

The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade

March 2016

Vernon Henderson, London School of Economics

Tim Squires, Amazon.com

Adam Storeygard, Tufts University

David Weil, Brown University and NBER¹

Abstract

We study the distribution of economic activity, as proxied by lights at night, across 250,000 grid cells of average area 560 square kilometers. We first document that nearly half of the variation can be explained by a parsimonious set of physical geography attributes. A full set of country indicators only explains a further 10%. When we divide geographic characteristics into two groups, those primarily important for agriculture and those primarily important for trade, we find that the agriculture variables have relatively more explanatory power in countries that developed early and the trade variables have relatively more in countries that developed late, despite the fact that the latter group of countries are far more dependent on agriculture today. We explain this apparent puzzle in a model in which two technological shocks occur, one increasing agricultural productivity and the other decreasing transportation costs, and in which agglomeration economies lead to persistence in urban locations. In countries that developed early, structural transformation due to rising agricultural productivity began at a time when transport costs were still relatively high, so urban agglomerations were localized in agricultural regions. When transport costs fell, these local agglomerations persisted. In late-developing countries, transport costs fell well before structural transformation. To exploit urban scale economies, manufacturing agglomerated in relatively few, often coastal, locations. With structural transformation, these initial coastal locations grew, without formation of more cities in the agricultural interior.

Key words: Agriculture, physical geography, development

JEL Codes: O13, O18, R12

¹ We thank Alex Drechsler, Joshua Herman, Young Min Kim, Kevin Proulx, Sameer Sarkar, Yang Shen, and Sanjay Singh for excellent research assistance, and seminar participants at Berkeley, George Mason, LSE, Williams, the Federal Reserve Bank of Philadelphia/NBER Conference on Macroeconomics Across Time and Space and the Brown University conference on Deep-Rooted Factors in Comparative Economic Development for helpful comments and suggestions. Storeygard thanks Deborah Balk, Marc Levy, Glenn Deane and colleagues at CIESIN for conversations on related work in 2004-2006, and LSE and UC-Berkeley for hospitality while this research was conducted. The authors acknowledge the support of the World Bank's Knowledge for Change Program and a Global Research Program on Spatial Development of Cities funded by the *Multi Donor Trust Fund on Sustainable Urbanization* of the World Bank and supported by the UK Department for International Development.

1. Introduction

The population of the world is distributed very unevenly. In some places, one can go for miles without seeing another living soul. In others, it is hard take a couple of steps without running into someone. The most obvious determinant of the spatial distribution of population is the physical setting, sometimes called “first nature.”² Some places are simply more amenable to human habitation and output production than others. A substantial literature describes how first nature characteristics affect location of economic activity.³

Natural characteristics are for the most part fixed over time.⁴ The effect that these characteristics have on the concentration of population may change, however. Our first interest is precisely in studying such changes. In some cases, the source of the change is obvious: Irrigation, air conditioning, and the discovery of new uses for subsoil minerals have all altered the values attached to particular aspects of first nature. In this paper, we take a systematic approach to analyzing changes in the effects of specific natural characteristics, focusing on what we believe to be the two areas in which the importance of first-nature characteristics has changed the most. These are, first, the suitability of a region for growing food, and, second, the suitability of a region for engaging in national and international trade. We show evidence that the weight attached to the first of these characteristics has declined over time, while the weight associated with the second has risen.

Over the past several centuries (the period of time in which most of the agglomeration in the

² As far as we are aware, this terminology was introduced to the economics literature by Krugman (1991), based on Cronon (1991).

³ Examples of this approach include Nordhaus (2006) and Nordhaus and Chen (2009), who look at the effect of a whole suite of geographic factors using coarse subnational data; Mellinger, Gallup, and Sachs (2000) and Rappaport and Sachs (2003), who investigate the role of coasts, for both productive and amenity reasons; and Nunn and Puga (2012), who look at the effect of terrain ruggedness. See also Gennaioli et al. (2013, 2014), who regress subnational income and growth on geographic factors along with institutions, population and human capital measures, for a sample that covers much of the world but largely excludes Africa. Related work in the trade literature (e.g. Allen and Arkolakis 2014) have used a more structural approach and focused on the United States, where data on subnational trade flows are available.

⁴ Exceptions include the drying of the Fertile Crescent, Iceland’s loss of its topsoil, and of course current global climate change.

world has taken place), the link from ease of food production to concentration of population has attenuated both because an increase in agricultural productivity has ensured that food represents a much smaller fraction of the consumption basket today than in the past, and because costs of transporting food have fallen dramatically. Thus on both the production and consumption sides, there is less need for most of the population to live near where food is produced. In a similar vein, access to international trade, via location on coasts, navigable rivers, or near natural harbors, has become more valuable as opportunities to reap gains from trade increased over the last several centuries.⁵

To assess a claim that a determinant of population concentration has become more or less important over time, it would at first appear that one would need to have data on population concentration itself over time. The rich data on the spatial distribution of lights that we harness in this paper has the downside of not being available over long periods. We show that using contemporary data we can nonetheless investigate how weights of different geographic characteristics have changed by taking advantage of a commonly observed phenomenon in the study of spatial economics: persistence. Agglomeration, balanced by congestion, the other great force determining population distribution beyond first nature, begets persistence. Large population agglomerations, once established, tend to stay put. This persistence, and our understanding of when agglomeration took place (specifically the fact that we can sort countries by how urbanized they were at a particular point in time) enables us to understand how the weights associated with natural characteristics have changed. Several economic studies have examined such persistence in more localized settings (i.e. specific regions, or in response to particular shocks).⁶ Our paper is the first to examine (and take advantage of) such persistence at

⁵ The historical changes in agricultural productivity and transport costs on which we focus are hardly the only ways in which technological change and economic development have impacted the spatial pattern of population. To mention two others: first, income growth has shifted the relative importance of natural characteristics associated with productivity and those associated with amenity. Second, the costs and benefits of agglomeration have also changed over time, for example due to improved medical and public-health technologies (which lowered the costs) and the use of more complex production processes (which raised the benefits).

⁶ Examples include Davis and Weinstein (2002), Bleakley and Lin (2012), and Michaels and Rauch (2013).

a global scale.⁷

Persistence leads to outcomes which might seem puzzling. Today, in developed countries that experienced the start of structural transformation releasing labor from agriculture two centuries ago, the location of economic activity is driven much more by factors determining agricultural productivity than in developing countries, where structural transformation is a post-1950's phenomenon and in much of Asia and Africa is still on-going. A primary goal of the paper is to explain this puzzle.

Although our primary interest is in examining the interaction of nature with history, we begin by examining the overall predictive power of first-nature characteristics for the distribution of economic activity in modern cross-sectional data. Our primary dependent variable is light at night, as observed from satellites, aggregated to roughly 250,000 quarter degree (longitude/latitude) grid squares, because unlike all global population datasets, it is sampled uniformly (Henderson, Storeygard and Weil 2012). Our measures of first nature include characteristics of the climate, land surface, natural water bodies, and plant life (temperature, precipitation, elevation and ruggedness, coasts, navigable rivers, natural ports, and biomes). We are particularly interested in the relative importance of characteristics related to the possibility of trade (such as being located near a natural harbor) versus those associated with agricultural productivity.

A significant advance we make over much of the current literature is that we focus on the distribution of activity *within* countries, in addition to between them. The most important reason for doing this is that our measure of density, lights observed from outer space, is a function of both population density and income per capita. Looking within countries, the variance of income

⁷ Mesbah et al. (2014) estimate the year in which a given half-degree grid cell passed various urbanization rate thresholds. Their urban and rural population data are gridded estimates for the past 2,000 years from Klein Goldenwijk et al. (2011). Mesbah et al. regress the date of urbanization on a cultivation suitability index, distance to coast, a river navigability proxy, frost, and elevation, finding significant predictive power for all of these variables except elevation. We view their work as complementary to ours, in that they examine the determinants of early urbanization and we examine the effect of early urbanization, along with other factors, on outcomes today.

per capita is greatly reduced. Additionally, institutions (for which countries are a convenient proxy) clearly matter for population density. While geographic factors may well play a significant role in shaping institutions, sorting out the effect of institutions versus geography in cross-country data is extremely difficult, if not impossible. Our approach of including country fixed effects throws away some geographic variation, but we show that there is still a very large amount of usable variation left.

Following this preliminary analysis, we present a simple theoretical model to capture the intuition of how, in the presence of geographical persistence, historical changes in the economic value of natural characteristics can be inferred from the modern mapping from characteristics to density. The model incorporates the key changes mentioned above: rising agricultural productivity lowers the fraction of the population who work producing food and so raises urbanization, while falling trade costs weaken the link between city locations and regions of high agricultural productivity. A key result is that the time pattern in which these changes arrive affects how first-nature characteristics are mapped into the eventual distribution of economic activity, which will explain the puzzle noted earlier.

Looking at historical data on urbanization and education (to proxy for agricultural productivity) and transport costs, we show that there is an important difference between today's developed countries where the process of agglomeration and structural transformation began early and those developing countries where it is more recent. In the former group, urbanization started at a time when transport costs were still relatively high, so urban agglomerations were more localized in different agricultural regions, due to high costs of trade and local home markets being able to support local manufacturing. In the latter, transport costs fell before structural transformation occurred. To exploit urban scale economies before structural transformation with a limited national urban labor force, manufacturing tended to agglomerate in relatively few, often coastal locations. With structural transformation, these initial coastal locations then grew, without formation of more cities in the agricultural interior. Thus the model predicts that agglomeration will be more tied to locations where food can easily be produced in the former than in the latter.

To test the model, we consider different ways of splitting the countries in the world in samples of “early agglomerators” and “late agglomerators.” Specifically, we let the data choose the cutoff values of two variables (urbanization and education in 1950) so that countries are sorted into one bin or the other. We show that in countries that agglomerated early, first-nature characteristics related to food production are more important in determining current settlement than are those associated with trade, relative to the case of countries that agglomerated later. From a static perspective this result is counterintuitive as noted above: highly urban Britain is far less dependent on agriculture and more dependent on trade than Mozambique. Yet early settlement patterns in present day rich countries have strongly persistent impacts.

The rest of this paper is organized as follows. Section 2 describes the data on lights and geographic characteristics that we use. In Section 3 we first discuss the interpretation of the lights data, and then turn to an examination of the ability of geographic factors in general to predict variation in observable lights, focusing on the fit as measured by R^2 or the increment to R^2 after country fixed effects are included. Section 4 presents the model of regional development, emphasizing the roles of rising agricultural productivity, falling transport costs, and the timing with which these two changes occur. In Section 5 we develop an empirical specification to assess the main implications of the model and discuss the division of countries into early and late agglomerators. Section 6 presents the main empirical results, while Section 7 discusses the effect of adjusting for spatial spillovers. Section 8 concludes.

2. Data

In order to carry out this exercise, we need measures of economic activity and several components of physical geography, all available on a global scale.

Our proxy for economic activity is night lights (Figure 1). Unlike Henderson, Storeygard and Weil (2012) and most quantitative work on lights, we use the radiance-calibrated version of the

data (Elvidge et al. 1999; Ziskin et al. 2010). In normal operations, the light detection sensor is very good at detecting low levels of light in small cities. However, the strong amplification that enables this detection also saturates the sensor in the most brightly lit places, including the centers of most of the largest 100 cities in the United States, so that their values are top coded. The 2010 Global Radiance Calibrated Nighttime Lights dataset we use combines the high magnification regime for low light places with a lower magnification regime for more brightly lit places. Thus nearly all topcoding is removed, with minimal loss of information about low light places. The lights data are distributed as a grid of pixels of dimension 0.5 arc-minute resolution (1/120 of a degree of longitude/latitude, or approximately 1 square kilometer at the equator).⁸

We use lights as the measure of economic activity because it is measured consistently worldwide at the same spatial scale. Alternatively, we could have considered population. There are three main sources of global population data. Landscan⁹ and Worldpop (Stevens et al. 2015) use other geographic data to interpolate population within census geographic units, which has the potential to bias our estimates. The Gridded Population of the World (GPW; CIESIN and CIAT 2005) uses population data exclusively, assuming uniform population density within enumeration units larger than its native (2.5 arc-minute) grid resolution. On average, this means that population estimates are more heavily smoothed in poorer countries with lower statistical capacity, as well as in more sparsely populated regions. This could also bias our results.

Apart from measurement issues, the lights data are conceptually different from population, because they reflect intensity of economic activity, which is a combination of population and income per capita. Assuming a reasonable degree of population mobility within countries in the long run, lights variation within countries will mostly reflect variation in the spatial distribution of population.

⁸ Available at <http://ngdc.noaa.gov/eog/dmsp.html>.

⁹ <http://web.ornl.gov/sci/landscan/>

Our other variables of interest are all reported at several different geographic scales, ranging from 1/120 of a degree to 1/2 degree. For analysis, we convert them all to a grid of 1/4-degree squares, with each square covering approximately 770 square kilometers at the equator, decreasing with the cosine of latitude. This scale is a compromise between the fine detail observed at the native resolution of several datasets and the computational practicality of coarser cells. It also allows us to be less concerned about spatial autocorrelation than we would be at finer scales, and to reduce true spillovers as well. At this resolution our sample is over 200,000 grid squares.

Variables originally reported at scales smaller than 1/4 degree are aggregated with an appropriate function. In the case of variables with continuous values, measures for our grid squares represent the mean or sum of all input cells falling within them, as appropriate. So for example, the night lights measure for each quarter-degree grid square is the sum of the 900 component raw lights pixels. In the case of categorical variables, we assign the modal value. For variables originally reported in 1/2 degree cells, each 1/4-degree grid square receives the value of the larger input cell into which it falls.

To analyze the determinants of variation in economic activity across locations, we define three sets of explanatory variables, which we refer to as agricultural, trade, and base covariates. The base covariates are two variables that arguably affect both trade and agriculture. These are malaria and ruggedness. Malaria affects human ability to live in an area regardless of the economic activities they perform, and ruggedness increases the cost of both trade and agriculture. The index of the stability of malaria transmission, based on species-specific measures of human biting rates and climate predictors of mosquito survival, is from Kiszewski et al. (2004). Ruggedness is based on Nunn and Puga (2012).¹⁰

¹⁰ We correct the Nunn and Puga measure to account for the fact that two east-west neighboring cells at high latitudes are closer than two east-west neighboring cells at low latitudes, biasing their measure downward at high latitudes. Applying this corrected measure to the main regression in Nunn and Puga (2012) leads to virtually no change in the point estimate of the variable of interest and an approximately 15% increase in its standard error. We also area-weight the average to follow Nunn and Puga. In practice, area weighting has minimal impact within our small units.

Our agricultural covariates comprise six continuous variables (temperature, precipitation, length of growing period, land suitability for agriculture, elevation, and latitude) as well as a set of 14 biome indicators. The temperature variable is a long run (1960-1990) average of UEA CRU et al. (2013) based on Mitchell and Jones (2005) and precipitation is the Wilmott and Matsuura (2012) measure averaged over the same period. Length of growing period, in days, is from FAO/IIASA (2011). Land suitability is the predicted value of the propensity of a given parcel of land to be under cultivation based on four measures of climate and soil, from Ramankutty et al. (2002).¹¹

Elevation, in meters, is from Isciences (2008). While high elevation locations often have poor transport, we believe that once distance to various types of water transport (see below) and ruggedness are controlled for, it is best interpreted as an agricultural variable. Furthermore, while ruggedness and malaria have similar effects on economic activity across different sample splits discussed below, elevation, like many trade and agricultural variables, does not. In practice, the choice of whether to place elevation in the agriculture category or the “both” category is inconsequential to the main results. Finally, we control for the absolute value of latitude which could affect agriculture even net of our climate controls.

Biomes are mutually exclusive regions encoding the dominant natural vegetation expected in an area, based on research by biologists. The distribution of 14 biomes is from Olson et al (2001). We combine “tropical and subtropical dry broadleaf forests” with “tropical and subtropical coniferous forests”, and also combine “tropical and subtropical grasslands and savannas and shrublands” with “flooded grasslands and savannas” because each pair is broadly similar, and

¹¹ Because several variables are only defined or reported for grid squares containing land, and different datasets have different effective definitions of the land surface, as noted below, values for some variables are imputed (or “grown”) as the mean (continuous) or mode (categorical) of their eight 1/4-degree grid square neighbors. This process is repeated up to two times until nearly all cells containing land based on our coastline dataset have values for all variables. Between the two iterations, interpolated values assigned to cells containing no land are dropped, so that imputation cannot occur across large water bodies. The only land cells without data following this spatial interpolation process are small islands. Land suitability, biomes, temperature and precipitation are grown twice, and length of growing season is grown once.

because the second member of each pair contains less than 1% of cells globally. We exclude areas covered by permanent ice.

