The Curse of Good Soil? Land Fertility, Roads, and Rural Poverty in Africa*

Leonard Wantchekon† Piero Stanig‡

September 20, 2016

Abstract

Using a global poverty map and standard soil productivity measures, we find that the poorest districts in Africa are more likely to have better (not worse) soil quality and that land fertility is higher in districts with worse roads. Our results are robust to a battery of controls and alternative measures of poverty and soil quality. The results indicate that transportation costs are the main drivers of poverty in Africa and that isolation might turn soil quality into a curse. More specifically, in districts with poor infrastructure, the poverty rate increases as soil quality gets better. We provide evidence for causality by using least-cost paths from mining areas to ports, and colonial road networks, as instruments for current transportation costs.

JEL: O13, O55, Q15, R42

*The authors would like to thank Nicole Mason, Harouna Kazianga, Doug Gollin, Jean-Francois Maystadt and participants of the PDL seminar at Princeton, ASE-CRED Workshop at the African School of Economics (Benin), the 2014 Annual Bank Conference on Africa in Paris, and the Political Economy Seminar Series at the Hertie School of Governance for comments and suggestions. Financial support from Open Society Foundations is gratefully acknowledged. Jonathan Liebman, Russell Morton, Sanata Sy-Sahande and Alexsandros Cavgias provided excellent research assistance. The usual caveat applies.

†Department of Politics, Princeton University.

‡Department of Policy Analysis and Public Management, Bocconi University.

1 INTRODUCTION

Soil quality and land fertility are considered key determinants of agricultural production and economic growth; soil degradation and drought are widely perceived to be associated with food insecurity and rural poverty (Barbier 2010; Nkonya et al. 2011). It is therefore not surprising that policy-makers and academics have recently focused on the adoption of modern inputs, including fertilizer use and improved seeds as a possible solution for rural poverty (Demery and Christiaensen 2007; World Bank 2008; Morris et al. 2007; Dercon and Christiansen 2011). For instance, Sanchez (2002) cites soil quality as a key driver of low agricultural production, while Scherr (1999) and Woomer et al. (1994) emphasize the role of soil degradation as a major concern for food security in Africa (See also Nkonya et al. 2011). In other words, the conventional wisdom is that soil quality is negatively correlated with rural poverty and that improving land fertility is crucial to poverty alleviation.

Using data on soil quality, poverty, transportation costs, and other potential determinants of poverty covering more than five thousand subnational units from forty-six sub-Saharan African countries we establish some important sets of empirical regularities that challenge the conventional wisdom that the main driver of rural poverty is soil fertility. First of all, there exists a positive correlation between soil quality and poverty in Africa, meaning that regions where land is most fertile are on average more likely to be impoverished than regions where soil is of lower quality. In addition, transportation costs and isolation are the main predictors of rural poverty. Furthermore, there exists a mismatch between soil quality and infrastructure: roads tend to be bad in areas with good soil, such as in hills and valleys, and good where the soil is of worse quality, such as in flat terrain close to the coast. Finally, when infrastructure is poorly maintained or non-existent, households are poorer in areas where the soil is quite fertile than in areas where the land is barren.

Several papers have already pointed to a weak association between soil quality and income (Drechsel et al. 2001, Ehui and Pender 2005, Okwi et al 2007). In particular, Okwi et al. (2007) use data from Kenya and show that if all soil were at the highest quality, it would only lead to a 1 percentage point decrease in poverty. While soil quality is found in some studies to have a limited effect on income, the literature finds a strong association between rural roads and poverty (Jacoby, 2000; Gibson and Rozelle 2003; Mu and van de Walle 2007; Jacoby and Minten, 2009; Khandker et al., 2009). From the theoretical point of view, Gollin and Rogerson (2014) propose a model that links transportation costs and the fraction of the rural workforce engaged in subsistence agriculture. Calibrating the model to match the features of sub-Saharan African countries, they make predictions about the role of transportation productivity in economic development and in the allocation of resources between manufacturing, subsistence agriculture and modern agriculture. In particular, they find that there are large interaction effects between agricultural productivity and improvements in transportation. Stifel and Minten (2008) find a strong relationship between poverty and isolation in Madagascar, highlighting four mechanisms at work: transportation costs, plot size and productivity, price variability and extensification onto less fertile lands, and insecurity. In other words, rural households choose to use larger plots with significantly fewer inputs, and invest in crops for their own consumption rather than for income because they are not sure they will be able to sell surplus crops to the market. Minten et al. (2013), in a case study of Northwestern Ethiopia, also find that transaction and transportation costs together add about 50 percent for the most remote farmer to the fertilizer prices charged at the input distribution center. However, as Stifel and Minten acknowledge, endogeneity of isolation may be a serious issue in their study, with superior land quality being associated with less isolation. We build upon these results by studying the relationship between isolation and soil quality as determinants of poverty, and exploiting sources of (plausibly) exogenous variation in transportation infrastructure on a comprehensive data set covering the almost-totality of sub-Saharan Africa.

The literature highlights several mechanisms through which rural roads decrease poverty and increase food security. Ali (2010) finds that households in Bangladesh invest in better technology once a road is built, replacing traditional local rice with a high-yield variety that requires more inputs. Households that only grow traditional rice increase their acreage instead. Additionally, the treatment effect is stronger for wealthier households, presumably better able to bear the costs of the inputs associated with the high-yield rice. This result is consistent with the general notion that rural households use more inputs once roads are built.

Bell and Dillen (2012) find that the introduction of rural roads in India leads households to invest more in education and health, with fewer days of school missed and much higher use of medical services. Yamano and Kijima (2010) find that soil quality is correlated with farm income only after controlling for isolation. They also find that both isolation and soil quality are associated with crop choice. The result suggests that isolated communities with poor soil quality may have stronger incentives to use non-farm income to supplement their agricultural output, while isolated agricultural households with higher quality soil are less likely to use non-farm income for consumption.

In a case study of Sierra Leone, Casaburi et al. (2010) apply a regression discontinuity design to a program where roads were built based on scores (based on population and economic value per kilometer, length and other considerations) and find heterogeneity in the treatment effect of roads on net returns for agriculture. In other words, the price impact of road construction depends upon other factors, specifically productivity and linkages with urban consumers, with higher prices observed in productive areas and vice-versa.

Some of the more recent research also highlights informational frictions and market aspects of agriculture in isolated areas. In a study of India, internet kiosks displaying current prices have been shown to raise prices in the regulated markets, and decrease price dispersion (Goyal, 2010). Additionally, the internet kiosks are associated with 19% higher production without increasing acreage, signaling either more intensive use of inputs and/or substitution to more profitable crops.

Our paper builds upon the notion of heterogeneity in the impact of roads, noting that roads have a much more positive benefit in areas with high soil quality, which is in line with the result that more productive areas receive higher prices, and, thereby, higher net returns due to roads.

On the methodological front, some recent constributions have advanced new identification strategies to tackle the issue of endogeneity of infrastructure. In two studies of China, Banerjee et al. (2012) use colonial-era railroads as an instrument for current transportation networks to show that proximity to transportation infrastructure promotes long-term growth, and Faber (2012) treats a large infrastructure-building project as a natural experiment. In a study of Sub-Saharan Africa, Jedwab and Moradi (2015) use the layout of colonial-era railways to study the effect of infrastructure investments on subsequent economic activity. They also instrument for presence of a railroad with a counterfactual Euclidean Minimum Spanning Tree railway network which is "the network that the colonial powers would have built if they had collaborated to optimally connect the initial cities while minimizing construction costs" (p.A10) assuming that cost depends only on distance. In some studies of Spain (Garcia-López et al. 2015; Holl 2011, 2016) and Italy (Licio and Pinna 2016), the layout of Roman roads and ancient postal routes are used as instruments for current transportation networks and contemporary trade costs. In a study of ethnic favoritism in Kenya, Burgess et al. (2015) construct a counterfactual road network based on the market potential of connecting a pair of settlement: in practice, the market potential is calculated as the sum of the populations of the settlement pair divided by Euclidean distance between them.[1]

We build on these intuitions, and exploit both variation in infrastructure induced by decisions made during the colonial period, and counterfactual colonial road networks based on shortest and least-cost paths to mining areas, to assess the role that transportation plays in determining rural poverty in contemporary Africa.

Specifically, we exploit information on the location of mines and quarries known during the colonial period to build instruments for the presence of legacy transportation infrastructure. [2] We then calculate both shortest paths (i.e., straight lines) from mines to cities and ports (similarly to Jedwab and Moradi 2015) and the least-cost paths taking into account the morphology of the terrain. We then use these to instrument for the presence of colonial and contemporary roads.

The main intuitions regarding our identification strategy are two. First of all, there is significant persistence in the transportation infrastructure from the colonial period to the post-independence period. Second, some mainly agricultural districts might be endowed with higher-quality transportation infrastructure just because of the priority, for colonial powers, to connect mining areas to seaports; other areas, albeit potentially similar in terms of soil quality and other characteristics, might find themselves isolated because they do not lie on the route that connects a mine to a seaport or city.

The use of least-cost paths allows us to exploit the fact that, in some cases, a piece of colonial transportation infrastructure might cross an area just because of the need to avoid a mountainous region along the shortest line connecting a mining area to a seaport or a large city. In this respect, our strategy is similar in flavor to the one adopted by Michaels (2009), who estimates the effects of reduced trade barriers on local economies exploiting the fact that the United States Interstate Highway System, while designed to connect cities and international borders, also crosses rural areas.

[1] See also Fujita et al. 1999.

[2] Glaeser et al. (2013) use the historical locations of mines to study entrepreneurship in the United States.

We show that the location of colonial infrastructure can be predicted based on geographic characteristics and, importantly, by the location of extractive resources (mines and quarries) but not by soil quality. This is in line with conventional wisdom in economic history: colonial powers were not after farmland but mostly after minerals. More specifically, we find that the probability of having a colonial road is significantly higher in districts located along the path from mining areas to cities and ports.[3] We use the presence of roads in the colonial period, as well as least-cost and shortest paths from mining areas to cities and seaports, as instruments for current infrastructure provision or transportation costs.

Our paper makes several important contributions to the literature. Foremost, in contrast to the conventional wisdom, we find a negative correlation between soil quality and poverty. This implies that poverty cannot be explained simply by invoking land degradation. In fact, as a stylized fact, a majority of poor households in Africa dwell on relatively high-quality soil.

Further, we find that the interaction between infrastructure and soil quality is positive and statistically significant. This can be interpreted as evidence of complementarity between soil quality and transportation infrastructure but it might also point to the existence of a "curse of good soil." We estimate flexible models to assess whether good soil without infrastructure may lead to worse economic outcomes. The data show that among bad-infrastructure districts the expected poverty rate is higher when soil quality is better. We also provide some evidence about the correlates of infrastructure and soil quality, and we find an element of "bad luck" in how infrastructure and soil are matched. In particular, the best soil is in districts with hilly terrain while the best infrastructure is found in flatter lands and at low altitude.

To sum up, we find a negative association between soil quality and income in isolated areas. The results are robust to the inclusion of measures of urbanization and the exclusion of urban districts. In other words, our results are not simply an indication that rural areas are poorer than urban areas due to agricultural productivity (see Lewis 1955, McMillan and Rodrik 2011). Instead, we relate the income differentials across rural areas to market access and high transportation costs and the dominance of subsistence agriculture, which is strongly associated with rural poverty. Our results suggest that the most important factor driving rural poverty and food security is the

[3] The results implicitly link food security to historical factors. Given that contemporary infrastructure availability is associated quite closely to the road network in colonial times, decisions made by colonizers for other reasons affect how land is used today, and therefore the current patterns of poverty.

provision of rural infrastructure and access to markets. In other words, to understand the relationship between factor endowment (e.g., soil) and rural development or rural poverty, it is crucial to take into account the complementarity between soil quality and infrastructure availability and market access. If one were to overlook this interaction, one would underestimate the return to infrastructure in poverty reduction. [4]

2 DATA

We construct a rich dataset about poverty, soil quality, and infrastructure, combining information from several different sources. Our dataset covers 5334 subnational units in 46 sub-Saharan African countries. The number of districts per country varies from a minimum of two districts (Sao Tome and Principe) to over 550 (in Nigeria) and its median is 80. The dataset provides measures of a wide array of phenomena, ranging from poverty to infrastructure to soil quality to population density.

