ij} (put simply, population is expected to have shifted towards land that was more productive than that of its ancestors).

Climate seems likely to have impacted migration. We acknowledge that impacts of source and destination country climate may have been complex—for example, number of frost-free days, rainfall, and seasonal variability among other details could matter, and the impact of temperature could be non-linear—but we control for it simply using average temperature by country, since it is not among our main interests or areas of expertise. Because agriculture was originally the principal activity of most migrants, we predict that migration will be positively related to temperature in our basic model. We conjecture that this relationship will be stronger for descendants of Africans than of Europeans because northern Europeans in particular may have avoided hotter climates, to which they might not have easily acclimated. In difference models, a larger difference between destination country and source country temperature will be associated with larger DS_{ij}.[24]

All else being equal, we expect to find more migrant descendants in countries with lower disease prevalence. In difference models, a larger difference between destination and source disease prevalence will be associated with lower DS_{ij}—or stated inversely, the less disease at destination and the more at origin, the greater the number of migrant descendants.

All else being equal, we expect to find more migrant descendants in countries with lower

[23] While average temperature is also probably a complementary factor in this illustrative comparison, we should note that the caloric yield of land, as measured, already incorporates some effects of climate.

[24] Obolensky et al. (2024) show that a tendency to minimize climate difference between origin and destination makes the similarity of climate of destination a significant predictor of where immigrants ended up living in the U.S., both in the 1850–1940 period and in 2011–2019. They also find that mortality of migrants and their children is an increasing function of climate difference.

state histories, partly because those with higher state histories tend to have been colonized later if at all (Ertan et al., 2016).[25] In difference models, we predict that the larger the difference between destination state history and that of the source country, the lower will be DS_{ij}.

All else being equal, we expect to find fewer migrant descendants from landlocked countries than from ones bordering on the sea, and to find fewer migrant descendants in landlocked countries than in ones bordering the sea. The era of European colonization began as one of exploration and navigation, almost all intercontinental migration was by sea until well into the 20th Century, and landlocked countries were significantly less likely to be colonized (Ertan et al., 2016). From the source standpoint, the idea of long-distance travel may have been less established in landlocked countries, with fewer examples of known emigrants to follow after.

All else being equal, we expect to find more migrant descendants in countries colonized by their home country, for instance British descendants in the U.S., Canada and Australia, Portuguese descendants in Brazil, Spanish descendants in Mexico and Argentina, etc. While we control for origin only for source country i itself with an indicator variable *colonizer of destination*, we also expect to find more descendants of countries that share cultural and other features with the colonizer than of countries with very different cultures. For simplicity, we proxy for these similarities by distance in a model variant used to explore this factor.

All else being equal, we expect to find more migrant descendants in countries colonized by the country that was the colonizer of their own country of origin. For example, descendants of migrants from India are more likely to be found in Belize and Guyana than in Panama or Brazil; descendants from Angola are more likely to be found in Brazil than in Venezuela. Note that we treat such connections as applicable even if full colonization of the source country came chronologically later than some of the migration in question, as is common for sub-Saharan African sources.

Since it is well known that some people from colonies migrated to and put down roots in the country that colonized their homeland, we also control for the destination being the colonizer of the source, as when people from French African countries settled in France,

[25] Countries with higher state histories are also more likely never to have been colonized, and to have higher population density, each of which we predict discourage migration. When both variables are directly controlled in our full model, only the effect by way of shorter colonial experience is in principle left to be controlled by *statehist1500* itself.

ones from India and Pakistan settled in Britain, and ones from Indonesia settled in the Netherlands. Although much migration of this kind probably occurred late in and even after the colonial period proper, we expect to find that DS_{ij} is higher (as calculated for 1960) if country j was the colonizer of country i during the colonial era.

As mentioned above, we control also for whether a destination country was colonized at all during the 1500 – 1960 era. It is important to include this control, because the *Colonizer of Destination* variable takes the same value, 0, when the destination country was colonized but not by the source as it does when the destination country was never colonized. Presence of both indicators resolves that ambiguity, as well as the similar one applying to the *Same Colonizer* variable. Although only some colonized countries, mainly ones in the greater New World, were large net recipients of migration, and numerous sub-Saharan countries were net "exporters" of population albeit themselves colonized late in the 1500 – 1960 era, we predict that the coefficient on *colonized* will on balance be positive. Note, for example, that in our main specifications, which use the dichotomous *Colonizer of Destination* variable rather than the broader *Distance to Colonizer* control used in a variant exercise, only the *Colonized* control would pick up the increased likelihood of, say, Irish settlement in the U.S. or Canada that is attributable to their having been colonies.

With respect to what accounts for values of the "stayer share," variables predicting more or less migration from countries in the main, dyadic migration estimates should show similar effects here. Thus, countries that were colonizers, that were more densely populated relative to arable land, that had smaller amounts of arable land, that had lower average land productivity (*avg. max-cal./ha.*), and that had lower average temperature and higher disease prevalence, are expected to have "sent off" more émigrés and accordingly have lower stayer shares. Because emigration was strongly influenced by regional factors, especially the impact of the slave trade on sub-Saharan Africa, we also include region dummy variables in the main "stayer share" estimates. In the main specification, which uses the *Colonizer of Destination* control rather than a measure of distance to the colonizer, we expect significantly negative coefficients on the Europe and SSA region dummies.

4 Results

4.1 Determinants of Who Migrated Where

PPML estimates of our model with destination minus source difference specifications for the six applicable explanatory factors, plus distance, source and destination landlocked indicators, and five colony and colonizer status dummies, are shown in Table 3. We enter variables sequentially in the order suggested by their respective contributions towards accounting for explained variation in DS_{ij} as determined by the Shapley method, with Shapley values themselves shown in Table A.6 in the Appendix. Exceptions are the inclusion of three low-impact variables, shown at bottom, in all specifications, exploratory entry of (*pop/arable*) difference and *statehist1500* difference separately from each other as mentioned earlier, and simultaneous addition of the colonization dummies as a group, to conserve space. Table 4 studies the relationships between the same explanatory and dependent variables, but relaxing the restriction that equation (2) imposes on source- and destination-side factors, as in equation (3). To gain insight into the potentially distinct effects of population density and state history, its estimate (1) includes only the latter, estimate (2) only the former, and (3) includes both. Table 4 also displays estimates that omit source-side variables in favor of source fixed effects (estimate (4)) or omit destination-side variables in favor of destination fixed effect (estimate (5)). Extensive robustness checks varying the multi-way clustering structures are provided in Appendix H. As emphasized by Larch et al. (2025), proper multi-way clustering in gravity estimations is a frontier requirement, and varying these structures proves that our standard errors are not artificially compressed. Finally, Table 5's regressions resemble those of Table 3 but are estimated separately for samples with European source countries (columns (1) - (3), sub-Saharan African source countries (columns (4) - (6)) or North African and Asian source countries (columns (7) - (9)).

Given the large number of independent variables, we find it expeditious to discuss their effects variable by variable with attention to Tables 3 - 5 as a whole. Of the total of about 34% of the variation in DS_{ij} values that the full model with differences (Table 3) appears to account for, according to the pseudo-R^2, five variables each account for between 3% and 10%, judging by the Shapley decomposition: arable area difference (9.2%), disease prevalence difference (5.0%), difference of year 1500 population per arable ha. (3.7%), and whether the destination was a colony (4.5%) and/or the source was its colonizer (3.7%). A Wald test rejects the restriction of source- and destination-side effects underlying the differences specification of Table 3, which means that the estimates of Table 4 can be treated as coming closer to accurately capturing the relationships in question. We nevertheless chose to include Table 3 in our discussion because it allows the main effects to be summarized with reference to a more compact set of determinants. We also find it noteworthy that all six factors lending themselves to entry in source-destination difference form obtain coefficients significant at the 1% level, although the highly correlated *population1500/arable ha.* and *state history* variables lose significance when entered jointly.

Consider, now, the variables representing arable land's absolute abundance, its abundance relative to population, and its productive potential, in each country. The expectations that population moved from countries having less to those having more arable land, from those more to those less densely populated relative to that land, and from those with lower to those with higher caloric potential of their land, are all supported in an overall sense. In the difference-form regressions, arable land difference and average maximum calories difference obtain highly significant positive coefficients in the complete specification, and explain 9.2% and 1.6% of variance respectively. The coefficient of *population1500/arable hectare* difference is similarly significant and negative as predicted, when *statehist1500* is omitted and it explains 3.7% of DS_{ij}'s variance according to Shapley analysis of the model as a whole. The non-difference specifications of Table 4 are strongly consistent for arable area, which obtains highly significant negative coefficients for source and positive coefficients for destination values, suggesting both push and pull (possibly strengthened by complementary effects on stayer and leaver fertility) were at work. For calories per ha., only the destination variable obtains positive and highly significant coefficients. *Population1500/arable ha.* returns significant coefficients only in the specifications that include the source- or the destination-side variable alone, but this remains broadly consistent with the idea that migrants were attracted to relatively productive land. For *population1500/arable hectare.*, neither its source nor its destination value are significant when *statehist1500* is omitted (col. (2)), and the destination value becomes positive and significant at the 5% level when *statehist1500* is present (col. (3)), contrary to hypothesis. The coefficient value could be misleading due to the high correlation with *statehist*; when that variable is omitted in variants of Table 4's regressions shown in Table A.16, neither source nor destination *population1500/arable hectare* obtains a significant coefficient. More importantly, we will see below that the sign reverses when accounting for the impact of the Colombian disease exchange on the population of the

Americas.

Table 3: Global Sample with Difference Variables

Dependent Variable: DS_{ij}	(1) Global	(2) Global	(3) Global	(4) Global	(5) Global	(6) Global	(7) Global	(8) Global
Arable Land Area Difference	0.390***	0.409***	0.292***	0.387***	0.300***	0.496***	0.453***	0.541***
	(0.094)	(0.091)	(0.068)	(0.084)	(0.071)	(0.072)	(0.064)	(0.058)
Disease Prevalence Difference		−0.781***	−0.634***	−0.784***	−0.656***	−1.438***	−1.586***	−2.304***
		(0.214)	(0.189)	(0.216)	(0.181)	(0.236)	(0.206)	(0.204)
Pop1500 / Arable Lands Difference			−0.287***		−0.248*	0.072	0.063	0.028
			(0.100)		(0.132)	(0.089)	(0.093)	(0.089)
Statehist 1500 Difference				−1.374***	−0.518	−0.536	−0.198	−0.316
				(0.420)	(0.654)	(0.693)	(0.732)	(0.745)
Colonizer of Destination						4.013***	4.111***	4.139***
						(0.498)	(0.576)	(0.511)
Destination Colony Dummy						3.046***	3.300***	2.966***
						(0.863)	(0.791)	(0.814)
Source Colony Dummy						−0.914*	−1.220**	−1.276**
						(0.524)	(0.532)	(0.515)
Same Colonizer						−0.497**	−0.402**	−0.302
						(0.227)	(0.204)	(0.192)
Destination is Colonizer						0.135	0.169	−0.010
						(0.706)	(0.730)	(0.741)
Pre-1500 Avg. Max.-Cal./Ha Difference							0.383***	0.363***
							(0.128)	(0.098)
Temperature Difference								0.069***
								(0.020)
Distance (Ln)	−0.190	−0.196	−0.353**	−0.232	−0.352**	−0.594***	−0.775***	−0.821***
	(0.246)	(0.230)	(0.175)	(0.220)	(0.177)	(0.182)	(0.165)	(0.165)
Source Landlocked	−1.370***	−1.265***	−1.121**	−1.144***	−1.098**	−1.340***	−1.194***	−1.422***
	(0.444)	(0.466)	(0.483)	(0.441)	(0.459)	(0.294)	(0.331)	(0.350)
Destination Landlocked	−3.567***	−3.584***	−3.569***	−3.630***	−3.588***	−2.496***	−2.559***	−2.530***
	(0.563)	(0.540)	(0.519)	(0.528)	(0.513)	(0.485)	(0.516)	(0.533)
Constant	−5.473**	−5.684***	−4.304***	−5.375**	−4.311**	−4.241**	−2.815*	−2.414
	(2.241)	(2.186)	(1.657)	(2.093)	(1.678)	(1.679)	(1.640)	(1.711)
Observations	20,083	20,083	20,083	20,083	20,083	20,083	20,083	20,083
Pseudo R^2	0.143	0.175	0.194	0.185	0.195	0.308	0.324	0.337

Notes: All regressions are estimated using PPML. The continuous explanatory variables are calculated as the destination value minus the source value. The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014), World Bank (2026), and authors' calculations. Independent variables include *State History* (Borcan et al., 2018 and authors' calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and authors' calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and authors' calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. Standard errors are two-way clustered at the destination and source levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In the separate estimates for each of three source super-regions (Table 5), results for source and destination arable area are qualitatively similar to those for the global sample in all three regions, with the exception that for European source observations, the coefficient on source arable area is positive, significant at the 5% or 10% level depending on specification. The magnitude of this coefficient is small relative to that on destination arable area, so the hypothesis that migration was on the whole associated with a positive difference in arable area still holds. The source-side coefficient indicates that outmigration did not come differentially from the European countries having the least cultivable land, which confirms that colonizer status was not driven by land scarcity.

Source region by source region results for land productivity (*avg. max-cal.*) are qualitatively different by region. As in the global estimates (Table 4), source country land productivity is insignificant (or in one case marginally significant) in two regions, but obtains highly significant positive coefficients, contrary to expectation, for North Africa and Asia. Land productivity at destination has the predicted positive coefficients for Europe and SSA region senders, but is highly significant for SSA only–suggesting that the slave trade directed labor differentially towards fertile destinations like Haiti, Cuba, Jamaica and Brazil. The corresponding coefficients are counter-intuitively negative and at most marginally significant for North African and Asia senders, perhaps because many of their emigrants were not seeking agricultural opportunities.

Estimates of coefficients on source- and destination-side population per unit of arable land are sensitive to whether *statehist1500* is included (Table 5) or omitted (Table A.17), as well as whether we add controls to account for post-contact population losses in the Americas (discussed in the next section). Including *statehist1500* and without such controls, Table 5 shows a strongly significant positive coefficient on the source side for Europe only, implying that more densely populated European countries (relative to cultivable land) placed more descendants abroad.

Population per arable hectare obtains strongly significant destination-side coefficients only for SSA and for North Africa & Asia, with the former having positive coefficients (contrary to hypothesis) and the latter negative ones (per hypothesis). Omitting *statehist1500* and thus the likely distortion of the coefficient on population per arable ha. due to the high correlation of *statehist* and density, Table A.17 finds no significant qualitative change in the coefficients of *population1500/arable hectare.* on the source side, but it finds significant effects on the destination side, for SSA and for North Africa and Asia sending countries, becoming insignificant when *statehist1500* is omitted.

We next consider other natural and geographic factors: disease prevalence, temperature, distance, and whether source and/or destination are landlocked. In the global sample tables, these all show the predicted correlations. Disease prevalence is the second most influential factor, accounting for 5.0% of variation in descendant share, with a highly significant negative coefficient in the difference specification, and highly significant positive and negative coefficients, respectively, for the separate source- and destination-side measures. While macro-level shifts toward warmer climates were insignificant for Europe as a whole in our data, this does not preclude the strong micro-level (sub-national) climate sorting that Obolensky et al. (2024) document. Insofar as migration towards warm climates may seem to be at odds with impressions that northern Europeans in particular sought out mild climates like those of the US and Argentina, we note that Table 5's disaggregation by source region shows that the effect is driven by movers from the SSA and North Africa and Asia regions, and that both coefficients are insignificant for European source countries. Distance is also one of the less important contributors to explaining variation in DS_{ij}, according to our Shapley decomposition, but it obtains highly significant negative coefficients, as predicted, in the more complete specifications of Table 3 and in all those of Table 4. Both source and destination *landlocked* indicators have highly significant negative coefficients in those tables also, as predicted.

The political factors influencing migration are represented by indicators of colonizer and colony status and relationships. Being the *colonizer of destination* country has a highly significant positive relationship with DS_{ij} in all specifications of Table 3 and Table 4, and in the Europe region estimates of Table 5, the only region in which it can be included. Being a colony (*destination colony*) also obtains a highly significant positive coefficient in all specifications of Table 3, but mostly insignificant or marginally significant coefficients in Table 4 and Table 5, which enter source- and destination-side variables separately. Countries that were colonized were also sources of more emigrants, as shown by positive coefficients on *source colony*. Although those coefficients are only marginally significant in Table 3 and insignificant in Table 4, they are significant at the 1% level for both the SSA source region and the North Africa and Asia source region, in Table 5. Our conjecture that countries sharing a colonizer, such as India and Belize or Angola and Brazil, will have had more bilateral migration, is strongly supported for the Africa region in Table 5, but supported with a significant positive coefficient ($p < 0.05$) in only one global specification in Table 4, and contradicted by negative coefficients meeting the same significance cut-off in two Table 3 estimates (but not the complete specification). The *destination is colonizer* indicator obtains no significant coefficients except in the estimate for the North Africa and Asia source region that includes only source-side measures. There, it is positive and significant at the 1% level, supporting our conjecture regarding cases such as Algerian migration to France.