Our five trade variables focus on access to water transport. We calculate Euclidean distances in kilometers from cell centroids to the nearest coast, navigable river, and major lake using the Fuller icosahedral map projection, and great circle distances to the nearest natural harbor.¹² Our specifications include indicators for the presence of each of these four features within 25 km of a cell centroid, as well as a continuous measure of distance to the coast. In each case, we take a more systematic approach to characterizing the universe of waterbodies than previous work. Vector coastline data are from NOAA (2011; “low” resolution), based on Wessell and Smith (1996). The same data are also gridded at 0.5 arc minutes in order to determine the fraction of these 0.5 minute cells in a quarter-degree grid square that fall on land. Our universe of rivers is those in size categories 1-5 (on a scale of 1-7) of the river and lake centerline dataset from Natural Earth (2012). We restrict to river segments that are navigable, having determined the navigability of each river using a variety of text sources.¹³ Lakes data are from the Global Lakes and Wetlands Database produced by the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel (Lehner and Döll 2004).¹⁴ We restrict consideration to the 29 lakes with a surface area greater than 5000 square kilometers, having excluded four that were wholly created by dams. Port locations are digitized from US Navy (1953). We restrict to ports defined there as natural harbors.

Columns 1 and 2 of Table 1 report summary statistics for all of these variables.

¹² All available GIS software of which we are aware performs this calculation in the plane, and thus requires choosing a projection (see Tobler (2002) for a critique). No projection preserves distance in general, and many, including the Plate Carrée implicitly used in most economics research, can induce substantial error. Spherical point-to-point distances, in contrast, can be calculated easily in many software packages. We use Fuller’s icosahedral projection, which we believe is relatively well-suited for the task, and has not previously been used for such quantitative purposes in any literature of which we are aware.

¹³ Full list available upon request

¹⁴ <http://www.worldwildlife.org/publications/global-lakes-and-wetlands-database-large-lake-polygons-level-1>, accessed 2014/8/14

3. The measure of economic activity and initial results

3.1 The measure of economic activity and specification

As seen in Figure 1, the lights data convey a great deal of information about the location of economic activity. At the country level, the cross-sectional correlation between emitted light and GDP in our data is 0.90. More importantly for our purposes, the lights map out the location of economic activity *within* countries. Note that lights represent total economic activity, which is a combination of the number of people and the activity level per person. Lights are comparably bright in northern India and the eastern United States, because while economic activity per person is lower in India, population density is higher in many places.

The land area falling within each grid square varies, both because some are partially covered by water or permanent ice, and because the surface distance between lines of longitude varies with the cosine of latitude. To deal with the first problem, we divide our lights measure by land area. Specifically, we sum the lights in each grid square and inflate by the inverse of the fraction of the 900 pixels in the grid square falling on land. This is equivalent to averaging over land pixels only.¹⁵

One notable limitation of the lights data is that 59% of our grid squares emit too little light for the satellite to detect. Since nearly all grid squares contain population and thus presumably emit some level of light, we consider this a censoring problem. The lowest non-zero values are generally interpreted as noise and recoded to zero at the pixel level in initial processing by NOAA. The lowest non-zero value of the sum of lights adjusted by fraction of pixels over land is 3.05. We assign this value to all measured zeroes to avoid artificially inducing variation between them and the smallest non-zero values.¹⁶

¹⁵ While cell area varies with latitude, the light readings are densities.

¹⁶ Alternatively, we could estimate a Tobit model, which is the traditional way to capture censoring. OLS avoids the Tobit error structure and provides a more intuitive measure of goodness of fit, which is our focus. Estimated

The base formulation for grid square i in country c is thus

$$\ln(\text{light}_{ic}) = X_{ic}\beta + \epsilon_{ic} \quad (1)$$

$$\text{where } \text{light}_{ic} = \begin{cases} \frac{900 \sum_{j \in i} \text{light}_{jc}}{\sum_{j \in i} 1(\text{land}_{jc})} & \text{if } \text{light}_{ic} \geq 3.05 \\ \ln(3.05) & \text{otherwise} \end{cases}$$

and j indexes pixels. We also consider the intensive and extensive margins separately. Figure 2 plots the distribution of the dependent variable excluding the bottom code.

We emphasize two further points about equation (1). First, it is a very simple functional form. With such a large number of covariates, a 2nd order Taylor series has hundreds of terms, which improves the fit but limits interpretation substantially. Second, we consider multiple error structures. We show all results with and without country fixed effects. This distinction is critical: the fixed effects regressions rely on within-country variation and account for the allocation of activity within a country, in a context where we expect a higher variance across countries than across grid cells within a country. Errors are clustered by 3-by-3 squares of cells to account for spatial autocorrelation. Conley standard errors, used in alternative specifications below, tend to be smaller.

3.2 Basic results

Columns 3 and 4 of Table 1 report the results of a regression of our lights variable on the full suite of physical geography characteristics without and with country fixed effects. The

coefficients from the analogous tobit models (with and without country fixed effects) on variables with significant coefficients are exclusively of the same sign and are mostly larger in magnitude.

coefficients with and without fixed effects are generally of similar magnitudes and are of the same sign for all covariates except the mangroves indicator, which applies to 0.4% of the sample. Because of the high potential for collinearity among the right hand side variables, it may be that looking at the change in a particular coefficient in comparing the specification with country fixed effects to the specification without them is not very informative. As an alternative, we created fitted values from the specifications in columns 3 and 4, in the latter case suppressing the country fixed effects (i.e. setting all the country dummies to zero when forming fitted values). The correlation of the fitted values is 0.860, suggesting that the two specifications provide very similar predictions of which regions have high light density. In other words, the geographic forces that drive the allocation of economic activity within and across countries are similar. In Figures 3a and 3b we plot the fitted values of lights under the two specifications. The absolute scales differ because when we omit the fixed effects themselves, the fixed effect predictions are all relative to the base country. Nonetheless, the relative variation in lights within continents and countries is similar in the two figures.

In columns 3 and 4, coefficients on covariates are largely in the expected direction. Most biomes have significantly more lights than deserts (the reference biome); only boreal forests, tundra, and perhaps surprisingly, tropical grasslands, have significantly less. Being near the coast, lakes, navigable rivers and natural harbors is associated with more lights, as is a longer growing season and higher agricultural suitability. Net of growing season, land suitability and biomes, higher temperatures and lower precipitation are associated with more lights, perhaps because of their residential consumer amenity value. In an alternative specification excluding growing season, land suitability, and the biomes (not shown), precipitation has a positive effect overall as might be expected based on agricultural productivity. When entered in quadratic form (not shown), both temperature and precipitation increase lights at a decreasing rate. In the base formulation, net of ruggedness and coastal distance, higher elevation is associated with more lights.

As reported in Table A1, using column 3 as an example, Conley (2008) standard errors using a kernel of radius 40 km (similar to clustering for immediate queen neighbors) are larger than the

ones in column 3 by 5-20%. We think our cluster design, which is substantially easier to compute for our many variants, yields reasonable estimates of standard errors, in a context where we are not particularly interested in the specific degree of significance of individual coefficients.

The most important numbers in Table 1, columns 3 and 4, for our exposition are the R^2 values. These 23 variables account for 46 percent of the variation in lights globally in column 3. We consider it remarkable that such a parsimonious specification can account for so much of the variation in global economic activity, without explicit regard to agglomeration or history. Of course to the extent that grid cells with better characteristics have neighbors with better characteristics (which we will see later is highly likely) and there are agglomeration forces, some of the light intensity associated with more lights in better places represents agglomeration. And of course there are within-grid square agglomeration forces. In short, these coefficients are reduced form estimates, which in part capture forces of agglomeration (see Section 7). In column 4 we add country fixed effects. While on their own these account for 34 percent of light variation, they only increase the column 4 R^2 by 11 percentage points relative to column 3. Country-level variation adds relatively little once physical geography factors are accounted for. Conversely, the geographic factors add 23 percentage points in explaining variation, on top of the fixed effects.

Table A2 reports OLS estimates of the effects of the same variables on the extensive and intensive margin of lights. Sign patterns for covariates with and without fixed effects are mostly the same across margins but there are some differences, especially for the 3 tropical biomes measures and elevation. Higher elevation increases the probability of being lit but is associated with lower light intensity, conditional on being lit.

Table 2 reports R^2 for a variety of specifications, exploring the role of different variable sets in explaining variation in lights in more detail. Column 1 excludes country fixed effects and column 2 includes them. The first row repeats the R^2 's from Table 1. Rows 2 and 3 report the extensive and intensive margin R^2 's, respectively. Although the R^2 values are not strictly comparable across

margins, it is nonetheless striking that the extensive margin, a linear probability model, has a relatively large R^2 value of 0.39 without fixed effects and 0.48 with them. As shown in rows 4-6, country fixed effects alone capture differing levels of economic development (and underlying cultural and institutional differences), explaining more than 20 percent of both extensive and intensive margin variation.

In rows 7-9, we start to explore the relative role of trade and agricultural variables. Row 7 shows that the two base controls on their own explain little and that fixed effects explain a lot. In row 8, agricultural variables on their own have much explanatory power, an R^2 of 0.44 without fixed effects, and 0.56 with them. Row 9 suggests that the short list of trade variables on their own explain much less of lights variation and add little to the explanatory power of country fixed effects. However, as we will show, these relative contributions vary importantly between early- and late-developing countries.

4. Model

As suggested in the introduction, the effect of physical geography is modulated by history. Changes in productive technologies have altered the importance of different first nature factors. We focus on two critical changes. First, over the last two centuries, the cost of transporting goods has fallen dramatically, both between and within countries. Among developing countries with low productivity agriculture, by 1950 and in many cases much earlier, transport costs had fallen with the building of colonial rails and roads as well as the use of trucks (Donaldson forthcoming; Jedwab and Moradi forthcoming; Jedwab, Kerby and Moradi forthcoming). Second, there has been an enormous increase in agricultural productivity, releasing labor to agglomerate in cities. In today's developed countries, this structural transformation began well before the major declines in transport costs (Desmet and Henderson, 2015). Structural transformation has been slower in today's developing countries, occurring in Latin America mostly after 1950 and still ongoing in most of Asia and Sub-Saharan Africa today.

We develop a model in which the order in which these two changes occurred may influence the spatial distribution of economic activity. As we will argue below, those countries that experienced the agricultural revolution before much of the dramatic drop in transport costs saw local agglomerations develop early in several regions. Higher agricultural productivity released farmers into the manufacturing cities, but since transport costs were high, cities developed in several agricultural regions at once. We will argue that these agglomerations in agricultural regions then persisted after the period of the most dramatic drop in transport costs. In contrast, today's developing countries experienced substantial drops in transport costs before they underwent structural transformation. Thus, their urban systems are more likely to be concentrated in one or a few places such as the coast, taking greater advantage of urban scale economies in the face of initial limited urban labor. Lowered transport costs allow trade of manufactures across the regions (as opposed to just local production). Once structural transformation starts, these initial agglomerations persist and grow, with little city development in the agricultural hinterland.

4.1 Setup

We consider a country consisting of two regions, which we call coast (c) and interior (i) for exposition. Workers in each region potentially produce food (f) with decreasing returns and a manufactured good (m) with external economies of scale subject to congestion. In both sectors, workers are paid their average product. We assume that (in the “long run”) workers are free to move between regions and among sectors such that utility is equalized. The economy is closed to the outside world.

For either region $r \in \{c, i\}$, in the food sector average product is $A_f L_{fr}^{-\beta}$ and total production is $A_f L_{fr}^{1-\beta}$, where $A_f > 0$ reflects productivity and L_{fr} is the amount of labor in the food sector in the region. Food sector productivity is the same across regions. Decreasing marginal productivity of labor in agriculture, due to a fixed supply of land, is reflected in the parameter

$1 > \beta > 0$. The urban sector produces the manufactured good. Average product per unit of labor in the urban sector is $A_{mr}(v + L_{mr}^\epsilon)$, where the v allows nonzero output by the marginal worker as $L_{mr} \rightarrow 0$, and $\epsilon > 0$ represents agglomeration economies increasing productivity in the presence of more workers. Each worker is endowed with one unit of time, to be used for labor and commuting in the city as in standard urban models (see Duranton and Puga, 2004, for a review), so labor supplied per worker is $1 - tL_{mr}$, where $0 < t \ll 1$ represents unit-distance commuting costs in the city.¹⁷ Average product per worker in the city is thus

$A_{mr}(v + L_{mr}^\epsilon)(1 - tL_{mr})$. The size of the manufacturing labor force that maximizes this is a solution to $\epsilon L_{mr}^{\epsilon-1} - (1 + \epsilon)tL_{mr}^\epsilon - vt = 0$. As long as $\epsilon < 1$, any L_{mr} fulfilling this expression will be a unique maximum, but further restrictions on the parameter space are required to guarantee the existence of an interior maximum.

Food, which is traded costlessly between regions as in standard new economic geography models, is the numeraire good. Preferences are such that each worker consumes a fixed amount of food γ , and spends the remainder of her value of net average product on the manufactured good. Welfare for any person in region r is then equivalent to consumption of the manufactured good, $(W_r - \gamma)/p_{mr}$, where W_r is net wage income and p_{mr} is the price of the manufactured good in region r .

A fixed national population of workers L is free to move between sectors and regions, so that

$$L_r = L_{fr} + L_{mr}, \quad r = c, i \tag{2a}$$

¹⁷ Following Duranton and Puga (2004), each worker is endowed with 1 unit of time, and lives on a lot of fixed size 1, with zero opportunity cost, in a two sided linear city. Working time is $1-4tu$ where u is distance from the city center and $4t$ unit commuting costs. Given a wage rate w , income after commuting is $w(1-4tu)$. Residents pay rents that differ by distance from the city center, and rental income is redistributed as an (equal) dividend to all city residents. Since the lots have zero opportunity costs, rent at the city edge ($u_{max}=L/2$) is zero. Net income of the person at the city edge before rent income transfers is $w(1-4tu_{max}) = w(1-2tL)$ and that equals net income of any other person $w(1-4tu)-R(u)$, where $R(u)$ is rent at location u . Thus $R(u) = 4tw(L/2-u)$, so integrating across the whole city, total rents are twL^2 , and rents remitted per worker are twL . Thus, total (labor plus rental dividend) income per person in the city net of rent and commuting time is $W=w(1-tL)$ and total income is $w(1-tL)L$. This corresponds to effective labor supply in the city being $L(1-tL)$.

$$L = L_c + L_i. \quad (2b)$$

Real income equalization across sectors within each region (assuming both sectors exist in the region) implies:

$$A_f L_{fr}^{-\beta} = p_{mr} A_{mr} (v + L_{mr}^\epsilon) (1 - t L_{mr}) , \quad r = c, i. \quad (3)$$

Free migration equalizes per person welfare (i.e. manufacturing consumption), across regions so that:

$$\frac{A_f L_{fc}^{-\beta} - \gamma}{p_{mc}} = \frac{A_f L_{fi}^{-\beta} - \gamma}{p_{mi}} . \quad (4)$$

The model is closed by imposing equilibrium in goods markets. How that is done depends on whether there is inter-regional trade or not and whether regions absolutely specialize or not. There are three different types of closure relating to three types of equilibria.

4.2 Autarkic equilibrium

If there is no trade between regions, clearing of the manufacturing good market in each region requires total regional demand equals regional supply, or:

$$L_r (A_f L_{fr}^{-\beta} - \gamma) = p_{mr} A_{mr} (v + L_{mr}^\epsilon) (1 - t L_{mr}) L_{mr} , \quad (5a)$$

or alternatively, using the agricultural market

$$\gamma L_r = A_f L_{fr}^{1-\beta} . \quad (5b)$$

Given $L, A_f, A_{mi}, A_{mc}, \beta, \epsilon, \gamma, t, v$, the eight equations implied in (2)-(4) and (5b) specify equilibrium in the distribution of labor and the price of the manufactured good wherever it is

produced ($L_{mc}, L_{mi}, L_{fc}, L_{fi}, L_c, L_i, p_{mc}, p_{mi}$).¹⁸

4.3 Trade equilibrium with both regions producing manufactures

If transport costs are sufficiently low, both regions can trade *and* produce manufactures if they have differential comparative advantage. If they are identical and have sufficient manufacturing scale (say, beyond the point where average product is maximized) then there will be no trade. We generally designate one region to be slightly better at manufacturing, in order to allow trade equilibria when trade costs are sufficiently low. We maintain the assumption that food can move costlessly between regions, and further assume that there is an iceberg cost τ that applies to movement of the manufactured good between regions. Trade will occur when the autarky price ratio of manufactured goods is outside the range $(1 - \tau, \frac{1}{1-\tau})$. When there is trade, and no absolute specialization in either region, the within-region goods market clearing conditions (5a) and (5b) are replaced by an inter-regional goods market clearing condition and an arbitrage condition. We specify food market equilibrium and leave the manufactured good as a residual:

$$\gamma L = A_f L_{fc}^{1-\beta} + A_f L_{fi}^{1-\beta}. \quad (6)$$

Assigning manufacturing comparative advantage to the coastal region, in practice it will always be the manufacturing exporter in this class of equilibria, although we check for equilibria where the interior is exporting manufactures as well. The prices of the manufactured good in the two regions are related by an arbitrage condition:

$$p_{mi} = p_{mc}(1 - \tau)^{-1}, \quad (7)$$

where τ is an iceberg trade cost between the two regions. Given $L, A_f, A_{mi}, A_{mc}, \beta, \epsilon, \gamma, t, \tau, v$, the eight equations embedded in (2)-(4), (6) and (7) specify an equilibrium in the distribution of

¹⁸ To see that these represent eight equations, note that (2a), (3) and (5b) each must be fulfilled for each region.

labor and the price of the manufactured good in the two regions.