The fact that our dataset is organized at the subnational level means that we are able to perform all of our analysis including country fixed effects. In this estimation approach, all the coefficients on the explanatory and control variables are identified exclusively by within-country variation. [5] The fixed effects account for all the features of the country that do not vary across districts (e.g., political regime, legal provisions regarding freedom of association, quality of governance at the national level, etc.) as well as the mean of all omitted variables that do vary across districts.

The backbone for constructing the data set is the GIS map of African sub-national entities at level 3, published by the UN's Food and Agriculture Organization (FAO). We overlay the shapefile of the subnational districts to the geocoded sources of soil quality, poverty, and many other variables described in the following. We compute district-level summaries, that we treat as the

[4] In fact, the estimated effect of infrastructure on poverty would be an average of the effect it has in places where roads would not matter much (because of low soil quality) and roads matter a lot (because the soil quality is high and the productive potential at high investment and high effort is large).

[5] The subnational districts at level 3 that we use are small enough to reflect sufficiently homogenous conditions, and large enough to sidestep the myriad of econometric and economic complications associated with performing analysis at the GIS cell level (which might be, for instance, just one km in side). For instance, even if an individual cell is not crossed by a road, this does not imply it is "isolated". Along the same lines, given that people do not necessarily live exactly where they cultivate the land, the association between poverty and soil quality at the cell level might not be informative of the underlying causal relationship we are after: many of the variables would have to be recast as distances from (rather than presence or absence of) a given feature, e.g., a colonial-era road or a mining area. In addition, being able to make claims in terms of districts (e.g., districts with such and such characteristics tend to have poverty rate around this given level) leads to straightforward substantive/economic interpretation of the results.

value of that given variable for the district.

When the resolution of the original measure is higher than that of the districts (so that one district encompasses many cells in the GIS grid) we compute the average of the median of the values of the cells. This is the case for the vast majority of variables, when the original data are released as gridded datasets or shapefiles (e.g., of rivers). In the case of the few measures that have lower resolution than the districts borders, we attribute to the district the value of the larger unit in which it is contained (or the area-weighted mean of the values of the larger units it spans, if the district belongs to more than one larger unit).

There is also an additional advantage coming from the use of district averages rather than individual cells: averaging leads to a reduction in measurement error if the values for individual cells are measured with error. Moreover, if some of the cell values are imputed (e.g., interpolated based on a sparser set of points) their exact value might not be very informative while district averages might be.

2.1 Main variables

Poverty The available source of poverty data with the highest resolution is the Global Poverty Map Derived From Satellite Data, published by the National Geophysical Data Center (NGDC). The map is released at the 30 arcseconds resolution (approximately one square kilometer at the equator). This dataset matches night lights visible from satellites with population density estimates from the LandScan dataset. These are then benchmarked with available poverty data at the national and subnational levels. See Elvidge et al. (2006) for a detailed description. The very high resolution of these data allows us to calculate a distinct value of the poverty rate for each subnational district.

Other measures of poverty that are georeferenced and are based on direct evidence (survey estimates, census data, etc.) are available. In particular, the Poverty Mapping Project: Global Subnational Prevalence of Child Malnutrition published by the Center for International Earth Science Information Network (CIESIN), Columbia University, provides poverty estimates based on "hard" data. It reports the percentage of children with weight-for-age z-scores that are more than two standard deviations below the median of the NCHS/CDC/WHO International Reference Population and aims at providing "a global subnational map of the prevalence of underweight children that can be used by a wide user community in interdisciplinary studies of health, poverty and the environment." The data refers, depending on the location, to the most recent available year between 1990 and 2002. In an analogous effort, HarvestChoice/International Food Policy Research Institute (IFPRI) publishes a geocoded map, referring to the year 2005, of sub-national poverty headcount ratios, derived from 23 nationally representative household surveys and population censuses. Poverty is defined at the $2/day level, expressed in 2005 international equivalent purchasing power parity (PPP) dollars. Rates are in percentages of total population.

The resolution of these alternative measures is, unsurprisingly, much coarser than the satellite-imaging data. This makes them less suitable for district-level estimation with country fixed effects. In fact, the CIESIN and the IFPRI data provide few distinct values per country, and, especially in smaller countries, this makes their use in our analysis problematic.[6] First of all, the variation across districts within country is smaller for these two measures than for the satellite data, as reflected by coefficients of variation for the satellite data that are in the overwhelming majority of cases quite higher.[7] In addition, the district-level values have within-country coefficient of variation exactly equal to zero (implying that the index takes the same value in all districts) in a quarter of the countries in the IFPRI data, and in two countries in the CIESIN data. These countries drop out of the sample in country-fixed-effects estimation.

The alternative measures of poverty are positively correlated with the satellite-based poverty rates (with correlations respectively .24 and .18). Linear regressions of the satellite values on the CIESIN and IFPRI data yield (positive, and substantively large) coefficients, with t-values respectively 24 and 28.8. These associations survive the inclusion of country fixed effects. In fixed effects models, in particular, the coefficient on the CIESIN measure is not statistically distinguishable from one. Given that both variables are measured as a percentage of total population (and hence on the same scale) this establishes an almost-perfect coincidence between changes in the expected value of the satellite measure and changes in the CIESIN measure.

In addition, CIESIN also releases some small-area estimates of poverty, for a few African coun-

[6]CIESIN releases the data in raster format at the nominal resolution of a quarter degree; yet, the data just replicates as a grid the few polygons per country available in the shapefile.

[7]The coefficient of variation discussed here is calculated by dividing the *within-country* standard deviation of the poverty value by the country-wise mean.

tries, and sub-national data on infant mortality. [8] While we cannot use these for our main analysis (the small-area estimates cover too few observations; the infant mortality data are about a slightly different concept) we can assess the validity of our main satellite-based data on poverty by comparing it with these other measures.

The IMR map subdivides Africa in 277 sub-regions, and for each of these sub-divisions, the Adjusted Infant Mortality Rate is reported.[9] We calculate the poverty rate, at the sub-region level, for each of the mapping units in the IMR dataset: this is calculated as the average satellite-based poverty rate within the mapping unit. We can then compare the ADJIMR and the satellite-based sub0national poverty rates. In a regression of the ADJIMR on the satellite-based poverty (and no additional controls), the coefficient is 0.66, with a standard error of 0.07 (and therefore a t statistic of 9.3). This points to the very strong relationship between poverty rates as calculated based on the satellite data, and a typical correlate of poverty, infant mortality. It is worth noting that the infant mortality rates are based on "hard" (census- or survey-based) information reported by the statistical offices of individual countries.

The Kenya Kajiado Case Study covers 121 sub-locations in the Kajiado district in Kenya. For the year 2003, the dataset reports, for each rural sub-location (excluding urbanized areas) the number of people residing in the sub-location, and the number of people below the poverty line in each sub-location. Based on these, we calculate the sub-location poverty rate (as 100 times the ratio of poor people over residents). We also overlay the sub-location map to the satellite-based gridded poverty data, and calculate, for each sub-location, the poverty rate as the mean of the cells in the grid that fall within the area of a given sub-location. When we regress the case-study poverty rates on the satellite-based poverty, the coefficient is 0.43, with a standard error of 0.049 (and therefore a t statistic of 8.634), pointing to the very strong association between the case-study poverty headcounts and the satellite-based measure.

The Nigeria case study reports, for almost 800 sub-national entities, estimates of income per capita. We can directly compare these with the estimates based on the satellite data. A regression of the income variable from the CIESIN Nigeria case study on the averages (by mapping unit in

[8] Center for International Earth Science Information Network (CIESIN), Columbia University; 2005 Global subnational infant mortality rates dataset. CIESIN, Palisades, NY, USA. Available at: `ttp://www.ciesin.columbia.edu/povmap/ds_global.tml`

[9] In these data, the sub-natonal infant mortality rate is adjusted to the national 2000 UNICEF rate.

the CIESIN data) of the satellite-based poverty rates yields a coefficient of -0.32, with a standard error of 0.12 (and a t statistic of -2.8), pointing to the strong negative association between income as reported by CIESIN and our estimate of poverty based on the satellite data.

All this evidence makes it clear that using the satellite-based poverty measure in the main analysis is warranted, given that it is strongly correlated with "hard" sub-national measures when these are available, and at the same time it provides a unique opportunity to calculate for every district in Africa a separate value of the poverty rate and, as we discuss below, carry out all the estimation with country fixed effects (effectively exploiting only within-country cross-district variation to econometrically identify the coefficients).

Infrastructure For infrastructure, we rely on one main variable, road cost, published by FAO. This measure is calculated as follows. First, the road network in Africa is classified according to the accompanying road type classification system. Then, the cost to travel from one cell to the next is estimated, assuming that "the time required to travel from one cell to another in absence of main roads is 5 times longer than the time needed on the main road." The information on the road network was derived from ArcWorld (ESRI, 1992). In addition, we employ the average of the class of the roads found in the district, based on the Roads of Africa dataset (also published by FAO). We also create variables based on information in the Global Roads Open Access Data Set, Version 1. We create dummies that take the value of one if a district is crossed respectively by a highway, a primary, or a secondary road as classified in the gROADS dataset.

Soil quality For soil quality we rely on several measures, from different sources. First of all, we collect the scores on seven dimensions, published by the International Institute for Applied Systems Analysis (IIASA) and part of the Harmonized World Soil Database (HWSD). These evaluate soil quality according to the following criteria: nutrient availability; nutrient retention capacity; rooting conditions; oxygen availability to roots; excess salts; toxicity; and workability. The data is released for small cells (30 arcseconds, less than one square kilometer at the equator) and the values in each cell can range from 0 to 7, with higher values meaning worse soil quality. For each subnational entity, we compute the average soil quality for each of the seven dimensions (as the simple average of the cells contained in the perimeter of the subnational entity). We then create a simple additive index as the average of the (standardized) values of the soil constraint variables. This is an index of soil constraints, with higher values indicating worse quality soil. We also calculate another measure from the same data: namely, the first principal component of the seven measures is used as a summary index of soil quality. For some robustness checks we also calculate the maximum for each district of the seven soil constraint measures. We also use the classification of problem land published by FAO. This classifies each cell, at high resolution, in one of several categories. We calculate the proportion of cells in the district that fall in a category characterized as "No problem soils > 30% of the mapping unit", to construct the variable "goodsoil". For the robustness checks, we create analogous variables with proportions for cells classified as "steep" or "infertile" by the problem land data set. We also check our results with the index of soil production published by FAO. This index "considers the suitability of the best adapted crop to each soil's condition in an area and makes a weighted average for all soils present in a pixel." (FAO 2007)

2.2 Control variables and instruments

As for the control variables and the instruments, these can be divided into pure geographic/geological variables, demographic variables, and historical variables.

Geographic variables The geographic controls are elevation and slope class of the terrain, distance from the coast, the presence of water bodies, the presence (and type) of rivers in the district, the length of the growing period, rainfall, forest cover, and classification into an agro-ecological zone.

Elevation data comes from the IIASA-LUC Global Terrain Slopes and Aspect Database, and is originally reported at the 30 arcseconds resolution. We compute the average by district. Elevation is reported in meters, and we rescaled it in hundreds of meters. The slope data comes from the FAO Geonetwork, originally from the FAO-UNESCO Soil Map of Africa: each cell is classified in one of three classes, and we compute the median, the mean and the standard deviation of the slope class within the district.