Table 4: Global Sample with Source & Destination Variables

Dependent Variable: DS_{ij}	(1) Global	(2) Global	(3) Global	(4) Global	(5) Global
Source Arable Land Area (Ln)	-0.264***	-0.313***	-0.235***		-0.205***
	(0.034)	(0.037)	(0.042)		(0.054)
Destination Arable Land Area (Ln)	0.960***	0.905***	1.013***	0.994***	
	(0.106)	(0.135)	(0.098)	(0.085)	
Source Disease Prevalence	1.514***	1.723***	1.469***		1.471***
	(0.422)	(0.530)	(0.384)		(0.229)
Destination Disease Prevalence	-1.483***	-2.013***	-2.081***	-1.913***	
	(0.316)	(0.746)	(0.389)	(0.344)	
Source Pop1500 / Arable Lands (ln)		-0.058	0.071		0.146**
		(0.054)	(0.070)		(0.069)
Destination Pop1500 / Arable Lands (ln)		0.011	0.378**	0.312**	
		(0.233)	(0.177)	(0.146)	
Source Statehist 1500 (Ln)	-0.051***		-0.062***		-0.059***
	(0.016)		(0.021)		(0.022)
Destination Statehist 1500 (Ln)	-0.268***		-0.300***	-0.264***	
	(0.079)		(0.080)	(0.057)	
Colonizer of Destination	3.134***	3.507***	3.043***	3.500***	2.877***
	(0.603)	(0.762)	(0.581)	(0.311)	(0.440)
Destination Colony Dummy	0.876	2.889**	1.462	0.763	
	(1.083)	(1.273)	(1.315)	(1.030)	
Source Colony Dummy	0.003	0.013	0.018		0.362
	(0.228)	(0.207)	(0.233)		(0.362)
Same Colonizer	-0.020	-0.082	-0.008	0.549**	0.205
	(0.353)	(0.341)	(0.365)	(0.217)	(0.247)
Destination is Colonizer	-0.050	-0.304	-0.512	-0.364	0.290
	(0.841)	(0.926)	(0.827)	(0.830)	(0.524)
Source Pre-1500 Avg. Max-Cal./Ha	-0.014	0.015	0.007		0.087
	(0.105)	(0.103)	(0.102)		(0.149)
Destination Pre-1500 Avg. Max-Cal./Ha	0.587***	0.854***	0.686***	0.517***	
	(0.117)	(0.277)	(0.213)	(0.174)	
Source Temperature	-0.102***	-0.103***	-0.101***		-0.119***
	(0.019)	(0.022)	(0.018)		(0.024)
Destination Temperature	0.085***	0.066	0.075*	0.080**	
	(0.028)	(0.050)	(0.040)	(0.032)	
Distance (Ln)	-1.759***	-1.404***	-1.809***	-1.310***	-2.779***
	(0.472)	(0.406)	(0.484)	(0.356)	(0.335)
Source Landlocked	-1.548***	-1.623***	-1.520***		-1.307***
	(0.378)	(0.345)	(0.363)		(0.361)
Destination Landlocked	-2.101***	-2.031***	-2.048***	-1.971***	
	(0.649)	(0.641)	(0.669)	(0.639)	
Constant	-5.139*	-6.073**	-5.669*	-13.035***	26.207***
	(2.845)	(2.839)	(3.194)	(2.861)	(3.032)
Observations	20,083	20,083	20,083	20,083	20,083
Source Fixed Effects	NO	NO	NO	YES	NO
Destination Fixed Effects	NO	NO	NO	NO	YES
Pseudo R^2	0.432	0.394	0.437	0.510	0.490

Notes: All regressions are estimated using PPML. The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014), World Bank (2026), and authors' calculations. Independent variables include *State History* (Borcan et al., 2018 and authors' calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and authors' calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and authors' calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. Standard errors are two-way clustered at the destination and source levels. *** ($p < 0.01$), ** ($p < 0.05$), * ($p < 0.1$)

Table 5: Source Region Samples with Source & Destination Variables

Dependent Variable: DS_{ij}	Europe			Sub-Saharan Africa			N. Africa & Asia		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Source Arable Land Area (Ln)	0.165**		0.213*	−0.347***		−0.263***	−0.358*		−0.301**
	(0.071)		(0.112)	(0.099)		(0.058)	(0.203)		(0.153)
Destination Arable Land Area (Ln)	1.147***	1.285***		0.998***	1.002***		0.811***	0.764***	
	(0.167)	(0.139)		(0.105)	(0.102)		(0.131)	(0.120)	
Source Pop1500 / Arable Lands (ln)	0.960***		0.965***	0.003		0.104	−0.072		0.011
	(0.134)		(0.287)	(0.099)		(0.070)	(0.294)		(0.233)
Destination Pop1500 / Arable Lands (ln)	0.293	0.320		0.442***	0.430**		−0.398***	−0.391**	
	(0.358)	(0.215)		(0.165)	(0.172)		(0.114)	(0.164)	
Source Statehist 1500 (Ln)	−0.881**		−0.831	0.008		−0.026	0.817		0.888
	(0.369)		(0.540)	(0.036)		(0.025)	(1.085)		(0.857)
Destination Statehist 1500 (Ln)	−0.255***	−0.255***		−0.312***	−0.302***		0.174***	0.176***	
	(0.075)	(0.040)		(0.092)	(0.098)		(0.063)	(0.052)	
Source Disease Prevalence	1.276**		1.341***	2.795***		2.726***	0.314		0.317
	(0.630)		(0.385)	(0.467)		(0.428)	(0.852)		(0.738)
Destination Disease Prevalence	−1.150	−1.689***		−2.374***	−2.397***		−2.197***	−1.916***	
	(0.759)	(0.376)		(0.530)	(0.543)		(0.530)	(0.596)	
Colonizer of Destination	2.389**	4.655***	2.222***						
	(1.017)	(0.532)	(0.575)						
Destination Colony Dummy	0.833	−0.091		0.721	0.736		1.846*	1.393*	
	(1.074)	(0.729)		(1.272)	(1.393)		(0.956)	(0.747)	
Source Colony Dummy				3.956***		4.099***	4.279***		4.057***
				(1.361)		(1.304)	(1.259)		(1.313)
Same Colonizer				0.407*	0.454***	0.544***	−0.392	0.060	−0.111
				(0.211)	(0.168)	(0.198)	(0.861)	(0.777)	(0.604)
Destination is Colonizer				−0.750	−0.837	0.157	1.431	1.759***	1.388
				(1.092)	(1.101)	(0.565)	(0.980)	(0.587)	(1.301)
Source Pre-1500 Avg. Max-Cal./Ha	−0.234		−0.008	0.351*		0.281	1.193***		1.238***
	(0.307)		(0.395)	(0.187)		(0.246)	(0.195)		(0.280)
Destination Pre-1500 Avg. Max-Cal./Ha	0.613	0.383		0.693***	0.669***		−0.219	−0.219*	
	(0.435)	(0.344)		(0.218)	(0.218)		(0.217)	(0.133)	
Source Temperature	−0.069		−0.113	−0.185*		−0.287***	−0.265***		−0.255***
	(0.059)		(0.078)	(0.104)		(0.069)	(0.074)		(0.056)
Destination Temperature	0.018	0.026		0.108***	0.112***		0.125***	0.105**	
	(0.068)	(0.049)		(0.035)	(0.036)		(0.045)	(0.043)	
Distance (Ln)	−1.279*	−0.655	−2.673***	−2.077***	−2.102**	−3.515***	1.020	1.604*	0.248
	(0.718)	(0.600)	(0.775)	(0.776)	(0.894)	(0.598)	(1.157)	(0.830)	(0.911)
Source Landlocked	−0.214		−0.181	−4.675***		−4.464***	−0.967		−0.674
	(0.226)		(0.375)	(0.938)		(0.751)	(1.191)		(0.930)
Destination Landlocked	−0.981	−0.950*		−4.725***	−4.719***		−2.554***	−2.543***	
	(0.616)	(0.525)		(0.961)	(0.991)		(0.721)	(0.531)	
Constant	−17.772***	−23.223***	19.316**	−3.380	−6.051	33.492***	−26.090**	−35.887***	−1.276
	(6.701)	(6.140)	(7.838)	(8.309)	(7.866)	(6.394)	(11.132)	(6.522)	(9.053)
Observations	4674	4674	4674	4914	4914	4914	5085	5085	5085
Pseudo R^2	0.523	0.587	0.557	0.454	0.465	0.518	0.392	0.432	0.448

Notes: All regressions are estimated using PPML.The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014), World Bank (2026), and authors' calculations. Independent variables include *State History* (Borcan et al., 2018 and authors' calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and authors' calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and authors' calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. For columns 1, 4, and 7, standard errors are clustered at the source and destination level, for columns 2, 5, and 8, standard errors are clustered at the destination level; for columns 3, 6, and 9, they are clustered at the source level. *** ($p < 0.01$), ** ($p < 0.05$), * ($p < 0.1$).

Finally we consider *statehist1500*, which is in a category of its own insofar as it measures institutional history rather than resources or geography, but is nonetheless related to population density and practice of agriculture. In Table 3's difference structured regressions, *statehist1500* (*difference*) obtains a significant negative coefficient when *population1500/arable ha.* is omitted, suggesting that population moved on balance from countries with long state histories to ones with a short or no state history before the colonial era. While it loses significance in the specifications that include *population1500/arable ha.*, the specifications that enter source and destination values separately see the *statehist1500* variable of both sides attaining negative and significant coefficients even when the density measure is included. The point value of *destination statehist1500* is much larger than that of the source-side variable, consistent with the negative coefficient on the difference version. The negative coefficient on the destination-side coefficient is consistent with (greater) New World destinations having no or low state histories, and with ones with lowest state history, such as U.S., Argentina and Haiti, receiving somewhat more immigrants than did higher *statehist* countries like Mexico and Peru.

The estimates by source region are qualitatively identical for source and destination disease prevalence, except for insignificance of the source-side measure for North Africa and Asia region descendants. It seems noteworthy that SSA descendants are not more numerous in more disease-prevalent destination countries, contrary to what might be expected based on more site-specific case studies like that for rice-growing coastal South Carolina (Wood, 1996). The latter suggests that slaves were set to work in disease-ridden environments that white slave-owners stayed away from, which leads one to expect that slaves were preferentially placed in more disease-prone places. The bivariate correlation between destination disease-prevalence and DS_{ij} is negative but statistically insignificant, and the variable's contribution to explained variation is small for the SSA regressions, according to its Shapley value. It appears that within-country examples like that from S. Carolina are not predictive of country-level correlations in this case, but the contrary impression given by the relevant regression coefficients in the SSA estimates of table 5 should be treated with caution. The significant negative coefficient on the source side conforms to expectation, since the source countries in SSA include some of the world's most disease prone.

Both source and destination temperature obtain insignificant coefficients for European source countries and display the negative sign on source temperature and positive sign on destination temperature of the global sample for the SSA source sample, although the coefficient on source temperature is only marginally significant. $\ln(Distance)$ obtains the predicted negative coefficient for Europe and SSA, being significant at 1% level in all SSA estimates, but for Europe, this is true only in the estimate using source-side variables and destination fixed effect. This suggests that the significant finding for the global sample rests heavily on the sourcing of slaves from the Atlantic coast of Africa and their overwhelming disembarkation on Caribbean islands and the Atlantic coast of North and South America. The *source landlocked* variable also obtains significant negative coefficients solely for SSA source countries, whereas *distination landlocked* is significantly negative for both SSA and North Africa and Asia region source countries. The most likely explanation rests on trade accessibility.

Coastal regions and navigable rivers foster economic development precisely because they allow integration into global trade, whereas hinterlands suffer from diminishing returns and high transport costs (Gallup et al., 1999). Being landlocked acted as a structural physical friction that isolated populations from the maritime networks.

We return to the Shapley decomposition of Table 3 specifications (see Table A.6) for a sense of the relative contributions of different sets of factors in accounting for who had descendants where. Of the 33.8% of variance in DS_{ij} accounted for by the full model according to the Pseudo R^2 measure, a total of 24.3% is associated with the resource and geography factors, 9.23 by the political (colonial relations) factors, and a negligible part by *statehist1500*. Similar decompositions for Table 4's and other specifications are shown in Table A.7, A.8 and A.9.

We also briefly discuss variant models shown in tables A.10, A.11 and A.12, where instead of capturing only the often high contribution of migrants to colonized countries by their colonizers, we confirm that neighbors of colonizers also contributed more to the population settling there. More specifically, the results demonstrate, perhaps for the first time so systematically, that taking Europe as a whole as a contributor of immigrants to the greater New World and other colonized countries, the closer within Europe was a non-colonizer (say Ireland, France) to the colonizer (say Britain), the more its descendants are present in the former colony's 1960 population. The further from each colonizer was the European source country (say Serbia or Ukraine), the less its population contribution to the 1960 populations of its colonies.

4.2 The Post-contact Demographic Collapse of Native Populations in the Americas

The prediction that global population was on the whole redistributed from more towards less densely populated parts of the world during the colonial era must grapple with the fact that the ratio of people to cultivable land shrank dramatically between 1492 and the early 17$^{\text{th}}$ Century in the Americas because the hemisphere's indigenous population lacked immunity to smallpox, measles, plague, and other Old World diseases. Since the mid-20$^{\text{th}}$ Century, scholarly estimates of the New World's pre-Colombian population have been revised upwards from totals of around 13 million, according to McEvedy and Jones (1978), to ones in the 90 - 100 million range Dobyns (1966), with the 57 million estimate of Denevan (1992), or more generally the 50 - 70 million range–more than four times the old estimate–now considered most plausible. Since the majority of colonial era migration was from Europe and Africa to the Americas, the idea that more populous lands were exporting population to less populous ones is more fully supported after the devastation wrought by Old World diseases had shifted the population balance decisively in the Old World's favor, following arrival of the first conquistadors.[26] Recent research goes so far as to assert that a further decline in global average temperatures in the late 16^{th} and early 17^{th} Centuries, exacerbating the "Little Ice Age", is largely attributable to the reversion of agricultural and other managed habitats to cover by CO_2-absorbing forest and other vegetation due to population "die off", some of it in areas that settlers assumed to have always been rainforest because the human devastation had spread through native populations well before colonial survey teams arrived (Koch et al., 2019).

The now-discredited lowness of McEvedy & Jones's population estimates for the Americas, used in initial estimates of our regressions (not shown), yields spuriously strong support for the prediction that migrant descendant numbers were significantly higher where population density (denominated by arable land) was lower in 1500,[27] especially when *statehist1500* was omitted due to its colinearity with the density variable. For the estimates thus far shown, we have adopted the 2023 version of the HYDE estimates of year 1500 population in all countries. HYDE's aggregate estimate of approximately 60 million for the Americas lies close to that of Denevan (1992).[28] According to Koch et al. (2019), the Americas suffered a staggering 90% depopulation rate (approximately 54.5 million deaths), which radically altered the land-to-labor ratio. This "Great Dying" in the Americas relieve the population pressures of a labor-abundant Old World by means of what Blanc and Wacziarg (2025) denote "Malthusian migrations".

To explore the effects of the sharp population contraction that followed contact, we report in Table 6 two alternative regression specifications for the global sample and for the samples for each of the three source regions studied in Table 5. While deaths that demographers attribute to absence of immunity to Old World diseases were occurring in some New World

[26]To be sure, violence, being driven off of previous farming and hunting territories, and forced labor also reduced indigenous descendants' numbers, but the estimates suggest that the effect of epidemics was at least an order of magnitude, or nearly two orders, greater, in accounting for the decline in numbers.

[27]That is, there was a highly significant negative coefficient on (log of) *source pop1500/arable area* in a preliminary version of Table 4.

[28]For a recent appraisal of Denevan's estimate, see Lovell (2021). Koch et al. (2019) use a slightly higher aggregate estimate, 60.5 million.

regions as late as the 17th or even 18th century, and while there are estimated to have been large but less extreme declines in the indigenous populations of Australia and New Zealand mostly after those in the Americas, we adopt the simplified approach of using population per arable hectare in 1600 for all countries as a way to capture the large majority of the indigenous population decline at least in the Americas.[29] For both years, population estimates are from HYDE which, as an example, estimates a population of 24 million within today's Mexican borders in 1500 versus 1 million in 1600. We report two alternative specifications for our global sample and for each of the three source region samples used above. In one specification, we retain year 1500 population density but add a control for the population growth rate of 1500 - 1600 in both source and destination country, including large negative growth rates capturing population losses in the Americas. In the other specification, we omit the population growth rate control but substitute population density in 1600 for that in 1500.

The population density of destination obtains a highly significant negative coefficient when our proxy "post die-off" density using population of 1600 is substituted for *pop1500* (columns (2), (4) and (6)), supporting the expectation that Old World population was drawn to the New World by its plenitude of land relative to population following the impact of the Colombian disease exchange. Absence of a significant coefficient in the case of North African and Asian source countries could be due to relative unimportance of farming at destination for many of that region's migrants. In the specifications that include it in Table 6, ln(*Pop1500*/*arable ha.*) also displays negative and significant coefficients, when population growth rate of 1500 - 1600 is included, but when *statehist1500* is included (as done in Table A.18, to better match Tables 4 and 5), the corresponding coefficients lose significance. In contrast, destination population growth rate of 1500 - 1600 obtains negative and highly significant coefficients in both Table 6 and Table A.18, suggesting that the large population declines (negative growth rates) in the Americas are associated with larger numbers of Old World descendants, as we conjecture.

[29] Note that our estimates of arable land area used in the denominator, whether using *pop1500* or *pop1600* in the numerator, are based on recent estimates, not estimates of the area actually cultivated at the time. We follow this approach because the fact that the area planted to crops contracted in parts of the Americas due to population loss (Koch et al., 2019) should not have affected the levels of migration and population growth of newcomers. That is, land that ceased to be cultivated due to population "die-off" should still be counted as cultivable in our calculations, assuming that potential migrants viewed it as potential farmland awaiting their development. An estimate of more-or-less current arable area can approximate the potential farmland as perceived during most of the colonial era.

4.3 Determinants of Share of 1960 Descendants Still in the Home Region or Country

Having explored the ability of the factors identified by us to explain in which other countries the descendants of year 1500 inhabitants of included countries were found in 1960 and in what proportions, we now separately examine the determinants of the proportion of 1960 descendants still residing in the country of their year 1500 ancestors. In our main analysis,

Table 6: Estimations with Pop. Growth Rate during 1500 - 1600 and Pop. 1600

Dependent Variable: DS_{ij}	Global		Europe		Sub-Saharan Africa		N. Africa & Asia	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Source Pop1500 / Arable Lands (ln)	0.003 (0.065)		0.786*** (0.161)		−0.000 (0.041)		0.123 (0.285)	
Destination Pop1500 / Arable Lands (ln)	−0.452*** (0.129)		−0.519** (0.225)		−0.440** (0.177)		−0.318** (0.159)	
Source Pop. Growth Rate 1500–1600	0.258 (0.247)		2.122*** (0.525)		−2.547** (1.257)		−0.106 (0.939)	
Destination Pop. Growth Rate 1500–1600	−3.003*** (0.340)		−2.724*** (0.379)		−3.143*** (0.384)		−0.190 (0.341)	
Source Pop1600 / Arable Lands (ln)		−0.006 (0.059)		0.712*** (0.095)		0.008 (0.044)		0.120 (0.279)
Destination Pop1600 / Arable Lands (ln)		−0.660*** (0.141)		−0.777*** (0.232)		−0.627*** (0.206)		−0.258 (0.163)
Source Arable Land Area (ln)	−0.258*** (0.044)	−0.276*** (0.040)	0.074 (0.080)	0.106 (0.075)	−0.298*** (0.062)	−0.345*** (0.051)	−0.277* (0.150)	−0.274* (0.158)
Destination Arable Land Area (ln)	0.623*** (0.091)	0.807*** (0.111)	0.702*** (0.127)	0.973*** (0.153)	0.557*** (0.133)	0.765*** (0.139)	0.728*** (0.115)	0.726*** (0.127)
Source Disease Prevalence	1.474*** (0.376)	1.547*** (0.424)	0.827 (0.536)	1.050 (0.687)	2.870*** (0.534)	3.047*** (0.586)	0.378 (0.858)	0.494 (0.520)
Destination Disease Prevalence	−1.570*** (0.340)	−1.504*** (0.448)	−0.372 (0.714)	−0.388 (0.638)	−1.928*** (0.601)	−1.890*** (0.559)	−1.250* (0.747)	−1.522** (0.724)
Colonizer of Destination	3.020*** (0.496)	3.172*** (0.574)	2.379*** (0.894)	2.233** (1.066)				
Destination Colony Dummy	1.333* (0.699)	1.264 (0.963)	0.862 (0.741)	0.736 (0.653)	0.809 (0.504)	0.519 (1.101)	0.702 (0.929)	0.791 (0.909)
Source Colony Dummy	0.177 (0.231)	0.098 (0.214)			4.016*** (1.281)	3.928*** (1.363)	4.386*** (1.278)	4.340*** (1.226)
Same Colonizer	−0.028 (0.360)	0.004 (0.361)			0.417** (0.208)	0.490** (0.244)	−0.526 (0.882)	−0.495 (0.877)
Destination is Colonizer	0.752 (0.720)	0.656 (0.911)			0.635 (0.811)	0.045 (1.041)	1.740 (1.080)	1.616 (1.098)
Source Pre-1500 Avg. Max-Cal./Ha	0.023 (0.098)	0.006 (0.098)	−0.188 (0.323)	−0.420 (0.311)	0.291* (0.151)	0.372** (0.183)	1.004*** (0.221)	0.995*** (0.223)
Destination Pre-1500 Avg. Max-Cal./Ha	0.403** (0.172)	0.521*** (0.128)	0.243 (0.284)	0.360** (0.178)	0.352 (0.215)	0.427** (0.199)	−0.201 (0.234)	−0.235 (0.253)
Source Temperature	−0.093*** (0.018)	−0.098*** (0.020)	−0.119* (0.064)	−0.109* (0.058)	−0.215** (0.092)	−0.187* (0.103)	−0.271*** (0.080)	−0.272*** (0.079)
Destination Temperature	0.099*** (0.036)	0.119*** (0.034)	0.019 (0.044)	0.055** (0.028)	0.123*** (0.045)	0.165*** (0.048)	0.086* (0.050)	0.092* (0.052)
Distance (ln)	−1.859*** (0.397)	−1.655*** (0.422)	−1.048* (0.556)	−1.181** (0.500)	−2.713*** (0.697)	−2.436*** (0.555)	1.088 (1.312)	1.056 (1.309)
Source Landlocked	−1.387*** (0.306)	−1.490*** (0.323)	−0.385 (0.246)	−0.289 (0.211)	−4.616*** (0.925)	−4.761*** (0.961)	−1.515* (0.780)	−1.465* (0.879)
Destination Landlocked	−1.527*** (0.484)	−1.820*** (0.553)	−0.851 (0.525)	−0.954* (0.505)	−3.824*** (0.872)	−4.344*** (0.917)	−2.115*** (0.548)	−1.993*** (0.510)
Constant	0.952 (3.264)	−4.094 (2.591)	−10.890** (5.440)	−15.235*** (3.521)	9.318 (6.480)	2.533 (5.493)	−26.651** (11.379)	−26.519** (11.097)
Observations	20,083	20,083	4,674	4,674	4,914	4,914	5,085	5,085
Pseudo R^2	0.448	0.422	0.531	0.520	0.466	0.424	0.380	0.379