4.4 Specialization equilibrium

Finally, there are equilibria where all manufactured goods are produced in one region. Since that one region can be either the coast or the interior, we consider the two corresponding types of specialized equilibria in the solution mechanism below. It can be defined by slightly adjusting the trade equilibria without specialization above, setting manufacturing employment in one region to zero and removing equation (3) for that region.

4.5 Solving the model

For any given set of parameters, we solve the model as follows. We have 3 types of possible equilibria: autarkic, trade without absolute specialization, and trade with absolute specialization, with each of the last two available in two variants, one for each region exporting manufactures. We pick an allocation of population to the interior region (with the coastal population being the remainder of national population) and suspend equation (4) (equalizing welfare across regions). We then use the remaining equations in each type of equilibria to solve for all remaining variables. From these we calculate the consumption per worker in each region (the LHS and RHS to (4)). Then, for each equilibrium type, we plot these two regional consumptions as a function of (say) interior population. Their intersections are equilibria.

We limit attention to stable equilibria, subject to two stability conditions. Type 1 stability is with respect to small changes in the population allocation across regions, assuming within-region labor markets and all goods markets always clear (“instantly”). Equilibria are stable as long as per-person manufacturing consumption in the interior (coast) is a declining (increasing) function of L_i (i.e., there are overall diseconomies to regional size). Type 2 stability is with respect to perturbations within regions, focused in particular on adding a small number of workers to a non-existent or small manufacturing sector within a region. For example, we perturb a small

number of workers out of food production in, say, the interior region and move them into manufacturing in the interior. We keep regional populations fixed, but allow intra- and inter-regional goods markets and coastal labor markets to clear (“instantly”). Equilibria are unstable if interior manufacturing workers then have higher consumption than interior food workers. They are stable if the reverse is the case. This condition implicitly assumes slower than instant adjustment in inter-regional labor markets. We note however that in practice in all examples we solved, in this type of experiment, under stability, interior food workers have the highest welfare (manufacturing consumption) of workers anywhere and interior manufacturing workers the lowest; and vice versa under instability.

Details of the solution method as applied to the examples below are given in the Appendix. In general, for any τ there will either be an autarky or non-specialization equilibrium but not both, with higher τ having autarky. There may or may not be specialization equilibrium in one or both regions, with the likelihood of stable specialization equilibria enhanced as τ falls.

4.6 Analysis of possible equilibria

As with many similar models, there is no closed form solution. We illustrate the relevant properties with several examples. Our baseline parameter set is

$$\{L = 10,000,000; \nu = 0.5; \epsilon = 0.08; t = 7 \times 10^{-8}; \beta = 0.25; \gamma = 0.018; A_f = 1; A_{mi} = 1; A_{mc} = 1.01\}.$$

With these parameters, average manufacturing product peaks at a city population of about 969,100. Note the regions are not precisely symmetrical, so that if trade is feasible, it will occur because the coast has a slight comparative advantage in manufacturing production (with a higher A_m). We consider all the specialization and non-specialization equilibria that exist and are stable for values of transport costs, τ , from 0.99 to 0.01. There are two types of specialization equilibria: the coast producing only food and the interior producing only food. Non-specialization implies autarky at high τ ; while at lower τ , when stable non-specialized equilibria exist, they are trade equilibria.

Our focus is on how these patterns change in the transition from low ($A_f = 1$) to high ($A_f = 1.5$) agricultural productivity. When A_f is low, at least with non-specialized manufacturing, there is insufficient manufacturing employment to support a city populous enough to exploit scale economies in any one region. When A_f is high, much less labor is needed to produce the required food, so there is a lot more manufacturing employment to allocate between the two regions.

Figure 4 shows the stable specialization and non-specialization equilibria when $A_f = 1$ for different costs of trade, τ , as graphed against the population of the interior region. The two outer prongs correspond to the two sets of specialization equilibria: one where the interior produces only food and one where the coast does that. When do these specialization equilibria exist and when are they stable? First they are (type 2) unstable when trade costs are high ($\tau > 0.4$). In that case, workers who begin manufacturing in the region with no existing manufacturing will be better off, because high trade costs make them profitable in their home market despite the limited scale. When τ is lower, the scale effect advantage of the existing manufacturing sector in the foreign region dominates any trade cost advantage in starting manufacturing in the home region to sell in the home market; starting a small scale manufacturing operation is not profitable for those workers. In our example with $A_f = 1$, the allocation of workers to manufacturing in the specialized region is less than the city size that maximizes average product. Thus stable specialization equilibria persist as τ falls to 0.

The middle prong represents non-specialized equilibria. At high τ , they are autarkic and stable. While manufacturing scale is low in both regions, trade is too costly for workers to profitably move to take advantage of scale economies in one region. As trade costs fall, it becomes potentially profitable to trade. However, once it is profitable to trade it is also profitable to enhance manufacturing scale in one region relative to the other. Thus at these parameter values, the only stable equilibria when τ is low have manufacturing located in a single region, with the other region producing only food.

In Figure 5 we turn to our second case, where agricultural productivity A_f has risen to 1.5, allowing more workers to enter manufacturing. As in the low agricultural productivity case, specialization equilibria are not stable at the highest τ . As in the previous case, when trade costs are high, the only non-specialized equilibrium is autarky. However, what is new in this case is that, as trade costs fall and autarky becomes unstable, a trade equilibrium without specialization now becomes stable. In this equilibrium, both regions have enough manufacturing scale that shifting a small number of manufacturing workers one way or another is not profitable for workers. Given our assumption of a slight coastal comparative advantage in manufacturing, the coast exports manufactured goods in this equilibrium, but such goods are produced in both regions.

Figure 6 shows the existence of the different classes of equilibria in (A_f, τ) parameter space for values of A_f from 1 to 2.3 and τ from 0.01 to 0.6. In the upper left of the figure, where agricultural productivity is high and trade costs are low, the only equilibrium is one in which manufacturing takes place in both regions (area G). At this equilibrium, the coast has a comparative advantage in manufacturing, so it exports this good. Maintaining low trade costs but lowering agricultural productivity, there are a series of different equilibrium configurations: in area B, there are two equilibria: one with trade and manufacturing taking place in both regions, and one with trade and manufactures produced only on the coast. In area C, there are three equilibria: the two just listed as well as one in which manufactures are produced only in the interior. For even lower agricultural productivity, maintaining low trade costs, are only two equilibria those where manufactures are produced in a single region (area D). Raising trade costs (i.e. moving from left to right in the figure) autarky appears as a possible equilibrium, although there are also possible equilibria where manufactures are produced in one or both regions (areas E and F). Finally, with high enough trade costs, the only possible equilibrium is autarky (area A).

Figure 6 demonstrates the path dependence we explore empirically in the next section. Circa

1800, all of the world was in the high trade cost and low agricultural productivity autarkic equilibrium at the lower right in area A. In the countries that developed early, the historical paths of rising agricultural productivity and transport costs that did not fall too quickly maintained the economy in areas A, E, or F, in all of which there was a stable autarkic equilibrium. Although we do not model persistence explicitly, our assumption is that in such areas, an economy that was formerly at the autarkic equilibrium will remain there. Finally, with further declines in transport costs, these early developing economies moved into areas B, C, or G, all of which lack an autarky equilibrium, but do feature an equilibrium with non-specialized trade. Again, in the absence of an explicit model of persistence, we assume that when the autarkic equilibrium disappeared, these economies naturally moved to the non-specialized trade equilibrium, which featured very similar distributions of population and economic activity between regions.

Today's developing world followed a different path, with the fall in transport costs relative to the rise in agricultural productivity occurring earlier. In this case, countries entered area D in the figure, where neither autarky nor non-specialized trade is a stable equilibrium. Manufactures will be produced in only one region, and it is natural to expect that this would be the coast, where productivity is higher. Once this specialization equilibrium has been entered, persistence would mean that further increases in agricultural productivity (moving into areas C and B) would not move the economy away from the equilibrium in which manufactures are produced in only one region. Thus, by the time of the mid- to late-20th century agricultural revolution in the developing world, urbanization (and manufacturing activity) is more concentrated in a smaller set of regions with better access to world markets.

5. Empirical specification

Operationalizing our model requires defining factors that affect trade and those that affect agriculture, as well as defining the set of early-agglomerating countries and the set of

late-agglomerating countries. Above in the data section we discussed the variables that relate primarily to trade and to agriculture.

To categorize countries that agglomerated late and those that agglomerated early, we rely primarily on human capital, which allowed farmers to take advantage of higher-yield technologies. Figure 7 plots adult literacy rates over time for the UK (an early adopter of advanced technology) and India. The UK had over 50% literacy by the 17th century, and following a rapid rise after 1820, over 75% by 1870. Thus the UK achieved a massive increase in human capital before the precipitous decline in the global freight cost index of the late 19th and early 20th century. In 1951 India's literacy rate was still under 20% and only then started to rise quickly, achieving 50% in the 1990s. More broadly, we argue that a set of "early agglomeration" countries, largely today's developed world, harnessed new agricultural technologies and urbanized before the radical decline in freight costs, while "late agglomeration" countries, today's developing world, only saw these new technologies and substantial urbanization much later, after the largest declines in transport costs.

We operationalize our human capital measure using national average years of schooling in the adult population in 1950, the earliest year with comprehensive data, from Barro and Lee (2010). Alternatively, we use an agglomeration outcome, the level of urbanization in 1950, again the earliest year with comprehensive data, from United Nations (2014). Figure 8 shows the cumulative distribution of these two measures in 1950 weighted by national populations. They closely track each other and indeed the partitions we make and the regression results are similar across the two measures.

Our theory provides no guidance on the precise distinction between early and late spatially transforming countries. We thus follow Durlauf and Johnson (1995), letting the data tell us the cutoff at which the overall unexplained variance, summed across the "early and "late" samples, is minimized. In general, we estimate the following equation, as well as use it to determine where to split the sample between early and late transformers:

$$\ln(\text{light}_{ic}) = X_{ic}\beta + D_{\text{split}_c} + D_{\text{split}_c}X_{ic}\beta_d + \epsilon_{ic} \quad (8)$$

where D_{split_c} is a dummy variable indicating, for example, whether a country is in the high education category. From Table 2, the overall R^2 's without splits are 0.459 and 0.569 without and with fixed effects, respectively. We carry out the sample split exercise out for four separate situations, one with and one without country fixed effects for each of our two proxies, education and urbanization.

Figure 9a provides an illustration of the approach for the education proxy without fixed effects. The vertical axis represents the sum of squared residuals (SSR), summed across two regressions carried out with the same specification on two separate samples. The horizontal axis specifies the cutoff level of education defining the early and late samples. SSR is minimized (and therefore explained variance is maximized) at a cutoff level of 3.6 years of education in 1950. In the fixed effects regressions in Figure 9b, explained variance is maximized at 3.0 years of education. Figures 9c and 9d show the analogous information for the urbanization proxy. A 1950 urbanization level of 44 percent is our cutoff without fixed effects; adding fixed effects reduces it slightly to 38 percent.

Regardless of the proxy we use, we end up with a similar split of the sample. Table 3 cross-tabulates the full sample and the lit sample using the two types of cutoffs, separately with and without country fixed effects. The off-diagonal cells, those that are characterized as being in early agglomeration countries using the urbanization criterion and late agglomeration countries using the education criterion, or vice versa, represent less than 5% of the full sample and less than 10% of the lit sample. It is also the case that the chosen cutoffs split the samples quite evenly. The countries and their categorization in each of the four variants are listed in Table A3.¹⁹

¹⁹ Because some countries lack a measure of urbanization in 1950, and several more lack an education measure, the sample sizes in these two differential exercises are smaller by 0.13% and 6.27%, respectively, than those in Tables 1 and 2. However, the overall R^2 changes by less than 1% with fixed effects, and 0.1% without fixed effects.

6. Differential results

6.1 Explanatory power

Table 4 reports our main results, the contribution of different blocks of variables in explaining lights variation within the early and late agglomeration samples, with and without country fixed effects, following equation (8). We discuss results for the education split shown in the left half of the table. Results for the urbanization split, in the right half, are very similar. Panel A shows each variable set's contribution to R^2 for low and high education countries. To highlight the comparison of interest, we can net out the contribution of the base variables. In the high education countries, the additional explanatory power of the agricultural variables is more than that of the trade variables. In the low education countries, it is the trade variables that offer relatively more explanatory power. So, for example, using the education split without fixed effects, agriculture adds 0.54 to explanatory power relative to the base for high education countries but only 0.27 for low education countries. In contrast, trade relative to the base adds 0.062 to explanatory power for high education countries but a higher 0.17 for low education countries. This pattern, consistent with our model, holds for all such pairwise comparisons in Panel A: for the education split with fixed effects and for the urbanization split with and without fixed effects. The last row of panel A summarizes this relationship, the relative advantage of agriculture over trade variables in explaining lights variation for high versus low education countries, in a double difference. The patterns are clear across both education and urbanization splits. Agriculture is relatively more important for early developing countries. Without fixed effects for both education and urbanization the double differential is over 0.31 and with fixed effects for both it is over 0.17.

Panels B and C are the analogs of Panel A splitting overall effects into the intensive and extensive margins, respectively. The distribution of effects between country sets and variable sets are similar across the two margins.

Finally in Panels D and E we address an issue so far put aside. This paper makes a split between developed and developing countries. A number of other splits have been suggested to us, one of which is New World versus Old World. Splitting the data (allowing differential slope coefficients of all covariates) in any way will improve the explanatory power of equation (8) over equation (1). A New-Old World split yields similar explanatory power as a high-low education or urbanization split.²⁰ What is critical to us is that the pattern we have shown in Panel A is not simply a proxy for the New World-Old World split. Panels D and E show that this is the case. The relative advantage of agriculture over trade variables in explaining lights variation for high versus low education countries (or high versus low urbanization) is present in both New and Old World countries. In other words, results are consistent with our model within the New World and within the Old World.

6.2 Marginal effects

Table 4 emphasized the overall explanatory power of groups of trade and agricultural variables in the two samples. We now consider the differential in their relative marginal effects. If marginal effects of trade variables, relative to marginal effects of agricultural variables, are stronger in late agglomerator countries than in early agglomerator countries, this is consistent with the explanatory power results. We first estimate equation (8) in general form with a full set of interactions. Results, shown in Tables A4 and A5 for education and urbanization, respectively, generally show a pattern analogous to the explanatory power results. The (uninteracted) marginal effects of the agriculture variables are usually of the same sign as their interactions with the early agglomerator indicator, implying that they have a stronger effect in the early agglomerator countries. Conversely, the (uninteracted) marginal effects of the trade variables are usually of the opposite sign as their interactions with the early agglomerator indicator, implying that they have a weaker effect in the early agglomerator countries.

²⁰ For example, low-high education, low-high urbanization, and new-old world splits yield R²s of respectively 0.492, 0.483, and 0.488 with no country fixed effects.

To test this idea more formally, we consider the following equation:

$$\ln(\text{light}_{ic}) = X_{ic}^B \beta_B + X_{ic}^A \beta_A + X_{ic}^T \beta_T + D_{\text{split}_c} (\alpha X_{ic}^A \beta_A + \gamma X_{ic}^T \beta_T) + \epsilon_{ic} \quad (9)$$

where “B” refers to the 2 base covariates, “A” to agriculture, and “T” to trade. The common (constrained) deviation of effects for early agglomerators (where $D_{\text{split}_c} = 1$) are α and γ for the sets of agricultural and trade variables, respectively. Table 5 reports non-linear least squares estimates of equation (9). Across the board (split according to either education or urbanization, with or without fixed effects) patterns are similar. The coefficient α is positive and significant, meaning that marginal effects of agricultural variables are larger in absolute value for early agglomerating countries than for late agglomerating countries. For trade the opposite is the case. The coefficient γ is negative and significant, meaning that marginal effects of trade variables are smaller in absolute value for the former group than for the latter. Thus not only is agriculture relatively more important than trade in explaining lights variation for high compared to low education or urbanization countries, but marginal effects of agriculture compared to trade variables are relatively stronger for high versus low education or urbanization countries.