We compute the distance of the district from the closest coast, based on the data on coastlines published by naturalearthdata.com. This takes the value of zero for a coastal district. For rivers, we rely on the World Rivers GIS file (from http://worldmap.harvard.edu). We compute several summaries: whether there is a river in the district; the distance of the district from the closest river (equal to 0 if a river flows in the district); the average rank class of the rivers present in the district; the rank class of the largest river in the district. We calculate the proportion of cells in the district classified as being water bodies. The data on water bodies also comes from the FAO Geonetwork. The data on forest cover and cultivation is from IIASA-LUC [10].

Demographic variables We collect several demographic variables. First of all, we compute the average and maximum level of urbanization in the district. The data on urbanization comes from IIASA and is released at the 30 arcseconds resolution; for each cell, the percentage of urban population is reported. We also collect data on travel times to towns of at least 20,000 thousand inhabitants, and to cities of more than half a million inhabitants. The data on which our measures are based are published by HarvestChoice/IFPRI in raster format.

In addition, we rely on the data on settlements published by CIESIN and now part of the Global Rural-Urban Mapping Project (CIESIN, IFRPRI, and CIAT 2011). This dataset provides detailed information about human settlements based on a variety of sources. For every settlement, the location, and the population as of 1995 (among other pieces of information) is provided. We restrict the analysis to settlements of at least 5000 inhabitants. We compute the number of settlements with more than 5000 and 10,000 inhabitants (both in absolute terms and relative to the area of the district) and the mean, the median, and the maximum size of the settlements in the district. If a district has no recorded settlement with a population of at least 5000 as of 1995, all of these take the value of zero.

We compute the average of the rural population density figures published by FAO at the resolution of 5 arc-minutes. Each pixel classified as "rural" by the urban area boundaries map has information about the number of persons per square kilometer, aggregated from the 30 arcsecond data layer. The measure of rural population density is missing by construction in cells (pixels) classified as urban by FAO. This fact has two consequences: first of all, the variable we compute (the mean by district) is unaffected by the presence of a high-density urban area in the district: urban areas are basically considered non existent when the district-wise average of this variable

[10] `http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/Landcover/`

is taken. In addition, districts that encompass only urban areas have a missing value on this variable: when we include rural population density in regression models, we are excluding from the analysis urban-only districts.

Historical/political variables Finally, we collect data on colonial era. Our instrumental variable strategy, whose results we present in Section 4, relies on the well-documented fact that colonial powers were mostly interested in mineral resources, and on the persistence of investment, and patterns in public goods provision from colonial to current times (Njoh 2000). The Mineral Resources Data System published by the U.S. Geological Survey provides a georeferenced map of mining sites across the world, including Africa. For our purposes, we need information on mines that were operating, or at least known, in the colonial era. In several cases, the data file reports the year in which the mine started operations. For the mines whose year of initial operation is not reported, we back out the year in which the location of the mineral resource was known based on the publication date of the report referenced in the MRDS for that specific location. In a large number of cases, the year of publication of the report is available: in these cases, we attribute the year to the mine, given that this established the year in which the mining site was *known* (albeit, possibly not yet exploited commercially). When a year of publication is not reported, a research assistant searched for full information about the reference and, whenever possible, attributed a date of publication to the report. Finally, we select only mines that were known or exploited before 1965. This leaves with a total of 408 mining areas, for which we have accurate geolocation information.

Based on the World database of large urban areas, 1950-2050 (SEI, Stockholm University), we identify all the cities in Africa that had at least fifty thousand inhabitants as of 1950 (the earliest date available). For the location of seaports, we rely on the World Port Index publication of the National Geospatial-Intelligence Agency of the United States government. [11] This reports current, rather than historical, ports. The instrumental variables, though, are based only on the connections between these ports and mining sites known during the colonial period.

Additional measures of colonial-era infrastructure come from a German road map of Africa as of 1941, found in the collection of the library at Princeton University and georeferenced for this

[11] `http://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_62&pubCode=0015`

study. For each road featured in the map, we code whether it is a “primary,” “secondary,” or lower-rank road. We then create district-level summaries about transportation infrastructure in the colonial era. The variables Primary colonial and Secondary colonial are dummy variables that take the value of one if the district was crossed by a (respectively, primary or secondary) road in the colonial era.

Credible evidence exists about the extent to which current patterns in the provision of public goods in sub-Saharan Africa reflect decisions made in the colonial period. For instance, Huillery (2009) documents the persistence of patterns of investment over the long haul in French West Africa. She also shows how pre-colonial characteristics are not predictive of colonial investment. One strong predictor, though, is distance from the coast, which in any case we account for in all of our models.

In order to build our instruments, we rely on the data on elevation, the location of mining sites known during the colonial period, the location of ports, and the locations of cities. We then calculate the shortest paths from mines to cities and ports, and the least-cost paths from mines to cities or to ports accounting for the cost of transportation induced by terrain features. Importantly, we restrict the paths from a mine to a port to only lie within one colonial-era geo-political unit. So, for a mine located in French West Africa, we require the shortest path to lead to a port or a city located in French West Africa and to never cross into another political unit (e.g., British Goald Coast). For colonial borders, we rely on the historical political maps published in geocoded format by the Harvard Geospatial Library.

The search for the least-cost path is based on the Dijkstra (1959) algorithm, as implemented in the package `gdistance` in the `R` statistical computing environment. We allow for moves in eight possible directions (the so-called “king” and “queen” moves in chess terminology). This means that the path leaving a given cell can head north, south, east, west, northeast, nortwest, southeast and southwest.

The relatively complicated part of this exercise has to do with choosing an appropriate cost function from which we can calculate the *conductance* (the reciprocal of the cost function) between two given contiguous geographic locations (two 30″ cells in the altitude grid). Research exists on optimal paths for human hikers (Tobler 1993; Whitley and Hicks 2003) but only a few papers have explored GIS-based cost functions for road construction (Collischonn and Pilar 2000; Yu et al.

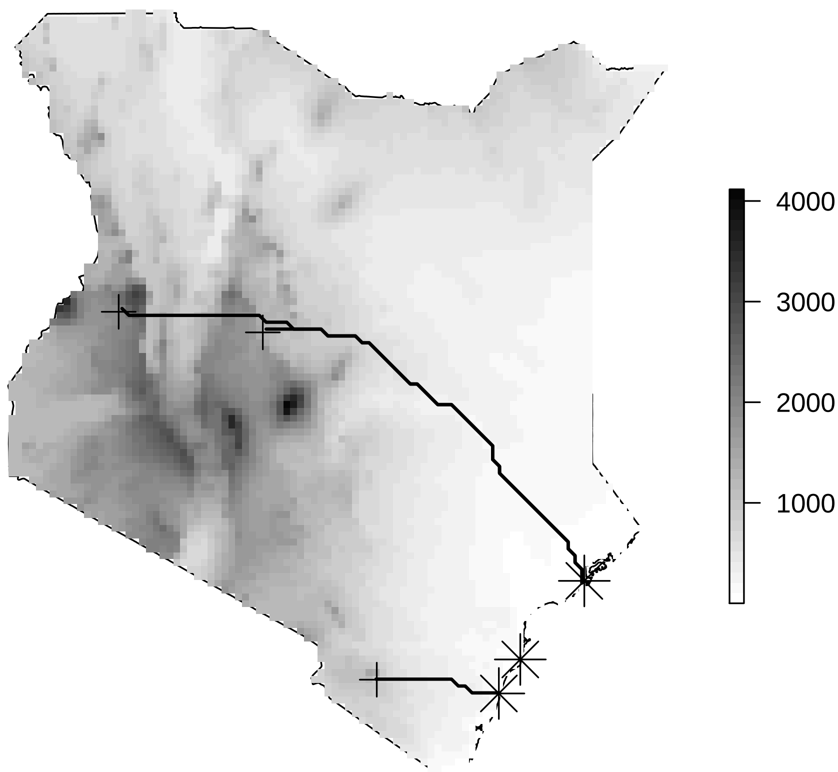

Figure 1: Illustration of the least-cost paths from mines to ports, for British Kenya. The shade of the grid represents elevation (with darker color meaning higher elevation). The cross-hairs are the locations of the mines in our data that were known in the colonial period. The star signs are seaports. The black lines are the least-cost paths from the mines to a seaport.

2003; Pingel 2010) To our knowledge, attempts to predict the paths of transportation infrastructure in colonial-era Africa based on this type of algorithm have not been made.

The main characteristic of the Tobler (1993) hiking function is that it is asymmetric, as a human hiker can be quite faster on mildly downward-sloping terrain than on flat terrain, but upward-sloping terrain offers much more resistance. The cost (and conductance) functions therefore reflect these aspects. It makes more sense, from our point of view, to adopt a symmetric and concave function of slope and altitude as a cost function: the considerations road planners might have made involve a certain degree of symmetry, given that roads have to be used in both directions, and probably construction cost is overwhelmingly important compared to other considerations (like uphill and downhill speed) in deciding the exact path.[12] The choice of function was made blindly with respect to the actual location of roads. As we show below, the ability of our shortest paths to approximate the actual trajectories of colonial-era roads reported in the 1941 German map is remarkable. In any case, the specific form for the conductance (the reciprocal of the cost function) between two (contiguous) locations i, j that we adopt is

$$C = \frac{1}{(1 + \log(1 + \alpha \max\{x_i, x_j\}) + \log(1 + |x_i - x_j|))} \tag{1}$$

where x_i and x_j are expressed in hundreds of meters and we set $\alpha = .2$. [13]

This simple function takes maximum value 1 (hence conductance is maximum, and cost is minimum) when the two locations i and j are both located at sea level. Conductance is decreasing (at a decreasing rate) in slope, and it is decreasing at a decreasing rate also in altitude. At the same time, altitude receives much lower weight than slope. This implies, mainly, that the least-cost path avoids high altitudes when indifferent between equally-sloping transition. In spite of the preference for flat terrain encoded in the function, it is possible that crossing a longer distance at sea level compensates an initial downward slope (so that, for instance, least-cost paths are located at the bottom of valleys rather than on hillsides).

The plot in figure 1 shows the least-cost paths and the altitude (with darker shades indicating

[12] Clearly a certain amount of asymmetry in the considerations made by the planners in the colonial period is possible – for instance because an empty truck or train can travel uphill faster than a fully loaded one – but it is beyond the scope of our exercise to take into account these "second order" considerations.

[13] As it is standard practice, the shortest path algorithm also takes into account the differential size of cells at different latitudes.

higher altitude) for three mining locations in British Kenya (as of 1950). The algorithm searches for the least-cost path from each mine to each port within colonial political unit.[14] We then select, among the various paths from a given mine to each of the ports, the one with lower cost (in a sense, the "shortest", or the one leading to the "closest" port).

After drawing all the shortest least-cost paths from each of the mines known before 1965 to all the cities of sufficiently large size as of 1950, and to all known seaports, we aggregate back at the district level, which is our unit of analysis. Using the language of instrumental variable models, we can categorize each district as being "encouraged" if it is crossed by one of the shortest paths from a mining area. Alternatively, we measure the distance between the centroid of the district and the closest predicted path. We can therefore define, for each district, four variables: two dummies for presence of a least-cost path (from a mine to respectively a city or a seaport) and two distances from a least-cost path.

To our knowledge, this is the first application of least-cost paths in economic studies of Sub-Saharan Africa. As we previewed, the identification strategy relies on the fact that some districts might have received a piece of colonial transportation infrastructure simply because the least-cost path passed through them, or, in other words, because of the presence of an obstacle (e.g., a high mountain range) in a neighboring area located on the shortest path from a mining region to a seaport or a city. Conditional on district morphology (e.g., elevation and slope of the terrain) the presence of a colonial road is plausibly "as if random" in this case. The crucial part of the instrumentation strategy that allows us to improve identification has to do with the fact that at least some districts find themselves endowed with a colonial-era piece of infrastructure just because an obstacle *elsewhere* forced a road to pass through the district. Conditional on geographic characteristics like altitude and slope, the layout of obstacles (e.g., mountain ranges) elsewhere on the trajectory connecting a mine to a seaport or a city should be plausibly independent of potential outcomes (i.e., poverty levels) in a given district.

