CEPREMAP
CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS

Document de travail (Docweb) n°2110

Take the Highway? Paved Roads and Well-Being in Africa

Elodie Djemaï
Andrew E. Clark
Conchite D'Ambrosio

Octobre 2021

Prendre la route : Routes bitumées et Bien-être en Afrique[1]

Elodie Djemaï [2], Andrew E. Clark[3], Conchite D'Ambrosio [4]

Résumé : La provision de biens publics vise à améliorer le bien-être individuel. Nous étudions les conséquences causales des routes sur le bien-être dans 24 pays africains, en instrumentant les routes bitumées par les lignes hypothétiques reliant les ports et les villes principales au 19ème siècle. Nous disposons des données d'enquêtes collectées auprès de 32 000 individus, et considérons le bien-être objectif et subjectif. Les routes réduisent la privation matérielle, en termes d'accès aux besoins de base, mais les conditions de vie ne sont pas perçues comme meilleures par les individus vivant à proximité des routes. Ceci suggère que le bénéfice des routes en termes de satisfaction des besoins de base est compensé par des conséquences défavorables dans d'autres domaines.

Mots-clés : Routes, Bien-être Subjectif, Besoins essentiels, Privation matérielle

Take the Highway? Paved Roads and Well-Being in Africa

Abstract : Public Goods aim to improve individual welfare. We investigate the causal consequences of roads on well-being in 24 African countries, instrumenting paved roads by 19th Century hypothetical lines between major ports and cities. We have data on over 32 000 individuals, and consider both their objective and subjective well-being. Roads reduce material deprivation, in terms of access to basic needs, but at the same time there is no relation between roads and subjective living conditions. The benefit of roads in providing basic needs then seems to be offset by worse outcomes in non basic-needs domains.

Keywords : Roads, Subjective Well-being, Basic Needs, Material Deprivation, Africa

[1]We are grateful to Matthias Kroenke, Anthony Lepinteur, David Margolis, Cecilia Poggi, Rong Zhu and participants at the DIAL conference (Paris, 2019), ECARES (Brussels), the London School of Economics, the University of Luxembourg, the ITEA annual congress (Paris, 2019) and the Workshop on The Distributional Impact of Social Protection (Addis Ababa, 2020) for helpful comments. Clément Baticle provided excellent research assistance in coding the road network. Data Availability Statement: The data used in this article can be obtained upon request to Afrobarometer (https://afrobarometer.org/fr/données/politique-dutilisation-des-données) and the Demographic and Health Surveys (www.dhsprogram.com). Additional replication materials will be provided in the Online Appendix if the manuscript is accepted for publication. Declarations of interest: none. Financial support from CEPREMAP and the Fonds National de la Recherche Luxembourg (Grant C18/SC/12677653) is gratefully acknowledged. Andrew Clark acknowledges financial support from the EUR grant ANR-17-EURE-0001.

[2]Université de Paris Dauphine

[3]Paris School of Economics - CNRS

[4]Université du Luxembourg

1 Introduction

We often think of social protection at the individual level, for example via transfers and social safety nets. We here instead consider the role of public goods in protecting individuals, by potentially improving the environment in which they live (De Janvry and Sadoulet 2015). Our public good here is paved roads.

There is by now a large body of literature in economics evaluating the effect of transportation infrastructure on economic outcomes, both in developed and developing countries. The traditional outcomes investigated in this literature are pecuniary: poverty, consumption, income and investments (see, for example, Aggarwal, 2018; Asher and Novosad, 2020; Banerjee *et al.*, 2020; Buys *et al.*, 2010; Dercon *et al.*, 2009; Dillon *et al.*, 2011; Donaldson, 2018; Gibbons *et al.*, 2019; Gibson and Rozelle, 2003; Jacoby, 2000; Jacoby and Minten, 2009; Khandker *et al.*, 2009; Straub *et al.*, 2008; Wang and Wu, 2015). Here roads have mostly been shown to be beneficial. However, only few contributions have focussed on the effects of roads on access to basic needs, confirming their positive effects (see Bucheli *et al.*, 2018).

We here estimate the effect of road infrastructure on both objective and subjective well-being, measured for the same individuals in Wave 5 of the Afrobarometer survey (2011-2013), covering most African countries. While existing work has suggested that access to roads reduces poverty and increases consumption, their effect on subjective well-being has remained largely unexplored in the literature. One exception is Dumas and Játiva (2020), who find a small negative life-satisfaction effect of road improvements in Tanzania. Although economists have become increasingly open to the use of subjective data, the robust analysis of life evaluations remains only rare in some areas of the discipline and, in particular, in the analysis of developing countries. The comparison of objective and subjective well-being has been suggested as a useful tool in the analysis of poverty (for instance, Clark and D'Ambrosio, 2019; Pradhan and Ravallion, 2000; Ravallion, 2014).