As a means of visualizing how the determinants of the location of agglomeration have changed over time, we examine the difference between fitted values generated using the estimates for early developing countries and those generated using estimates for late developers. We can generate both sets of these fitted values for every country, regardless of whether it developed early or late. The higher is this difference, the more that grid square is favored by the coefficients that governed early developers relative to those that govern late developers. In practice, this is equivalent to looking at the values of $X_{ic} \hat{\beta}_d$ in equation (8). This can be done either with or without country fixed effects.²¹ Figure 10 shows this difference in fitted values for Europe, Africa, and parts of west Asia, using the education split.

²¹ The actual estimates for equation (8) are reported in Table A4. For these figures we use education as the splitting variable between early and late developers. In the case of estimates using country fixed effects, these fixed effects are left out in forming fitted values. After forming the difference in fitted values, we normalize by subtracting the mean difference.

In Africa, for example, interior areas such as the Congo basin and the Ethiopian highlands would have had higher light density under the early development regime than under the late development regime (which is in fact what applied to them). And similarly, in Africa, the areas around navigable rivers, particularly the Nile and Niger, have higher densities than if the region had developed early. Within Europe, coastal areas, which of course already have particularly high density, would have had even higher density if Europe had developed late instead of early. It is also interesting to note that Europe has a predominantly negative values for the difference between predicted lights using early developer coefficients and predicted lights using late developer coefficients. This means that Europe is particularly rich in characteristics that favor population density in late developers, despite the fact that it developed early.

7. Spatial spillovers

Both the lights and the physical geography characteristics predicting them are highly spatially correlated. To the extent that this is manifested in spatially correlated errors, we have accounted for this by clustering errors and, alternatively, calculating Conley standard errors. However, direct agglomeration of lights and spillover effects of the X's, such as a natural harbor's impact on its hinterland, are also possible. We have tried to minimize the extent to which these affect our results by aggregating individual light pixels to much larger grid squares. While minimizing spillovers across cells, this essentially internalizes the agglomeration externalities. The estimated coefficients are thus reduced form, reflecting not just raw agricultural and trade effects but also endogenous agglomeration. For example, a natural harbor may represent a natural trade advantage, which induces clustering around it, but because of scale economies, the intensity of lights in a cell containing a harbor reflects both the basic harbor advantage and the induced agglomeration from scale economies focused on a harbor.

Separating these three phenomena (correlated errors, spillovers, and agglomeration) is notoriously difficult (e.g. Gibbons, Overman, and Patacchini 2015). One solution is to focus on the reduced form, adding as covariates the trade and agriculture determinants of neighbors' lights. Alternatively, one may try to uncover spillover effects from lights themselves (pure agglomeration), although it is hard to do counterfactuals with the results (by exogenously varying an endogenous variable). One way to estimate these effects, common in the literature, is to add neighbors' lights as a covariate and instrument for it, using second order neighbors' trade and agriculture determinants. This method relies on spillovers attenuating fully beyond immediate neighbors.

However, the physical geography characteristics we use are highly spatially correlated. The simple correlation coefficient between cells and their neighbors for each variable are given in Table 6, with neighbors defined on a rook basis (the N, S, E and W neighbors of a grid square sharing a finite border) and on a queen basis, which adds the NE, NW, SE and SW neighbors sharing a corner. In either case, fifteen of these have a correlation coefficient in excess of 0.958 and several are over 0.99. Given these correlations, we cannot credibly separate own and neighbor effects let alone instrument for them.

As an alternative, we experimented with imposing the structure of the spatial autoregressive model

$$\ln L_i = \rho W_i \ln L_i + X_i \beta + \epsilon_i \quad (10)$$

where the weights matrix W_i is 1 for queen neighbors and 0 for all other squares. Note in this we assume neighbors' X and ϵ have no effect on outcomes. We then estimate this model in the traditional fashion where

$$\ln L_i = (I - \rho W_i)^{-1} X_i \beta + \epsilon_i \quad (11)$$

Results are in Appendix Table A6. The coefficient of interest is ρ , the effect of neighbor's lights on own lights. The estimates for the full sample in columns 1 and 2 of Table A6 are 0.99 with rook contiguity and 1.001 with queen contiguity, implying extreme spillovers. This is in part a symptom of the huge patches of contiguous unlit areas. When we run the same specification on lit areas, the coefficient is still above 0.6, whereas the agglomeration literature suggests a value of 0.10 at most. Our attempts at instrumenting yielded similarly implausible results. We did not further pursue this line of investigation.

8. Conclusion

In this paper we have explored the role of natural characteristics in determining the locations of population agglomerations, with a focus on the within-country distribution of population.

Natural characteristics have a surprisingly high degree of overall explanatory power, but when we divide these natural characteristics into those associated with agricultural productivity and those associated with ease of trade, a puzzle emerges: agricultural characteristics have relatively more explanatory power in countries that developed early, despite the fact that these countries, which tend to be wealthy, have much smaller agricultural sectors than countries which developed later. We see the resolution of this puzzle in the intersection of three forces. The first is persistence, the strong tendency for spatial patterns of agglomeration, once established, to remain in place. The second is the changing weights on different characteristics natural as economies develop. The two most important changes, in our view, are a reduction in the weight of characteristics associated with agricultural productivity and an increase in the weight of characteristics associated with trade. Finally, the third force is that early- and late-developing countries experienced changes in the weights associated with sets of natural characteristics in a different order. In today's developed countries the process of agglomeration and structural transformation began early, when transport costs were still relatively high, so urban agglomerations were more localized in different agricultural regions. High costs of trade protected local markets. In later developing countries, transport costs fell well before structural

transformation started. To exploit urban scale economies before structural transformation with a limited national urban labor force, manufacturing tended to agglomerate in relatively few, often coastal, locations. With structural transformation these initial coastal locations grew, without formation of more cities in the agricultural interior. Thus agricultural fundamentals drove the location of economic activity in developed countries, while cost of trade fundamentals played a much bigger role in developing countries.

References

- Allen, Treb and Costas Arkolakis. 2014. “Trade and the Topography of the Spatial Economy.” *Quarterly Journal of Economics* 129(3): 1085-1140.
- Barro, Robert and Jong-Wha Lee. 2010. “A New Data Set of Educational Attainment in the World, 1950-2010.” *Journal of Development Economics*, 104: 184-198, April.
- Bleakley, Hoyt and Jeffrey Lin. 2012. “Portage and Path Dependence.” *Quarterly Journal of Economics* 127: 587-644.
- Broadberry, Stephen and O’Rourke, Kevin H. (2010) The *Cambridge Economic History of Modern Europe: Volume 1, 1700-1870*. Cambridge: Cambridge University Press.
- Buringh, Eltjo and Jan Luiten Van Zanden (2009). Charting the “Rise of the West”: Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries. *The Journal of Economic History*, 69, pp 409-445.
- Central Intelligence Agency. 2013. The World Factbook 14. Washington: Central Intelligence Agency. <https://www.cia.gov/library/publications/the-world-factbook/> Accessed 9 August 2015
- Chen, Xi and William Nordhaus. 2011. “Using luminosity data as a proxy for economic statistics.” *Proceedings of the National Academy of Sciences* 108(21): 8589-8594.
- Center for International Earth Science Information Network - CIESIN - Columbia University, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4XK8CG2>.
- Conley, Timothy G. 2008. Spatial Econometrics. *The New Palgrave Dictionary of Economics*, Second Edition, Eds .Steven N. Durlauf and Lawrence E. Blume.
- Costinot, Arnaud, Dave Donaldson and Cory Smith. 2016. “Evolving Comparative Advantage and the Impact of Climate Change on Agricultural Markets: Evidence from 1.7 Million Fields Around the World.” *Journal of Political Economy* 124(1): 205-248, February.
- Cronon W. 1991. *Nature’s Metropolis: Chicago and the Great West*. New York: W.W. Norton.

- Davis, Donald R. and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review* 92(5): 1269-1289, December.
- Desmet Klaus and J.V. Henderson 2015. The geography of development within countries. In G. Duranton, J. V. Henderson and W. Strange (eds.), *Handbook of Regional and Urban Economics*, 5.
- Donaldson, Dave (forthcoming) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*.
- Durlauf, Steven N & Johnson, Paul A, 1995. "Multiple Regimes and Cross-Country Growth Behaviour," *Journal of Applied Econometrics*, 10(4): 365-384, Oct.-Dec.
- Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W, Davis, E.R. 1997. "Mapping of city lights using DMSP Operational Linescan System data." *Photogrammetric Engineering and Remote Sensing* 63: 727-734.
- Elvidge, C.D., Baugh, K.E., Dietz, J., Bland, T., Sutton, P.C., Kroehl, H.W. 1999. "Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements." *Remote Sensing of Environment* 68(1): 77-88.
- FAO/IIASA, 2011. Global Agro-ecological Zones (GAEZ v3.0). FAO Rome, Italy and IIASA, Laxenburg, Austria. <http://gaez.fao.org>
- Gollin, Douglas, and Richard Rogerson. 2014. "Productivity, transport costs and subsistence agriculture." *Journal of Development Economics* 107: 38-48, March.
- Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics* 128(1): 105-164.
- Gennaioli, Nicola, Rafael La Porta, Florencio Lopez De Silanes, and Andrei Shleifer. 2014. "Growth in Regions." *Journal of Economic Growth* 19: 259-309
- Gibbons, S., H.G.Overman and E. Patacchini. 2015. Spatial methods. In G. Duranton, J. V. Henderson and W. Strange (eds.) *Handbook of Urban and Regional Economics* Volume 5.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102(2): 994-1028.

- ISciences LLC (2008) Elevation and Depth v2 2000 (dataset)
<http://geoserver.isciences.com:8080/geonetwork/srv/en/resources.get?id=208&fname=ElevationGridv2.zip&access=private>
- Jedwab, Remi and Alexander Moradi (forthcoming) “The Permanent Economic Effects of Transportation Revolutions in Poor Countries: Evidence from Africa,” *Review of Economics and Statistics*.
- Jedwab, Remi, Edward Kerby and Alexander Moradi (forthcoming) History, Path Dependence and Development: Evidence from Colonial Railroads, Settlers and Cities in Kenya. *The Economic Journal*.
- Klein Goldewijk, Kees, Arthur Beusen, Gerard Van Drecht, Martine De Vos. 2011. “The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years” *Global Ecology & Biogeography* 20(1):73–86, January.
- Krugman, Paul. 1991. “First Nature, Second Nature, and Metropolitan Location” NBER Working Paper 3740.
- Lehner, B. and Döll, P. 2004. “Development and validation of a global database of lakes, reservoirs and wetlands.” *Journal of Hydrology* 296(1-4): 1-22.
- Mellinger, Andrew, Jeffrey D. Sachs, and John Luke Gallup. 2000. “Climate, Coastal Proximity, and Development,” in Clark, Gordon L., Maryann P. Feldman, and Meric S. Gertler, eds., *The Oxford Handbook of Economic Geography*, 169-194.
- Mesbah J. Motamed, Raymond J.G.M. Florax and William A. Masters. 2014. “Agriculture, Transportation and the Timing of Urbanization: Global Analysis at the Grid Cell Level.” *Journal of Economic Growth* 19(3): 339-368, September.
- Michaels, Guy and Ferdinand Rauch. 2013. “Resetting the Urban Network: 117-2012.” BREAD Working Paper 405, November.
- Ministry of Human Resource Development. 1987. A Hand Book of Educational and Allied Statistics. New Delhi.
- Mitchell, Timothy D. and Philip D. Jones. 2005. “An improved method of constructing a database of monthly climate observations and associated high-resolution grids.” *International Journal of Climatology* 25(6): 693–712.

- Mohammed, S.I.S. and J.G. Williamson 2004. Freight rates and productivity gains in British tramp shipping 1869–1950. *Explorations in Economic History* 41: 172–203
- Natural Earth. 2012. Natural Earth version 2.0.0. <http://www.naturalearthdata.com/>, Accessed 2013.
- NOAA National Geophysical Data Center (2011). Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG) Version 2.2.0.
<http://www.ngdc.noaa.gov/mgg/shorelines/shorelines.html>.
- Nordhaus, William. 2006. “Geography and Macroeconomics: New Data and New Findings,” *Proceedings of the National Academy of Sciences* 103(10): 3510-3517, March.
- Nordhaus, William and Xi Chen. 2009. “Geography: Graphics and Economics.” *The B.E. Journal of Economic Analysis and Policy* 9(2): 1-12.
- Nunn, Nathan and Diego Puga. 2012. “Ruggedness: The Blessing of Bad Geography in Africa.” *The Review of Economics and Statistics* 94(1): 20-36.
- Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'Amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, Y. Kura, J.F. Lamoreux, W.W. Wettenberg, P. Hedao, and K.R. Kassem. 2001. “Terrestrial ecoregions of the world: A new map of life on earth.” *Bioscience* 51(11): 933-938.
- Ramankutty, Navin, Jonathan A. Foley, John Norman and Kevin McSweeney. 2002. “The global distribution of cultivable lands: current patterns and sensitivity to possible climate change.” *Global Ecology and Biogeography* 11(5): 377-392.
- Rappaport, Jordan and Jeffrey Sachs. 2003. “The United States as a Coastal Nation.” *Journal of Economic Growth* 8: 5-46.
- Rauch J. 1993. “Does History Matter Only When it Matters Little?” *Quarterly Journal of Economics* 108(3): 843-67
- Stevens, Forrest R., Andrea E. Gaughan, Catherine Linard, and Andrew J. Tatem. 2015 “Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data” *PLoS One* 10(2): e0107042.
- UNESCO. 1957. *World Illiteracy at Mid-Century: A Statistical Study*. Paris, UNESCO

United States Navy (1953) *World Port Index 1953*. Hydrographic Office, US Navy
University of East Anglia Climatic Research Unit; Jones, P.D.; Harris, I. (2013): CRU TS3.20:
Climatic Research Unit (CRU) Time-Series (TS) Version 3.20 of High Resolution Gridded
Data of Month-by-month Variation in Climate (Jan. 1901 - Dec. 2011). NCAS British
Atmospheric Data Centre.

<http://catalogue.ceda.ac.uk/uuid/2949a8a25b375c9e323c53f6b6cb2a3a>

United Nations. 2014. *World Urbanization Prospects, the 2014 revision*. New York: United
Nations.

Wessel, P., and W. H. F. Smith. 1996. "A Global Self-consistent, Hierarchical, High-resolution
Shoreline Database." *Journal of Geophysical Research* 101(B4): 8741-8743.

Willmott, C. J. and K. Matsuura (2012) Terrestrial Precipitation: 1900-2010 Gridded Monthly
Time Series (V 3.02),
http://climate.geog.udel.edu/~climate/html_pages/Global2011/Precip_revised_3.02/README.GlobalTsP2011.html. Accessed 25 January 2013

World Bank. 2015. World Development Indicators,
<http://data.worldbank.org/data-catalog/world-development-indicators> Accessed 9 August
2015.

Ziskin, Daniel, Kimberly Baugh, Feng Chi Hsu, Tilottama Ghosh, Chris Elvidge. 2010.
"Methods Used For the 2006 Radiance Lights." *Proceedings of the 30th Asia-Pacific
Advanced Network Meeting* 131-142.

Appendix. Algorithm to solve the model

The algorithm begins by creating a vector of all possible interior populations and a corresponding vector of all possible coastal populations based on L and L_i such that:

$$L_c = L - L_i$$

To find equilibria, we cycle through these vectors in a loop. As a result, the following steps are carried out for a fixed population allocation between the interior and coast.

First, we create another vector of all possible L_{fi} values, ranging from 0 (no agriculture in the interior) to the entire interior population (everyone is employed in agriculture in the interior). Then, a corresponding vector is created of L_{fc} values. This vector is calculated based on food needs of the entire population, solving the following equation based on the text:

$$L_{fc} = \left(\frac{L\gamma - A_f L_{fi}^{1-\beta}}{A_{fc}} \right)^{\frac{1}{1-\beta}}$$

We subsequently cycle through these vectors in another loop, nested within the previous one. Consequently, the following steps are carried out for fixed regional agricultural labor forces *and* regional population allocations.