As a simpler measure of the "encouragement" related to the need to connect mining areas to cities and ports, we also trace the shortest path (as a straight line) from each mine to a city or a port. This strategy is similar in nature to the one used by Jedwab and Moradi (2015) for their

[14]In order to prevent the path from leaving the political unit, we set altitude *outside* that political unit equal to a very large value, so that no least-cost path would ever pass through areas outside the political unit.

instrumental variable. The district map is overlayed on the shortest paths (the straight segments that connect mines and ports or cities overlooking terrain morphology). We then create a binary variable that takes the value of one if the district lies on the path from a mine to a large city, and zero otherwise. We also calculate the distance between the centroid of the district and the shortest path

From the colonial roads map, we also compute dummy variables equal to one if the district is crossed by a colonial-era road, and the (point to set) distance between the geometric center of the district, and the closest road of each class in the colonial period. In the case of these latter measures, the road does not necessarily pass through the district and it is not required to lie within the country to which the district belongs in the post-colonial period. Roads in neighboring districts and, for that matter, in different countries, are also included in the computation. Very high values mean that not only infrastructure was underprovided in that district, but also in the overall region in which the district is located.[15]

3 BASIC ANALYSIS

The plot in Figure 2 displays the level of poverty (percentage of the population living on less than two dollars a day) according to the IFPRI data, averaged by soil category, after controlling for country fixed effects. Higher values of the soil production index reflect better soil. From the raw data, then, it appears that, without conditioning on other district characteristics, poverty is more of an issue in districts with better land.

To explore the pattern rigorously, we regress the measure of poverty on the measures of soil quality. All the models include country fixed effects (omitted from the tables to improve readability). The fixed effects capture the effect of all features of the districts that do not vary within country. In addition, the standard errors are clustered at the country level: doing so accounts for the fact that there might be remaining correlation in the errors at the district level even after the inclusion of fixed effects to account for the nesting of the districts within countries. (Angrist and

15

As long as contemporary infrastructure is affected by the long-term patterns set in the colonial era, the fact that the distance measures are calculated also based on colonial roads located very far from the district does not affect our identification strategy directly. Above a certain level of remoteness, distance might not matter much. But this does not violate the exclusion restriction.

Pischke 2009, chapter 8; see also Arellano 1987)

Model 1 measures soil quality with the first principal component of the 7 IIASA measures; Model 2 uses the additive index of soil constraints; Model 3 includes the Soil production index; Model 4 uses the proportion of the district that is classified as having good soil according to FAO's problem soil classification. In all the models, the coefficient on the measures of soil quality is estimated as positive, and substantively large; the association is statistically significant at conventional levels with the exception of the soil production index. According to the estimate of Model 1, if we compare a district with soil quality at the first quartile with one at the third quartile, the poverty rate in the latter is expected to be three percentage points higher than in the former. This points to the fact that, on average, districts that have better land tend to have higher poverty rates than districts with worse soil. This result is highly counterintuitive. Notice that the index of soil constraints has opposite polarity than the other three, with higher values implying worse soil.

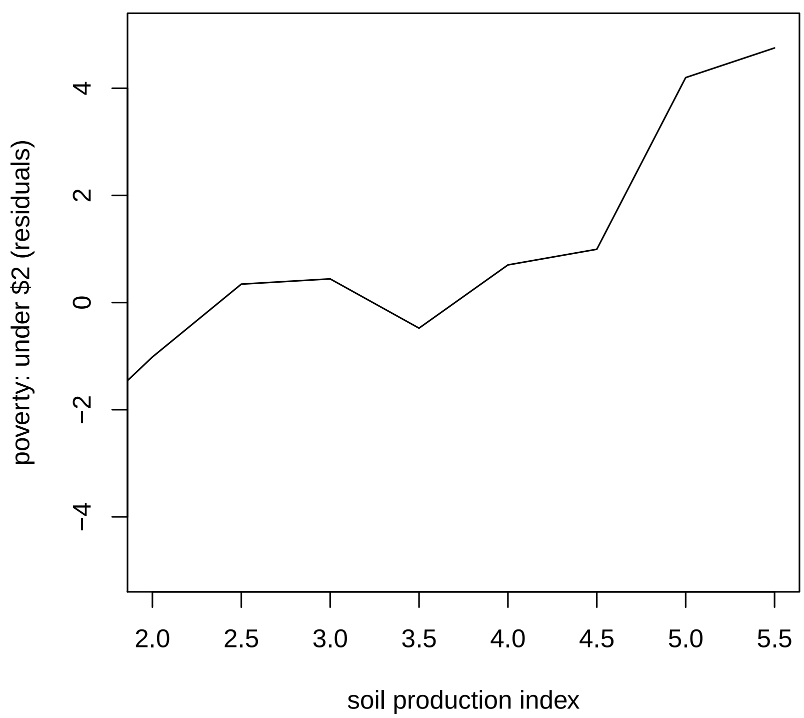

Figure 2: Average poverty by category of soil productivity. The poverty figures are residuals from a regression on country fixed effects.

It is interesting to note that, working on smaller datasets at different levels of aggregation and on smaller samples of countries, the literature in agricultural economics has not been able to establish a clear correlation between soil quality and development outcomes such as poverty alleviation. If anything, the literature has found that the two tend not to be correlated in clear and statistically detectable ways, and when systematic relationships are detectable, they are at times counterintuitive. For instance, Yamano and Kijima (2010a), studying a sample of households in rural Uganda, find that soil quality (measured at the the household level) is associated with higher crop income but *lower* non-crop income; Yamano and Kijima (2010b) find that soil quality has a positive association with income in Kenya, but not in Uganda and Ethiopia. In a study of rural Kenya, Okwi et al. (2007) estimate that improving soil fertility (from poor to good soil) would "reduce poverty by up to one percentage point."

Using a much larger dataset, we are able to detect this counterintuitive positive correlation. The result, as we show in the following subsection, survives the inclusion of controls for many of the possible determinants of poverty. It is also worth bearing in mind that the model includes fixed effects at the country level, so that all the features of the country that do not vary across districts are accounted for. Hence, this correlation (or lack thereof) is not driven by the fact that poor countries have better soil than rich countries. To estimate the coefficients we only exploit variation in soil quality and in poverty within each country.

The second set of models in Table 1 adjusts for some basic characteristics of the district: urbanization, cultivation (average percentage of the district that is cultivated), percent of terrain that is forest, and quadratic terms for elevation and for slope class. Poverty tends to be lower in more urbanized districts, and higher in more cultivated districts [16]; in addition, poverty increases (at a decreasing rate) with elevation and with steepness. Most importantly, the magnitude (and statistical significance) of the coefficients on the soil quality measures are left unchanged when we include these covariates.

The coefficient on soil quality should not, obviously, be interpreted as an estimate of a causal effect. All we establish with these models is that, on average, districts with better soil tend to be poorer, statistically significantly so in most cases. This is a snapshot of the existing situation,

[16] That urban districts are less poor is not surprising but an explanation of the positive correlation is left for future research.

Table 1: Basic models for poverty, soil quality, and transportation infrastructure

basictable	1	2	3	4	5	6	7	8
(Intercept)	78.97**	79.07**	79.98**	79.37**	71.39**	70.99**	55.46**	34.94**
	(0)	(0)	(0)	(0)	(5.92)	(5.88)	(6.39)	(5.33)
Comp.1	4.02**				4.08**			
	(0.29)				(0.31)			
Soil constraints		-3.05**				-3.09**		
		(0.22)				(0.23)		
st.var(soil.prod)			2				3.19	
			(1.54)				(2.13)	
st.var(goodsoil)				2.22*				2.26*
				(1.01)				(0.88)
urbanization					-1.14**	-1.14**	-1.1**	-1.05**
					(0.12)	(0.12)	(0.11)	(0.11)
forests$_{pcent}$					0.04	0.05	0.07	0.11*
					(0.05)	(0.05)	(0.05)	(0.05)
cultivation					0.2**	0.2**	0.28**	0.3**
					(0.02)	(0.02)	(0.02)	(0.02)
elevation					1.91**	1.89**	1.8**	1.78**
					(0.43)	(0.43)	(0.48)	(0.47)
I(elevation^2)					-0.07**	-0.07**	-0.07**	-0.06*
					(0.02)	(0.02)	(0.02)	(0.02)
median.slope					-1.08	-0.81	14.51**	32.29**
					(5.05)	(5.03)	(5.3)	(4.92)
I(median.slope^2)					-0.17	-0.19	-3.92**	-8**
					(1.08)	(1.07)	(1.14)	(1.13)

Basic models with poverty rate as dependent variable. Standard errors clustered by country in parentheses. $*$: statistically significant at the 5% level.$**$: statistically significant at the 1% level.

which, as we will show, depends on several observable factors. In the remainder, we explore this counterintuitive correlation, first by showing how infrastructure is a stronger driver of rural poverty than soil quality, and second by showing that, in cases of extreme isolation, soil of good quality can be a curse in terms of development.

3.1 The role of transportation infrastructure and access to markets

The literature on rural infrastructure has shown the importance of roads for development (See Ayogu (2007) for a review). This literature has not, however, explored to the fullest the possible complementarities between agricultural factor endowments and infrastructure provision. Measures of transportation infrastructure or market access are included in models of rural poverty in additive fashion, overlooking the fact that the role of soil quality for development and poverty reduction depends itself on the availability of infrastructure and the accessibility of markets. For instance, Radeny and Bulte (2012) find that distance from the nearest market and the nearest town are negatively associated with per capita income in Kenya; similarly, Okwi et al.(2007) include measures of soil quality and of access to markets (distance from towns).

Our empirical models take more seriously the complementarity between factor endowment (and soil quality in particular) and market access. We estimate econometrically the variation in the soil/poverty association that is driven by the availability of transportation infrastructure. Here and in the next section, we establish that the effect of soil on poverty depends on transportation infrastructure, and that, more surprisingly, isolation (i.e., lack of transportation infrastructure) might turn high soil quality into a curse.

In the models reported in Table 2, we first include soil quality and the measures of transportation cost in isolation; then, we include their interaction. The variable "road cost" measures the transportation cost (averaged over the district), with higher values reflecting worse roads. Positive values of the coefficient mean that worse roads are associated with more poverty. Transportation cost is centered to have mean zero (and scaled so it has standard deviation one half), hence the main effect of soil quality captures the effect of soil quality on poverty in a district with average road quality, and symmetrically the main effect of road quality captures the effect of road quality on poverty for a district with average soil quality. The interaction term captures how the association between soil and poverty varies across different levels of road quality. In Subsection 3.2, we model non-parametrically the interaction between infrastructure provision and soil quality to understand whether good soil can be a curse in extremely isolated areas.

In line with the conventional wisdom and results in the literature, lack of transportation infrastructure is systematically associated with poverty. The effect of infrastructure on poverty outcomes has large economic significance too. According to the estimates of Model 1, if one compares two districts in a given country, with the same soil quality, and respectively one standard deviation below and one above the mean of road quality, their poverty rates are expected to differ by 13 percentage points.

A positive coefficient on the soil quality measure means that better soil is associated with more poverty. Analogously, a negative coefficient on the soil constraints index implies that worse soil is associated with less poverty. According to these estimates, in a district that is average in terms of road cost, higher soil quality is associated with a higher poverty rate.

The interaction between the measures of soil quality and the measure of infrastructure capture how the association between soil quality and poverty varies depending on the quality of transportation infrastructure.