Notes: All regressions are estimated using PPML.The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014), World Bank (2026), and authors' calculations. Independent variables include *State History* (Borcan et al., 2018 and authors' calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and authors' calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and authors' calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. Standard errors are clustered at the source and destination level. *** ($p < 0.01$), ** ($p < 0.05$), * ($p < 0.1$)

shown in Table 7, we enter the independent variables in a mainly sequential fashion, similar to the regressions of Tables 4 and 5, but in an order guided by each factor's relative importance according to a separate Shapley decomposition exercise. The factors considered are the same as with the previous analyses, but since we are focusing on each country as a source only, we include only the source country values (similar to col. (5) of Table 4) of the variables entered in difference form, in Table 7. Variables defined only for country dyads—distance, and the dyadic colony and colonizer indicators—are omitted. Indicators for the country having ever been a colony or a colonizer are retained. Because these are simple cross sectional regressions with a single observation for each of 153 includable countries, OLS models can be used, and there can be no country fixed effects to control for omitted country-specific factors. Because patterns might differ among regions, columns (6) to (10) reprise the specifications of columns (1) to (5) but adding regional dummy variables. Omitted regions Americas and Oceania are controlled for by the regression constant. More detailed sets of estimates that build up to the full model more gradually are shown in Table A.14 and A.15.[30]

Figure 3 shows the distribution of Stayer Share (w/in Region) over the 153 countries covered, with the upper panel representing the full distribution and the lower panel zeroing in on the 124 countries having 70% or more of their year 1500 people's descendants residing in the home region in 1960, including separate break-outs of the top five percentiles. In addition to the several countries identified by name on the figure, for illustrative purposes, the reader can find a detailed list of stayer shares for all included countries in Table A.13. These exhibits show that while only about a sixth of covered countries projected a third or more of their descendants abroad during the colonial era, there remains considerable variation among countries having smaller diasporas, for example a French diaspora constituting 25% of 1500 CE France's descendants, a Greek diaspora of 17%, a Syrian diaspora of 13%, a Mexican diaspora of 10%, a Turkish diaspora of 7%, an Iranian diaspora of 3%, and Indian and Chinese diasporas of slightly over 2%.

[30]While each half of Table 7 builds the full model by adding two variables per step, Table A.14 and A.15 show the intermediate steps when adding one variable at a time, by always excluding the region dummies in Table A.14 and always including them in Table A.15. An exception is that column (5) adds *statehist1500* instead of *pop1500/arable* prior to including both variables in column (6), so that we can observe the impact of each without concern for their correlation with one another.

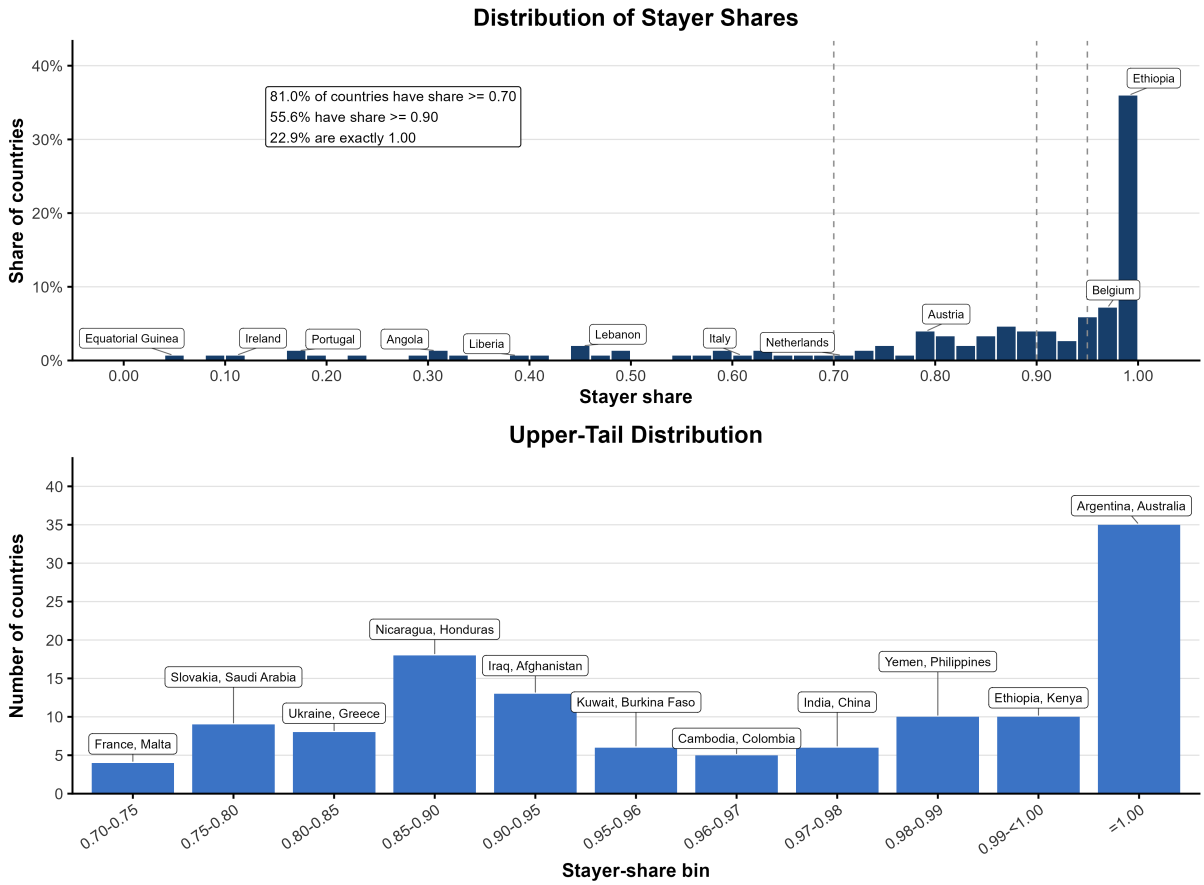


Figure 3: Distribution of Stayer Share

Notes: The upper panel shows the full distribution and the lower panel the upper 30% of the distribution of Stayer Share (w/in Region), the share of descendants of year 1500 population within the contemporary country borders who are estimated, according to Chanda et al.'s migration matrix (Chanda et al., 2014) and U.N. population data (United Nations, 2024), to have lived within the home region in 1960. The lower panel's data are presented in 5% bands for stayers shares between 70% and 95%, and in 1% bands plus a bar for countries at the maximum level, within the top 5%.

The regressions of Table 7 indicate that the full model can explain about 31% of the variation in stayer share without region terms and about 44% with them. Of the region dummies, only the coefficients for SSA and Europe are significant in the full specification, presumably due to the negative impacts on stayer share of the slave trade and on voluntary emigration by people of colonizing powers and their neighbors, respectively, on share of descendants still in the regions. Five of the nine explanatory factors are significant at the 5% level or better, including several at the 1% level, in specifications (5) and (10).

The rankings of shares of variance explained by each non-region factor, shown in Tables A.8 and A.9, are generally similar for both the specification with and the one without region dummies, but they differ, in some ways substantially, from those calculated for Table A.6 (which we used to order variables in Tables 3 – 6). In both the *stayer share* analysis without and that with region dummies, the top three variables by explanatory power are *colonizer*, *disease prevalence* and *landlocked*, together explaining 18 – 20% of the variance. Of the three, only *disease prevalence* is among the three most important variables in bilateral analyses (see Table A.6). Whereas *destination landlocked* was only the sixth most important, and *source landlocked* the 9$^{\text{th}}$, in Table 5, explaining 2.0% and 1.1% of DS_{ij}, respectively, *landlocked* is the most important predictor of *stayer share* when region dummies are included, accounting for 8.7% of overall variation (or 20% of the explained variation), and it is third in importance without region dummies, explaining 4.9% of variation.

Focusing on the Shapley analysis for the full specification of col. (5) of Table 7, without region dummies, the three directly resource-related factors—*pop1500/arable hectare*, *arable area*, and *pre-1500 avg. max-cal/ha.*—account for about 6% (a 19% share of the total), the other "natural" or "efficiency-related" factors of being *landlocked*, of *disease prevalence*, and *temperature*, account for about 15% (a 49% share of the total), the power factors *colonizer* and *colony* account for 10.1% (a 33% share of the total), and the mixed political, history and social development indicator *statehist1500* accounts for a negligible remainder. The last finding may be explained by the already mentioned fact that the colonizing powers (Spain, Britain, etc.) were above average but below the top-ranking countries (Iraq, Turkey, etc.) in terms of state history, while the countries losing a substantial share of residents to the slave trade (Ghana, Ivory Coast, etc.) were generally lower but not on average lowest in state history, and thus the countries that were affected most extensively by outgoing colonial era migration display little linear correlation of emigration and state history.

Table 7: Stayer Share (within Region)

Stayer Share (w/in Region)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Colonizer	−0.372***	−0.334***	−0.343***	−0.353***	−0.355***	−0.272***	−0.215*	−0.228**	−0.262**	−0.265**
	(0.100)	(0.106)	(0.103)	(0.100)	(0.099)	(0.102)	(0.112)	(0.107)	(0.112)	(0.110)
Disease Prevalence	−0.073**	−0.065**	−0.069**	−0.199***	−0.198***	−0.095***	−0.071**	−0.083**	−0.156***	−0.163***
	(0.031)	(0.030)	(0.029)	(0.049)	(0.049)	(0.033)	(0.034)	(0.034)	(0.045)	(0.050)
Landlocked		0.085***	0.088***	0.122***	0.124***		0.147***	0.148***	0.169***	0.159***
		(0.023)	(0.024)	(0.030)	(0.031)		(0.034)	(0.035)	(0.037)	(0.036)
Pop1500 / Arable Lands (ln)		−0.013		0.005	0.006		−0.016		−0.006	−0.003
		(0.015)		(0.018)	(0.017)		(0.015)		(0.017)	(0.017)
Statehist 1500 (Ln)			−0.001	−0.000	0.001			0.002	−0.001	−0.001
			(0.003)	(0.003)	(0.004)			(0.004)	(0.004)	(0.004)
Arable Land Area (Ln)				0.035***	0.034***				0.031***	0.031***
				(0.012)	(0.011)				(0.011)	(0.011)
Temperature				0.012***	0.012***				0.011***	0.011***
				(0.003)	(0.003)				(0.004)	(0.004)
Colony Dummy					0.031					−0.031
					(0.039)					(0.036)
Pre-1500 Avg. Max-Cal./Ha					0.023*					0.017
					(0.012)					(0.018)
Central Asia						−0.037**	−0.166***	−0.185***	−0.097	−0.098
						(0.016)	(0.038)	(0.053)	(0.060)	(0.067)
East Asia						−0.028	−0.017	−0.062	0.076	0.047
						(0.029)	(0.051)	(0.057)	(0.078)	(0.084)
Europe						−0.205***	−0.216***	−0.248***	−0.136**	−0.172**
						(0.044)	(0.046)	(0.061)	(0.062)	(0.069)
Middle East and North Africa						−0.045	−0.022	−0.059	0.015	0.026
						(0.031)	(0.033)	(0.045)	(0.042)	(0.056)
South Asia						0.013	−0.013	−0.040	−0.029	−0.034
						(0.019)	(0.044)	(0.046)	(0.043)	(0.043)
Southeast Asia						0.011	0.003	−0.013	−0.048	−0.056
						(0.018)	(0.031)	(0.039)	(0.039)	(0.041)
Sub-Saharan Africa						−0.143***	−0.192***	−0.194***	−0.184***	−0.160***
						(0.040)	(0.046)	(0.047)	(0.044)	(0.052)
Constant	0.915***	0.877***	0.886***	0.183	0.206	1.013***	0.972***	1.019***	0.324	0.353*
	(0.015)	(0.029)	(0.024)	(0.204)	(0.197)	(0.010)	(0.029)	(0.040)	(0.201)	(0.209)
Observations	153	153	153	153	153	153	153	153	153	153
R^2	0.140	0.178	0.173	0.293	0.305	0.285	0.364	0.357	0.431	0.435

Notes: All regressions are estimated using OLS. The dependent variable is Stayer Share (w/in Region) calculated by the authors based on Chanda et al. (2014) and World Bank (2024). Independent variables include *State History* (Borcan et al., 2018 and authors' calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Hectare* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Hectare, 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and authors' calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and authors' calculations). See Section 2 and Appendix B for further details. Columns (1)–(5) report estimates without region fixed effects; columns (6)–(10) are estimations including them. The omitted (reference) category is defined as *New World*, which includes the Americas and Oceania regions. Robust (Huber–White) standard errors are reported in parentheses. *** ($p < 0.01$), ** ($p < 0.05$), * ($p < 0.1$)

The relationship between the "natural" factors and *stayer share* is strongly consistent with our explanatory framework: people left places with more disease and stayed in places with warmer climates and more arable land. Year 1500 population density (represented by *pop/arable*) shows a negative significant relation to *stayer share*, as would be expected, in columns (2) and (7) of Table 7, but thereafter loses significance, though retaining its sign and order of magnitude. The land quality indicator *avg. max-cal/ha.* has a positive coefficient but low explanatory power, hence it is added in columns (5) and (10) only, and is only marginally significant in column 5, which omits region dummies. The fact that all region dummies other than those for SSA and Europe obtain insignificant coefficients with small absolute values is consistent with the fact that only those two regions experienced substantial out-migration relative to their base populations.

The relative prominence of *landlocked* in explaining the "home bound" descendant share is interesting. The influential work of Gallup et al. (1999) found that landlocked countries were greatly economically disadvantaged, with respect to late 20th Century economic growth, by their lower access to international trade given the lower cost of sea than of land transport. In the present analysis, they are found to differ significantly from others, after controlling for region, in retaining (according to the point estimates in columns (5) and (10) of Table 7) roughly 12% more (without region controls) or almost 16% more (controlling for regions) of their global descendants (on average) than otherwise comparable countries. It is remarkable that, although it would probably have required at most a few weeks of travel to reach a port from most locations in the sample's 37 landlocked countries, so many fewer residents ended up departing from them over the course of almost five centuries—arguably an indication of cultural inertia as much as of cost and benefit considerations *per se*. However, considerable caution is warranted since, while the remark just made may apply well to cases such as Switzerland, Hungary, or Nepal, the reason for having fewer descendants abroad in sub-Saharan African cases like Niger and Chad stems principally from logistics of the slave trade (Eltis et al., 1999). Also, emigration from the two landlocked countries of the Americas (Bolivia and Paraguay) differed little from that of the region's other countries, most of which had exceedingly little recorded emigration from their region. Furthermore, our one sub-region consisting entirely of landlocked countries—Central Asia—obtains a significant negative region coefficient in models (6) – (8). The imprecisely estimated coefficient - 0.100 on Central Asia combined with the highly significant estimate 0.157 on landlocked imply that Central Asia kept about 5.7% more descendants at home than the average, while direct calculations show that average descendants outside the home region are slightly above 1% of overall region descendants.[31]

Taken together, the fact that just over two-thirds of the cross country variation in stayer share that our analysis explains is due to the "natural" or "efficiency-related" correlates of migration reaffirms our general theme. The fact that the colonizing countries "propelled" substantially more of their populations abroad, *ceteris paribus*, is also clearly indicated by these estimates. The full specification of column (10) suggests that more than a quarter fewer of the colonizing countries' year 1960 descendants were on average to be found in the home country than in otherwise identical countries.[32] The coefficient on the Europe region dummy implies that colonizers lost an additional 16.3% of descendants to emigration, while non-colonizers within the region also sent off about 16% on average. We discuss impacts on the relative population sizes of Europe and of the Asian population giants in the final section.

5 Discussion and Conclusion

Some of the largest scale migrations with the most profound impacts on the dispersion of global population in world history were those that followed the exploration of the Atlantic, Indian and Pacific Oceans by Spanish and Portuguese navigators beginning in the 15$^{\text{th}}$ Century, and European empires' conquests of the Americas, Australia and New Zealand, and much of the rest of the world, which followed in the next four and a half centuries. Although members of our species spread from Africa to the breadth of Eurasia by fifty thousand years ago, and although founder populations reached Australia by forty thousand and the Americas by twenty thousand years ago, the populations of the three "Old World" continents lost ongoing contact with those of the other continents and islands until west Europeans honed Eurasia-wide progress in navigation and ship-building to undertake increasingly regular voy-

[31]Readers familiar with Central Asian (in some respects equivalent to "Silk Road") history, may recall that the region probably did "export" large numbers of migrants, during earlier centuries, into the Middle East, Europe, South Asia and China, as indicated by accounts of the Mongol, Turkic, Timurid and other conquests and population shifts between the late first millennium and the mid second millennium CE. But descendants of those émigrés would be counted as peoples of Russia, India, Turkey, etc., by 1500, and by the sources on which our own source Chanda et al. (2014) relies.