Within these two loops, we begin to find equilibria. If $L_{fc} < L_c$ and $L_{fi} < L_i$ (both regions have some manufacturing labor force), then we calculate L_{mc} and L_{mi} using the following equation:

$$L_{mr} = L_r - L_{fr}$$

Now that we have L_i , L_c , L_{fi} , L_{fc} , L_{mi} , and L_{mc} , we calculate prices in each region based on the average product of agriculture and manufacturing in each region (so that wages are equalized across sectors within each region):

$$p_{mr} = \frac{A_f L_{fr}^{-\beta}}{A_{mr} L_{mr}^{\epsilon} (1-t L_{mr})^{1+\epsilon}}$$

Next, we determine which of the two regions is exporting manufactured goods. This can be determined by checking which of the two regions produces less food than its population requires. Then, we check if the inter-regional goods market clears by checking if prices in the exporting region are equal to prices in the importing region, adjusted for the iceberg trade cost τ . For most

allocations of L_{fc} and L_{fi} this condition is not met, and the algorithm simply ends at this point and starts at the next allocation of L_{fc} and L_{fi} .

However, if this condition is met, manufacturing consumption per capita is calculated for each region. In the exporting region, manufacturing consumption is calculated by subtracting the quantity of manufactured goods that are exported from the total quantity of manufactured goods produced in the region, divided by the region's population. The quantity of exported manufactured goods is determined utilizing the fact that the inter-regional goods market clears. As a result, exported manufactured goods necessarily equals the quantity of imported food divided by the price of manufactured goods in the region. The quantity of imported food is determined by the gap in the region's food needs and food production in the region.

In the importing region, manufacturing consumption per capita is equal to total manufactured goods produced plus the quantity of imported manufactured good, divided by regional population. Analogously to the previous case, the value of imported manufactured goods is determined by the quantity of exported food divided by price of manufactured goods in that region.

Manufacturing consumption in each region is not necessarily equal at this point. As such, this data point is recorded as a “possible equilibrium,” where every equilibrium condition is met *except* that manufacturing consumption is equal across regions. If manufacturing consumption is also equal across regions, then this data point is recorded as an “equilibrium.”

If $L_{fc} = L_c$ or $L_{fi} = L_i$, then we have a corner solution where one region has no manufacturing labor force. In this case, L_{mc} and L_{mi} are calculated just like before. Prices in the region that has a manufacturing labor force are calculated using the average products of agriculture and manufacturing just like above. However, prices in the region that has no manufacturing labor force are now determined solely by adjusting the other region's prices by the iceberg trade cost. Next, manufacturing consumption per capita is calculated for each region. The region that has a manufacturing labor force obviously exports manufactured goods in this case. Manufacturing consumption per capita in this region is equal to total manufactured goods minus exported manufactured goods (determined just as before) divided by regional population.

In the region with no manufacturing labor force, manufacturing consumption is just equal to imported manufactured goods divided by regional population.

This data point is recorded as a possible equilibrium. If manufacturing consumption is also equal across regions, then this data point is recorded as an equilibrium. This ends the loop through possible values of L_{fc} and L_{fi} .

Next we address the endogenous no-trade equilibria where, as the name might imply, there is no trade between the two regions. First we check if the current (fixed) population allocation is feasible in that each region can feed itself without any trade. Then, we calculate L_{fc} and L_{fi} based on each region's individual food needs, remembering that there is no trade between regions. As a result:

$$L_{fi} = \left(\frac{L_i \gamma}{A_f} \right)^{\frac{1}{1-\beta}}$$

L_{mc} and L_{mi} are then calculated using regional population and agricultural labor force. Prices for each region are also calculated based on the average product of agriculture and manufacturing so that wages are equalized across sectors in a region. Next, we check if prices are in the “no-trade band” where no amount of trade is profitable, or:

$$\frac{p_{mi}}{1-\tau} \geq p_{mc} \text{ and } \frac{p_{mc}}{1-\tau} \geq p_{mi}$$

If this condition is met, then there is no incentive for trade between regions. We then calculate manufacturing consumption per capita in each region as total manufactured goods divided by regional population. Since the regional populations are still fixed (i.e. no mobility between regions), manufacturing consumption is not necessarily equal across regions. This data point is recorded as a possible equilibrium; every equilibrium condition is met *except* that manufacturing consumption is equal across regions. If manufacturing consumption is also equal across regions, then this data point is recorded as an equilibrium. This ends the loop through possible values of L_i .

This ends the procedure for calculating equilibria for a set of parameters. Next, we check the stability of all “full equilibria,” where all markets clear and manufacturing consumption is equal across regions. We define two types of stability. “Type 1 stability” occurs when there is no

incentive to move between regions. “Type 2 stability” occurs when there is no incentive to move industries *within* regions (i.e. move from agriculture to manufacturing).

To check type 1 stability, we take 100 people from the coast and move them to the interior. We allow for all other markets to clear, but manufacturing consumption is not equal between regions (i.e. calculate the resulting “partial equilibrium”). If the people who moved have lower consumption than before, then the equilibrium passes the stability check.

To check type 2 stability, we take 100 people from agriculture and move them into manufacturing *in a single region*, which puts the labor market in that particular region in disequilibrium. We then calculate the agricultural labor force in the other region based on the food needs of the entire population. We hold regional populations constant, so manufacturing labor forces are determined by regional population minus agricultural labor force. Next, since we still allow the labor market in the other region (where people did not initially change sectors) to be in equilibrium, we calculate prices in that region using the average product of manufacturing and agriculture like in earlier steps. Prices in the region where the labor market is in disequilibrium are then determined by prices in the other region adjusted for the iceberg trade cost. Wages in agriculture and manufacturing are finally calculated for that region. If the manufacturing consumption of manufacturing workers rises above that of their neighbors in agriculture within the region, then the equilibrium fails the stability check. The stability test is passed if manufacturing consumption of manufacturing workers in, for example, the interior is below food workers in the interior. We note that under stability in this experiment in all examples, food workers in the interior have the highest consumption nationally and manufacturing interior workers the lowest, and vice versa for instability.

Figure 1. Global nighttime lights in 2010

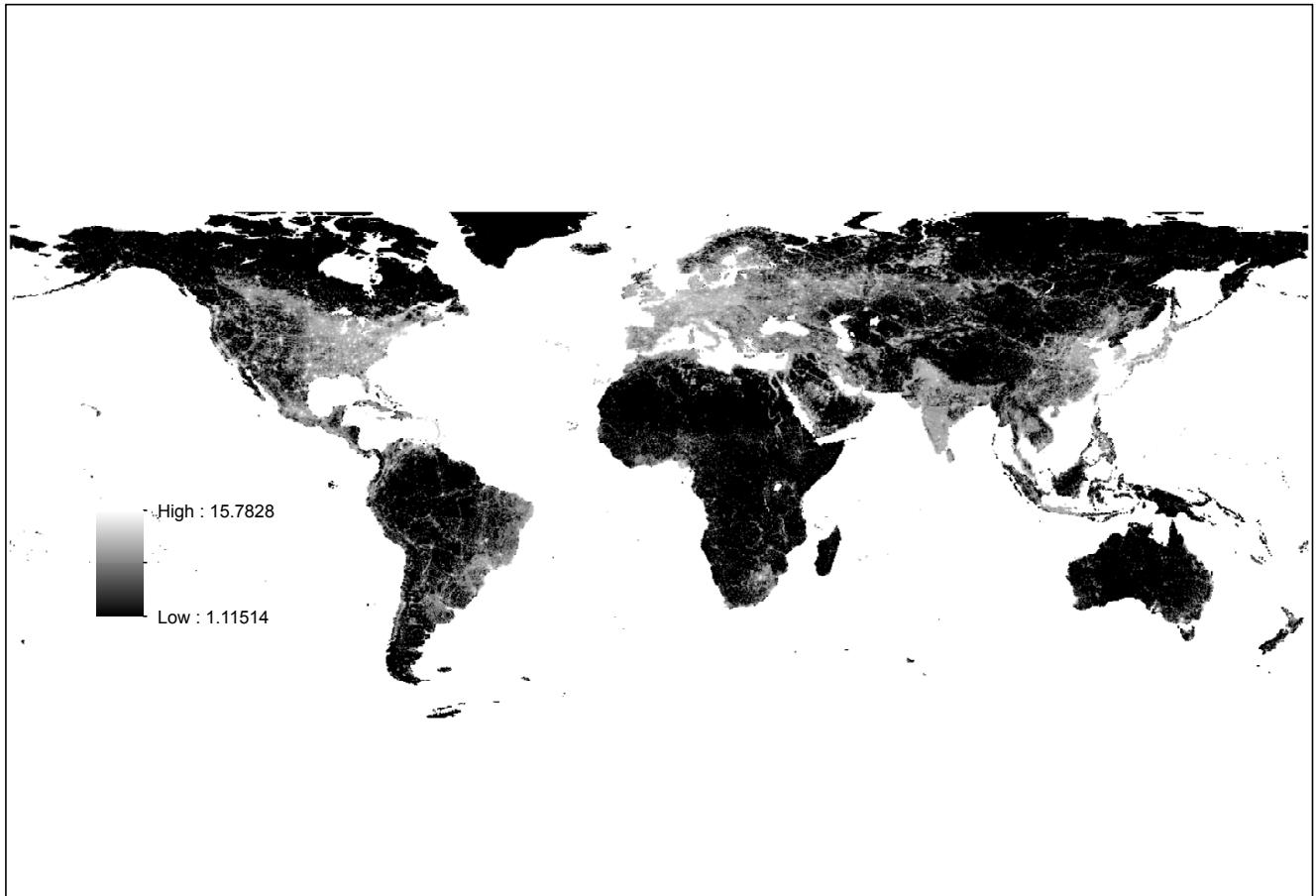


Figure 2. Distribution of lights in lit areas

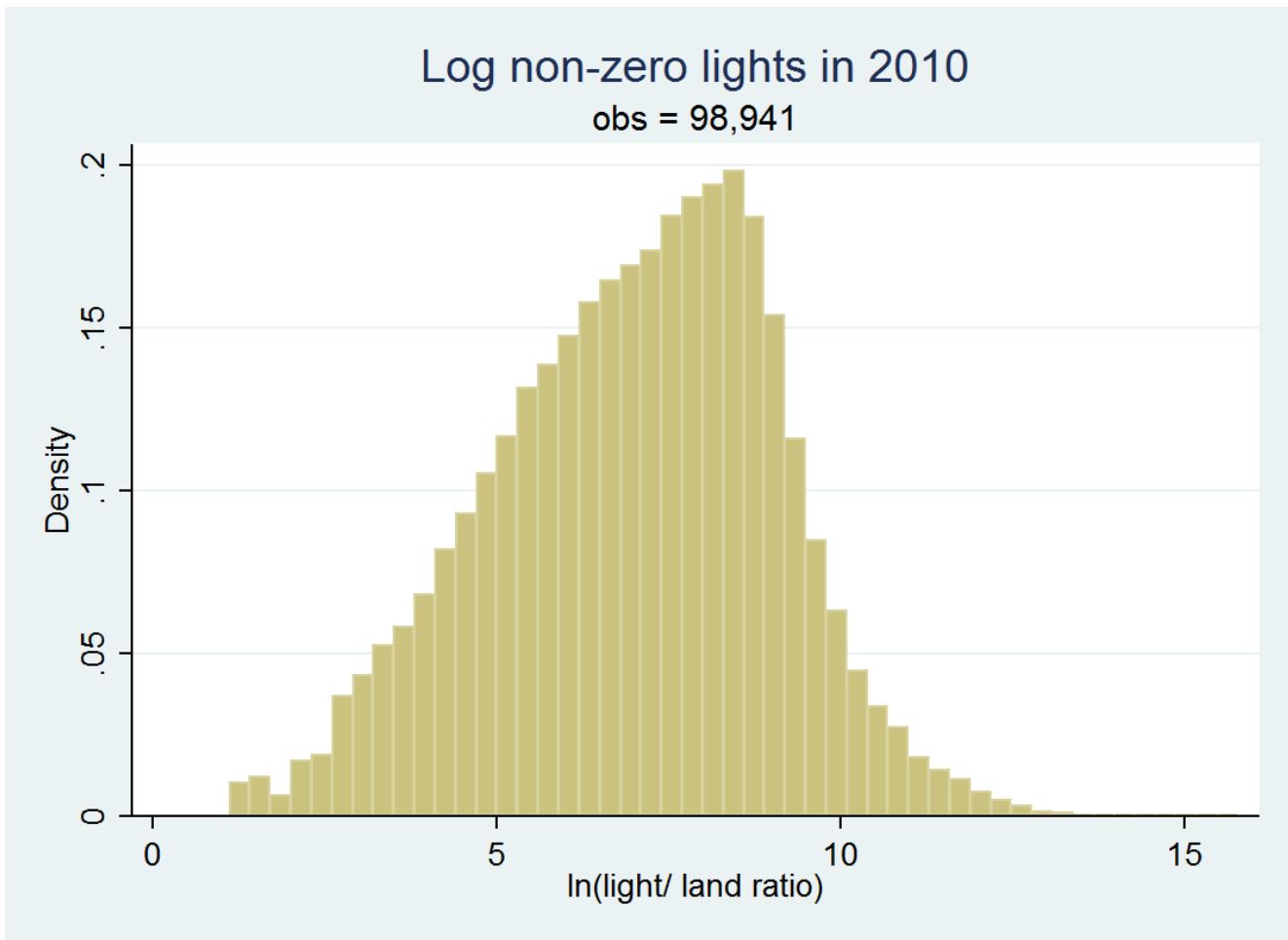
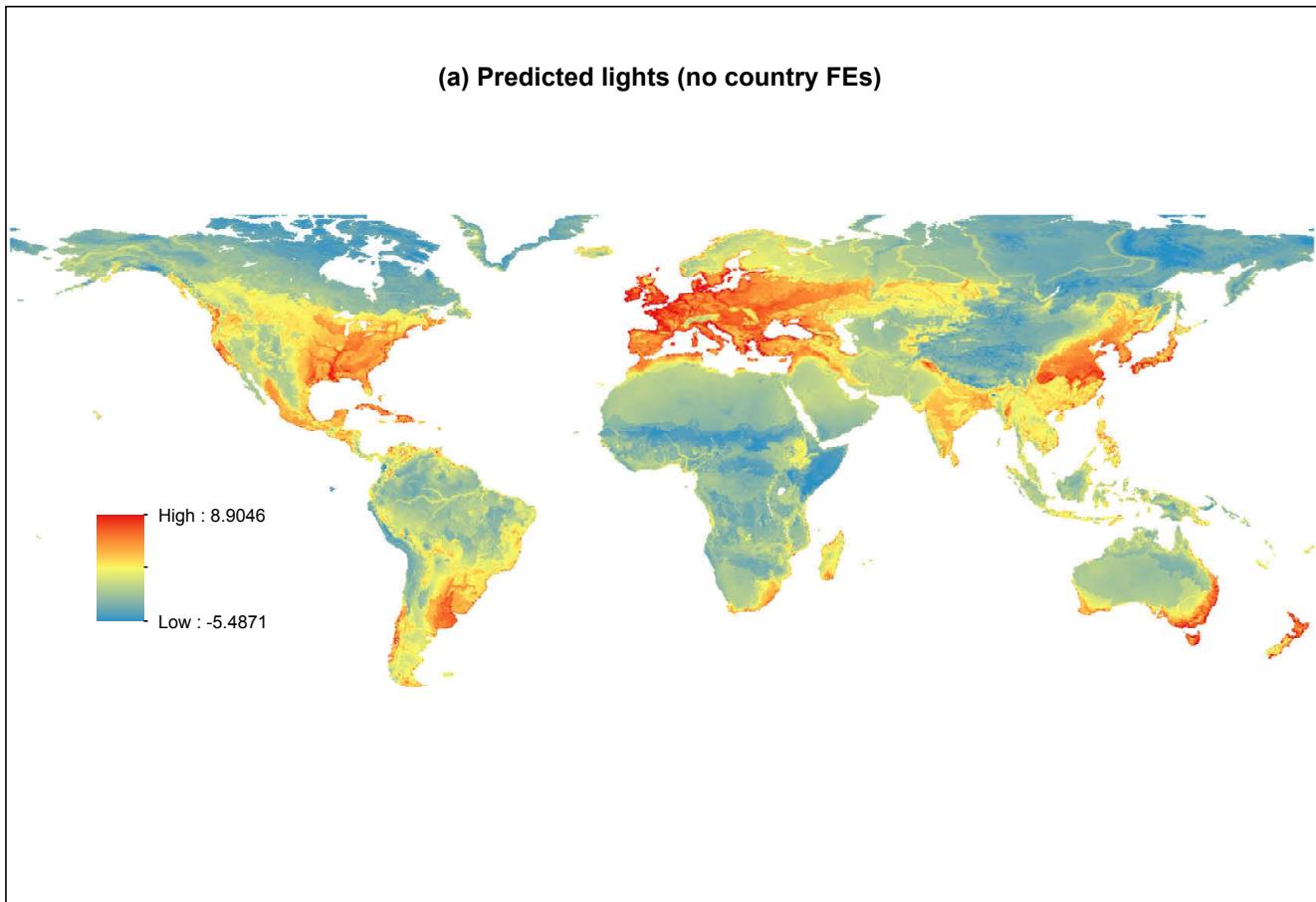