The relationships with objective and subjective well-being might differ as roads also bring problems, for example via disease and accidents leading to worse health (see Djemaï, 2018 for HIV, and Riley-Powel *et al.*, 2018 for qualitative evidence on Dengue fever in Peru), environmental deterioration through reduced forest cover and biodiversity (Asher *et al.*, 2020; Damania *et al.*, 2018), pollution, crime, and lower social capital. In addition, transport infrastructure may affect individuals' evaluations of their lives through its effect on individual expectations, aspirations and the salience of different reference groups via the exposure to richer individuals.

Our objective well-being measure is an index of material deprivation in terms of the individual's access to basic needs (food, water, fuel and medical care). Material deprivation is a special case of multidimensional poverty when the dimensions considered are only related to material aspects of life. The subjective well-being measure we look at is the respondent's evaluation of her current living conditions. The same two individual well-being measures to which we appeal here were analysed in Clark and D'Ambrosio (2019), who used five rounds of Afrobarometer data over the 2004-2016 period to explore the association between the two. They find that, as expected, the more deprived in basic needs the individual is, the lower their evaluation of their current life, controlling for standard socio-economic and demographic variables, wave and region dummies. The estimated relationship is such that a one standard-deviation rise in the material-deprivation index is associated with a lower evaluation of current living conditions by around one quarter of a standard deviation.

Our goal in the current paper is to investigate the relationship between both objective and subjective well-being indicators and the distance to the nearest paved road, where we are able to instrument the latter to address problems of endogeneity. This, to the best of our knowledge, is the first to analyse the causal relationship between roads and multiple measures of well-being. We complement existing work that has taken an event-study approach (typically in a single country) with an IV analysis of road location in 24 countries across Sub-saharan Africa.

Wave 5 of the Afrobarometer survey is geo-coded and, when matched to the road network from Bing Maps using ArcGIS, allows us to measure the distance between the residential location of the respondent and the nearest paved road. The correlation between this distance and well-being is negative: people living further from roads are more materially-deprived and report lower satisfaction with current living conditions (although the latter effect is only small in size).

These correlations do not however reveal the causal effect of roads on well-being, as road placement is partly endogenous. We consider the main instruments used in the current infrastructure-location literature: the location of historical routes (Agrawal *et al.*, 2017; Baum-Snow *et al.*, 2017; Duranton and Turner, 2012; Martincus *et al.*, 2017), the straight lines connecting historical population centres (Atack *et al.*, 2010; Banerjee *et al.*, 2020; Bird and Straub, 2020; Donaldson, 2018; Faber 2014; Jedwab and Moradi, 2016), and land characteristics such as slope (Batzilis *et al.*, 2016; Dinkelman, 2011; Djemaï, 2018; Duflo and Pande, 2007; and Lipscomb *et al.*, 2013).

Our IV results confirm that material deprivation rises with distance to roads, but there is now no effect on current living conditions. We suggest that this contrast may reflect that roads reduce deprivation but also bring problems in other domains of life (for example, health, environment and crime); along the same lines, exposure to roads, and to new goods and richer people, may change the benchmark to which individuals compare when evaluating their lives. The IV road coefficient for deprivation is larger than that in OLS: roads were placed where individuals were more deprived.

We conclude by considering the potential role of migration, in that certain types of individuals may move closer to roads. As the Afrobarometer does not include migration information, we here turn to the Demographic and Health Surveys, in which only objective well-being appears. Our results show a consistent negative effect of road distance on the latter. When we split our sample based on migration status, we find that migration does not lie behind our core results: the effects of road distance on objective well-being is negative and significant (and of similar size) for both migrants and non-migrants.

The remainder of the paper is organised as follows. Section 2 describes the Afrobarometer survey and the geographical data. The naïve regression results appear in Section 3, the IV results in Section 4, and the robustness checks in Section 5. The potential role of mediators is investigated in Section 6, and Section 7 uses the Demographic and Health Surveys to explore the role of migration. Last, Section 8 concludes.