[32]Note that this colonizer emigration estimate averages the effects for countries like Spain, Portugal and Britain, which sent large numbers to their colonies, and ones like the Netherlands and Belgium, for which numbers were considerably smaller. Interestingly, Blanc and Wacziarg (2025), show that the early timing of its demographic transition caused France to have far fewer descendants abroad than did other major colonizers.

ages between them, initiating the first centuries of fully global contacts. There had been migrations of considerable scale within Eurasia and North Africa (viz. migrations of Indo-European, Turkic and Arabic language speakers), within sub-Saharan Africa (viz. those of Bantu language speakers), and within the Americas (viz. Tupi-Guarani, Athabaskan, and Arawak/Carib language group migrations). But between the time of the populating of the Americas and Australia and the voyages of Columbus and successors, known movements of people among these continents were exceedingly small in scale. For this reason, the cultures, languages, belief systems, historical knowledge, and technological capabilities of each region had developed in radical isolation from one another.

During the centuries after 1500, however, millions of Europeans migrated to the Americas, Australia and New Zealand, millions of sub-Saharan Africans were brought involuntarily to the Americas and islands nearer to Africa, and long distance migration also took place between various parts of Asia and the Americas, sub-Saharan Africa, and Australia. In all, a little under 13% of global population in 1960 lived outside the region its ancestors had lived in 1500.[33]. If we instead counted the millions of Mexicans, Colombians, Brazilians, and others of mixed ancestry among those alive in 1960 or today whose location and descent reflects the colonial era migrations, we are likely to find that the proportion of global population that had substantial (say 1/4 or more) ancestry from a different global region and are thus directly affected by the world population redistribution of 1500 - 1960 approaches twice the 13% share. The shares of descendants of Eurasian and African countries that were living outside their ancestral country as of 1960 was quite low in some landlocked parts of Eurasia and Africa, but rises to levels as high as 88% of the descendants of the year 1500 Irish, 67% of descendants of year 1500 Spanish, and 54% of descendants of year 1500 Lebanese.

In our paper, we used data organized at the level of countries of year 2000 to examine the roles that roughly a dozen basic factors play in accounting for who moved to and whose descendants ended up living where outside of their home region over this four and a half century period. Although a substantial part of the population movement that occurred was coerced, although military and political factors such as who colonized whom played a large role in explaining who moved where, and although indigenous populations were decimated by unfamiliar diseases, forced off their lands, and in numerous cases worked to death, forced

[33]For example the estimate requires treating some 25 million mixed-ancestry Mexicans of 1960 as "fractional (approximately one third) Spaniards". The 25 million Mestizos are thus counted, among the 13% of global "movers", as if they were approximately "9 million Mexicans of Spanish descent"

to bear children for or robbed of their cultures by colonizers, some patterns consistent with an efficiency-enhancing "redistribution" of post-"Great Dying" world population turn out to be discernible.[34] On the whole, migration occurred from places with higher to places with lower (post "die-off") ratios of population to cultivable land, from places with less to places with more cultivable land, from places with less productive to places with more productive land, from places with cooler to places with warmer temperatures, and from places with higher to places with lower disease prevalence. Also consistent with efficiency, narrowly understood, is that bilateral migrations tended to be from countries bordering on one of the world's oceans or ocean-draining seas to other countries also bordering on an ocean, leaving landlocked countries significantly less although not entirely uninvolved.

While the majority of the population movement is consistent with one or more of these efficiency-related or natural factors being at work, a substantial share of the explanatory power in our models still resides in political factors: colonizing countries and countries sharing their continent with and relatively nearer to colonizers, plus countries in the sub-continent from which colonizers obtained slaves, account for the lion's share of those whose descendants were to be born in a distant region, and countries that at the time of colonization had no large scale polity, or ones in which the history of polities was shorter than that of the more developed parts of Eurasia, were the main recipients of movers. Political weakness or lack of political unity of countries with less state experience, e.g. in parts of sub-Saharan Africa, in Southeast Asia, and even in the Indian subcontinent in which empires had ruled much of but rarely the entirety of the subcontinent, probably contributed to ease of colonial conquest. Differences between source regions with respect to the extent to which efficiency-fostering population movements hold, as seen in Table 5, also provide evidence of interference of political power factors with those movements that might have been expected had free mobility of labor been the universal driving force. Europe contained the overseas colonizing powers, people from most of that continent were able to emigrate to its colonies, and the patterns predicted find the most support there. In contrast, there are more insignificant coefficients, and a contrary-to-pattern significant coefficient for source land productivity, for North Africa and Asia, from which emigration to the New World was relatively difficult, especially before the late 19th and early 20th centuries. Much of this large potential source

[34] If one assumes a pre-Columbian population of 57 million (per Denevan) and a contraction by 93% (a rough consensus estimate) in the Americas, the number who died there alone due to the introduction of Old World diseases to it would be 53 million, almost 2/3 as many people as the entire population of Europe at the time (per Maddison, McEvedy & Jones or HYDE estimates).

region was part of the land-based Ottoman, Russian, and Chinese empires, which fell behind the European technological frontier in the second half of the centuries studied, and many of its politically weaker countries became colonies.

By contributing to the disparity of population density and wealth between western Europe and most of Asia, as of the 18th century, the colonization and migration process may have helped to launch the Industrial Revolution in Europe (Acemoglu et al., 2005), and it certainly contributed to the result that Asian countries like China and India had larger and more impoverished populations as they began concerted economic catch-up in the late 20th Century. Our data indicate that, under the simple counterfactual scenario in which the 1960 descendants of European emigrants lived in Europe rather than the Americas and other regions, while descendants of migrants from other regions were removed from its count, Europe's 1960 population would have been 903 million rather than 603 million in 1960, which would have made it 35 percent larger than that of China and 107 percent larger than that of India in 1960, rather than being 10 percent smaller than that of China and only 38 percent larger than that of India. Impacts of assuming that both sub-Saharan African and European 'leavers' had stayed in their home regions (and symmetric treatment of in-migrants) on the comparative population sizes of Europe, sub-Saharan Africa and China and India in 1960, 2000, and 2025, are shown in Table 9. These calculations make the extreme simplifying assumptions that emigrant populations would have grown at the rates they registered abroad until 2000, then grown at the average observed rate of each home region from 2000 onward. Among other things, they show that in the counterfactual world lacking the major global migrations of 1500 - 1960, Europe and sub-Saharan Africa might each have been rough peers in population to each of the world's largest countries today, whereas there are in fact only about half as many residents of Europe as of China and India today.

To be sure, the assumption that the European "lineages" in question would have grown in number to an equal degree, had they remained at home, is at odds with demographic theory, and there are other reasons the numbers shown should be taken as crude illustrations, only. The long-run impact of the African exodus is smaller partly because it lost less of its population but even more because the growth rate of SSA population within the continent exceeded those of China especially, during 1960 - 2025, and also that of SSA's diaspora.[35]

[35]That is, the world's modern-era African-descended population would be considerably larger had the African diaspora experienced the same population growth rates as did the inhabitants of SSA at home.

Table 8: Hypothetical scenarios of Europe and Sub-Saharan Africa Populations

	1960			**2000**			**2025**		
	Pop.	**% China**	**% India**	**Pop.**	**% China**	**% India**	**Pop.**	**% China**	**% India**
Europe – actual	602.8	90.4	138.3	723.9	57.3	68.3	742.4	52.4	50.7
Europe – hypothetical	902.9	135.4	207.1	1,300.7	103.0	122.7	1,333.9	94.2	91.1
SSA – actual	224.7	33.7	51.5	662.2	52.4	62.5	1,247.3	88.1	85.2
SSA – hypothetical	269.7	40.4	61.9	750.9	59.5	70.9	1,414.3	99.9	96.6
China	667.1	–	–	1,262.6	–	–	1,416.0	–	–
India	436.0	–	–	1,059.6	–	–	1,464.0	–	–

Notes: Actual populations by year shown for Europe, sub-Saharan Africa (SSA), China and India are taken from United Nations (2024). Europe and SSA totals are calculated for the same countries as in the rest of our analysis; see Appendix for details. Hypothetical populations of Europe and of SSA in 1960 and 2000 are computed by calculating the total descendants of immigrants and of emigrants to and from other regions (using the WMM and UN population numbers), adding back the former and subtracting the latter from actual population. Hypothetical population of 2025 is calculated by multiplying the hypothetical populations of 2000 by the actual population growth rate of each region. "% China" ("% India") shows the population of the region in the row heading as a percentage of the population of China (of India) in the year listed in the nearest column heading to the left. Population growth rates of 2000–2025, as imputed from UN population statistics, are 0.10% p.a. for Europe, 2.57% p.a. for sub-Saharan Africa, 0.46% p.a. for China, and 1.30% p.a. for India.

One motivation for adopting a macro-regional stance in parts of our discussion is the importance of the knock-on effects of migratory and other trends of the 1500 – 1960 era for post-1960 migration. In particular, the disease effects of Old World-New World contact and the migrations of 1500 – 1960 left Europe, North America and Australia among the highest income but less densely populated parts of the world, even after the substantial net migration to the latter regions, studied by us. Combined with further reduced transportation and political barriers, this pairing of global income and population disparities has been a leading contributor to the South-to-North migration that emerged as a top source of political, social and cultural tensions during the past few decades, and that has lately led to doubts about the sustainability of the West's liberal political order (Alesina et al., 2023; Gidron and Hall, 2017; Guiso et al., 2024).

Our analyses point to the spatial frictions that allowed the "Persistence of Fortune" for people (Chanda et al., 2014) to occur amidst the "Reversal of Fortune" for territories (Acemoglu et al., 2002), as illustrated, e.g., by the contrast between the U.S. and Canada, on the one hand, and Mexico and Central America, on the other. Whereas some ancient, densely populated Old World states successfully resisted both colonization and European settlement, New World empires and states were easily overthrown due to technological and institutional disadvantages and lack of disease resistance, and the Europeans who settled in the least thickly populated parts of the New World carried their technological and institutional advantages and connections with them. Our results indicate important ways in which both New World development gaps and the divergence between Europe and the rest of the Old

World were conditioned by pre-existing differences of state history and population density.

Late 20^{th} and early 21^{st} century migration differs in some respects from that of most of the period that we cover, but appears to be similar in others. While detailed analysis lies outside the scope of our paper, modern migration shares striking similarities with these historical flows. Today, culturally Western, high-income countries in Europe and North America remain prime destinations. In both cases, these destinations tend to possess lower population densities and more productive land than the regions sending migrants, which currently include North Africa, Asia, Sub-Saharan Africa and Latin America. Differences in standards of living and wages, and in availability of social services, may be the primary attractors of migration, but considerable additional migration appears unlikely to be stopped except by dramatic tightening of borders, and enforcement efforts that can only with much effort be kept from exceeding the bounds of humane treatment.

The shifts in global population attributable to migration and the slave trade, between 1500 and 1960, went a small way, at most, towards bringing global population distribution towards an equilibrium, as they were relatively small compared to other forces, especially the massive increases in global income inequality, and the immense and unequal increases in world population. Contemporary demand for right-wing populism—driven by the economic insecurity and subjective social status threats among native working classes (Gidron and Hall, 2017; Guiso et al., 2024)—is a direct, delayed reaction to the immense global inequalities and core-periphery divides that were originally forged by the 1500–1960 population shifts on which we focused in this paper. If the people of the richer countries want to maintain their standards of living while accepting only modest inflows of migration and also stopping short of grossly inhumane treatment of would-be migrants, finding ways to foster political stability and economic growth in the rest of the world will have to be prioritized.

References

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, *91*(5), 1369–1401.

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

Acemoglu, D., Johnson, S., and Robinson, J. A. (2005). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American Economic Review*, *95*(3), 546–579.

Alesina, A., Harnoss, J., and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, *21*, 101–138.

Alesina, A., Miano, A., and Stantcheva, S. (2023). Immigration and redistribution. *The Review of Economic Studies*, *90*(1), 1–39.

Ashraf, Q., and Galor, O. (2013). The "out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, *103*(1), 1–46.

Beckert, S. (2025). *Capitalism: A global history*. Allen Lane.

Beine, M., Bertoli, S., and Moraga, J. F.-H. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, *39*(4), 496–512.

Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, *95*(1), 30–41.

Beine, M., and Parsons, C. (2015). Climatic factors as determinants of international migration. *The Scandinavian Journal of Economics*, *117*(2), 723–767.

Bertoli, S., Goujon, M., and Santoni, O. (2016). *The CERDI-Seadistance Database* (Working Paper No. 2016/07). CERDI. https://shs.hal.science/halshs-01288748

Blanc, G., and Wacziarg, R. (2025). *Malthusian migrations* (tech. rep. No. 33542). National Bureau of Economic Research.

Bockstette, V., Chanda, A., and Putterman, L. (2002). States and markets: The advantage of an early start. *Journal of Economic growth*, *7*(4), 347–369.

Bolt, J., and Van Zanden, J. L. (2025). Maddison-style estimates of the evolution of the world economy: A new 2023 update. *Journal of Economic Surveys*, *39*(2), 631–671.

Borcan, O., Olsson, O., and Putterman, L. (2018). State history and economic development: Evidence from six millennia. *Journal of Economic Growth*, *23*(1), 1–40.

Borcan, O., Olsson, O., and Putterman, L. (2021). Transition to agriculture and first state presence: A global analysis. *Explorations in Economic History*, *82*, 101404.

Chanda, A., Cook, C. J., and Putterman, L. (2014). Persistence of fortune: Accounting for population movements, there was no Post-Columbian Reversal. *American Economic Journal: Macroeconomics*, *6*(3), 1–28.

Chen, S., and Kung, J. K. (2016). Of maize and men: The effect of a new world crop on population and economic growth in China. *Journal of Economic Growth*, *21*(1), 71–99.

Conte, M., Cotterlaz, P., Mayer, T., et al. (2022). The CEPII gravity database.

Correia, S., Guimarães, P., and Zylkin, T. (2020). Fast poisson estimation with high-dimensional fixed effects. *The Stata Journal*, *20*(1), 95–115.

Denevan, W. M. (1992). *The native population of the Americas in 1492*. University of Wisconsin Press.

Diamond, J. (1997). *Guns, germs and steel: The fates of human societies*. New York: W. W. Norton.

Dobyns, H. F. (1966). An appraisal of techniques with a new hemispheric estimate. *Current Anthropology*, *7*(4), 395–416.

Eltis, D., Behrendt, S. D., Richardson, D., and Klein, H. S. (Eds.). (1999). The trans-atlantic slave trade: A database on CD-ROM. https://www.cambridge.org/de/universitypress/subjects/history/regional-history-after-1500/trans-atlantic-slave-trade-database-cd-rom

Ertan, A., Fiszbein, M., and Putterman, L. (2016). Who was colonized and when? A cross-country analysis of determinants. *European Economic Review*, *83*, 165–184.

Fincher, C. L., Thornhill, R., Murray, D. R., and Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1640), 1279–1285.

Foubert, K., and Ruyssen, I. (2024). Global migration and the role of terrorist attacks. *Journal of Economic Behavior & Organization*, *220*, 507–530.

Gallup, J. L., Sachs, J. D., and Mellinger, A. D. (1999). Geography and economic development. *International Regional Science Review*, *22*(2), 179–232.

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

Gidron, N., and Hall, P. A. (2017). The politics of social status: Economic and cultural roots of the populist right. *The British Journal of Sociology*, *68*, S57–S84.

Guiso, L., Herrera, H., Morelli, M., and Sonno, T. (2024). Economic insecurity and the demand for populism in Europe. *Economica*, *91*(362), 588–620.

Hariri, J. G. (2012). The autocratic legacy of early statehood. *American Political Science Review*, *106*(3), 471–494.

Head, K., and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. *Handbook of International Economics*, *4*, 131–195.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of Colonial trade linkages after independence. *Journal of International Economics*, *81*(1), 1–14.

HYDE. (2023). History database of the global environment (HYDE v 3.3) [From Utrecht University/PBL Netherlands Environmental Assessment Agency - History Database of the Global Environment, Based on Klein Goldewijk, Beusen, Doelman and Stehfest (2017)]. https://doi.org/10.24416/UU01-AEZZIT

Klein Goldewijk, C. G. M., Beusen, A., Doelman, J., and Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth System Science Data*, *9*, 927–953.

Koch, A., Brierley, C., Maslin, M. M., and Lewis, S. L. (2019). Earth system impacts of the European arrival and Great Dying in the Americas after 1492. *Quaternary Science Reviews*, *207*, 13–36.

Lander, K., and Pritchett, J. (2009). When to care: The economic rationale of slavery health care provision. *Social Science History*, *33*(2), 155–182.

Larch, M., Shikher, S., and Yotov, Y. V. (2025). Estimating gravity equations: Theory implications, econometric developments, and practical recommendations. *Review of International Economics*.

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., Dobney, K., et al. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, *111*(17), 6139–6146.

Lovell, W. G. (2021). Historical demography introducer: W. george lovell. In *Forest, field, and fallow: Selections by william m. denevan* (pp. 1–39). Springer.

McEvedy, C., and Jones, R. (1978). *Atlas of World population history.* Penguin Books Ltd, Harmondsworth, Middlesex, England.

McNeill, W. H. (1976). *Plagues and peoples.* New York: Anchor Press/Doubleda.

Murray, D. R., and Schaller, M. (2010). Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. *Journal of Cross-Cultural Psychology*, *41*(1), 99–108.

Nunn, N., and Qian, N. (2010). The columbian eexchange: A history of disease, food, and ideas. *Journal of Economic Perspectives*, *24*(2), 163–188.

Obolensky, M., Tabellini, M., and Taylor, C. (2024). *Migration, climate similarity, and the consequences of climate mismatch* (tech. rep.). National Bureau of Economic Research.

Our World in Data. (2026). Population-hyde 3.3 [HYDE (2023) – with major processing by Our World in Data. Original data: PBL Netherlands Environmental Assessment Agency, "History Database of the Global Environment 3.3". Archived on April 2, 2026]. https://archive.ourworldindata.org/20260402-163007/grapher/population.html

Özden, Ç., Parsons, C. R., Schiff, M., and Walmsley, T. L. (2011). Where on Earth is everybody? the evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, *25*(1), 12–56.

Putterman, L., and Weil, D. N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *The Quarterly Journal of Economics*, *125*(4), 1627–1682.

Rodenwaldt, E., and Jusatz, H. J. (Eds.). (1961). *Welt-seuchen-atlas: Weltatlas der seuchenverbreitung und seuchenbewegung* [3 volumes]. Falk-Verlag.

Silva, J. S., and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 641–658.

Silva, J. S., and Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, *112*(2), 220–222.

Silva, J. S., and Tenreyro, S. (2022). The log of gravity at 15. *Portuguese Economic Journal*, *21*(3), 423–437.

Simmons, J. S., Whayne, T. F., Anderson, G. W., and Horack, H. M. (1944). *Global epidemiology: A geography of disease and sanitation. volume 1: India and the Far East. part 2: The Pacific Area.* J. B. Lippincott Company.