Figure 3. Worldwide predicted lights



(b) Predicted lights (with country FEs)

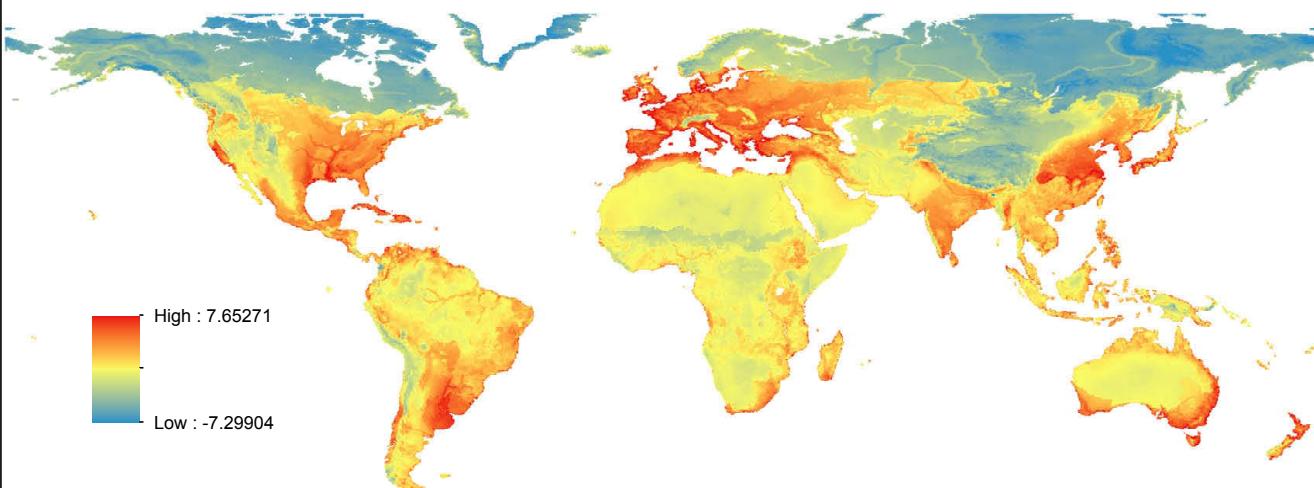


Figure 4. Equilibria with low agricultural productivity ($A_f = 1$)

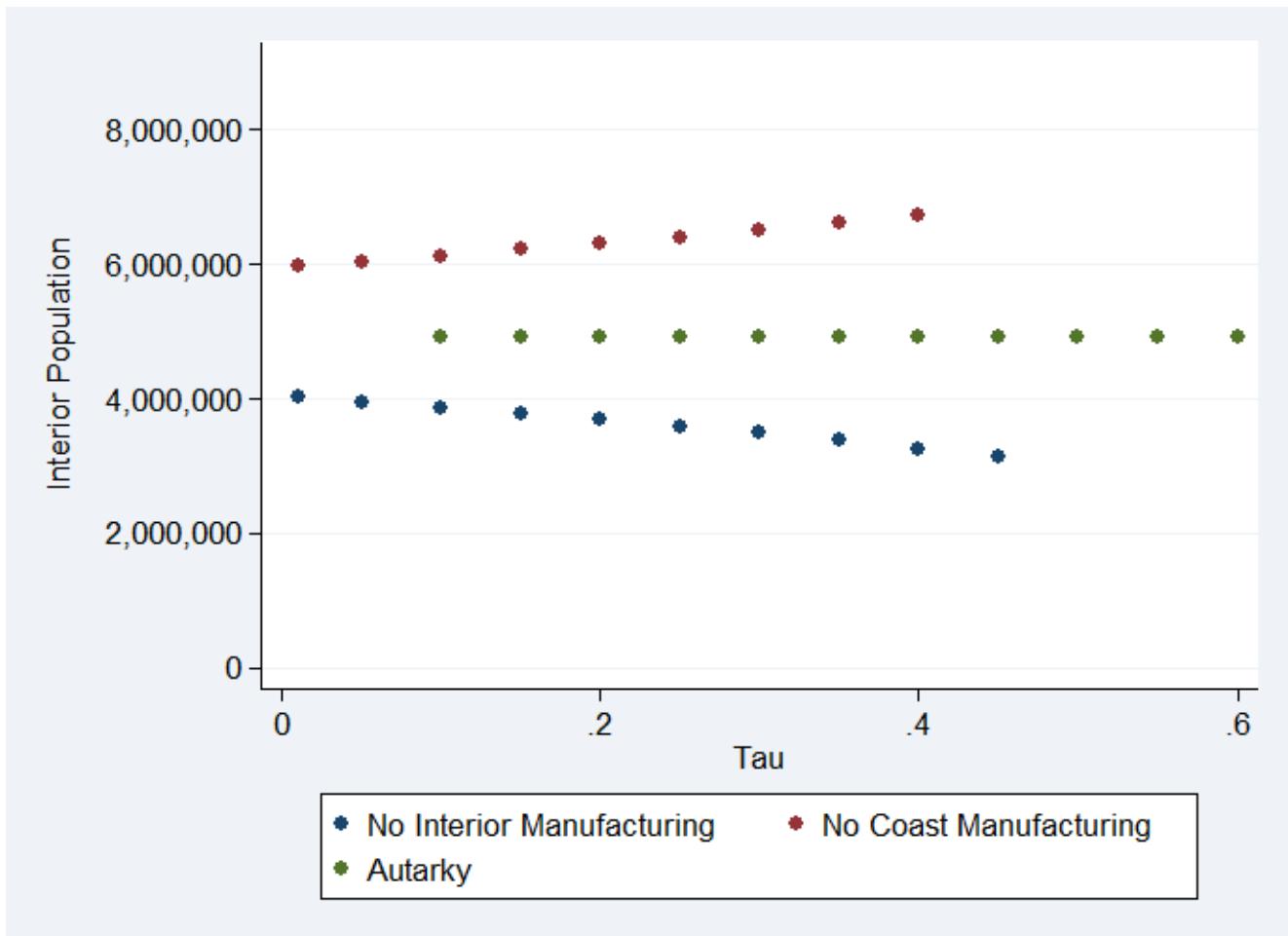


Figure 5. Equilibria with high agricultural productivity ($A_f = 1.5$)

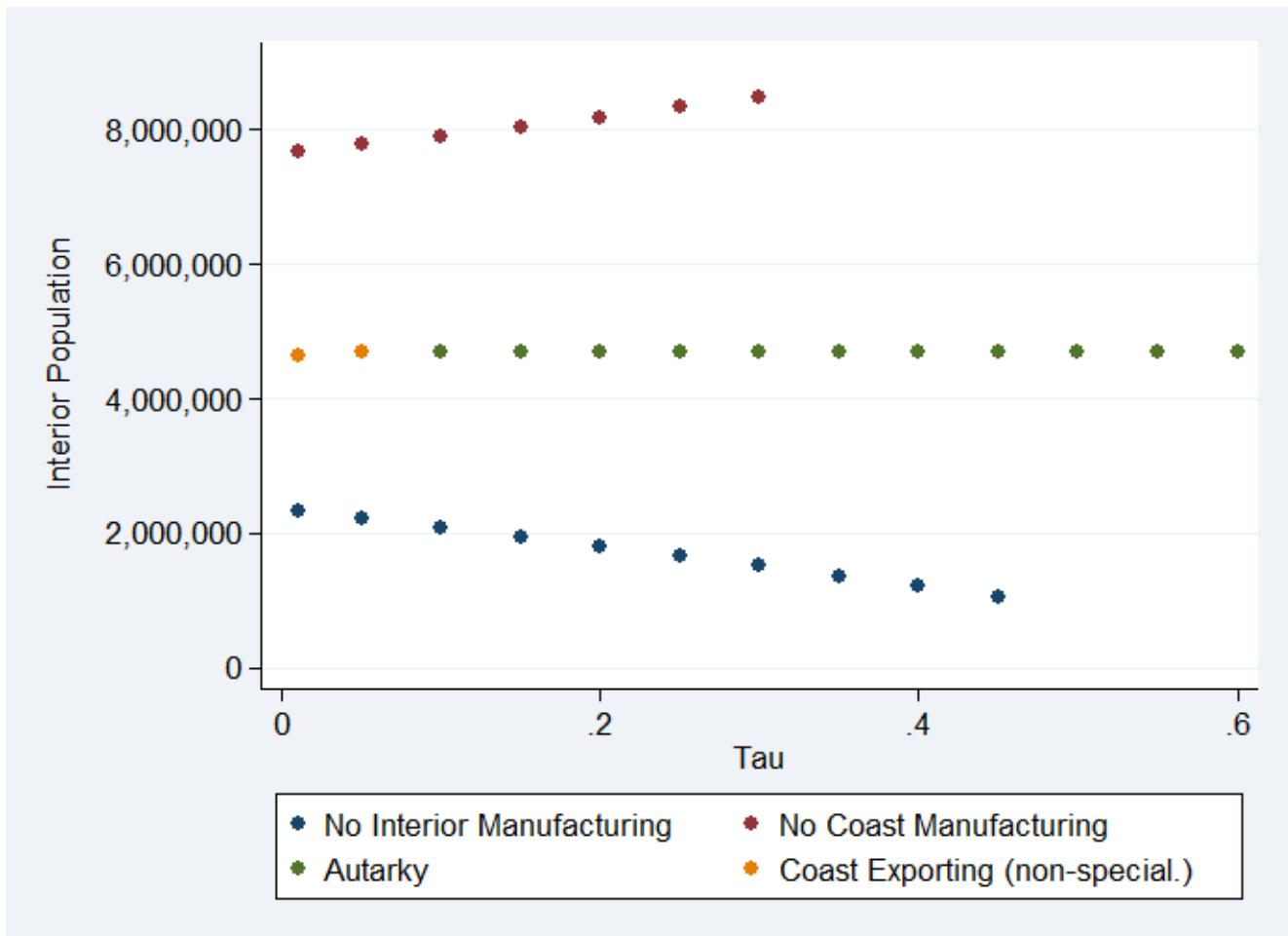
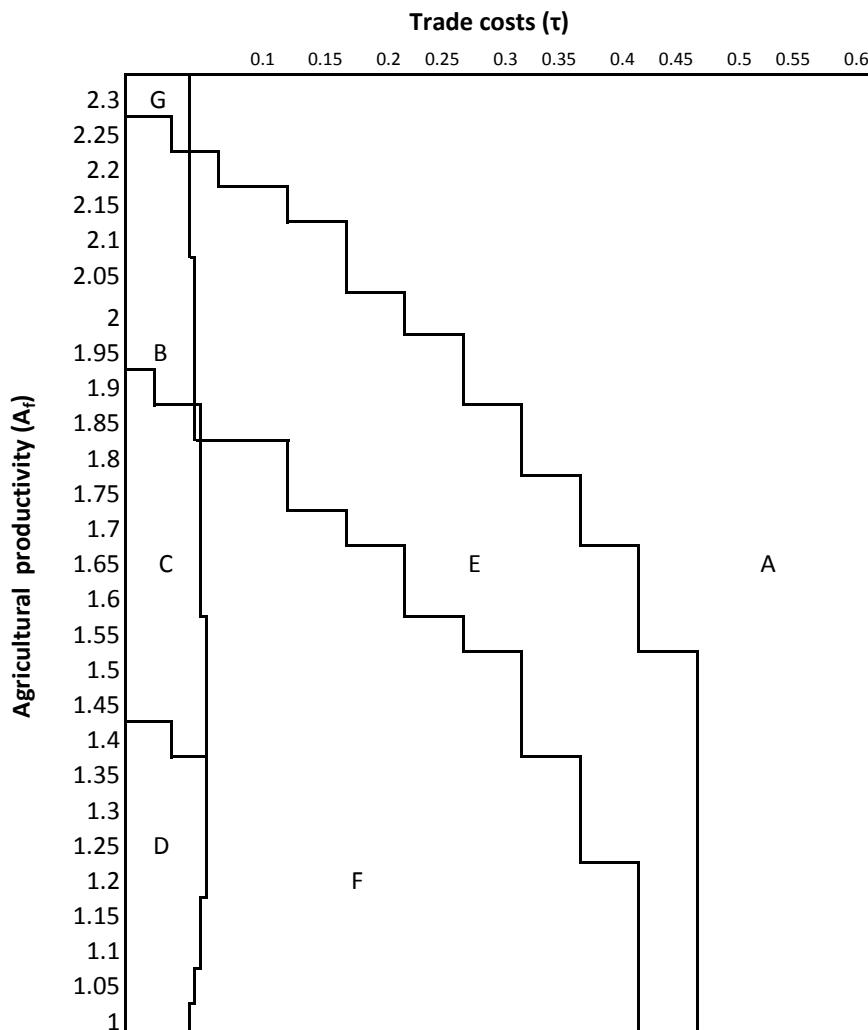


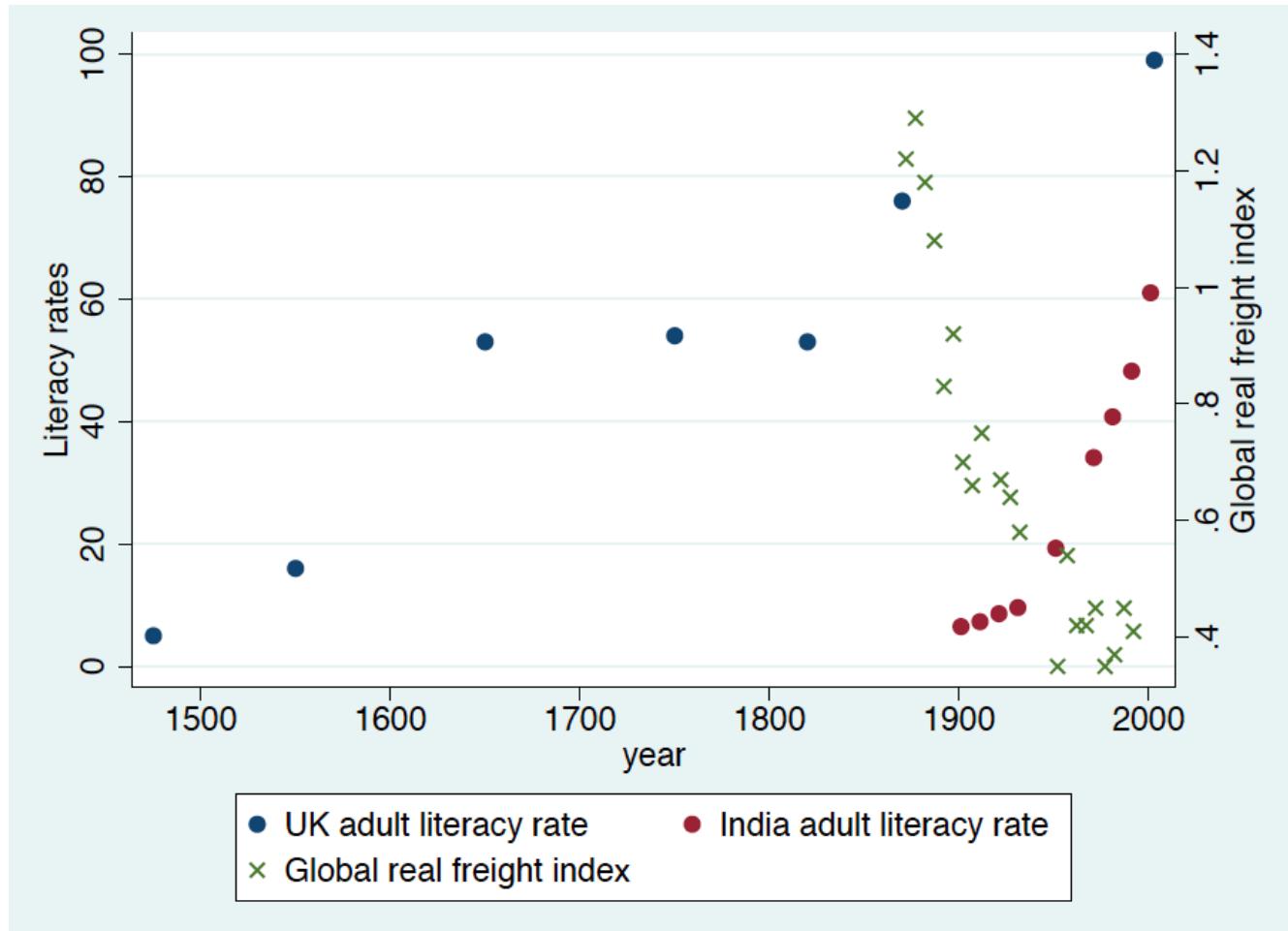
Figure 6. Equilibria in Agricultural productivity-transport costs ($A_f - \tau$) phase space



Key (number of equilibria in bold)

- | | |
|---|---|
| A | 1 : autarky |
| B | 2 : trade with manufacturing in both regions or solely in coast |
| C | 3 : trade with manufacturing in both regions, solely in coast, or solely in interior |
| D | 2 : trade with manufacturing solely in coast or solely in interior |
| E | 2 : autarky or manufacturing solely on coast |
| F | 3 : autarky or trade with manufacturing solely in coast or solely in interior |
| G | 1 : trade with manufacturing in both regions |

Figure 7. Timing of transport cost changes versus changes in effective technology



Note: Freight rates exclude periods including world war years. Sources: Freight rates: Mohammed and Williamson (2004)' UK Literacy 1475-1750 (midpoints of reported periods 1451-1500, 1501-1600, 1601-1700, 1701-1800): Broadberry and O'Rourke (2010), 1820-1870: Buringh and Van Zanden (2009), 2003: CIA (2015), India: 1901-1951: UNESCO (1957), 1971: Ministry of Human Resource Development (1987), 1981-2001: World Bank (2015).