Table 2: Models with interaction between soil quality and transportation infrastructure

interactiontable	1	2	3	4	5	6	7	8
(Intercept)	75.83**	76.01**	71.57**	71.26**	75.96**	76.14**	72.35**	72.02**
	(0.02)	(0.02)	(6.31)	(6.27)	(0.09)	(0.09)	(6.44)	(6.39)
cen(Comp.1)	4.45**		4.17**		4.17**		3.78**	
	(0.28)		(0.31)		(0.36)		(0.36)	
st.var(road.cost)	12.8**	12.81**	4.6**	4.62**	12.52**	12.55**	4.28**	4.31**
	(2.13)	(2.13)	(1.5)	(1.51)	(2.13)	(2.14)	(1.54)	(1.54)
index.soil.quality)		-3.38**		-3.16**		-3.18**		-2.88**
		(0.21)		(0.24)		(0.28)		(0.28)
urbanization			-1.02**	-1.02**			-1.01**	-1.01**
			(0.12)	(0.12)			(0.12)	(0.12)
forests$_{p}cent$			0.04	0.04			0.03	0.03
			(0.05)	(0.05)			(0.05)	(0.05)
cultivation			0.21**	0.21**			0.21**	0.21**
			(0.02)	(0.02)			(0.02)	(0.02)
elevation			1.85**	1.82**			1.85**	1.83**
			(0.44)	(0.43)			(0.44)	(0.44)
I(elevation^2)			-0.07**	-0.07**			-0.07**	-0.07**
			(0.02)	(0.02)			(0.02)	(0.02)
median.slope			-1.73	-1.46			-2.15	-1.87
			(5.24)	(5.21)			(5.37)	(5.34)
I(median.slope^2)			-0.03	-0.05			0.08	0.05
			(1.11)	(1.1)			(1.14)	(1.14)
cen(Comp.1):ROAD					1.19$^+$		1.58*	
					(0.68)		(0.76)	
ADDIN:ROAD						-0.84		-1.14$^+$
						(0.51)		(0.57)

Models with poverty rate as dependent variable and interaction between measures of soil quality and measures of infrastructure. Standard errors clustered at the country level in parentheses. $+$: statistically significant at the 10% level.$*$: statistically significant at the 5% level. $**$: statistically significant at the 1% level.

The coefficient on the interaction between road cost and soil quality is positive, and the coefficient on the interaction between road cost and soil constraints is negative. This is far from counterintuitive, pointing at a complementarity between soil and infrastructure.

In areas with average infrastructure, soil quality does not have a poverty-reducing role. The coefficient on the main effect of soil quality (that captures the effect of soil quality in a district with average quality infrastructure) is positive, and analogously the coefficient on soil constraints is negative. The coefficients on these main effects are statistically significant in all specifications. According to these estimates soil quality starts having a poverty-reducing effect only when road cost is sufficiently low. At the same time, the results for the interaction might potentially provide evidence (whose robustness we probe below) that good soil can be a *curse* in the absence of infrastructure.

These are not causal regressions (or estimates of structural parameters) but in terms of conditional expectations (and therefore, as descriptive summaries) they point to the fact that infrastructure in most of Africa is insufficient for the available resources, in terms of cultivable soil, to be used to significantly reduce rural poverty.

3.2 CAN GOOD SOIL REALLY BE A CURSE?

The estimates reported above show that infrastructure has a stronger association with reduced poverty in areas with good soil, and, symmetrically, soil quality has a stronger association with reduced poverty in areas with good infrastructure. One possible interpretation is that when the quality of infrastructure is sufficiently bad, soil quality is associated with *increased* poverty. This would point to the existence of a "curse" of good soil: if we were to compare two districts, with equally insufficient infrastructure, the one with the better soil would be expected to have a higher poverty rate than the one with the worse soil.

The linear multiplicative interaction is symmetric, and it cannot discriminate between a "curse" effect from simple complementarity. The interaction effect between soil and infrastructure might reflect exclusively the increase in returns to soil quality when infrastructure is better rather than a curse that affects locations that are poorly served in terms of infrastructure but "sit" on good soil.

We address this issue by turning the soil quality measure and the infrastructure availability measure into categorical variables and then estimating a model fully saturated in these categorical variables. In practice, we create dummies for all the possible combinations of soil quality (turned into a six-category variable by rounding to the nearest integer the value for the district) and road cost (turned into a three-category variable, grouping in turn the bottom two levels and the middle two levels of road cost). There is a total of 18 possible combinations in which both road cost and soil quality are observed, plus those cases in which one of the two variables is not observed. We exclude from the analysis the districts for which one of the two (or both) measures is not observed.

We then estimate models of the form

$$y_i = \alpha_{j(i)} + \beta X_i + \sum_r \sum_s \gamma_{r,s} 1(r_i = r \& s_i = s) + \epsilon_i \tag{2}$$

where the α terms are country fixed effects for country j to which district i belongs, and the $\gamma_{r,s}$ is the parameter for districts that belong to the r road cost category and the s soil quality category. In other words, we include one intercept for each combination of soil and road cost. We estimate two variants of the model. In the first, we include the dummies for the combinations of soil and infrastructure: these are fixed effects for the groups defined by a given soil and infrastructure combination. In the second variant, we model the γ as draws from a normal distribution with variance estimated from the data: the γ coefficients are random effects for given soil-infrastructure combinations. The advantage of this approach, as opposed to estimating them as fixed effects is that, whenever (or if) they are estimated imprecisely for a given category (for instance, because there are few districts that belong to the category) they are shrunk towards zero, and the expected level of poverty is shrunk towards the grand mean for the country (see Gelman et al. 2008; Ghitza and Gelman 2013).

The plots in Figure 3 display the estimates from these models. For each category of road cost and soil quality, we display the expected level of poverty based on the estimates of the saturated model. Each line is for a given level of road cost (with darker lines indicating higher road cost –hence worse infrastructure). On the horizontal axis is the Soil quality index, and on the vertical axis the poverty level (re-centered). The plot on the top left is for a model that only includes fixed effects for soil-road combination and country fixed effects, while the plot on the top right is for a model that also includes controls: urbanization, the cultivation measure, the dummy for the presence of a river in the district, distance from the capital, distance from towns of at least 20,000 inhabitants, distance from the coast, distance from the border, and quadratic polynomials for elevation and terrain slope.

From the inspection of the two plots, one can infer that while poverty increases (all else equal) with road cost, the association between soil quality and poverty is far from straightforward. In the top two categories of road cost (the most remote districts with the worst transportation infrastructure) poverty turns out to be higher in districts with relatively high soil quality. While poverty is lower on average in the districts with the very best soil (category 6), categories 3 to 5 of soil quality have on average higher poverty than districts with soil in the lowest categories. The evidence suggests that poverty is at its worst when we have a combination of good soil and very poor infrastructure.

The bottom two plots in Figure 3 probe this relationship further, by displaying the results from estimation that models soil/road combination categories as random effects. The left plot only includes the random effects for soil and road combination and the country fixed effects, while the right plot includes the controls listed above.

Table 3 reports in tabular form the same information. For each combination of soil quality and road cost, we report the expected value of poverty according to the model estimates. The coefficients on the controls are, unsurprisingly, not different in any substantial way from those in the models reported in the main tables, hence we omit them to save space. Each row reports all the estimates and, for the fixed effects estimations, also the standard error (in the columns labeled "se"). So for instance, the first row in the table reports the estimates for road category one (districts with very low road cost) and soil category one (districts with very bad soil).

3.3 Excluding urban districts from the analysis and including additional controls

In all of the analysis above, we include all the districts in each country, but we account for level of urbanization in the regressions. Districts are often large enough to contain both urban and rural areas: excluding all the districts that contain also urban areas might be overkill. We show that the basic results are unaffected if we drop from the analysis districts that can be classified as "urban" according to some criteria. The first criterion is that the district lies in the top 2.5 percent of most

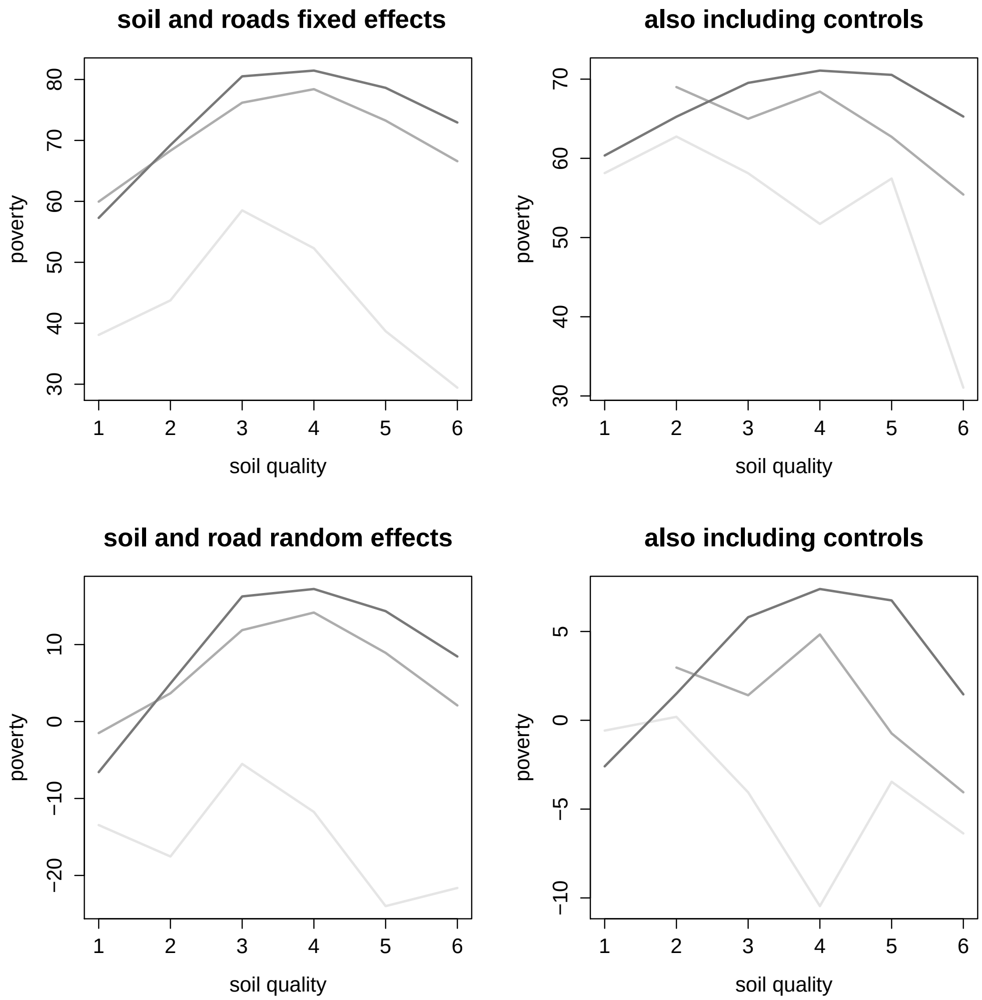

Figure 3: Saturated models for poverty. Darker lines represent districts with worse transportation infrastructure.

urbanized[17]; the second, that it lies in the bottom 2.5 percent of distance from the capital; the third is that every cell in the district is classified as urban according to the FAO data on rural population density. This leads to the exclusion of around 1100 districts from the analysis. Notice that these exclusion criteria are quite stringent: the median excluded district has around 15 percent of its surface classified as cultivated, and is only 6 percent urban overall. In addition, in columns 3 and 4 we add the main geographic controls (elevation and slope) and additional controls for distance from the coast, from the capital, from a town, and from the border, and a dummy for districts with a river. Importantly, the patterns we detect cannot be considered to be driven by comparisons we

[17]Specifically, we consider as urban a district that contains at least one cell with urbanization rate higher than 41 percent.