United Nations. (2024). World population prospects 2024. https://population.un.org/wpp/

Wood, P. H. (1996). *Black majority: Negroes in colonial South Carolina from 1670 through the Stono rebellion*. WW Norton & Company.

World Bank. (2026). World Development Indicators. Retrieved March 25, 2026, from https://databank.worldbank.org/source/world-development-indicators

Yotov, Y. V., Piermartini, R., Larch, M., et al. (2016). *An advanced guide to trade policy analysis: The structural gravity model*. WTO iLibrary.

APPENDIX

Land, Power and Population Density in the Global Population Shifts of 1500 – 1960

Mehmet Burak Akren[36] Arhan S. Ertan[37] Louis Putterman[38]

July 22, 2026

[36]University of Duisburg-Essen, Department of Economics, 45141 Essen, Germany. e-mail: mehmet.akren@vwl.uni-due.de

[37]Boğaziçi University, Department of International Trade, 34342 Istanbul, Turkey. e-mail: arhan.ertan@bogazici.edu.tr

[38]Brown University, Department of Economics, 02912 Providence RI, USA. e-mail: louis_putterman@brown.edu

A Countries and Regions

Table A.1: Regional Classification of Countries

Region	Countries
Central Asia	Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan
East Asia	China, Hong Kong, Japan, Korea, Dem. Rep. (North), Korea, Rep. (South), Mongolia
MENA	Algeria, Azerbaijan, Bahrain, Egypt, Iran, Iraq, Israel/Palestine, Jordan, Kuwait, Lebanon, Libya, Malta, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, Turkey, United Arab Emirates, Yemen
South Asia	Bangladesh, Bhutan, India, Nepal, Pakistan, Sri Lanka
Southeast Asia	Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam
Europe	Albania, Armenia, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia, Moldova, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom
Oceania+	Australia, Fiji, New Zealand, Papua New Guinea
The Americas	Argentina, Belize, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Trinidad and Tobago, United States, Uruguay, Venezuela

Continued on next page

Region	Countries
Sub-Saharan Africa	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Dem. Rep. of Congo, Rep. of Congo, Cote d'Ivoire, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe

Table A.2: Super-regions and regions

Super-region	Region(s) included
Europe	Europe
Sub-Saharan Africa	Sub-Saharan Africa
N.Africa & Asia	Central Asia East Asia South Asia Southeast Asia MENA
New World	The Americas Oceania+

B Descriptive Statistics

The table below reveals estimated Pearson correlation coefficients for the explanatory variables. To assess potential multicollinearity issues, VIF diagnostics have been computed. All of them are below ten and the results are provided in the supplementary tables.

Table A.3: Pairwise Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Arable Land Area (Ln)	1.000						
(2) Pop / Arable 1500 (Ln)	-0.280***	1.000					
(3) Statehist 1500 (Ln)	0.093	0.386***	1.000				
(4) Pre-1500 Avg. Max-Cal./Ha	0.109	-0.083	-0.130	1.000			
(5) Temperature	-0.234***	0.089	-0.279***	-0.166**	1.000		
(6) Historical Disease Prevalence	0.096	0.083	-0.190**	-0.141*	0.723***	1.000	
(7) Colony Dummy	-0.151*	-0.106	-0.457***	-0.142*	0.710***	0.572***	1.000

Notes: Pairwise correlation coefficients are reported. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A.4: Pairwise Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Diff. Arable Land Area (ln)	1.000								
(2) Diff. Pop / Arable 1500 (ln)	-0.348***	1.000							
(3) Diff. Statehist 1500 (ln)	0.059***	0.406***	1.000						
(4) Diff. Pre-1500 Avg. Max-Cal./Ha	0.123***	-0.120***	-0.122***	1.000					
(5) Diff. Temperature	-0.208***	0.110***	-0.121***	-0.179***	1.000				
(6) Diff. Historical Disease Prevalence	0.093***	0.071***	-0.061***	-0.153***	0.735***	1.000			
(7) Same Colonizer	-0.021***	-0.004	-0.005	0.014**	-0.003	-0.006	1.000		
(8) Colonizer of Destination	-0.038***	-0.029***	-0.037***	-0.018***	0.065***	0.069***	-0.022***	1.000	
(9) Destination is Colonizer	0.040***	0.029***	0.038***	0.018***	-0.065***	-0.066***	-0.021***	-0.004	1.000

Notes: Pairwise correlation coefficients are reported. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Note that the sample size in Table A.4 (N = 24786) is substantially larger than in Table A.3 (N = 162), accounting for the higher statistical significance of correlation coefficients with similar magnitudes.

Table A.5: Descriptive Statistics

Variable	Obs.	Mean	S.D.	Min	Max
DS (Descendant Share)	20,194	0.001	0.014	0.000	0.796
Destination Arable Land Area (Ln)	20,194	14.228	2.146	6.908	18.983
Destination Pop / Arable 1500 (Ln)	20,194	-1.153	1.399	-5.688	2.377
Destination Statehist 1500 (Ln)	20,194	-5.115	5.460	-13.816	-0.275
Destination Pre-1500 Avg. Max-Cal./Ha	20,194	0.001	1.015	-2.098	3.729
Destination Historical Disease Prevalence	20,194	0.182	0.599	-1.180	1.200
Destination Temperature	20,194	18.315	8.233	-7.929	28.639
Source Arable Land Area (Ln)	20,194	14.455	1.930	7.606	18.983
Source Pop / Arable 1500 (Ln)	20,194	-1.093	1.308	-5.302	2.377
Source Statehist 1500 (Ln)	20,194	-4.478	4.292	-11.513	-0.275
Source Pre-1500 Avg. Max-Cal./Ha	20,194	-0.016	0.984	-2.098	3.729
Source Historical Disease Prevalence	20,194	0.185	0.615	-1.180	1.200
Source Temperature	20,194	17.906	8.333	-7.929	28.639
Colonizer of Destination Dummy	20,194	0.005	0.071	0.000	1.000
Same Colonizer Dummy	20,194	0.083	0.276	0.000	1.000
Destination is Colonizer Dummy	20,194	0.005	0.068	0.000	1.000
Source Colony Dummy	20,194	0.604	0.489	0.000	1.000
Destination Colony Dummy	20,194	0.624	0.485	0.000	1.000
Difference in Arable Land Area (Ln)	20,194	-0.228	2.896	-12.075	11.377
Difference in Pop / Arable 1500 (Ln)	20,194	-0.061	1.935	-8.065	7.679
Difference in Statehist 1500 (Ln)	20,194	-0.637	7.275	-13.541	11.238
Difference in Pre-1500 Maximum Calories	20,194	0.018	1.486	-5.827	5.827
Difference in Historical Disease Prevalence	20,194	-0.003	0.930	-2.380	2.380
Difference in Temperature	20,194	0.410	12.631	-36.569	36.569
Distance, ln	20,194	9.312	0.487	5.631	10.246
Source Landlocked Dummy	20,194	0.240	0.427	0.000	1.000
Destination Landlocked Dummy	20,194	0.232	0.422	0.000	1.000

C Notes on Variables

Coding of countries missing *Statehist 1500* values in Borcan *et al.* (2018)

Values of the state history index are missing for 12 countries included in the present paper. We developed values for them using the same principles, in order to include these countries in our study. This involves assigning a value in the 0,1 interval to each of three dimensions for each relevant half century: z1 = presence of a state-level polity (0 = none, 0.5 = paramount chiefdom or proto-state, 1 = full state); z2 = home-based versus colony (0.5 = colony, 1 = home-based); and z3 = share of territory of present-day country ruled and unity of polity (0.3 = 10% or less of area ruled or multiple separate polities; 0.5 = 10 – 25% of area ruled or a few main polities; 0.75 = 25 – 50% of area ruled or two main polities; 1 = 50 – 100% of area ruled by one main polity). The three dimension scores are multiplied by one another (z1*z2*z3) to obtain a combined score s1 for each 50 year period, or if there is a shift of regime mid-way, scores are obtained for each sub-period and the 50 year period is assigned their weighted average, with weights proportionate to duration. When there was no state presence at all, i.e. z1 = 0, then z2 and z3 are irrelevant, so where we write z1 = 0 below, we don't specify z2 and z3 values. Values for 50 year periods are summed using 1% discount for each half century earlier in the past, taking the 1451 – 1500 period as the undiscounted present. See Borcan *et al.* for details on method. Here, we provide a brief narrative account of our coding.

Arabia and Persian Gulf

Bahrain: Was central to the Dilmun civilization which E. Britannica says began in the late 4th millennium BCE. Although Dilmun may mostly have been more of a proto-state (z1 = 0.5), we choose, as also in the case of Qatar, to treat Dilmun as a fall state, although only beginning at a later time; and because the capital of Dilmun seems to have moved around, we treat it as being a local state for Bahrain as well as for Qatar, not a colonizer. We treat Bahrain as being under the Dilmun state from 2050 BCE until 1251 BCE, when it came under Assyrian control. Thereafter, it appears that Bahrain was part of many empires including the Achaemenid, Seleucid, Parthian, Sasanian, and the early Muslim caliphates, and its only period of true local rule was when it was the headquarters of the Qarmatians who were a religiously extreme mid-east political force based in the area; we estimate 950 – 1051

CE as the period of Qarmatian rule based in Bahrain and for a time vying with the Abassids. After that, like Qatar, Bahrain was under regional Muslim rulers based in mainland Arabia until the Portuguese arrived around 1500. So we code z1 = 0, thus no state until 2050 BCE, then z1 = 1, z2 = 1, z3 = 1 from 2050 – 1251 BCE, then z1 = 1, z2 = 0.5, z3 = 1 from 1250 BCE to 950 CE, then z1 = 1, z2 = 1, z3 = 1 from 951 to 1050 CE, then z1 = 1, z2 = 0.5, z3 = 1 from 1051 – 1500 CE.

Kuwait: was part of Dilmun together with Qatar and Bahrain from 2050 BCE until conquest by the Middle Assyrian kingdom around 1365, and thereafter was probably part of empires of Mesopotamia, Persia, Greece, etc. It is debatable whether to think of it as a colony since it can alternatively be viewed as simply having been a part of Mesopotamia, hence integral to the large Mesopotamian states. Its early association with Persian Gulf neighbors as part of Dilmun, and cultural distinction from early Sumer, lends some support to treating it as distinct. We assign to Kuwait z2 = 0.75 as a compromise between independent and colonized, whenever it is ruled by a Mesopotamian empire (such as Babylonian and Assyrian), and we assign to Kuwait z2 = 0.5 when it is part of an empire that includes Iraq itself being counted as colonized in Borcan *et al.* Hence: z1 = 0 until 2050 BCE. From 2050 to 1365 BCE, z1 = 1, z2 = 1, z3 = 1. From 1365 BCE to 1500 CE, z1 = 1, z3 = 1, and for those same years z2 = 0.75 whenever z2 = 1 or 0.75 for Iraq and z2 = 0.5 whenever z2 = 0.5 for Iraq, in Borcan *et al.*

Oman: Magan, ca. 2600 – 2001 BCE, was a proto-state located in present-day Oman that traded with Mesopotamian cities. For 2600 – 2001, we can set z1 = 0.5, z2 = 1, z3 = 0.3, using the low z3 (for under 10% of territory controlled) to capture that the polity was probably able to control some trade goods but that it is doubtful that it directly controlled or governed very much territory. This is followed by z1 = 0 until the Achaemenid empire followed by Seleucid, Parthian and Sasanian empires, which begin around 550 BCE. These empires claimed control of much of the coast, especially the northern coast, of present day Oman, including establishing some fortresses. So for 2000 BCE – 551 BCE, z1 = 0. From 550 BCE to 300 CE, z1 = 1, z2 = 0.5, and z3 = 0.75 (this assumes that only 25 – 50% of territory, i.e. mainly the coast, was controlled) due to this series of empires. Then during 301 – 650 CE, z1 = 0 again, since the Sasanians became inactive in the area. Then 651 – 750 CE, the Rashidun Caliphate attained control, so z1 = 1, z2 = 0.5, z3 = 1. Then 751 – 1154, the local Ibadi Imamate ruled the interior of Oman, while the Umayyad and Abassid caliphates and successor entities ruled the coast. Then from 1154 to 1500 there was a similar situation with the local Nabhani dynasty replacing the Ibadi one, so the values for 1154 – 1500 should be the same as the ones from 750 – 1154. Specifically, from 751 – 1500, we set z1 = 1, z2 = 0.75, z3 = 0.5.

Qatar: What is now Qatar was probably sparsely populated, and was visited, traded with, or directly exploited, by visitors from the Dilmun civilization based in Bahrain, then Assyria, then the Seleucid Empire. Each of the last two claimed territory on the Persian Gulf coast of Arabia that included present day Qatar, but there is little hard evidence of settlements. We consider a plausible compromise to be the tentative conclusion that there was not enough settled population in Qatar experiencing enough oversight by outside forces so that it is unreasonable to say that any of Qatar was under a state, until the Sasanian period in which there was presence of a Christian bishopric of the northeastern Arabian coast including Bahrain and Qatar called "Beth Qatraye". There was also an early established Muslim fort at Umm al-Ma'a (a name indicating presence of reliable potable water) on the northwest edge of the Qatar peninsula, though the exact dates of its construction appear not to be well established. Sasanian control over northeast Arabia is dated by Wikipedia to 224 AD, and a bishop's travel to the region is dated to 325. But the indications of definite settlement remain few for a long time, so we judge 400 CE to be a good rough approximation for arrival of state presence. We adopt z1 = 1, z2 = 0.5, z3 = 0.5 for the years 401 – 650 CE (661 is the founding year of the Umayyad Caliphate, but there was already early Islamic activity in the area before that), then from 651 to 1500 we raise z3 to 1 (so the three values are 1, 0.5, 1). That long period had rule by the Umayyad caliphate, the Abbasid caliphate, and some more regional groups, the Usfurids and Ormus.

United Arab Emirates: Bordering on Oman, with a similar history, Amm an-Nar Culture was closely linked to Oman's Magan culture or considered by some to be the same, a bronze age trading partner of Mesopotamian cities, and important during roughly 2600 – 2001 BCE. For those years we set z1 = 0.5, z2 = 1, z3 = 0.3, using the low z3 (representing that under 10% of the territory of today's country was controlled by this state) to capture the idea that it is doubtful that the polity directly controlled very much territory. Sources say there was continuity of this culture during 2000 – 1300 BCE, but there are no clear signs of large settlements or of a state, including lack of reports of trade with the region in Mesopotamia, so we can set z1 = 0 in these years. Then, unlike Oman, there are in this case excavations demonstrating that larger settlements and fortified towns existed during 1300 – 551 BCE, so we set z1 = 0.5, z2 = 1, z3 = 0.3 for that period. From 550 BCE to 650 CE, there was control of parts of the coast by the Achaemenid, Parthian and Sasanian empires, with a likely break during 330 – 247 between the Achaemenid and Parthian period and lack of evidence that the Seleucids then ruling the same Persian region were politically active in what is now UAE. So we set z1 = 1, z2 = 0.5, and z3 = 0.75 (the last value reflects an assumption that 25 – 50% of the territory was controlled by the sequential empires) from 550 – 330 BCE and again from 247 BCE – 650 CE due to this series of empires. For 330 BCE – 247 BCE, we set z1 = 0.5, z2 = 1, z3 = 0.5 based on an assumption that some local polities partly filled the gap left by Achaemenid dominance until the Parthian empire arrived. Then 651 CE – 900 CE, z1 = 1, z2 = 0.5 and z3 = 1 due to incorporation into the Rashidun, Umayyad and Abbasid Caliphates. Then in 900 – 1500 CE, a period of weak caliphate presence in southeastern Arabia, there were only small city-state-like powers in what is today UAE, the UAE coast was referred to at the time as "the pirate coast" due to lack of safety for foreign traders, so we set z1 = 0.5, z2 = 1, z3 = 0.3 for this 600 year period.

Americas

Puerto Rico: there was no state-like polity on the island prior to Spanish conquest, so we code no state up to 1500, i.e. z1 = 0 throughout the centuries considered.

Belize: what is today Belize was part of the same Mayan sphere as Guatemala, with which it shares a long land border reflecting the resolution of conflicting claims by Spain and England, not differences in indigenous societies. We thus adopt the identical state history as for Guatemala.

Asia

Bhutan based on The History of Bhutan, by Karma Phuntsho, Vintage Books, 2016, such local political authority as existed was concentrated in the hands of numerous monasteries and religious elite families, without a state to speak of, before the 1700s. The descriptions indicate that there existed local cultural identities distinct from those of Tibet, and some important leading personalities over the centuries, but that the area now called Bhutan interacted primarily with Tibetan religious and state authorities, as a kind of stateless or 'wild' southern periphery. We thus set z1 = 0 and thus Statehist as of 1500 is 0.

North Korea the state history index values used for S. Korea in Borcan *et al.* (2018), updated and extended from those in Bockstette *et al.* (2002), are for the Korean peninsula as a whole. The same values therefore apply to both S. Korea and N. Korea.

Europe

Luxembourg: could be said to have z1 = 0 until conquest by Rome (Julius Cesar) in 53 BCE.

Then z1 = 1, z2 = 0.5, z3 = 1 under first Rome to 406 CE, then under Merovingia, then the Carolingian Empire, then Middle Francia, then Lotharingia, then as part of Upper Lorraine within the Holy Roman Empire, then as the House of Luxembourg within Upper Lorraine, until it became an independent duchy during 1352 – 1482. It then came under Habsburg rule. So we set z1 = 0 until 54 BCE, then from 53 BCE until 1351 CE we set z2 = 0.5, we then set z2 = 1 from 1352 – 1482, and then set z2 = 0.5 again from 1482 – 1500. We set z3 = 1 whenever z1 = 1, since Luxembourg has a small territory that we assume to have been fully incorporated within each of the polities that ruled it.

Malta could be said to have z1 = 0 until conquest by Rome in 218 BCE. Malta stayed under Roman rule to 395 CE, when it was given to the control of Constantinople (the Eastern Roman Empire) until 870 CE, then came under Arab rule during 870 – 1090, then under Norman and Swabian rulers under the Kingdom of Sicily during 1091 – 1266, then Angevin (1266 -82) and Spain (1282 – 1530). So we can consider Malta's state history as beginning in 218 BCE, and its status is the same from 218 BCE until 1500 CE: z1 = 1, z2 = 0.5, z3 = 1 (the latter because the whole island was always within the same polity).

Sub-Saharan Africa

Guinea Bissau: according to Emilio Depetris-Chauvin's African state history data at country level (constructed for his unpublished doctoral thesis chapter titled "State History and Contemporary Conflict: Evidence from Sub-Saharan Africa," Brown University, 2015), covering 1000 – 1800 CE, Guinea Bissau had no territory ruled by states until his last two periods, denoted 1750 and 1800. However, Encyclopedia Britannica and Wikipedia say "much of Guinea Bissau." was under the Mali Empire in the 14th century, with a perhaps more local empire present in the 15th century. We settle on a state history value using positive codings for those two centuries, assuming control by the Mali Empire and perhaps another state external to today's Guinea Bissau. We code z1 = 0 until 1300 CE, then z1 = 1, z2 = 0.5 and z3 = 0.75 (25%-50% of country-of-today's territory was under a state) from 1301 - 1500.