Figure 8. Cumulative distribution of education and urbanization in 1950

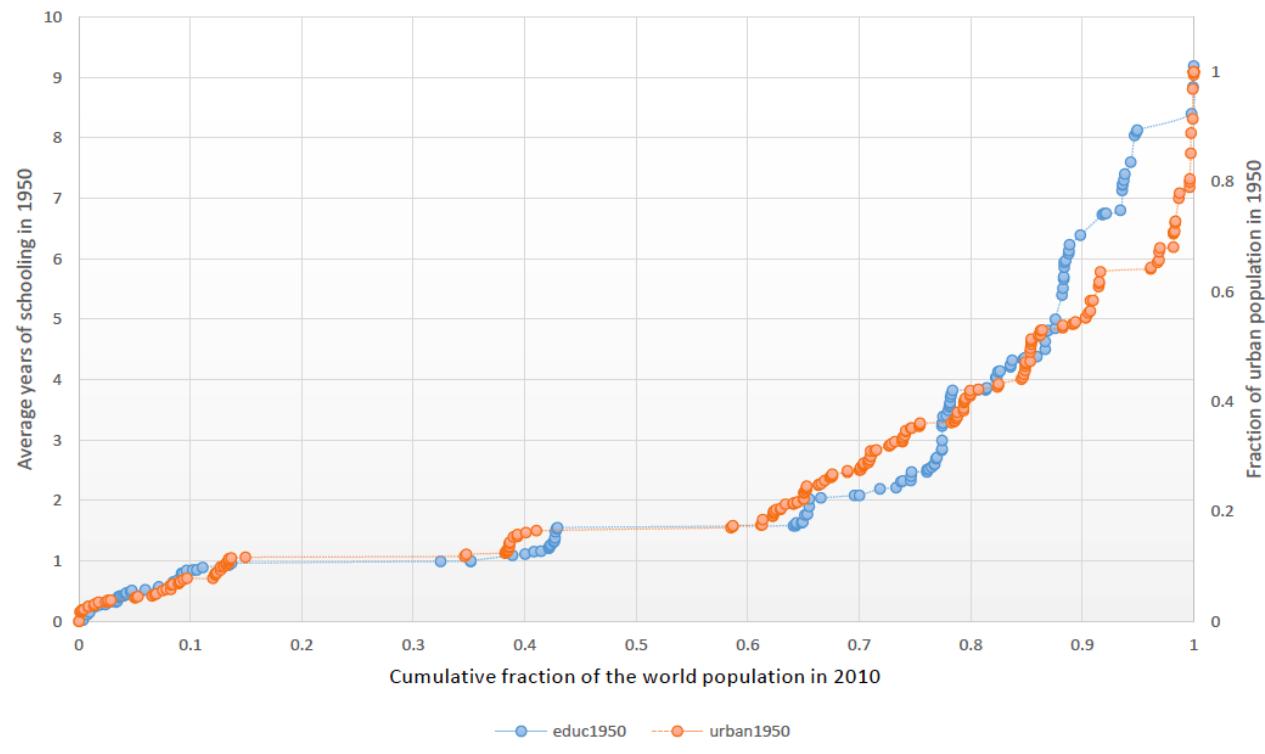
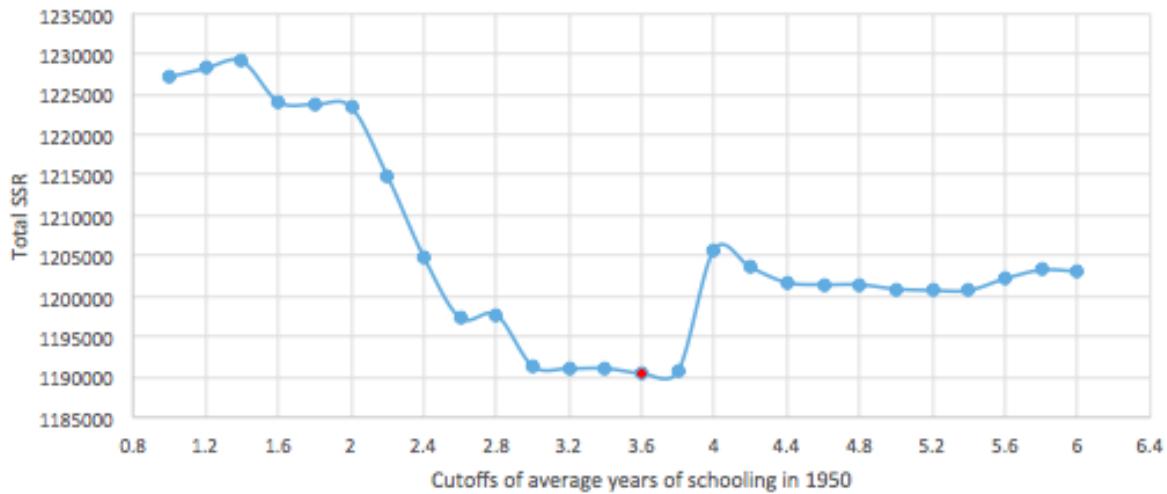
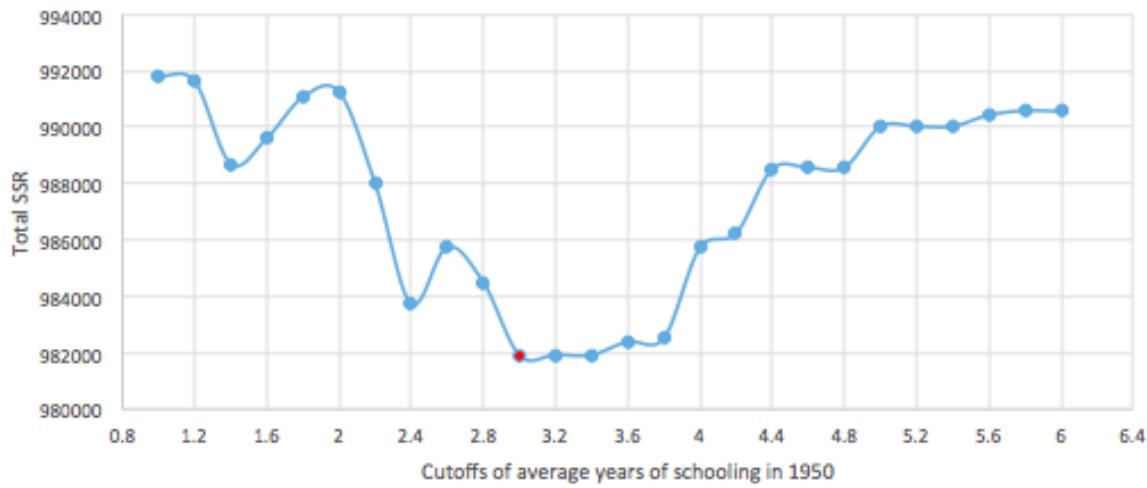


Figure 9. Cutoffs to minimize sum of squared residuals

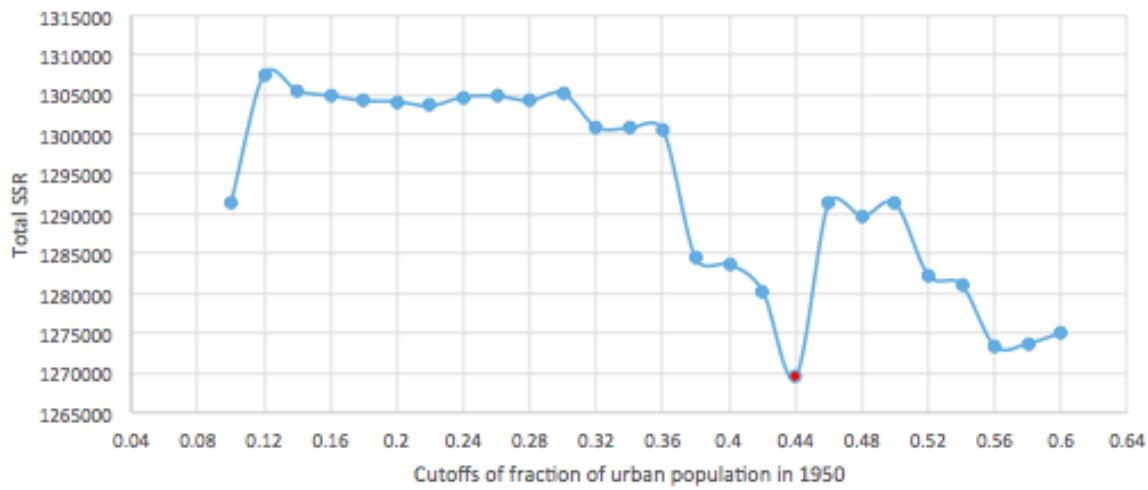
(a) Education: Total SSR (no country FEs)



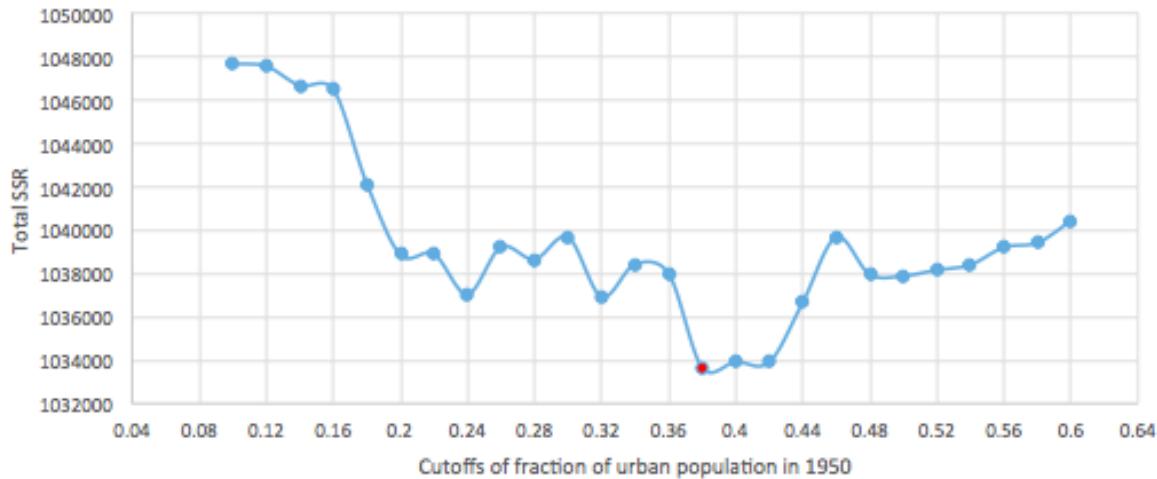
(b) Education: Total SSR (with country FEs)



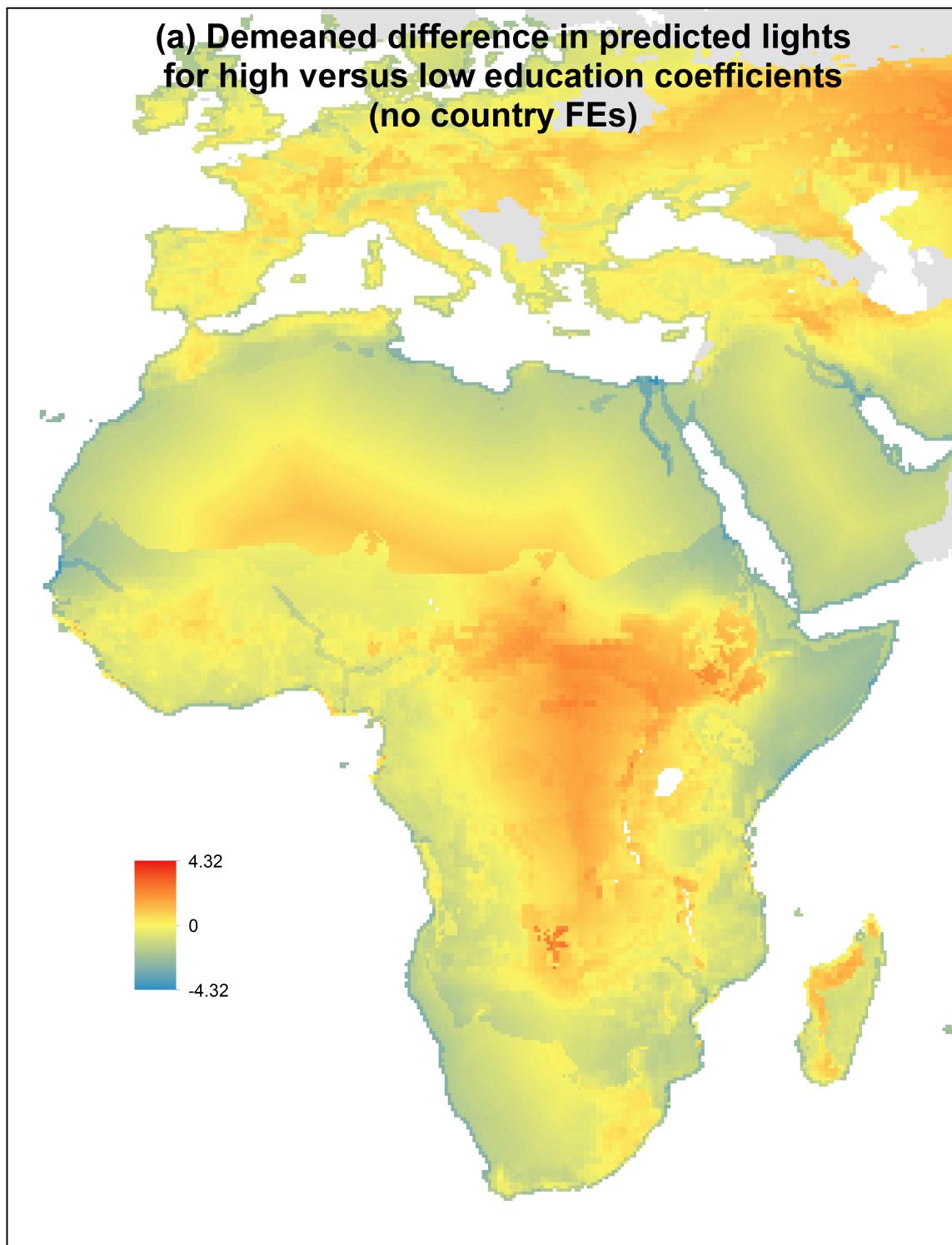
(c) Urbanization: Total SSR (no country FEs)



(d) Urbanization: Total SSR (with country FEs)