Table 3: Non-parametric interactive models for poverty rate

road cost	soil	fe	se	fe+controls	se	re	re+controls
1	1	38.09	13.96	58.14	17.01	-13.45	-0.58
1	2	43.75	6.05	62.75	12.17	-17.53	0.19
1	3	58.53	3.06	58.12	4.65	-5.51	-4.06
1	4	52.31	2.30	51.72	4.19	-11.75	-10.46
1	5	38.70	3.97	57.45	5.60	-23.98	-3.46
1	6	29.43	11.26	31.04	12.08	-21.63	-6.37
2	1	59.96	19.52			-1.50	
2	2	68.31	5.26	69	6.43	3.67	2.97
2	3	76.19	2.10	64.99	3.93	11.87	1.41
2	4	78.41	1.80	68.43	3.86	14.16	4.83
2	5	73.26	2.52	62.72	4.18	8.92	-0.74
2	6	66.59	5.83	55.43	6.83	2.09	-4.05
3	1	57.29	3.91	60.36	4.66	-6.58	-2.59
3	2	69.24	2.36	65.25	3.92	4.95	1.50
3	3	80.52	1.68	69.53	3.75	16.25	5.80
3	4	81.46	1.56	71.08	3.79	17.22	7.39
3	5	78.63	1.82	70.54	3.89	14.34	6.75
3	6	72.94	3.02	65.27	4.42	8.45	1.46

Estimates from the saturated models, with district-level poverty rate as dependent variable. In the columns labeled “fe” the estimates come from the fixed-effects estimates, in the columns labeled “re” from random effects. The columns labeled “se” report the standard errors from the fixed-effects estimation.

are carrying out between rich, mostly urban areas (for which soil quality does not matter) and poor, generally rural, areas. The results are reported in Table 4.

3.4 Understanding the distribution/allocation of infrastructure

In order to understand the phenomenon we highlight, we must examine which systematic patterns can be detected in the allocation of infrastructure. For this purpose, we regress the measures of road quality on geographic characteristics of the districts. In the first three columns of Table 5, we report models that regress infrastructure on quadratic terms for elevation and terrain slope, on the dummy for river districts, and on the measures of distance from rivers, distance from the coast, and distance from current borders. The measure of infrastructure we use is such that higher values reflect higher transportation costs, hence worse transportation infrastructure. Positive coefficients therefore mean that a given variable is associated with worse roads. First of all, river

Table 4: Excluding urban districts

robustness	1	2	3	4
(Intercept)	79.25**	79.05**	79.98**	80.02**
	(0.26)	(0.24)	(3.39)	(3.4)
cen(additive.index.soil.quality)	-3.18**		-2.33**	
	(0.42)		(0.3)	
st.var(road.cost)	3.54	3.46	5.33**	5.29**
	(2.38)	(2.31)	(1.68)	(1.61)
cen(additive.index.soil.quality):st.var(road.cost)	-2.96**		-3.32**	
	(0.8)		(0.57)	
cen(Comp.1)		4.09**		3**
		(0.5)		(0.36)
cen(Comp.1):st.var(road.cost)		4.12**		4.54**
		(0.97)		(0.69)
distance$_{coast}$			0.89**	0.87**
			(0.11)	(0.11)
river$_{d}istrictTRUE$			0.86	0.87
			(0.74)	(0.75)
elevation			1.79**	1.82**
			(0.37)	(0.37)
I(elevation^2)			-0.07**	-0.07**
			(0.02)	(0.02)
median.slope			4.68	4.5
			(3.14)	(3.13)
I(median.slope^2)			-1.45*	-1.44*
			(0.68)	(0.68)
distance.town			-1.25**	-1.25**
			(0.24)	(0.24)
distance.border			-0.87	-0.87
			(1.51)	(1.51)
distance.capital			-0.25	-0.25
			(0.17)	(0.17)

More urbanized districts are excluded from the analysis. Standard errors clustered at the country level in parentheses. $+$: statistically significant at the 10% level. $*$: statistically significant at the 5% level. $**$: statistically significant at the 1% level.

districts, and districts closer to rivers, tend to have significantly worse infrastructure. In addition, infrastructure is worse in districts farther from the coastline, in districts closer to borders, and in districts farther from the current capital. Finally, elevation and slope have non linear effects. In the case of elevation, both terms are positive, implying that districts at sea level have better infrastructure than districts at higher altitude.

Table 5: The mismatch between good soil and good infrastructure

DV:	Road cost (1)	Road cost (2)	Road cost (3)	Soil prod. (4)	Soil prod.(5)	Soil prod. (6)
Intercept	4.73^{**}	4.73^{**}	4.79^{**}	-8.6^{**}	2.29^{**}	2.35^{**}
	(0.11)	(0.1)	(0.11)	(0.99)	(0.36)	(0.35)
River district	0.12^{**}	0.12^{**}		0.13	0.08	
	(0.03)	(0.03)		(0.09)	(0.05)	
Elevation	0.01	0.01	0.02	-0.04	-0.04^{*}	-0.03^{*}
	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.01)
Elevation^2	0	0	0	0.01^{*}	0^{+}	0^{+}
	(0)	(0)	(0)	(0)	(0)	(0)
Median slope	-0.14	-0.13	-0.13	8.75^{**}	1.38^{**}	1.37^{**}
	(0.11)	(0.1)	(0.1)	(0.92)	(0.34)	(0.34)
Median slope^2	0.03	0.03	0.04^{+}	-2.05^{**}	-0.25^{**}	-0.24^{**}
	(0.02)	(0.02)	(0.02)	(0.21)	(0.07)	(0.07)
Distance coast	0.07^{*}	0.07^{*}	0.05	0.12	0.08^{+}	0.07
	(0.03)	(0.03)	(0.04)	(0.08)	(0.05)	(0.05)
Distance capital	0.42^{**}	0.41^{**}	0.44^{**}	-0.52^{**}	-0.15^{**}	-0.12^{**}
	(0.05)	(0.05)	(0.05)	(0.13)	(0.05)	(0.04)
Distance border		-0.04^{+}	-0.07^{**}			-0.01
		(0.02)	(0.02)			(0.03)
Distance river			-0.1^{**}			-0.09
			(0.03)			(0.07)
Distance 1950 border			0.04^{**}			
			(0.01)			

In columns 1-3 the dependent variable is road cost; in columns 4-6 the dependent variable is soil quality (the soil production index). Standard errors in parentheses. $+$: statistically significant at the 10% level. $*$: statistically significant at the 5% level. $**$: statistically significant at the 1% level.

Where is the good land? Here we describe in general terms, some patterns of association between soil quality and observable geographic characteristics of the districts. We regress the measures of soil quality on the quadratic terms for elevation and slope, the dummy for river district, another measure of presence of rivers (the class rank of the largest river in the district), and distance from the coast. The best land is located in river districts (even if the evidence is not strong enough to be statistically significant at conventional levels), and in districts farther from the coast (also in this case not statistically significant). In addition, soil quality is not correlated with distance from the border, while it is worse farther from the capital. Again, we estimate quadratic polynomial terms in elevation and terrain slope. According to the estimates, soil quality is maximum (all else equal) at an altitude of 200 meters. In addition, the coefficients on terrain slope also point to the fact that soil quality is highest in "hilly" areas: soil quality is at a maximum when the median slope of the district is in class three. We explore these non-linear effects in Figure 4.

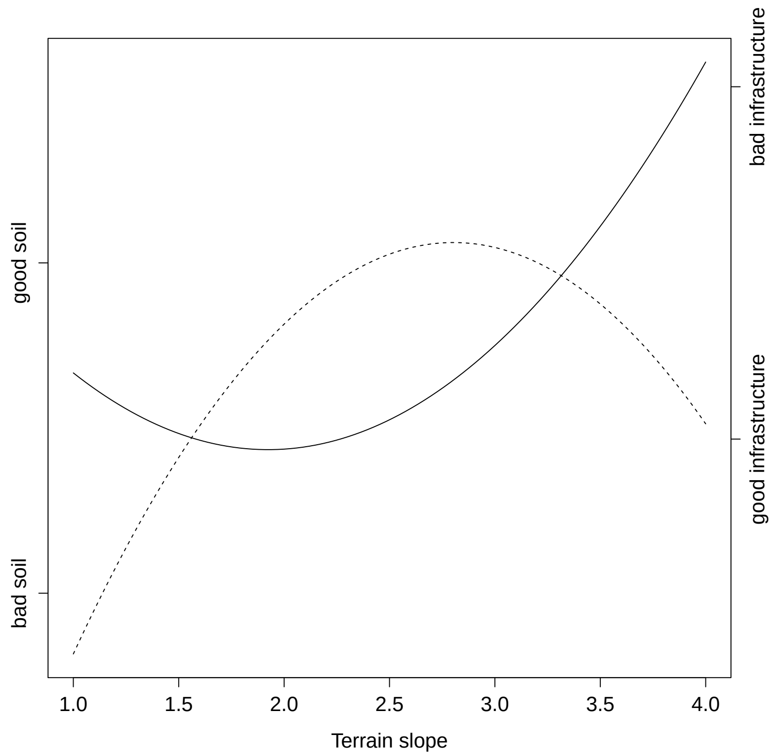

Figure 4: Expected values of soil quality (dashed) and road cost (solid) as a function of median terrain slope in the district, from the models respectively columns 2 and 5 of Table 5.

The disconnect between good land and infrastructure The plot in Figure 4 displays the expected value of soil quality (the dashed line) and of road cost (the solid line) as a function of the median terrain slope of the district. Higher values of road cost reflect lower transportation infrastructure provision. The best soil is estimated to be located, all else equal, in hilly districts, while the worst is in flat areas and, to a lesser extent, in mountain areas. The worst infrastructure is found in mountainous areas with very steep terrain (i.e., right side of the plot). There is some mismatch between soil quality and infrastructure provision, on average: the best land from the point of view of soil quality is not the tract that attracts the best transportation infrastructure.

Our results implicitly suggest that rugged terrain is associated with poverty in Africa through its effect on infrastructure provision. From a superficial point of view, this appears to be at odds with the results of Nunn and Puga (2012) which indicate that ruggedness has a positive effect on economic development in Africa, since more rugged African countries have experienced less slave exports. However, we should note that their measure of ruggedness captures “small-scale terrain irregularities, such as caverns, caves, and cliff walls, that afforded protection to those being raided during the slave trades” (p. 21), while our geographic controls (average elevation and average slope in the district) capture more “macro” features of a given territory. In addition, once quality of institutions (measured by an index of rule of law) is included in their model, Nunn and Puga (2012) find that their measure ruggedness has no association with level of development. Our models include country fixed effects, that capture all the institutional features of the country that do not vary across districts: our results for geographic characteristics then have to be interpreted net of (country-specific) indirect effects through quality of institutions.

3.5 Accounting for spatial correlation

One obvious concern one might have is that there is remaining correlation in the errors across contiguous districts. In order to assess the robustness of the findings to potential spatial correlation of the errors, we estimate the basic specification with spatially-modeled errors. Namely, we use spatial simultaneous autoregressive error models and autoregressive spatial lag models, estimated via maximum likelihood. The contiguity matrix is built starting from the map of African districts described above. Importantly, the results are robust to direct modeling of spatial correlation. Notice also that we probe the robustness of the barebones models: not including many control regressors increases the chances that there is remaining correlation between errors for neighboring districts, but our results still carry through. The results of two specifications, one additive and one interactive, estimated using the two methods, are reported in table 6.

Table 6: Models with spatial errors

gigi	1	2	3	4
(Intercept)	63.05**	21.45**	63.44**	21.46**
	(3.55)	(1.36)	(3.56)	(1.36)
st.var(additive.index.soil.quality)	-16.61**	-13.03**	-14.78**	-11.6**
	(0.53)	(0.44)	(0.58)	(0.5)
st.var(road.cost)	7.03**	8.29**	6.59**	8**
	(0.51)	(0.47)	(0.5)	(0.47)
st.var(additive.index.soil.quality):st.var(road.cost)			-7.83**	-5.92**
			(1.02)	(1.03)

4 Endogeneity of infrastructure

One might be more concerned that the associations we report are driven by possible endogeneity of the amount of infrastructure to rural poverty. In other words, current infrastructure provision might be driven by current rural poverty, or by other (unobserved) contemporary features that also affect poverty levels. We want to show that the systematic relationship between infrastructure provision and rural poverty that we document survives if temporally pre-determined instruments are used to correct for possible endogeneity of current transportation infrastructure to rural poverty.