2 Data on Roads and Well-Being in Africa

2.1 The Afrobarometer Survey

We use data from the 5^{th} round of the Afrobarometer surveys collected in 2011-2013 in 30 countries in Sub-Saharan Africa.[1] Afrobarometer data has been widely used in analyses of development and resources in Africa, such as Nunn and Wantchekon (2011) and Cagé and Rueda (2016). This Afrobarometer round is the closest in time to the data on the current paved-road network (which comes from 2013), is geo-referenced, and includes questions on both self-assessed living conditions and access to basic needs.[2] We exclude from our sample the islands (Cape Verde, Madagascar and Mauritius), and the countries for which we do not observe the location of hypothetical lines as they were not included in Jedwab and Moradi (2016) (these are Lesotho, South Africa and Swaziland).

The resulting sample consists of around 38,000 individuals living in 307 administrative

[1] See http://afrobarometer.org/ for more details.

[2] We use data from the 5^{th} wave only as access to geo-coded Afrobarometer data is restricted: it is possible to obtain either geo-coded data on all countries in one wave only, or for one country over all Afrobarometer waves.

regions in 24 different countries (see Appendix Table A1). The number of respondents per country is around 1,200 in 16 of these 24 countries, and twice this size in the remaining eight. There are on average 125 observations per region.

As described in Afrobarometer (2015), the sample is nationally-representative and randomly-defined following a clustered, stratified, multi-stage design. The stratification is based on the largest sub-national administrative unit and urban-rural location. If the strata is rural, secondary sampling units are randomly selected within each strata, and two Primary Sampling Units, also called Enumeration Areas (EA), are drawn from each secondary sampling unit. If the strata is urban, the EA are selected in a direct way, that is without the preliminary selection of secondary sampling units. This is also the case for rural areas in some countries (e.g. Cameroon). Once the EAs have been drawn, eight households are randomly-selected by starting from a certain geographical point in the EA and then walking in turn in four directions that are 90 degrees from each other, and sampling the 5th and 10th dwelling in each of the four directions. The design of the survey is for eight households to be interviewed per enumeration area: in over 90% of EAs, this exact figure was reached. Most of the remaining EAs include either 16 or 24 interviews, reflecting over-sampling in some EAs. There are 4,105 EAs in our data.

One individual per sampled household is interviewed. Interviewers alternate interviews between men and women to reach a gender balance. All household members age 18 or older are eligible to be interviewed. If there is more than one household member of eligible sex and age, one is randomly-selected from the list of members.

The questionnaire for each individual includes an initial section, part of which indicates whether the EA is urban or rural: we thus typically have eight urban/rural evaluations per enumeration area. In 97% of cases, all of the urban/rural location evaluations are consistent with each other. For the remaining 3% (114 enumeration areas out of 4,105), some questionnaires in the enumeration area report this latter as being rural and some as urban. In these cases of disagreement, we designate an enumeration area as urban if 50% or more of the completed questionnaires in that area refer to the area as being urban. We end up with 36% of enumeration areas being urban, and the rest rural.

The Afrobarometer data was geo-coded after the data collection, using a double-blind methodology where two coders assigned latitude and longitude co-ordinates to each EA (in case of disagreement, there was an arbitration round: see BenYishay *et al.*, 2017). There are fewer GPS points than EAs: in a third of cases, GPS points covered households in more than one EA. There are 4,105 EAs in the Afrobarometer data, but 3,396 GPS points. All observations are assigned to a GPS point. We will call these points communities below, and it is at this level that we calculate the distance to the nearest paved road. We have information on the presence of various public goods, such as health centres and schools, for each EA: this is provided by the interviewer in conjunction with the field supervisor. With more than one EA per GPS point, it is thus possible for households in the same GPS point to have different levels of public goods.

2.2 Descriptive Statistics

Our two well-being measures are of subjective living conditions and objective material deprivation.[3] The data on self-assessed current living conditions come from the question "*In general, how would you describe your own present living conditions?*" The answers are on a five-point scale from "very bad" to "very good".[4] 4% of respondents reported that their living conditions were very good, 26% fairly good, 19% neither good nor bad, 31% fairly bad and 20% very bad. This distribution of answers is plotted in Panel a) of Figure A1. The means of all of our analysis variables appear in Table 1 (and Appendix Table A1 gives the means by country). The mean value of current living conditions, on the 1-5 scale, is 2.6. We also create a dummy variable for replying "very good or fairly good" to this question, which covers 30% of respondents.