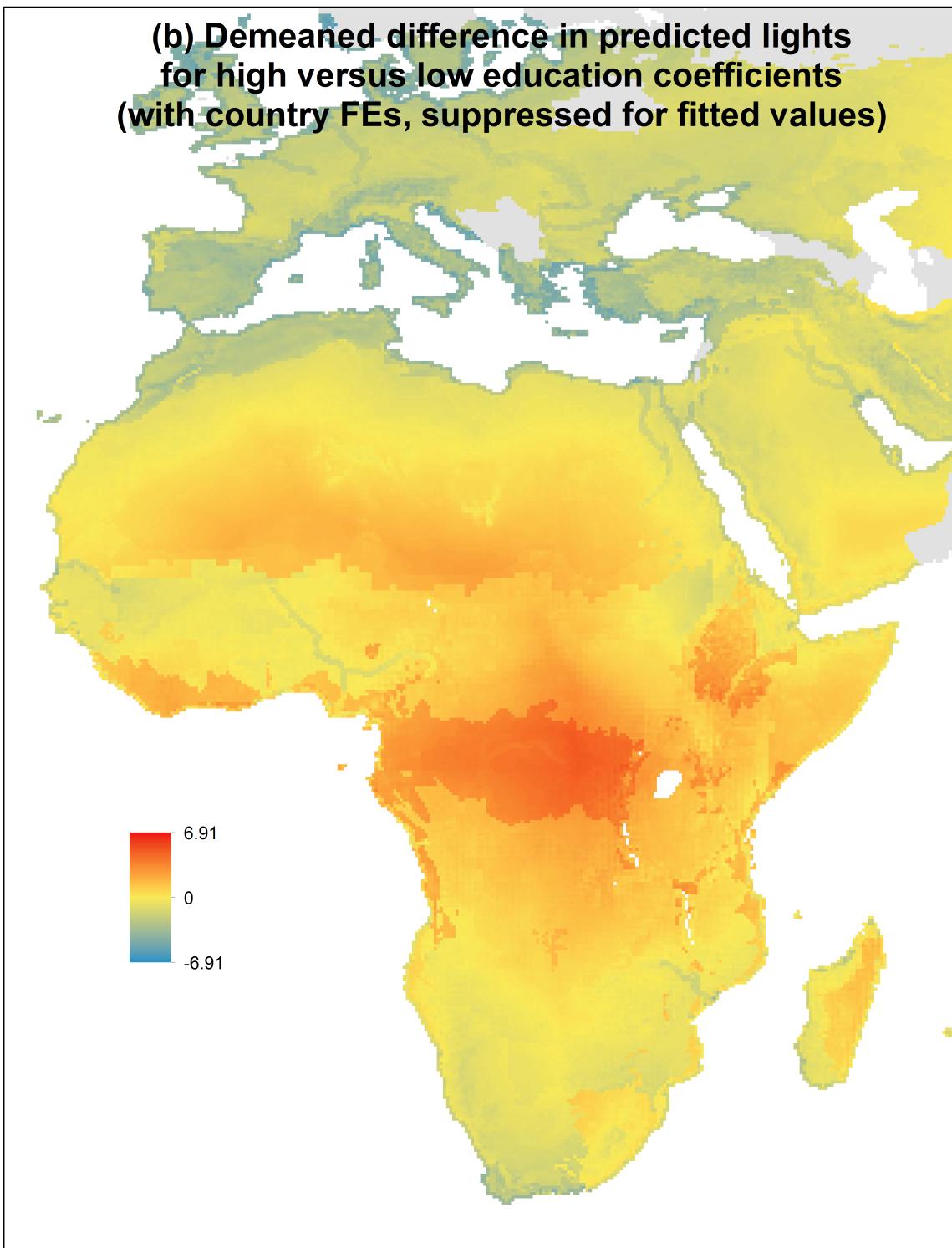


**Figure 10. Changes from having or being given high education coefficients
(over low education ones)**



Note: While 1950 education levels are unavailable for the following modern-day countries, for the purposes of this map we assume that they are below 3.6 years: Angola, Burkina Faso, Chad, Comoros, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Guinea, Guinea-Bissau, Madagascar, Nigeria, São Tomé and Príncipe, and Somalia.

**(b) Demeaned difference in predicted lights
for high versus low education coefficients
(with country FEs, suppressed for fitted values)**



Note: While 1950 education levels are unavailable for the following modern-day countries, for the purposes of this map we assume that they are below 3 years: Angola, Burkina Faso, Chad, Comoros, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Guinea, Guinea-Bissau, Madagascar, Nigeria, São Tomé and Príncipe, and Somalia.

Table 1: Summary Statistics and Baseline Regression Results

	Summary Statistics		Baseline Regression Results	
	mean, (sd) (1)	min, max (2)	w/o country FEs (3)	w/ country FEs (4)
Dependent variable				
ln(light/area)	3.501 (3.176)	1.115 15.783		
Base Covariates				
ruggedness	2766.746 (4838.898)	0 95814.383	-8.70e-06*** (1.99e-06)	-1.82e-05*** (1.68e-06)
malaria index	1.912 (5.279)	0 38.081	-0.0348*** (0.00260)	-0.0486*** (0.00243)
Agricultural Covariates				
tropical moist forest	0.116 (0.321)	0 1	-0.0406 (0.0756)	-0.203*** (0.0661)
tropical dry forest	0.022 (0.147)	0 1	0.936*** (0.0943)	0.228*** (0.0804)
temperate broadleaf	0.104 (0.305)	0 1	1.771*** (0.0707)	1.294*** (0.0648)
temperate conifer	0.033 (0.178)	0 1	0.769*** (0.0822)	0.159** (0.0776)
boreal forest	0.167 (0.373)	0 1	-0.442*** (0.0769)	-1.246*** (0.0812)
tropical grassland	0.12 (0.325)	0 1	-0.862*** (0.0561)	-0.0548 (0.0487)
temperate grassland	0.077 (0.267)	0 1	0.710*** (0.0648)	0.922*** (0.0568)
montane grassland	0.033 (0.179)	0 1	0.631*** (0.0806)	0.762*** (0.0721)
tundra	0.124 (0.329)	0 1	-0.874*** (0.0863)	-1.438*** (0.0895)
Mediterranean forest	0.024 (0.153)	0 1	0.837*** (0.0926)	1.373*** (0.0877)
mangroves	0.004 (0.064)	0 1	0.418** (0.180)	-0.0312 (0.154)
desert	0.176 (0.381)	0 1	(reference)	(reference)
temperature (° C)	9.970 (13.788)	-22.286 30.366	0.174*** (0.00337)	0.124*** (0.00387)
precipitation (mm/month)	60.678 (59.29)	0.387 921.909	-0.00896*** (0.000446)	-0.0112*** (0.000468)
growing days	139.306 (99.067)	0 366	0.00985*** (0.000284)	0.00856*** (0.000283)

Continued on next page

Table 1 – *Continued from previous page*

	Summary Statistics		Baseline Regression Results	
	mean, (sd)	min, max	w/o country FEs	w/ country FEs
	(1)	(2)	(3)	(4)
land suitability	0.273 (0.319)	0 1	2.690*** (0.0552)	2.184*** (0.0528)
abs(latitude)	38.34 (20.923)	0 75	0.114*** (0.00251)	0.0371*** (0.00336)
elevation (m)	601.505 (788.097)	-187.341 6169.01	0.000501*** (2.41e-05)	8.02e-05*** (2.58e-05)
Trade Covariates				
1 (coast)	0.098 (0.297)	0 1	0.421*** (0.0396)	0.414*** (0.0323)
distance to coast (km)	485.226 (480.554)	0 2273.801	-0.000665*** (2.81e-05)	-0.000677*** (3.24e-05)
1 (harbor<25km)	0.027 (0.163)	0 1	1.625*** (0.0677)	1.414*** (0.0575)
1 (river<25km)	0.027 (0.163)	0 1	0.764*** (0.0662)	0.690*** (0.0601)
1 (big lake<25km)	0.127 (0.333)	0 1	0.317*** (0.0272)	0.179*** (0.0244)
Observations	243,985	243,985	243,985	243,985
Non-zero observations	98,941	98,941	98,941	98,941
R-squared			0.459	0.568

Notes: Standard errors in parentheses in columns (3) and (4) are clustered by square blocks of 9 cells. *** p<0.01,
** p<0.05, * p<0.1.

Table 2: R^2 for selected regressions predicting $\ln(\text{light}/\text{area})$

	(1)	(2)
	No country FEs	With country FEs
(1) All variables, both margins ($N = 243,985$)	0.459	0.568
(2) All variables, extensive margin (LPM)	0.385	0.475
(3) All variables, intensive margin (OLS)	0.262	0.359
(5) Country fixed effects, both margins		0.338
(5) Country fixed effects, extensive margin		0.269
(6) Country fixed effects, intensive margin		0.219
(7) Base variables (malaria, ruggedness)	0.020	0.350
(8) Agriculture variables, both margins (plus base)	0.437	0.554
(9) Trade variables, both margins (plus base)	0.072	0.367

Table 3: Cell counts in 4 way splits: Education vs. Urbanization

No country FEs (ed cutoff = 3.6; urban cutoff = 0.44)		With country FEs (ed cutoff = 3; urban cutoff = 0.38)			
Whole Sample	high urban low urban		Whole Sample	high urban low urban	
	high urban	low urban		high urban	low urban
high education	119,824	7,399	high education	123,908	3,879
low education	1,681	99,713	low education	6,214	94,616

Lit Sample	high urban low urban		Lit Sample	high urban low urban	
	high urban	low urban		high urban	low urban
high education	43,444	5,840	high education	46,614	3,196
low education	961	44,360	low education	3,790	41,005

Table 4: R^2 differentials of trade and agriculture variables in regressions predicting $\ln(\text{light}/\text{area})$ for high/low education & urbanization countries

	Education				Urbanization			
	No FEs		With FEs		No FEs		With FEs	
	High	Low	High	Low	High	Low	High	Low
Panel A: Both margins								
Base	0.008	0.052	0.379	0.291	0.008	0.058	0.352	0.346
Agriculture + base	0.546	0.326	0.638	0.445	0.529	0.356	0.611	0.489
Trade + base	0.070	0.218	0.418	0.399	0.073	0.212	0.389	0.438
High - Low differential	0.368		0.174		0.312		0.171	
Panel B: Extensive margin LPM								
Base	0.006	0.046	0.281	0.248	0.006	0.047	0.264	0.285
Agriculture + base	0.489	0.268	0.555	0.371	0.475	0.280	0.533	0.401
Trade + base	0.058	0.171	0.328	0.324	0.065	0.163	0.306	0.350
High - Low differential	0.334		0.180		0.293		0.176	
Panel C: Intensive margin OLS								
Base	0.011	0.050	0.253	0.171	0.008	0.063	0.240	0.214
Agriculture + base	0.251	0.184	0.366	0.255	0.239	0.213	0.344	0.298
Trade + base	0.082	0.163	0.281	0.262	0.078	0.166	0.271	0.297
High - Low differential	0.148		0.092		0.114		0.072	
Panel D: New world, both margins								
Base	0.006	0.193	0.243	0.239	0.007	0.208	0.249	0.267
Agriculture + base	0.485	0.302	0.601	0.346	0.488	0.300	0.578	0.398
Trade + base	0.139	0.284	0.302	0.326	0.132	0.286	0.295	0.352
High - Low differential	0.328		0.279		0.342		0.237	
Panel E: Old world, both margins								
Base	0.016	0.049	0.473	0.342	0.052	0.056	0.442	0.388
Agriculture + base	0.648	0.385	0.683	0.518	0.627	0.412	0.664	0.541
Trade + base	0.099	0.213	0.505	0.437	0.140	0.210	0.477	0.467
High - Low differential	0.377		0.097		0.285		0.113	

Notes: Education cutoffs are 3.6 (no FEs) and 3 (with FEs), and urbanization cutoffs are 0.44 (no FEs) and 0.38 (with FEs). High - Low differential is high[(Agriculture + base)-(Trade + base)] - low[(Agriculture + base)-(Trade + base)]

Table 5: Differential group marginal effects

Dependent variable: ln(lights/area)

	Education		Urbanization	
	No FEs	With FEs	No FEs	With FEs
	(1)	(2)	(3)	(4)
ruggedness	-9.44e-06*** (2.09e-06)	-1.46e-05*** (1.78e-06)	-9.53e-06*** (1.97e-06)	-1.70e-05*** (1.69e-06)
malaria index	-0.0429*** (0.00261)	-0.0443*** (0.00262)	-0.0403*** (0.00252)	-0.0436*** (0.00235)
tropical moist forest	-0.406*** (0.0703)	-0.440*** (0.0660)	-0.346*** (0.0676)	-0.326*** (0.0629)
tropical dry forest	0.757*** (0.0890)	0.140* (0.0815)	0.717*** (0.0882)	0.182** (0.0777)
temperate broadleaf	1.253*** (0.0640)	1.044*** (0.0633)	1.359*** (0.0618)	1.184*** (0.0633)
temperate conifer	0.610*** (0.0703)	0.118 (0.0722)	0.719*** (0.0697)	0.158** (0.0731)
boreal forest	-0.354*** (0.0684)	-1.049*** (0.0826)	-0.274*** (0.0664)	-1.061*** (0.0846)
tropical grassland	-0.760*** (0.0500)	-0.103** (0.0483)	-0.727*** (0.0479)	-0.0583 (0.0457)
temperate grassland	0.464*** (0.0543)	0.671*** (0.0558)	0.538*** (0.0540)	0.776*** (0.0562)
montane grassland	0.562*** (0.0756)	0.737*** (0.0757)	0.502*** (0.0730)	0.714*** (0.0721)
tundra	-0.548*** (0.0745)	-1.086*** (0.0891)	-0.462*** (0.0723)	-1.139*** (0.0922)
Mediterranean forest	0.548*** (0.0746)	1.086*** (0.0807)	0.655*** (0.0758)	1.195*** (0.0823)
mangroves	-0.930*** (0.180)	-1.042*** (0.154)	-0.327** (0.166)	-0.532*** (0.149)
temperature	0.139*** (0.00338)	0.113*** (0.00411)	0.135*** (0.00306)	0.117*** (0.00411)
precipitation	-0.00709*** (0.000370)	-0.00996*** (0.000437)	-0.00704*** (0.000356)	-0.0101*** (0.000436)

Continued on next page

Table 5 – *Continued from previous page*

	Education		Urbanization	
	No FEs	With FEs	No FEs	With FEs
	(1)	(2)	(3)	(4)
growing days	0.00901*** (0.000263)	0.00821*** (0.000289)	0.00853*** (0.000251)	0.00810*** (0.000284)
land suitability	2.348*** (0.0501)	1.955*** (0.0559)	2.313*** (0.0485)	1.971*** (0.0532)
abs(latitude)	0.107*** (0.00265)	0.0380*** (0.00310)	0.101*** (0.00235)	0.0382*** (0.00313)
elevation	0.000429*** (2.34e-05)	5.03e-05* (2.57e-05)	0.000380*** (2.14e-05)	8.67e-05*** (2.53e-05)
1 (coast)	1.127*** (0.0737)	1.109*** (0.0667)	0.861*** (0.0670)	0.946*** (0.0604)
distance to coast	-0.00149*** (3.65e-05)	-0.00126*** (4.72e-05)	-0.00133*** (3.37e-05)	-0.00119*** (4.49e-05)
1 (harbor<25km)	2.379*** (0.115)	2.292*** (0.102)	2.160*** (0.104)	2.088*** (0.0901)
1 (river<25km)	1.244*** (0.109)	1.172*** (0.101)	1.218*** (0.103)	1.051*** (0.0968)
1 (big lake<25km)	0.454*** (0.0495)	0.377*** (0.0441)	0.501*** (0.0459)	0.314*** (0.0405)
above cut	-3.954*** (0.142)	-74.26*** (0.158)	-3.780*** (0.134)	24.30*** (2.468)
α (agriculture deviation)	0.368*** (0.0209)	0.167*** (0.0258)	0.375*** (0.0191)	0.119*** (0.0237)
γ (trade deviation)	-0.761*** (0.0164)	-0.712*** (0.0177)	-0.740*** (0.0188)	-0.644*** (0.0201)
Observations	228,690	228,690	243,661	243,661
R-squared	0.486	0.574	0.474	0.571

Notes: Above cut is an indicator, which equals 1 if average years of education is greater than or equal to 3.6 (column 1) or 3 (column 2), or the fraction of urban population is greater than or equal to 0.44 (column 3) or 0.38 (column 4). Agriculture and trade deviation represent the constrained deviation of each variable set for early agglomerator countries (α and γ , respectively, in equation (12)). Standard errors in parentheses are clustered by square blocks of 9 cells. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Correlations with the average value of neighbors
as defined by rook or queen contiguity

Variable	(1)	(2)
	Rook	Queen
mangroves	0.6336	0.6486
ruggedness	0.7550	0.7611
$\mathbb{1}(\text{harbor} < 25\text{km})$	0.8222	0.7843
$\mathbb{1}(\text{big lake} < 25\text{km})$	0.8815	0.8585
$\mathbb{1}(\text{river} < 25\text{km})$	0.8914	0.8735
$\mathbb{1}(\text{coast})$	0.9050	0.8772
tropical dry forest	0.9057	0.9001
temperate conifer	0.9201	0.9120
montane grassland	0.9291	0.9235
tropical moist forest	0.9595	0.9586
Mediterranean forest	0.9674	0.9639
temperate grassland	0.9683	0.9635
tropical grassland	0.9721	0.9699
temperate broadleaf	0.9737	0.9699
tundra	0.9753	0.9701
boreal forest	0.9761	0.9709
malaria index	0.9876	0.9799
land suitability	0.9892	0.9818
elevation	0.9923	0.9895
precipitation	0.9943	0.9903
growing days	0.9989	0.9985
temperature	0.9994	0.9988
abs(latitude)	1	1
distance to coast	1	1