Distance

The CERDI-SeaDistance database was used to construct the distance variable (Bertoli et al., 2016). It covers 227 countries and territories, including all those in our sample. The distance variable is defined as the length of the shortest existing sea route between the two relevant ports, measured in nautical miles. Because the best sea routes between many country pairs were shortened by the opening of the Suez Canal in 1869 and the Panama Canal in 1914 and we are studying population movements over the long 1500 - 1960 period, we needed to choose whether to treat the routes through these canals as being available, as does CERDI. We have adopted CERDI-SeaDistance numbers thereby assuming the routes to be open, because inter-regional migration was rising over time. Although the Panama Canal opened towards the end of some of the heaviest migration from Europe to the Americas, distance numbers which assume it to be available seem acceptable because a significant share of migration followed its opening and because for countries bordering the Pacific between Mexico and Colombia, overland passage from Atlantic ports allowed accessiblity from Europe and Africa prior to the canal's opening.

For coastal countries, CERDI identifies the relevant port by dividing the coastline into grid cells and selecting the coastal cell that contains the highest number of shipping links. For landlocked countries, the country is linked to the foreign port with the shortest land distance to its capital city. Three large countries, Canada, Russia, and the United States, are assigned two ports to account for their Atlantic and Pacific coasts, while this is not done for smaller countries bordering two oceans, such as Nicargua and Colombia. The key advantage of the CERDI-SeaDistance variable over the standard great-circle (geodesic) distance commonly used in gravity models is that it captures the actual navigable path ships must take, for instance the sea distance from London to Alexandria involves westward and southward legs before traversing the Mediterranean, whereas the great-circle distance is an arc in a strictly southeasterly direction. As another example, the great-circle distance from Bremen, an Atlantic-facing port in Germany, to Tokyo, is approximately 5 thousand nautical miles of great-circle stance, whereas the sea-distance, even assuming passage via the Suez Canal, is approximately 11 thousand nautical miles (around Europe's Atlantic coast, through the Mediterraean and Suez Canal, and through the Indian and western Pacific oceans).

The CERDI database also provides us with land distances from each capital to its relevant port, necessary for modeling full logistics cost when a country's economic and political centers of gravity are located inland. Specifically, for each country i, the variable d_i^{port} measures the shortest land distance from the contemporary country's capital city to the country's relevant port.[39] For landlocked countries, this land segment extends across borders to the nearest foreign port. Using these components, we construct the bilateral distance variable used in

[39] Admittedly, the capital city of today might not be the best possible representation of each country's economic or demographic center during the 1500 - 1960 era. For a rough indicator of bilateral distance on a global scale, we judge use of this available data series to be acceptable.

our analysis as follows:

$$distance_{ij} \; = \; d_{ij}^{\text{sea}} \; + \; d_i^{\text{port}} \; + \; d_j^{\text{port}},$$

where d_{ij}^{sea} is the shortest maritime distance between the relevant ports of countries i and j, and d_i^{port} and d_j^{port} are the respective land distances from each country's capital to its relevant port. This composite measure thus reflects the full origin-to-destination distance. In cases where a country is coastal, d_i^{port} captures a domestic leg only; where a country is landlocked, it additionally incorporates the cross-border land segment to the nearest accessible port.

For certain pairs of landlocked countries, the overland distance between their capitals is shorter than the combined land distances to their respective relevant ports—as is the case, for example, between Bangladesh and Mongolia. When the direct land distance between two capitals, d_{ij}^{land}, is shorter than the sum of their respective port distances, $d_i^{\text{port}} + d_j^{\text{port}}$, the database uses d_{ij}^{land} as the bilateral distance of the dyad:

$$distance_{ij} \; = \; d_{ij}^{\text{land}}$$

We acknowledge that use of current capital cities introduces arbitrariness insofar as some countries have capital cities that may poorly reflect their economic or population centers early in the 1500 - 1960 period. But the countries as such often did not exist, or existed with very different boundaries, in the early centuries of the era. In the majority of cases, bilateral distances measured by our source are likely to be a good approximation, since we are dealing with global distances.

Population 1500

Values of the Population in 1500 and in 1600 are missing in our main source, HYDE, for 2 countries included in the present paper.

Mauritius: The earliest HYDE-based estimate, available for 1700, places the population at approximately 401 individuals. McEvedy and Jones (1978) report that the island was nearly uninhabited at the time of Dutch settlement in 1638, noting only a few hundred inhabitants before the colony was evacuated in 1710. Subsequent HYDE figures for 1800 and 1850 are broadly consistent with the McEvedy and Jones (1978) series. And Maddison Population Database (Bolt and Van Zanden, 2025) offers no estimate before 1950. Given that all available sources point to the idea that the island was either uninhabited or negligibly populated before European contact, we set the population for both 1500 and 1600 at zero.

Ireland: For Ireland the HYDE dataset cannot be used for our estimates, as it covers the entire island (including Northern Ireland) and provides figures only from 1800 onward (HYDE, 2023). Instead, we rely on the Maddison database (Bolt and Van Zanden, 2025), which reports population figures for Ireland within its present-day borders, placing the population at approximately 800,000 in 1500, a figure which aligns with McEvedy and Jones (1978). For 1600, we consistently assign a population of 1 million from Bolt and Van Zanden (2025).

D Shapley Decomposition

Table A.6: Shapley Decomposition of Difference Variables - Global Sample

DS_{ij}	**Dominance Stat.**	**Exp. Var. Shr.**	**Rank**
Arable Land Area Difference	0.0915	0.2696	1
Disease Prevalence Difference	0.0498	0.1467	2
Destination Colony Dummy	0.0449	0.1324	3
Colonizer of Destination	0.0373	0.1098	4
Pop1500 / Arable Lands Difference	0.0369	0.1089	5
Destination Landlocked	0.0197	0.0579	6
Pre-1500 Avg. Max.-Cal./Ha Difference	0.0164	0.0483	7
Temperature Difference	0.0146	0.0430	8
Source Landlocked	0.0106	0.0313	9
Source Colony Dummy	0.0085	0.0251	10
Distance (Ln)	0.0038	0.0113	11
Statehist 1500 Difference	0.0037	0.0109	12
Same Colonizer	0.0015	0.0043	13
Destination is Colonizer	0.0001	0.0004	14

Notes: The dependent variable is DS_{ij}, and the sample includes only observations of pairs i and j located in different regions. General dominance statistics (i.e., Shapley values) are computed from PPML estimations over all possible subsets of the listed covariates. Distance (Ln), source landlocked, and destination landlocked are included in all subsets. As throughout the analysis, all decompositions use the same fixed-effects specification and standard-error structure as the main estimations. The fixed effects are included in every Shapley decomposition and are therefore partialled out rather than treated as separate explanatory components. Exp. Var. Shr. denotes the explained variance share and is calculated from each covariate's average contribution to the regression Pseudo R^2 across all subset models.

Table A.7: Shapley Decomposition of Source & Destination Variables - Global Sample

DS_{ij}	**Dominance Stat.**	**Exp. Var. Shr.**	**Rank**
Source-Side Covariates			
Colonizer of Destination	0.0206	0.0447	1
Source Disease Prevalence	0.0121	0.0263	2
Source Arable Land (Ln)	0.0047	0.0103	3
Source Temperature	0.0043	0.0094	4
Source Pop 1500 / Arable Lands (Ln)	0.0030	0.0066	5
Source Statehist 1500 (Ln)	0.0017	0.0038	6
Source Colony Dummy	0.0007	0.0016	7
Source Pre-1500 Avg. Max.-Cal./Ha	0.0002	0.0004	8
Same Colonizer	0.0001	0.0002	9
Destination Is Colonizer	0.0001	0.0000	10
Destination-Side Covariates			
Destination Arable Land (Ln)	0.1412	0.3004	1
Destination Statehist 1500 (Ln)	0.0527	0.1121	2
Destination Pop 1500 / Arable Lands (Ln)	0.0375	0.0797	3
Destination Temperature	0.0213	0.0454	4
Destination Pre-1500 Avg. Max.-Cal./Ha	0.0212	0.0450	5
Colonizer of Destination	0.0203	0.0433	6
Destination Disease Prevalence	0.0178	0.0378	7
Destination Colony Dummy	0.0138	0.0293	8
Same Colonizer	0.0031	0.0066	9
Destination Is Colonizer	0.0001	0.0002	10

Notes: The dependent variable is DS_{ij}, and the sample includes only observations of pairs i and j located in different regions. General dominance statistics (i.e., Shapley values) are computed from PPML estimations over all possible subsets of the listed covariates. Distance (Ln), source landlocked, and destination landlocked variables are included in all subsets. Exp. Var. Shr. denotes the explained variance share and is calculated from each covariate's average contribution to the regression Pseudo R^2 across all subset models. The two panels differ in fixed effects and clustering. The Destination-Side panel controls for origin fixed effects with clustering at the destination level, while the Source-Side panel controls for destination fixed effects with clustering at the source level. The fixed effects are included in every Shapley decomposition and are therefore partialled out rather than treated as separate explanatory components.

Table A.8: Shapley Decomposition for Stayer Share (w/in Region) Regressions (without Region FE)

Variable	**Dominance Stat.**	**Exp. Var. Shr.**	**Rank**
Colonizer	0.0827	0.2679	1
Disease Prevalence	0.0694	0.2247	2
Landlocked	0.0492	0.1593	3
Arable Land (Ln)	0.0360	0.1166	4
Temperature	0.0318	0.1031	5
Colony	0.0181	0.0586	6
Pre-1500 Avg. Max-Cal./Ha (Ln)	0.0116	0.0376	7
Pop 1500 / Arable Land (Ln)	0.0089	0.0288	8
Statehist 1500	0.0011	0.0035	9

Notes: The dependent variable is Stayer Share (w/in Region) .General dominance statistics (also called Shapley values) are computed from OLS estimations over all possible subsets of the listed covariates. Exp. Var. Shr. denotes the explained variance share and is calculated from each covariate's average contribution to the regression R^2 across all subset models. Robust standard errors are used in the underlying regressions. Region fixed effects are not included.

Table A.9: Shapley Decomposition for Stayer Share (w/in Region) Regressions (with Region FE)

Variable	**Dominance Stat.**	**Exp. Var. Shr.**	**Rank**
Landlocked	0.0870	0.1971	1
Disease Prevalence	0.0495	0.1123	2
Colonizer	0.0428	0.0971	3
Arable Land (Ln)	0.0313	0.0710	4
Colony	0.0135	0.0306	5
Temperature	0.0124	0.0282	6
Pop 1500 / Arable Land (Ln)	0.0110	0.0250	7
Pre-1500 Avg. Max-Cal./Ha	0.0052	0.0118	8
Statehist 1500 (Ln)	0.0005	0.0012	9

Notes: The dependent variable is Stayer Share (w/in Region) .General dominance statistics (also called Shapley values) are computed from OLS estimations over all possible subsets of the listed covariates. Exp. Var. Shr. denotes the explained variance share and is calculated from each covariate's average contribution to the regression R^2 across all subset models. Robust standard errors are used in the underlying regressions. Region fixed effects are included. The fixed effects are included in every Shapley decomposition and are therefore partialled out rather than treated as separate explanatory components.

E Colonization

Colonizers: Our colonizer dummy variables are constructed using the colonial linkage data from Head et al. (2010) and Conte et al. (2022), from which we identify six European powers as colonizers: Belgium, Spain, France, Great Britain, the Netherlands, and Portugal. The restriction to these six reflects the historical reality that these countries were responsible for the overwhelming majority of overseas territorial control and settlement during the 1500–1960 period. When the source data lists multiple colonizers for a territory, we keep only the primary one (ranked first in the original dataset) and treat that as the sole colonizer-colony link. As exceptions, Eritrea and Namibia are coded as colonized territories of Italy and Germany, respectively, even though Italy and Germany are not among the six colonizer countries listed above. For these two cases, the colonial relationship is defined at the specific dyad level using the *Colonizer of destination* indicator, rather than through the *Colonizer* country classification.

Colonized: Our colony dummy variables are constructed using the colonial linkage data from Head et al. (2010) and Conte et al. (2022). Colonized country list: Algeria, Angola, Argentina, Australia, Bahrain, Bangladesh, Belize, Benin, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Central African Republic, Chad, Chile, Colombia, Comoros, Congo Dem. Rep., Congo Rep., Costa Rica, Cote d'Ivoire, Cuba, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Fiji, Gabon, Gambia, Ghana, Guatemala, Guinea, Guinea-Bissau, Guyana, Honduras, India, Indonesia, Israel/Palestine, Jordan, Kenya, Kuwait, Laos, Lebanon, Lesotho, Liberia, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Mexico, Morocco, Mozambique, Myanmar, Namibia, New Zealand, Nicaragua, Niger, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Puerto Rico, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, Sri Lanka, Sudan, Syrian Arab Republic, Tanzania, Togo, Tunisia, Uganda, United Arab Emirates, United States, Uruguay, Venezuela, Vietnam, Yemen, Zambia, and Zimbabwe.

Colonization Decisions for Liberia, Eritrea, Ethiopia and Namibia: Liberia was not colonized by the European colonial powers primarily because a resettlement colony for African-born and African-descended freed American slaves was set up there under the auspices of the American Colonization Society beginning in the 1820s. Since the vast majority of freed slaves who volunteered to settle in Liberia were not descended from the indigenous ethnic groups of Liberia itself, since they established an American style, Christian, English-speaking settler society which treated the indigenous people as an inferior and politically impotent class for more than a century, and since Liberia was nevertheless the source of several times more enslaved, involuntary migrants to the Americas than the number of ex-slaves who entered it as settlers, we think that the territory which became Liberia resembles neighboring countries subjected to colonization during the European colonial era more than it does an indigenously ruled region such as Ethiopia that was largely spared the colonial experience. While it could justifiably be placed in a category of its own, the value of keeping to a dichotomous classification as either "colonized" or "not colonized" and the facts mentioned lead us to chose the "colonized" designation as the better of the two imperfect options for Liberia. Regarding Eritrea and Ethiopia, the former was colonized by Italy from 1890 to 1941 and administered by Britain from 1942 to 1952, whereas Ethiopia was only ruled by Italy during 1936 to 1941, a half decade. The duration of Eritrea's colonial status resembles that of African regions colonized by European countries during the "scramble for Africa," so we think it appropriate to treat it as having been colonized, whereas we agree with Ertan et al. (2016) that Ethiopia was under Italian rule for so short a period that it can be treated as "never colonized," a judgment adopted in other studies as well. Although Italy's colonization of Eritrea alone leads us to omit Italy from our general status *Colonizer*, assigned only to the six countries listed above, nonetheless in our bilateral dataset, the *Colonizer of Destination* dummy takes the value of one for the Italy–Eritrea dyad. Namibia is coded as colonized in our dataset, with the sole bilateral colonizer relationship assigned to Germany, reflecting the formal German South West Africa colonial period from 1884 until World War I. Accordingly, the *Colonizer of Destination* dummy takes the value of one for the Germany–Namibia dyad. Note that in other former German colonies in Africa that passed into British or French hands after World War I, we treat the successor power as the colonizer. We do not apply this rule to Namibia, because its post-war administrator was South Africa—itself a former colony—making a successor-colonizer assignment problematic. Like Italy in relation to Eritrea, we treat Germany as a colonizer in its specific bilateral relationship to Namibia but do not count it among the six countries given the general *Colonizer* designation.[40]

[40]The case of Papua New Guinea bears some resemblance to that of Namibia. Its northern half was ruled by Germany and its southern half by Great Britain until Germany's defeat in World War I, after which Britain claimed the entirety it what is today PNG (that is, the eastern half of the main and outlying islands, with the western half having been absorbed by Indonesia). Most administration was carried out by neighboring Australia from 1902 onwards, but Australia's own independence from Britain was only

F Distance to Colonizer

We construct a continuous proxy for cultural, linguistic, and geographic proximity that supplements the binary *Colonizer of Destination* indicator by capturing tendencies for people of countries neighboring a colonizer to also send migrants to that colonizer's overseas colonies. Let $\mathcal{C}(j)$ denote the colonizing power of destination j, identified from the primary colonizer field in our colonization dataset.[41] For each dyad (i, j) where j is a colonized destination, we define:

$$\text{Distance to Colonizer}_{ij}^{\text{std}} = \frac{d(i, \mathcal{C}(j))}{\max_k \; d(k, \mathcal{C}(j))} \tag{4}$$

where $d(i, \mathcal{C}(j))$ is the bilateral distance between source country i and the colonizer of destination j, and the maximum in the denominator is taken over all source countries k in the estimation sample.[42]

The standardization is performed separately for each colonizer c, rescaling the variable to the unit interval $[0, 1]$. The colonizer itself receives a value of 0, and the source country most distant from the colonizer receives a value of 1, with all other sources scaled linearly in between. For instance, among all source countries paired with Spanish colonies, Spain receives 0, Portugal receives a value close to 0 reflecting its geographic proximity to Spain, while New Zealand receives a value 1, being the farthest source country from Spain in our sample.

For the European source subsample regressions, we construct a separate version in which the maximum distance in the denominator is recalculated using only European source countries. This yields a wider spread of standardized values within the European subsample, as the most distant European country from each colonizer anchors the upper bound at 1, rather than a distant non-European country.

gradually formalized. Since Britain is judged a major colonizing power and Germany is not, since Australian autonomy was incomplete, and since we prefer to leave the long-time colony Australia, like S. Africa, out of our colonizing power category, we assign Britain as colonizing power. This matches the assignment by our main source for colony-related data, Head et al. Papua New Guinea became fully independent from Australia in 1975.

[41] We restrict attention to the six major overseas colonial powers: Belgium, Britain, France, the Netherlands, Portugal, and Spain. Germany and Italy are included only for their respective special-case colonies (Namibia and Eritrea). The variable is undefined for never-colonized destinations, reducing the estimation sample from 20,194 to approximately 12,826 observations for global analyses, and to 3,800 for the European source country sample.

[42] The bilateral distance measure is taken from the CERDI-seadistance database (Bertoli et al., 2016).