4.1 Mining regions, least-cost paths, and first-stage relationship

We first want to assess to what extent the layout of colonial roads can be predicted based on our variables calculating least-cost paths (and shortest paths – i.e., straight lines) from mining areas to ports and cities. Then, we want to examine the extent to which the encouragement (i.e., the fact that a district lies on the least-cost or shortest path) is related to current road cost (which can be considered a measure of "take-up" of the encouragement). This is, in practice, the "first stage" in a 2SLS estimation where the variables based on the mining information are used as instruments for current roads providion. In addition, we also show that current transportation infrastructure is closely related to the presence of colonial era infrastructure.

The location of colonial-era roads As we discussed in subsection 2.2, our identification strategy relies on the fact that a good deal of the colonial-era transportation infrastructure was built to connect mining areas to the coast. We now show that, in fact, districts that lie on the least-cost and shortest path from a mine to a port or a city are more likely to have a colonial-era road. The first four columns of Table 7 report the estimates of models in which respectively an indicator for presence of a first-class colonial-era road, and (log) distance from a first-class colonial-era road, are regressed (in the order) on an indicator for whether the district lies on the least-cost path, distance from the least cost path, an indicator for whether the district lies on a straight line from a mine, and distance from the straight line. In all the models, we control for the presence of a mine in the region given that, trivially, the least-cost and the shortest path pass through districts where mines are located. As one can appreciate, the least-cost and shortest paths have high predictive power for the layout of colonial-era roads.

Contemporary infrastructure and colonial legacy The fifth and sixth columns of Table 7 show the estimates of models that predict contemporary road cost as a function of the presence of colonial transportation infrastructure. In column 5, we report a model with road cost regressed on (log) distance from a colonial-era first-class road. The coefficient on distance from colonial-era roads is positive, and statistically significant, indicating that road cost is higher (transportation infrastructure is worse) in districts farther from colonial-era infrastructure, also after accounting for the geographic features of the district. In column 6, we estimate a logit model with, as dependent variable, a dummy equal to one if the district is crossed by a highway or a first-class road according to the gRoads dataset discussed above. This provides a more direct indication of whether there is a high-quality road in the district. The coefficient on the indicator for the presence of a colonial-era first class road in the district is positive and statistically significant, showing that districts with colonial-era infrastructure are still much more likely to have a higher-quality road in current times. These estimates show that there is remarkable persistence from the layout of roads in the colonial era and contemporary road networks.

Contemporary infrastructure and least-cost paths Finally, we want to look at the extent to which contemporary infrastructure is predicted by least-cost paths from mines to ports and cities, and, most importantly, after controlling for geographic characteristics of the district. The last two columns in Table 7 report the estimates of regressions of the measure of road cost on the distance from a least-cost path and distance from the straight line connecting port and mine. As we said, these are the first stages of the two-stage least squares estimates we report in the next subsection. As one can appreciate, contemporary road cost is strongly predicted by the least-cost paths from mines.

4.2 IV REGRESSION RESULTS

Having established that there is a strong relationship between the least-cost paths and current transporation infrastructure, we re-estimate models analogous to those in section 3 instrumenting for current transportation infrastructure with the variables based on least-cost, shortest-path, and presence of colonial-era transportation infrastructure discussed in subsection 2.2. The results of the estimations are shown in Table 8. In the first column, we instrument road cost with the distance from a least-cost path from a mine and a port; in the second column, we also add distance from the straight line connecting a mine and a port; in the third column, we also include the distance from a colonial-era first class road as an instrument. All these models use the the additive measure of soil constraints (with higher values, as usual, indicating worse soil), the standard measure of road cost, and and their interaction. As in standard practice, we use the interaction between the instruments and soil quality as instruments for the interaction between road cost and soil quality.

In the next three columns, we add the set of geographic controls: polynomials for elevation and slope, distance from the coast, and a dummy for the presence of a river. Again we instrument for road cost with distance from a least-cost path (columns 4-6), distance from the straight line (added only in columns 5 and 6) and distance from a first-class colonial-era road (added only in column 6).

The models include country fixed effects, and the standard errors are clustered at the country level to account for any remaining correlation across shocks to observations within the same country.

The results of these IV estimations support the conclusions of the basic analysis: in particular, the interaction between road cost and the soil constraints measure is negative, implying that better soil (a lower value of the additive measure) is associated with more poverty when roads are sufficiently bad. In the models in the third and sixth columns of table 8 the interaction term is not statistically significant, but road cost has a strong effect (not driven by endogeneity) on rural poverty rate.

5 Conclusion

In this paper, we detect a surprising empirical regularity, the positive correlation between soil quality and poverty in Africa, and then provide rigorous evidence about how this pattern emerges by linking it to resource under-utilization due to lack of infrastructure and limited access to markets. We exploit plausibly exogenous variation in the quality of roads induced by the fact that the transportation network in Sub-Saharan Africa during the colonial period was designed with the main aim of connecting mineral resources to coastal cities and ports.

Our main priority, in this paper, is to highlight that infrastructure plays a larger role than soil quality in determining rural poverty. In addition, we also show that good soil, if combined with severe isolation, can generate a "curse of good soil", in that among isolated regions, having better soil might lead to higher rates of rural poverty. A full exploration of the mechanisms behind the "curse of good soil" that we detect requires a separate study. Yet, our findings are compatible with a mechanism that involves both fertility decisions (and therefore population density) and human capital accumulation. Human capital might provide a possible mechanism mediating the relationship between soil quality, isolation, and rural poverty, and potentially leading to the "curse" of good soil. In a framework akin to the one proposed by Becker et al. (1994), areas with higher quality soil but no roads (and no access to markets) might be prone to be stuck in a Malthusian trap. To put it simply, isolated districts endowed with high quality soil might produce enough food to sustain a relatively larger population than equally isolated districts with lower-quality soil. This leads to higher fertility rates, and at the same time, lower human capital accumulation. Dwelling on good soil means that there's enough food for large families to sustain themselves at basic subsistence level. At the same time, isolation and lack to access to markets makes it impossible to mobilize the available resources, in terms of farmland, to start development: the returns to human capital investment, in isolated areas with very limited access to markets, are small.

In light of these results, the current emphasis on fertilizer use among a number of policy-makers and agricultural economists could be misguided; more attention should be paid to rural infrastructure and human capital investment. Roads facilitate access to markets, and the presence of schools locally provides incentives for investment in human capital. The low rate of adoption of new technologies and the reluctance to use fertilizer in many rural communities might be driven, paradoxically, by the combination of poor rural infrastructure and the availability of relatively high-quality land. Therefore the growing interest among development agencies on rural infrastructure and local public goods in general is a promising avenue for rural development and poverty alleviation in Africa.

In addition, it is important to point out the link between rural poverty and the "mismatch" between road quality and soil quality that we detect. This is likely a consequence of the priorities of the colonial powers, that were arguably more interested in connecting mining areas to seaports rather than providing transportation from and to farmland. Due to the persistence over time of the layout of transportation infrastructure, there is a mismatch today between where the good land is and where the transportation network lies. This leads to the isolation of areas with good soil, which might therefore be stuck in a Malthusian trap with low human capital accumulation.

Importantly, persistence over time of the layout of transportation networks is not unique to Sub-Saharan Africa: for instance, the layout of the current highway network of France can be traced back to the Royal roads network, which itself followed the Roman network (Rodrigue et al. 2006, p.11). Carreras and De Soto (2013) observe that the Roman network constitutes "a precedent" for the integration of European mobility. Transportation networks in Europe and Asia, though, were most commonly designed with trade and transportation of foodstuffs in mind –rather than mineral extraction like in Sub-Saharan Africa. Moving and storing food, and guaranteeing food security to the capital and other large cities, was a central concern in China; for instance, integration of grain markets at the provincial level, supported by transportation infrastructure, is well documented for the Qing period (Li 2000). Grain trade, and guaranteeing the food supply of large cities, was also a central priority in the Roman Empire (Kessler and Temin 2007). One of the issues that Sub-Saharan Africa might face, then, is that its transportation network, inherited in part from the colonial era, is designed based on completely different considerations than those that shaped networks in Europe and Asia.

At the same time, one can also speculate that the extent to which the road network is not matching the quality of agricultural soil is a consequence (and therefore an indirect measure) of quality of governance in the post-colonial period. Better governed countries, indeed, should face a less serious mismatch, if they have invested more resources to improve the layout of their transportation networks to integrate farmland in local, national, and international markets. Understanding the political economy considerations that might explain variation in how much the "mismatch" has been amended constitutes a promising avenue for future research.

Table 7: Understanding the layout of colonial and current road networks

firststage	1	2	3	4	5	6	7	8
(Intercept)	0.73^{**}	0.04	0.63^{**}	0.16^{**}	4.61^{**}	-24.13	4.33^{**}	4.68^{**}
	(0.15)	(0.02)	(0.16)	(0.01)	(0.12)	(1371.07)	(0.1)	(0.1)
Least-cost path (to port)	0.47^{**}							
	(0.14)							
Least-cost path (to city)	0.14							
	(0.13)							
Mining district	0.43^{*}	0.01	0.3^{+}	0				
	(0.18)	(0.02)	(0.18)	(0.03)				
Distance least-cost path (to port)		0.02^{+}					0.08^{**}	
		(0.01)					(0.02)	
Distance least-cost path (to city)		0.06^{**}					0.14^{**}	
		(0.01)					(0.03)	
On shortest path (dummy)			0.48^{**}					
			(0.09)					
Distance from shortest path				0.04^{**}				0.09^{**}
				(0.01)				(0.03)
Distance from colonial road					0.59^{**}			
					(0.09)			
Has colonial first-class road (dummy)						1.41^{**}		
						(0.17)		
distance coast					0.02^{**}	-0.07^{+}	0.03^{**}	0.03^{**}
					(0)	(0.04)	(0)	(0.01)
elevation					0.01	0.18^{**}	0.01	0.02
					(0.02)	(0.05)	(0.02)	(0.02)
I(elevation^2)					0	0^{*}	0	0
					(0)	(0)	(0)	(0)
median.slope					-0.07	1.16^{+}	-0.12	-0.11
					(0.11)	(0.63)	(0.1)	(0.1)
I(median.slope^2)					0.02	-0.26^{+}	0.03	0.02
					(0.02)	(0.14)	(0.02)	(0.02)
Country F.E.	Y	Y	Y	Y	Y	Y	Y	Y

In columns 1 and 3 the dependent variable is a dummy for whether the district had a first class road in the colonial era; in columns 2 and 4 the dependent variable is (log) distance from a colonial-era first class road. In columns 5, 7, and 8 the dependent variable is the road cost index, and in column 6, the dependent variable is a dummy for whether the district has a higher-quality road in current times. Standard errors clustered by country in parentheses. $+$: statistically significant at the 10% level. $*$: statistically significant at the 5% level. $**$: statistically significant at the 1% level.

Table 8: Instrumental variables estimation

ivproper	1	2	3	4	5	6
(Intercept)	74.67**	74.85**	77.62**	80.88**	78.37**	76.43**
	(0.7)	(0.63)	(0.3)	(4.9)	(5.87)	(5.6)
st.var(road.cost)	21.01**	19.9**	6.69*	23.14**	0.08	6.25*
	(2.56)	(2)	(2.93)	(1.64)	(1.64)	(2.86)
additive.index.soil.quality	-1.54	-1.78+	-2.93**	-1.09	-1.03	-2.34**
	(1.14)	(1.01)	(0.5)	(0.94)	(0.97)	(0.66)
st.var(road.cost):additive.index.soil.quality	-8.43*	-7.33*	-1.5	-8.48**	-7.65*	-2.76
	(3.84)	(3.52)	(1.61)	(2.95)	(2.96)	(2.2)
elevation				1.91**	2.02**	1.94**
				(0.55)	(0.44)	(0.46)
I(elevation^2)				-0.08*	-0.09**	-0.08**
				(0.03)	(0.02)	(0.02)
median.slope				-2.27	-0.71	1.45
				(4.4)	(5.3)	(4.75)
I(median.slope^2)				-0.01	-0.36	-0.93
				(0.95)	(1.13)	(1)
distance.capital				-1.15**	-0.68**	-0.78**
				(0.12)	(0.16)	(0.14)
distancecoast				1.02**	1.17**	1.08**
				(0.16)	(0.21)	(0.18)
riverdistrictTRUE				0.05	1.99*	1.94+
				(1.03)	(0.97)	(1)

REFERENCES

Ali, Rubaba. 2010. "Impact of rural road improvement on high yield variety technology adoption: Evidence from Bangladesh." Unpublished manuscript, University of Maryland Department of Agricultural and Resource Economics.