Table A.10: Global Sample with Difference Variables and Distance To Colonizer

Dependent Variable: DS_{ij}	(1) Global	(2) Global	(3) Global
Distance to Colonizer		−5.914***	−3.112
		(2.040)	(2.045)
Europe			3.734***
			(1.018)
Europe × Distance to Colonizer			−5.817*
			(3.496)
Arable Land Area Difference	0.569***	0.567***	0.553***
	(0.055)	(0.057)	(0.060)
Pop1500 / Arable Lands Difference	0.060	−0.004	0.078
	(0.091)	(0.101)	(0.098)
Statehist 1500 Difference	−0.392	0.093	−0.218
	(0.801)	(0.939)	(0.926)
Disease Prevalence Difference	−2.278***	−2.040***	−1.983***
	(0.209)	(0.216)	(0.268)
Colonizer of Destination	4.243***	2.989***	2.332***
	(0.529)	(0.858)	(0.731)
Source Colony Dummy	−1.330***	−0.981***	1.166*
	(0.509)	(0.358)	(0.616)
Same Colonizer	−0.340*	−0.052	−0.191
	(0.182)	(0.165)	(0.193)
Pre-1500 Avg. Max.-Cal./Ha Difference	0.344***	0.441***	0.505***
	(0.097)	(0.109)	(0.137)
Temperature Difference	0.062***	0.047**	0.030
	(0.019)	(0.023)	(0.029)
Distance (Ln)	−0.784***	−0.519*	−0.497
	(0.180)	(0.309)	(0.314)
Source Landlocked	−1.382***	−0.556	−0.593
	(0.347)	(0.441)	(0.412)
Destination Landlocked	−3.327***	−3.380***	−3.366***
	(0.537)	(0.525)	(0.526)
Constant	0.193	−0.987	−4.114*
	(1.972)	(2.525)	(2.403)
Observations	12,592	12,592	12,592
Pseudo R^2	0.334	0.367	0.378

Notes: All regressions are estimated using PPML. The continuous explanatory variables are calculated as the destination value minus the source value. The dependent variable is Descendant Share (DS_{ij}), based on Chanda *et al.* (2014) and World Bank (2024). Independent variables include *State History* (Borcan et al., 2018 and author's calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and author's calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and author's calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. Standard errors are two-way clustered at the destination and source levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.11: Global Sample with Source & Destination Variables and Distance To Colonizer

Dependent Variable: DS_{ij}	(1) Global	(2) Global	(3) Global	(4) Global	(5) Global
Distance to Colonizer	−1.832	−1.890**	−2.582**	−5.185	−2.009**
	(1.117)	(0.960)	(1.193)	(5.282)	(1.011)
Europe	2.594***	2.790***	2.092***		1.863**
	(0.728)	(0.695)	(0.709)		(0.808)
Europe × Distance to Colonizer	−5.462**	−6.065**	−4.126*	−1.119	−4.061
	(2.664)	(2.391)	(2.455)	(5.804)	(2.825)
Source Arable Land Area (Ln)	−0.233***	−0.263***	−0.167***		−0.161***
	(0.033)	(0.043)	(0.036)		(0.056)
Destination Arable Land Area (Ln)	0.955***	0.930***	1.003***	0.998***	
	(0.101)	(0.125)	(0.089)	(0.086)	
Source Disease Prevalence	1.612***	1.759***	1.537***		1.514***
	(0.428)	(0.466)	(0.396)		(0.254)
Destination Disease Prevalence	−1.348***	−1.918**	−1.978***	−1.978***	
	(0.319)	(0.760)	(0.375)	(0.347)	
Source Pop1500 / Arable Lands (ln)		0.005	0.173***		0.198***
		(0.060)	(0.058)		(0.073)
Destination Pop1500 / Arable Lands (ln)		0.131	0.416*	0.380**	
		(0.279)	(0.212)	(0.190)	
Source Statehist 1500 (Ln)	−0.053***		−0.089***		−0.081***
	(0.016)		(0.016)		(0.029)
Destination Statehist 1500 (Ln)	−0.198***		−0.227***	−0.213***	
	(0.053)		(0.048)	(0.038)	
Source Pre-1500 Avg. Max-Cal./Ha	−0.041	−0.010	−0.009		0.083
	(0.108)	(0.102)	(0.107)		(0.162)
Destination Pre-1500 Avg. Max-Cal./Ha	0.552***	0.852**	0.687**	0.567**	
	(0.118)	(0.351)	(0.281)	(0.252)	
Source Temperature	−0.097***	−0.091***	−0.097***		−0.114***
	(0.033)	(0.034)	(0.029)		(0.027)
Destination Temperature	0.074***	0.044	0.057	0.062	
	(0.028)	(0.054)	(0.048)	(0.043)	
Source Colony Dummy	1.692***	1.803***	1.526***		1.585***
	(0.480)	(0.467)	(0.498)		(0.609)
Destination Colony Dummy	1.892**	1.523*	2.108**	1.690*	
	(0.749)	(0.845)	(0.865)	(0.869)	
Colonizer of Destination	1.870***	2.040***	1.821***	2.952***	1.846***
	(0.502)	(0.694)	(0.532)	(0.378)	(0.536)
Same Colonizer	0.087	0.043	0.193	0.560***	0.363
	(0.315)	(0.300)	(0.303)	(0.212)	(0.269)
Distance (Ln)	−1.417***	−1.228***	−1.475***	−1.221***	−2.074***
	(0.391)	(0.363)	(0.428)	(0.402)	(0.323)
Source Landlocked	−1.036***	−0.989***	−0.999***		−0.997**
	(0.392)	(0.362)	(0.384)		(0.391)
Destination Landlocked	−2.978***	−2.780***	−2.862***	−2.802***	
	(0.609)	(0.537)	(0.634)	(0.587)	
Constant	−10.082***	−8.260*	−9.965***	−12.617***	18.383***
	(2.631)	(4.238)	(3.846)	(4.161)	(3.089)
Observations	12,592	12,592	12,592	12,592	12,592
Source Fixed Effects	NO	NO	NO	YES	NO
Destination Fixed Effects	NO	NO	NO	NO	YES
Pseudo R^2	0.449	0.424	0.456	0.517	0.492

Notes: All regressions are estimated using PPML. The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014) and World Bank (2024). Independent variables include *State History* (Borcan et al., 2018 and author's calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and author's calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and author's calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. For columns 1, 2 and 3, standard errors are clustered at the destination and source levels, for column 4 at destination level, and for column 5 at source level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.12: Europe Sample with Source & Destination Variables and Distance To Colonizer

Dependent Variable: DS_{ij}	(1) Europe	(2) Europe	(3) Europe	(4) Europe	(5) Europe
Distance to Colonizer (Europe)	-3.458^{***}	-3.294^{***}	-3.122^{***}	-2.697^{***}	-2.895^{***}
	(0.774)	(0.821)	(0.717)	(0.907)	(0.889)
Source Arable Land Area (Ln)	0.066	0.072	0.156^{**}		0.177^{*}
	(0.071)	(0.118)	(0.065)		(0.106)
Destination Arable Land Area (Ln)	1.004^{***}	1.017^{***}	0.989^{***}	1.162^{***}	
	(0.099)	(0.159)	(0.086)	(0.122)	
Source Disease Prevalence	1.538^{**}	2.104^{**}	1.830^{***}		1.706^{***}
	(0.746)	(1.032)	(0.675)		(0.439)
Destination Disease Prevalence	-0.380	-0.939	-1.002	-1.429^{***}	
	(0.626)	(1.166)	(0.762)	(0.341)	
Source Statehist 1500 (Ln)	0.578^{***}		-0.471^{**}		-0.433
	(0.155)		(0.238)		(0.615)
Destination Statehist 1500 (Ln)	-0.229^{***}		-0.248^{***}	-0.270^{***}	
	(0.066)		(0.053)	(0.041)	
Source Pre-1500 Avg. Max-Cal./Ha	-0.232	0.180	0.099		0.125
	(0.446)	(0.648)	(0.426)		(0.437)
Destination Pre-1500 Avg. Max-Cal./Ha	0.510^{***}	0.688^{**}	0.613^{**}	0.439^{*}	
	(0.124)	(0.341)	(0.243)	(0.248)	
Source Temperature	-0.153^{**}	-0.230^{*}	-0.172^{**}		-0.182^{**}
	(0.078)	(0.135)	(0.074)		(0.092)
Destination Temperature	-0.004	-0.011	-0.021	-0.019	
	(0.032)	(0.062)	(0.041)	(0.042)	
Destination Colony Dummy	1.191	0.634	1.583^{**}	1.021^{**}	
	(0.773)	(1.218)	(0.772)	(0.493)	
Colonizer of Destination	1.347^{*}	1.817^{*}	1.682^{**}	3.733^{***}	1.565^{***}
	(0.762)	(1.054)	(0.694)	(0.631)	(0.552)
Distance (Ln)	-0.013	-0.147	0.268	0.629	-0.503
	(0.442)	(0.481)	(0.475)	(0.521)	(0.616)
Source Landlocked	0.134	0.063	0.088		0.069
	(0.257)	(0.271)	(0.284)		(0.445)
Destination Landlocked	-2.440^{***}	-2.067^{***}	-2.426^{***}	-2.122^{***}	
	(0.621)	(0.511)	(0.612)	(0.498)	
Source Pop1500 / Arable Lands (ln)		0.573^{***}	0.786^{***}		0.786^{**}
		(0.094)	(0.143)		(0.336)
Destination Pop1500 / Arable Lands (ln)		-0.012	0.355	0.415^{**}	
		(0.359)	(0.260)	(0.175)	
Constant	-22.412^{***}	-18.742^{***}	-26.522^{***}	-32.416^{***}	1.939
	(4.850)	(6.392)	(5.679)	(5.709)	(6.032)
Observations	3,838	3,838	3,838	3,838	3,838
Source Fixed Effects	NO	NO	NO	YES	NO
Destination Fixed Effects	NO	NO	NO	NO	YES
Pseudo R^2	0.530	0.514	0.539	0.593	0.565

Notes: All regressions are estimated using PPML. The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014) and World Bank (2024). Independent variables include *State History* (Borcan et al., 2018 and author's calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and author's calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and author's calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. For columns 1, 2 and 3, standard errors are clustered at the destination and source levels, for column 4 at destination level, and for column 5 at source level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

G Detailed Tables of Stayer Share w/in Region:

Table A.13: Stayer shares by country, ordered from lowest to highest.

Country	Share	Country	Share	Country	Share
Equatorial Guinea	0.053	Croatia	0.854	Algeria	0.984
Gabon	0.088	Turkmenistan	0.856	Egypt, Arab Rep.	0.985
Ireland	0.117	Switzerland	0.858	Rwanda	0.985
Cuba	0.171	Uganda	0.859	Nepal	0.986
Portugal	0.172	Czech Republic	0.871	Philippines	0.989
Guyana	0.184	Syria	0.872	Yemen	0.989
Congo, Rep.	0.229	Nigeria	0.873	Sri Lanka	0.990
Guinea-Bissau	0.292	Macedonia	0.875	Latvia	0.991
Gambia	0.306	Slovenia	0.875	Somalia	0.991
Angola	0.310	Finland	0.876	Thailand	0.994
Spain	0.331	Bosnia	0.878	Indonesia	0.996
Liberia	0.398	Tanzania	0.882	Malawi	0.996
Senegal	0.409	Moldova	0.882	Malaysia	0.996
Norway	0.457	Russia	0.890	Brazil	0.997
Lebanon	0.459	Pakistan	0.891	Kenya	0.998
Mali	0.460	Honduras	0.899	Ethiopia	0.999
Ghana	0.468	Nicaragua	0.899	Argentina	1.000
United Kingdom	0.484	Mexico	0.903	Australia	1.000
Sierra Leone	0.499	Uzbekistan	0.906	Bahrain	1.000
Israel/Palestine	0.548	Guatemala	0.907	Belarus	1.000
Armenia	0.566	South Africa	0.910	Belize	1.000
Benin	0.592	Chad	0.916	Bhutan	1.000
Cameroon	0.597	Bangladesh	0.916	Bolivia	1.000
Italy	0.617	Zambia	0.921	Botswana	1.000
Togo	0.622	Oman	0.922	Chile	1.000
Germany	0.624	Laos	0.931	Comoros	1.000
C. African Rep.	0.641	Turkey	0.932	Costa Rica	1.000
Sweden	0.661	Ecuador	0.942	Cyprus	1.000
Sudan	0.693	Afghanistan	0.945	Eritrea	1.000
Netherlands	0.706	Iraq	0.947	Estonia	1.000
Guinea	0.723	South Korea	0.950	Fiji	1.000
Malta	0.733	Vietnam	0.951	Georgia	1.000
France	0.747	Bulgaria	0.952	Kazakhstan	1.000
Albania	0.755	Burundi	0.957	North Korea	1.000
Denmark	0.758	Burkina Faso	0.958	Kyrgyzstan	1.000
Poland	0.771	Kuwait	0.958	Lesotho	1.000
Cote d'Ivoire	0.782	Niger	0.961	Luxembourg	1.000
Mozambique	0.785	Belgium	0.963	Mongolia	1.000
Austria	0.785	Romania	0.965	Myanmar	1.000
Mauritania	0.795	Colombia	0.969	Namibia	1.000
Saudi Arabia	0.795	Cambodia	0.969	New Zealand	1.000
Slovakia	0.800	Jordan	0.971	Panama	1.000
El Salvador	0.800	Iran, Islamic Rep.	0.972	Papua New Guinea	1.000
Congo, Dem. Rep.	0.803	Morocco	0.975	Paraguay	1.000
Lithuania	0.806	Peru	0.976	Puerto Rico	1.000
Hungary	0.817	China	0.978	Tajikistan	1.000
Dominican Republic	0.818	India	0.979	UAE	1.000
Azerbaijan	0.821	Canada	0.980	United States	1.000
Greece	0.831	Libya	0.980	Uruguay	1.000
Ukraine	0.836	Tunisia	0.982	Venezuela	1.000
Madagascar	0.852	Japan	0.982	Zimbabwe	1.000

Notes: This table reports Stayer Share (within Region), prepared by authors, based on Chanda *et al.* (2014) and World Bank (2026), for all countries, ordered from lowest to highest. As discussed in Section 2, values of 1.000, signifying that the year 1500 population within the borders of the present-day country had almost no descendants outside the home region, are likely to reflect imprecision arising from the method of construction of the PW and CCP tables. The number of cases at the upper bound of Stayer Share's range is thus likely to be inaccurately high, though the measurement error is probably small in most, if not all, cases.

Table A.14: Stayer Share (within Region) without Region FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Colonizer	−0.330***	−0.372***	−0.346***	−0.334***	−0.343***	−0.334***	−0.366***	−0.353***	−0.346***	−0.355***
	(0.111)	(0.100)	(0.102)	(0.106)	(0.103)	(0.107)	(0.108)	(0.100)	(0.100)	(0.099)
Source Disease Prevalence		−0.073**	−0.067**	−0.065**	−0.069**	−0.065**	−0.080**	−0.199***	−0.202***	−0.198***
		(0.031)	(0.030)	(0.030)	(0.029)	(0.029)	(0.032)	(0.049)	(0.050)	(0.049)
Source Landlocked			0.088***	0.085***	0.088***	0.085***	0.088***	0.122***	0.124***	0.124***
			(0.024)	(0.023)	(0.024)	(0.023)	(0.024)	(0.030)	(0.031)	(0.031)
Source Pop1500 / Arable Lands (ln)				−0.013		−0.013	0.006	0.005	0.005	0.006
				(0.015)		(0.017)	(0.018)	(0.018)	(0.018)	(0.017)
Source Statehist 1500 (Ln)					−0.001	0.000	−0.003	−0.000	0.000	0.001
					(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Source Arable Land Area (Ln)							0.023**	0.035***	0.035***	0.034***
							(0.011)	(0.012)	(0.012)	(0.011)
Source Temperature								0.012***	0.011***	0.012***
								(0.003)	(0.003)	(0.003)
Source Colony Dummy									0.026	0.031
									(0.036)	(0.039)
Source Pre-1500 Avg. Max-Cal./Ha										0.023*
										(0.012)
Constant	0.899***	0.915***	0.892***	0.877***	0.886***	0.878***	0.559***	0.183	0.186	0.206
	(0.016)	(0.015)	(0.019)	(0.029)	(0.024)	(0.029)	(0.160)	(0.204)	(0.204)	(0.197)
Observations	153	153	153	153	153	153	153	153	153	153
R^2	0.094	0.140	0.172	0.178	0.173	0.178	0.209	0.293	0.295	0.305

Notes: All regressions are estimated using OLS. The dependent variable is stayer Share (w/in Region) prepared by authors, based on Chanda *et al.* (2014) and World Bank (2026). To represent the so-called Old World regions more clearly, we merged the Americas and Oceania+ into a single fixed effect category for this estimation. Robust (Huber–White) standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.15: Stayer Share (within Region) with Region FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Colonizer	-0.285^{**}	-0.272^{***}	-0.228^{**}	-0.215^{*}	-0.228^{**}	-0.215^{*}	-0.242^{**}	-0.262^{**}	-0.263^{**}	-0.265^{**}
	(0.117)	(0.102)	(0.107)	(0.112)	(0.107)	(0.112)	(0.111)	(0.112)	(0.112)	(0.110)
Disease Prevalence		-0.095^{***}	-0.080^{**}	-0.071^{**}	-0.083^{**}	-0.075^{**}	-0.098^{**}	-0.156^{***}	-0.153^{***}	-0.163^{***}
		(0.033)	(0.033)	(0.034)	(0.034)	(0.035)	(0.038)	(0.045)	(0.045)	(0.050)
Landlocked			0.148^{***}	0.147^{***}	0.148^{***}	0.147^{***}	0.152^{***}	0.169^{***}	0.166^{***}	0.159^{***}
			(0.035)	(0.034)	(0.035)	(0.035)	(0.034)	(0.037)	(0.037)	(0.036)
Pop1500 / Arable Lands (ln)				−0.016		−0.019	0.001	−0.006	−0.005	−0.003
				(0.015)		(0.015)	(0.015)	(0.017)	(0.017)	(0.017)
Statehist 1500 (Ln)					0.002	0.004	0.001	−0.001	−0.001	−0.001
					(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Arable Land Area (Ln)							0.023^{**}	0.031^{***}	0.030^{***}	0.031^{***}
							(0.010)	(0.011)	(0.011)	(0.011)
Temperature								0.011^{***}	0.011^{***}	0.011^{***}
								(0.004)	(0.004)	(0.004)
Colony Dummy									−0.033	−0.031
									(0.035)	(0.036)
Pre-1500 Avg. Max-Cal./Ha										0.017
										(0.018)
Central Asia	−0.009	-0.037^{**}	-0.170^{***}	-0.166^{***}	-0.185^{***}	-0.193^{***}	-0.199^{***}	−0.097	-0.126^{*}	−0.098
	(0.009)	(0.016)	(0.038)	(0.038)	(0.053)	(0.055)	(0.057)	(0.060)	(0.064)	(0.067)
East Asia	-0.017^{**}	−0.028	−0.046	−0.017	−0.062	−0.042	−0.078	0.076	0.044	0.047
	(0.008)	(0.029)	(0.044)	(0.051)	(0.057)	(0.060)	(0.057)	(0.078)	(0.084)	(0.084)
Europe	-0.146^{***}	-0.205^{***}	-0.232^{***}	-0.216^{***}	-0.248^{***}	-0.243^{***}	-0.250^{***}	-0.136^{**}	-0.166^{**}	-0.172^{**}
	(0.034)	(0.044)	(0.047)	(0.046)	(0.061)	(0.060)	(0.062)	(0.062)	(0.066)	(0.069)
Middle East and North Africa	-0.058^{**}	−0.045	−0.044	−0.022	−0.059	−0.047	−0.025	0.015	0.003	0.026
	(0.029)	(0.031)	(0.031)	(0.033)	(0.045)	(0.045)	(0.043)	(0.042)	(0.047)	(0.056)
South Asia	-0.007^{***}	0.013	−0.028	−0.013	−0.040	−0.031	−0.054	−0.029	−0.039	−0.034
	(0.002)	(0.019)	(0.040)	(0.044)	(0.046)	(0.048)	(0.045)	(0.043)	(0.044)	(0.043)
Southeast Asia	-0.020^{**}	0.011	−0.001	0.003	−0.013	−0.017	−0.023	−0.048	−0.052	−0.056
	(0.009)	(0.018)	(0.031)	(0.031)	(0.039)	(0.039)	(0.038)	(0.039)	(0.040)	(0.041)
Sub-Saharan Africa	-0.206^{***}	-0.143^{***}	-0.192^{***}	-0.192^{***}	-0.194^{***}	-0.196^{***}	-0.183^{***}	-0.184^{***}	-0.188^{***}	-0.160^{***}
	(0.044)	(0.040)	(0.045)	(0.046)	(0.047)	(0.048)	(0.045)	(0.044)	(0.043)	(0.052)
Constant	1.000^{***}	1.013^{***}	1.000^{***}	0.972^{***}	1.019^{***}	1.000^{***}	0.681^{***}	0.324	0.361^{*}	0.353^{*}
	(0.000)	(0.010)	(0.011)	(0.029)	(0.040)	(0.044)	(0.139)	(0.201)	(0.206)	(0.209)
Observations	153	153	153	153	153	153	153	153	153	153
R^2	0.249	0.285	0.356	0.364	0.357	0.367	0.394	0.431	0.432	0.435