Amaral, Pedro S., and Quintin Erwan. 2006. "A competitive model of the informal sector." *Journal of Monetary Economics* 53(7):1541-1553.

Angrist, Joshua. D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Arellano, Manuel. 1987. "Computing Robust Standard Errors for Within-groups Estimators." *Oxford Bulletin of Economics and Statistics* 49(4)431-434.

Ayogu, Melvin. 2007. "Infrastructure and Economic Development in Africa: A Review." *Journal of African Economies* 16: 75–126.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, 2012. *On the Road: Access to Transportation Infrastructure and Economic Growth in China.* National Bureau of Economic Research Working Paper No. 17897.

Barbier, E.B. 2010. "Poverty, development and environment." *Environment and Development Economics* 15 (Special issue 06): 635-660.

Becker, G.S., Murphy, K.M. and Tamura, R., 1994. Human capital, fertility, and economic growth. In Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition) (pp. 323-350). The University of Chicago Press.

Bell, Clive, and Susanne van Dillen. 2012. "How does India's rural roads program affect the grassroots? Findings from a survey in Orissa." Policy Research Working Paper Series 6167, The World Bank.

Blimpo, M. P., Harding, R., and Wantchekon, L. 2013. Public Investment in Rural Infrastructure: Some Political Economy Considerations. *Journal of African Economies*, 22(2), 57-83.

Boone, C. 2009. Electoral populism where property rights are weak: land politics in contemporary sub-Saharan Africa. *Comparative Politics*, 183-201.

Burgess, R., Jedwab, R., Miguel, E. and Morjaria, A., 2015. The value of democracy: evidence from road building in Kenya. The American Economic Review, 105(6), pp.1817-1851.

Carreras, C. and De Soto, P., 2013. "The Roman transport network: a precedent for the integration of the European mobility." Historical Methods: A Journal of Quantitative and Interdisciplinary History, 46(3), pp.117-133.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri. 2010. "Rural roads and intermediated trade: Regression discontinuity evidence from Sierra Leone." Working Paper.

Christiaensen L. and L. Demery. 2007. "Down to earth: Agriculture and Poverty Reduction in Africa, Directions in development," World bank: Washington D.C.

Collischonn, W. and Pilar, J.V., 2000. "A direction dependent least-cost-path algorithm for roads and canals." International Journal of Geographical Information Science, 14(4), pp.397-406.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly exogenous." *The Review of Economics and Statistics* 94 (1):260-272.

Dercon S. and L. Christiansen. 2011. "Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia." *Journal of development Economics* 96(2): 159-173.

Dijkstra, E. 1959. "A Note on Two Problems in Connexion with Graphs." Numerische Mathematik, 1(1), 269271.

Dreschel, Pay Lucy Gyiele, Dagmar Kunze, and Olufunke Cofie. 2001. "Population density, soil nutrient depletion, and economic growth in sub-Saharan Africa." *Ecological Economics* 38(2):251-258.

Ehui, S. and Pender, J. 2005. "Resource degradation, low agricultural productivity, and poverty in sub-Saharan Africa: pathways out of the spiral." *Agricultural Economics*, 32: 225-242.

Ephraim Nkonya, Nicolas Gerber, Philipp Baumgartner, Joachim von Braun, Alex De Pinto, Valerie Graw, Edward Kato, Julia Kloos, and Teresa Walter. 2011. “The Economics of Land Degradation." Peter Lang Internationaler Verlag der Wissenschaften, 2011.

Elvidge, Christopher D., Paul C. Sutton, Tilottama Ghosh, Benjamin T. Tuttle, Kimberly Baugh, Buhendra Bhaduri, and Edward Bright. 2009. “A global poverty map derived from satellite data.” *Computers & Geosciences* 35: 1652-60.

Environmental Systems Research Institute, Inc. 1992. *Esri ArcWorld Database* [computer file]. Redlands, CA: Environmental Systems Research Institute.

Faber, Benjamin. 2012. “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System.” Unpublished manuscript, London School of Economics.

FAO. 2007. *Soil Production Index* [computer file]. Rome, Italy: FAO.

`http://www.fao.org/geonetwork/srv/en/resources.get?id=30577`

FAO-UNESCO Soil Map of the World. 1974. 1:5.000.000. UNESCO, Paris.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables. 1999. The Spatial Economy: Cities, Regions, and International Trade. Cambridge, MA: MIT Press.

Garcia-López, Miquel-Angel, Adelheid Holl, and Elisabet Viladecans-Marsal. 2015. “Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities.” Journal of Urban Economics 85: 52-67.

Gelman, Andrew, Aleks Jakulin, Maria G. Pittau, and Yu-Sung Su. 2008. “A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models.” *The Annals of Applied Statistics* 2(4):1360-1383.

Gelman, Andrew, David Park, Boris Shor, Joseph Bafumi, and Jeronimo Cortina. 2008. *Rich State, Poor State, Red State, Blue State: Why Americans Vote the Way They Do.* Princeton: Princeton University Press.

Ghitza, Yair, and Gelman, Andrew. 2013. “Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups”. American Journal of Political Science 57(3):762-776.

Gibson J. and S. Rozelle. 2003. “Poverty and Access to roads in Papua New Guinea.” *Economic Development and Cultural Change* 52(1): 159-185.

Glaeser, E.L., Kerr, S.P. and Kerr, W.R., 2015. Entrepreneurship and urban growth: An empirical assessment with historical mines. Review of Economics and Statistics, 97(2), pp.498-520.

Gollin Douglas, and Richard Rogerson, 2014. “Productivity, Transport Costs and Subsistence Agriculture”. Journal of Development Economics 107: 38-48.

Goyal, Aparajita. 2010. “Information, direct access to farmers, andrural market performance in central India.” Policy Research Working Paper Series 5315, The World Bank.

Holl, Adelheid. 2011. “Market potential and firm-level productivity in Spain." Journal of Economic Geography (2011): lbr030.

Holl, A., 2016. “Highways and productivity in manufacturing firms.” Journal of Urban Economics, 93, pp.131-151.

Jedwab, Remi, and Alexander Moradi, 2016. “Colonial Investments and Long-Term Development in Africa: Evidence from Ghanaian Railways." The Review of Economics and Statistics 98(2): 268â“284

Jacoby, H. 2000. “Access to markets and the benefits of rural roads.” *Economic Journal* 465:713-737.

Jacoby, H. and Minten, B. 2009. “On Measuring the benefits of lower transport costs.” *Journal of Development Economics*, 89:28-38.

Kessler, David, and Peter Temin. 2007. “The organization of the grain trade in the early Roman Empire.” The Economic History Review 60, no. 2: 313-332.

Khandker, S., Bakht, Z., and Boolwal, G. 2009. “The Poverty Impact of Rural Roads: Evidence from Bangladesh.” *Economic Development and Cultural Change*, 57(4):685-722.

Khemani, S. (2010). “Political economy of infrastructure spending in India.” World Bank. `http://openknowledg`

Kramon, E., and Posner, D. N. (2013). Who benefits from distributive politics? How the outcome one studies affects the answer one gets. *Perspectives on Politics*, 11(02):461-474.

Lewis, W. A. (1955). *The Theory of Economic Growth.* Homewood, IL: Richard Irwin.

Li, Lillian M. 2000. “Integration and Disintegration in North China’s Grain Markets, 1738â“1911.” The Journal of Economic History 60, no. 03: 665-699.

Licio, Vania Manuela, and Anna Maria Pinna. 2016. “The long-term effects of the historical Roman road network: Trade costs of Italian provinces.” Working paper, University of Cagliari and CRENoS.

McMillan, M.S., and Rodrik, D. 2011. “Globalization, Structural Change and Productivity Growth.” NBER Working Paper Series 17143.

Michaels, Guy. 2008. “The effect of trade on the demand for skill: Evidence from the interstate highway system." The Review of Economics and Statistics 90, no. 4: 683-701.

Minten, B., Koru, B. and D. Stifel. 2013. “The last mile(s) in modern input distribution: Pricing, profitability, and adoption.” *Agricultural Economics* 44: 1-18.

Morris, M., Kelly, V.A., Kopicki, R.J. and R.J. Byerlee. 2007. “Fertilizer Use in African Ariculture - Lessons Learned and Good Practice Guidelines, Directions in Development, Agriculture and Rural development.” World Bank: Washington D.C.

Mu, R. and D. van de Walle. 2011. “Rural Roads and Local Market development in Vietnam.” *Journal of Development Studies* 47(5): 709-734.

Njoh, A.J. 2000 . “Transportation infrastructure and economic development in sub- Saharan Africa.” Public Works Management and Policy, 4(4), 286-296

Nunn, N., and D. Puga. 2012. “Ruggedness: The Blessing of Bad Geography in Africa.” *Review of Economics and Statistics* 94(1): 20-36.

Okwi, Paul O., Godfrey Ndeng'e, Patti Kristjanson, Mike Arunga, An Notenbaert, Abisalom Omolo, Norbert Henninger, Todd Benson, Patrick Kariuki, and John Owuor. 2007. "Spatial determinants of poverty in rural Kenya." *Proceedings of the National Academy of Sciences of the USA (PNAS)* 104 (43):16769-16774.

Pingel, Thomas J. 2010. "Modeling slope as a contributor to route selection in mountainous areas." Cartography and Geographic Information Science 37, no. 2: 137-148.

Radeny, Maren, and Bulte, Erwin. 2012. "Determinants of Rural Income: The Role of Geography and Institutions in Kenya." *Journal of African Economies* 21(2):307-341.

Sanchez, Pedro A. 2002. "Soil fertility and hunger in Africa." *Science* 295(5562):2019-2020.

Scherr, Sara J. 1999. "Soil degradation: A threat to developing country food security in 2020?" Food, Agriculture and the Environment Discussion Paper 27. Washington DC: IFPRI.

Stifel, David, and Minten, Bart. 2008. "Isolation and agricultural productivity." *Agricultural Economics* 39(1):2019-2020.

Tobler, W. 1993. Non-isotropic geographic modeling (Technical Report No. 93-1). Santa Barbara, CA: National Center for Geographic Information and Analysis.

Whitley, T. G., Hicks, L. M. 2003. A Geographic Information Systems Approach to Understanding Potential Prehistoric and Historic Travel Corridors. Southeastern Archeaology, 22(1), 77-91.

Woomer, Paul L., and Michael J. Swift. 1994. "The Biological Management of Tropical Soil Fertility." *Tropical Soil Biology and Fertility Programme.* John Wiley, New York.

World bank. 2008. World Development Report: Agriculture for development. World Bank, Washington D.C.

Yamano, Takashi, and Yoko Kijima. 2010a. "The associations of soil fertility and market access with household income: Evidence from rural Uganda." *Food Policy* 35(1):51-59.

Yamano, Takashi, and Yoko Kijima. 2010b. "Market Access, Soil Fertility, and Income in East Africa." GRIPS Discussion Papers 10-22, National Graduate Institute for Policy Studies.

Yu, C., Lee, J., Nunro-Stasiuk, M. J. 2003. Extensions to least-cost path algorithms for roadway planning. International Journal of Geographic Information Science, 17(4), 361- 376.