Notes: All regressions are estimated using OLS. The dependent variable is stayer Share (w/in Region) prepared by authors, based on Chanda *et al.* (2014) and World Bank (2026). To represent the so-called Old World regions more clearly, we merged the Americas and Oceania+ into a single fixed effect category for this estimation. Robust (Huber–White) standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

H Replications of Tables 4, 5 and 6 - comparing results with & without State History 1500

Table A.16: Global Sample with Source & Destination Variables (Replication of Table 4, Omitting statehist1500)

Dependent Variable: DS_{ij}	(1) Global	(2) Global	(3) Global
Source Arable Land Area (Ln)	-0.313***		-0.239***
	(0.037)		(0.056)
Destination Arable Land Area (Ln)	0.906***	0.931***	
	(0.135)	(0.126)	
Source Disease Prevalence	1.723***		1.483***
	(0.530)		(0.238)
Destination Disease Prevalence	-2.016***	-2.045***	
	(0.746)	(0.560)	
Source Pop1500 / Arable Lands (ln)	-0.058		0.074
	(0.054)		(0.071)
Destination Pop1500 / Arable Lands (ln)	0.011	0.015	
	(0.234)	(0.213)	
Colonizer of Destination	3.507***	3.909***	2.901***
	(0.763)	(0.470)	(0.419)
Destination Colony Dummy	2.889**	2.177*	
	(1.274)	(1.118)	
Source Colony Dummy	0.012		0.427
	(0.207)		(0.365)
Same Colonizer	-0.082	0.597**	0.171
	(0.340)	(0.254)	(0.253)
Destination is Colonizer	-0.305	-0.351	0.303
	(0.926)	(0.973)	(0.536)
Source Pre-1500 Avg. Max-Cal./Ha	0.015		0.073
	(0.103)		(0.156)
Destination Pre-1500 Avg. Max-Cal./Ha	0.855***	0.728***	
	(0.277)	(0.251)	
Source Temperature	-0.103***		-0.111***
	(0.022)		(0.024)
Destination Temperature	0.066	0.081*	
	(0.050)	(0.044)	
Distance (Ln)	-1.406***	-1.280***	-2.731***
	(0.406)	(0.341)	(0.328)
Source Landlocked	-1.623***		-1.240***
	(0.344)		(0.352)
Destination Landlocked	-2.029***	-1.981***	
	(0.641)	(0.592)	
Constant	-6.077**	-12.236***	26.213***
	(2.844)	(2.989)	(3.024)
Observations	20,194	20,194	20,194
Source Fixed Effects	NO	YES	NO
Destination Fixed Effects	NO	NO	YES
Pseudo R^2	0.395	0.477	0.489

Notes: All regressions are estimated using PPML. The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014) and World Bank (2024). The same set of independent variables is used, see Section 2 and Appendix B for further variable details. In every estimation, sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows.For Column 1, standard errors are two-way clustered by source and destination country. In Column 2, they are likewise clustered by source and destination, whereas in Column 3 they are clustered by source country. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.17: Regional Sample with Source & Destination variables (Replication of Table 5, Omitting statehist1500)

Dependent Variable: DS_{ij}	Europe			Sub-Saharan Africa			N. Africa & Asia		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Source Arable Land Area (Ln)	0.072		0.171*	−0.389***		−0.297***	−0.275*		−0.230*
	(0.133)		(0.098)	(0.038)		(0.046)	(0.165)		(0.127)
Destination Arable Land Area (Ln)	1.095***	1.223***		1.017***	1.099***		0.763***	0.709***	
	(0.168)	(0.112)		(0.155)	(0.154)		(0.136)	(0.133)	
Source Pop1500 / Arable Lands (ln)	0.582***		0.605***	0.007		0.063	0.125		0.191
	(0.105)		(0.161)	(0.052)		(0.048)	(0.299)		(0.211)
Destination Pop1500 / Arable Lands (ln)	0.002	0.035		0.196	0.238		−0.283*	−0.275	
	(0.395)	(0.306)		(0.264)	(0.299)		(0.161)	(0.237)	
Source Disease Prevalence	1.252		1.188***	3.454***		2.762***	0.432		0.554
	(0.955)		(0.385)	(0.830)		(0.397)	(0.628)		(0.489)
Destination Disease Prevalence	−1.081	−1.901***		−2.592***	−2.857***		−1.327*	−1.006	
	(1.094)	(0.643)		(0.628)	(0.601)		(0.779)	(0.711)	
Colonizer of Destination	2.420*	4.795***	2.046***						
	(1.400)	(0.409)	(0.632)						
Destination Colony Dummy	2.070**	1.177*		2.617	2.722		0.819	0.321	
	(1.041)	(0.663)		(1.757)	(1.943)		(0.879)	(0.825)	
Source Colony Dummy				4.016***		4.063***	4.383***		4.079***
				(1.451)		(1.285)	(1.240)		(1.230)
Same Colonizer				0.322	0.456**	0.534***	−0.528	−0.069	−0.179
				(0.238)	(0.207)	(0.196)	(0.871)	(0.775)	(0.584)
Destination is Colonizer				−1.032	−1.269	0.172	1.688	1.997***	1.296
				(1.279)	(1.281)	(0.557)	(1.077)	(0.739)	(1.112)
Source Pre-1500 Avg. Max-Cal./Ha	−0.480		−0.233	0.434*		0.290	0.993***		1.040***
	(0.350)		(0.411)	(0.262)		(0.241)	(0.232)		(0.256)
Destination Pre-1500 Avg. Max-Cal./Ha	0.809*	0.584		1.012***	0.988***		−0.181	−0.175	
	(0.443)	(0.442)		(0.233)	(0.243)		(0.235)	(0.160)	
Source Temperature	−0.102		−0.134**	−0.185		−0.281***	−0.272***		−0.262***
	(0.068)		(0.067)	(0.117)		(0.074)	(0.080)		(0.054)
Destination Temperature	0.002	0.027		0.122*	0.141**		0.089*	0.066	
	(0.073)	(0.063)		(0.065)	(0.066)		(0.052)	(0.044)	
Distance (Ln)	−0.994*	−0.677	−2.683***	−2.606***	−3.101***	−3.278***	1.113	1.797*	0.235
	(0.536)	(0.526)	(0.794)	(0.485)	(0.637)	(0.574)	(1.297)	(0.985)	(0.907)
Source Landlocked	−0.272		−0.197	−5.016***		−4.445***	−1.491*		−1.161
	(0.229)		(0.399)	(1.080)		(0.732)	(0.880)		(1.074)
Destination Landlocked	−0.853	−0.864	0.000	−4.793***	−4.770***		−2.120***	−2.100***	
	(0.640)	(0.611)	(.)	(0.938)	(1.000)		(0.549)	(0.422)	
Constant	−16.316***	−21.831***	21.308***	1.008	1.215	31.768***	−27.569**	−36.205***	−2.943
	(5.103)	(6.008)	(7.411)	(6.544)	(5.603)	(6.555)	(11.147)	(7.554)	(8.810)
Observations	4,712	4,712	4,712	4,914	4,914	4,914	5,130	5,130	5,130
Source Fixed Effects	NO	YES	NO	NO	YES	NO	NO	YES	NO
Destination Fixed Effects	NO	NO	YES	NO	NO	YES	NO	NO	YES
Pseudo R^2	0.493	0.567	0.554	0.401	0.419	0.518	0.380	0.425	0.444

Notes: All regressions are estimated using PPML.The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014), World Bank (2026), and author's calculations. Independent variables include *State History* (Borcan et al., 2018 and author's calculations); *Disease Prevalence* (Murray and Schaller, 2010); *Pre-1500 Avg. Maximum Calories per Ha.* (Galor and Özak, 2016); *Temperature* (Ashraf and Galor, 2013); *Arable Land* (World Bank, 2026); *Population per Arable Ha., 1500* (Klein Goldewijk et al., 2017; World Bank, 2026, and authors' calculations); *Colonial* and *Landlocked* variables (Head et al., 2010; Conte et al., 2022); and *Distance* (Bertoli et al., 2016 and author's calculations). See Section 2 and Appendix B for further details. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. For columns 1, 4, and 7, standard errors are clustered at the source and destination level, for columns 2, 5, and 8, standard errors are clustered at the destination level; for columns 3, 6, and 9, they are clustered at the source level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.18: Estimations with Pop. Growth Rate of 1500 - 1600 & Pop. 1600 (Replication of Table 6, adding statehist1500)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Source Pop1500 / Arable Lands (ln)	0.103		1.033***		−0.017		−0.078	
	(0.074)		(0.179)		(0.089)		(0.306)	
Destination Pop1500 / Arable Lands (ln)	−0.238		−0.288		−0.166		−0.569**	
	(0.150)		(0.237)		(0.283)		(0.226)	
Source Pop. Growth Rate 1600-1500	0.367		1.843***		−2.742**		0.560	
	(0.257)		(0.509)		(1.203)		(0.938)	
Destination Pop. Growth Rate 1500 - 1600	−2.418***		−2.040***		−2.361***		−0.727	
	(0.352)		(0.262)		(0.683)		(0.480)	
Source Pop1600 / Arable Lands (Ln)		0.096		1.080***		0.003		−0.017
		(0.074)		(0.149)		(0.096)		(0.246)
Destination Pop1600 / Arable Lands (Ln)		−0.378**		−0.484**		−0.021		−0.444**
		(0.170)		(0.208)		(0.328)		(0.188)
Source Statehist 1500 (Ln)	−0.066***	−0.064***	−0.683**	−0.935***	0.000	0.009	0.954	0.813
	(0.020)	(0.021)	(0.328)	(0.328)	(0.034)	(0.037)	(1.091)	(1.080)
Destination Statehist 1500 (Ln)	−0.124***	−0.197**	−0.107***	−0.148***	−0.104	−0.271**	0.193***	0.164***
	(0.048)	(0.081)	(0.037)	(0.028)	(0.084)	(0.138)	(0.067)	(0.059)
Source Arable Land Area (Ln)	−0.212***	−0.227***	0.125	0.156**	−0.269***	−0.344***	−0.349*	−0.328*
	(0.043)	(0.042)	(0.086)	(0.074)	(0.095)	(0.098)	(0.201)	(0.180)
Destination Arable Land Area (Ln)	0.665***	0.839***	0.778***	0.984***	0.643***	0.905***	0.690***	0.738***
	(0.088)	(0.097)	(0.145)	(0.116)	(0.133)	(0.180)	(0.110)	(0.120)
Source Disease Prevalence	1.406***	1.446***	0.928*	1.114**	2.768***	2.891***	0.582	0.254
	(0.353)	(0.381)	(0.503)	(0.563)	(0.406)	(0.506)	(1.163)	(0.698)
Destination Disease Prevalence	−1.625***	−1.312***	−0.642	−0.491	−1.986***	−1.692***	−1.900***	−2.232***
	(0.364)	(0.309)	(0.707)	(0.568)	(0.602)	(0.506)	(0.536)	(0.538)
Colonizer of Destination	2.891***	2.999***	2.420***	2.377**				
	(0.510)	(0.589)	(0.812)	(0.929)				
Destination Colony Dummy	1.127*	0.605	0.695	0.301	0.708	0.071	1.330	1.481
	(0.614)	(0.939)	(0.823)	(0.711)	(0.529)	(0.948)	(0.978)	(0.947)
Source Colony Dummy	0.087	0.016			4.110***	3.903***	4.179***	4.288***
	(0.238)	(0.222)			(1.311)	(1.378)	(1.326)	(1.165)
Same Colonizer	0.021	0.056			0.405*	0.484**	−0.364	−0.358
	(0.367)	(0.367)			(0.239)	(0.218)	(0.863)	(0.845)
Destination is Colonizer	0.625	0.514			0.266	−0.317	1.671	1.468
	(0.681)	(0.827)			(0.866)	(1.057)	(1.026)	(1.043)
Source Pre-1500 Avg. Max-Cal./Ha	0.018	−0.007	−0.065	−0.156	0.302*	0.340*	1.190***	1.189***
	(0.095)	(0.099)	(0.288)	(0.293)	(0.168)	(0.180)	(0.205)	(0.188)
Destination Pre-1500 Avg. Max-Cal./Ha	0.379**	0.495***	0.284	0.382**	0.427*	0.488**	−0.294	−0.294
	(0.154)	(0.101)	(0.273)	(0.151)	(0.232)	(0.207)	(0.198)	(0.239)
Source Temperature	−0.099***	−0.101***	−0.090	−0.074	−0.202*	−0.193*	−0.265***	−0.264***
	(0.018)	(0.018)	(0.069)	(0.062)	(0.107)	(0.104)	(0.073)	(0.073)
Destination Temperature	0.088***	0.099***	0.023	0.048	0.109**	0.121***	0.117***	0.125***
	(0.030)	(0.029)	(0.045)	(0.029)	(0.044)	(0.040)	(0.043)	(0.047)
Distance (Ln)	−1.952***	−1.836***	−1.055	−1.205**	−2.418***	−2.263***	1.017	0.945
	(0.434)	(0.467)	(0.657)	(0.593)	(0.857)	(0.722)	(1.134)	(1.208)
Source Landlocked	−1.431***	−1.500***	−0.409	−0.314	−4.564***	−4.682***	−0.782	−0.912
	(0.349)	(0.362)	(0.284)	(0.248)	(0.890)	(0.939)	(1.044)	(1.182)
Destination Landlocked	−1.546***	−1.923***	−0.827	−0.945*	−4.032***	−4.717***	−2.551***	−2.339***
	(0.525)	(0.616)	(0.520)	(0.498)	(0.915)	(0.991)	(0.702)	(0.628)
Constant	0.496	−3.511	−13.926**	−17.360***	4.608	−0.311	−24.103**	−24.647**
	(3.246)	(3.002)	(6.862)	(5.288)	(8.850)	(7.111)	(10.697)	(10.936)
Observations	20,194	20,194	4,712	4,712	4,914	4,914	5,130	5,130
Pseudo R^2	0.456	0.437	0.536	0.530	0.468	0.445	0.394	0.391

Notes: All regressions are estimated using PPML. The dependent variable is Descendant Share (DS_{ij}), based on Chanda et al. (2014), World Bank (2026), and author's calculations. The same set of independent variables is used, see Section 2 and Appendix B for further variable details. Population of 1600 is from HYDE 2023 accessed in Our World in Data (2026). Population growth rates for the 1500 - 1600 period are calculated based on the same source. In every estimation, the sample excludes intra-regional migration pairs (e.g., Europe-to-Europe) to focus on inter-regional flows. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

I Population and Population Density for Super Regions and Countries in The Americas

Table A.19: Population per Arable Land by Super Regions

Region	Pop/Arable 1500	Pop/Arable 1600	Pop 1500	Pop 1600	Arable Land
Europe	0.275	0.354	78,321,818	100,900,000	284,900,000
N. Africa & Asia	0.598	0.661	311,100,000	343,900,000	520,000,000
Sub-Saharan Africa	0.274	0.313	50,288,809	57,548,415	183,600,000
Greater New World	0.167	0.031	61,682,841	11,689,502	368,877,000
Americas Only	0.174	0.029	60,162,380	10,054,217	344,800,000

Notes: Population per hectare of arable land is calculated as historical population in 1500 and 1600 from the 2023 update of the HYDE database (HYDE, 2023; Klein Goldewijk et al., 2017), accessed via Our World in Data (2026), divided by arable land area in hectares from the World Bank (World Bank, 2026). "Greater New World" includes the Americas together with associated New World territories, while "Americas Only" excludes Papua New Guinea, Australia and New Zealand.

Table A.20: Population per Arable Land in The Americas

Country	Pop/Arable 1500	Pop/Arable 1600	Pop 1500	Pop 1600	Arable Land
Argentina	0.011	0.011	300,000	300,000	27,640,000
Belize	0.108	0.117	6,885	7,511	64,000
Bolivia	0.286	0.254	900,000	800,000	3,144,000
Brazil	0.412	0.020	18,738,040	887,751	45,448,000
Canada	0.005	0.005	200,000	200,000	40,971,000
Chile	0.343	0.343	600,000	600,000	1,750,000
Colombia	1.418	0.447	3,994,975	1,258,417	2,818,000
Costa Rica	2.143	0.038	450,000	8,000	210,000
Cuba	0.018	0.023	63,845	79,806	3,504,000
Dominican Republic	0.224	0.109	200,000	97,966	894,750
Ecuador	0.371	0.309	600,000	500,000	1,616,000
El Salvador	0.146	0.160	91,746	100,087	627,100
Fiji	0.440	0.488	50,898	56,554	115,800
Guatemala	0.473	0.355	800,000	600,000	1,690,720
Guyana	0.945	0.945	90,000	90,000	95,276
Haiti	0.222	0.044	200,000	39,747	900,000
Honduras	1.022	0.292	1,091,464	311,847	1,068,000
Jamaica	2.046	1.023	286,449	143,224	140,000
Mexico	1.068	0.044	24,460,394	1,000,000	22,905,000
Nicaragua	0.170	0.085	325,097	162,548	1,917,000
Panama	0.359	0.180	196,849	98,424	548,000
Paraguay	0.066	0.066	200,000	200,000	3,020,000
Peru	0.960	0.312	4,000,000	1,300,000	4,168,576
Puerto Rico	0.077	0.116	4,600	6,900	59,700
Trinidad and Tobago	0.003	0.007	115	231	35,000
United States	0.011	0.004	1,887,028	778,503	175,500,000
Uruguay	0.017	0.019	23,995	26,701	1,373,000
Venezuela	0.154	0.154	400,000	400,000	2,595,000

Notes: Population per hectare of arable land is calculated using historical population in 1500 and 1600 from the 2023 update of the HYDE database (HYDE, 2023; Klein Goldewijk et al., 2017), accessed via Our World in Data (2026), divided by arable land area in hectares from the World Bank (World Bank, 2026).