Our second dependent variable is material deprivation, where respondents are asked about their difficulty in satisfying their basic needs in five dimensions. The five questions here are "*Over the past year, how often, if ever, have you or anyone in your family gone without enough food to eat, clean water for home use, medicines or medical treatment, fuel to cook your food, a cash income?*" The answers to each of these are given on a five-point scale: Never (0), Just once or twice (1), Several times (2), Many times (3) and Always (4). Due to concerns about collinearity between cash income and the other four elements of material deprivation, we drop the former and calculate a deprivation index as the sum of the answers to the first four questions. We then invert this figure in order to provide an index of access to basic needs, or lack of material deprivation, where higher values refer to better outcomes. This index thus ranges from zero (for respondents who are "always" deprived in these four dimensions) to 16 (for respondents who are never deprived in any dimension): this is our objective measure of well-being. In Table 1 the average value of this index is 11.4. In terms of the individual elements of the index, 60% of respondents declared that they have never or just once or twice gone without food, with analogous figures of 59%, 56% and 69% for water, medical care and cooking fuel respectively. The distributions

[3]The Afrobarometer survey does not include income or consumption information.

[4]The "Don't Knows" are recoded as missing values. These represent only 0.3% of the sample.

of the answers to the four retained basic-needs questions appear in Panels b)-e) of Figure A1 and Table A2. These are fairly similar across the domains, but with somewhat more material deprivation regarding medical care and less deprivation with respect to cooking fuel.

In terms of the exogenous control variables, Table 1 shows that average age is 36 and 50% of respondents are women (reflecting the sampling design). The remainder of the table refers to Afrobarometer variables that may themselves be affected by roads. Apart from urban, described above, these are education, labour-market status, health, public goods, the use of mobile phones and internet, crime, and social capital (trust and participation in social or religious groups). In our sample, 12% of respondents have higher education as their highest education level, 34% secondary education, 32% primary education and 22% no education. Labour-force status is measured as being a part- or full-time paid employee (which applies to just under one-third of our respondents). Health status is not collected as part of the survey, but we do know whether the respondent visited a Public Health Centre at least once over the past 12 months (85% did so).

A number of the other mediating variables are aggregate. Access to public goods within the enumeration area or within walking distance is reported by the interviewers and field supervisors. First, they report the presence of an electricity grid, piped water, sewage system and cell-phone service within the EA that most houses could access. 54% of the respondents live in a EA where there is electricity, 47% in an EA with piped water, 21% with a sewage system and 92% with cell-phone coverage. They then indicate whether there is a Post Office, school, Police Station, Health Clinic, and market stalls (selling groceries and/or clothing) in the EA or within walking distance to the EA. The lowest percentage here refers to Post Offices, where only 18% have access. 59% have access to a clinic, 88% to a school, 34% to a Police Station and 66% to a market.

In addition to cell-phone coverage at the EA level, respondents are asked about their own use of mobile phones, internet and computers. There are separate questions on the frequency of computer and internet use, with replies on a five-point scale from Never (0) to Every Day (4). For mobile phones, individuals report how often they normally use a mobile phone to make or receive calls, send or receive text messages, and send or receive money or pay a bill on a five-point scale from Never (0) to Five or More Times per Day (4). Computer and internet use in Africa in the early 2010's was only rare, with only one in seven respondents using a computer or internet. Mobile-phone use was more widespread: only 17% never used a mobile phone to receive or make a call, with analogous figures of 40% for text messages, and 80% for receiving or paying money. One-third of respondents use the mobile phone over 5 times a day to receive or make calls.

Social capital may well also act as a mediator, being affected by access to roads and also affecting both self-assessed living conditions and objective well-being (if social support helps satisfy basic needs when the individual lacks them). We measure social capital as involvement in associations and trust. Respondents report whether they belong to a religious group, and other voluntary associations or community groups (0 = No, 1 = Inactive member, 2 = Active member and 3 = Official leader): 53% belong to a religious group, and 40% to another form of association. The general trust question is: "*Generally speaking, would you say that most people can be trusted or that you must be very careful in dealing with people?*": 20% declared that most people can be trusted. Last, respondents are asked "*How much do you trust your relatives/ your neighbours/ other people you know?*" on a scale of 0 (Not at all), 1 (Just a little), 2 (Somewhat) and 3 (A lot). Average trust is highest for relatives (2.4), followed by neighbours (1.8) and others (1.4).

The last mediator refers to feelings of insecurity and experience of crime. For the latter there are two questions: "*During the past year, have you or anyone in your family (No, Once, Twice, 3 or more) had something stolen from your house?*" and "*During the past year, have you or anyone in your family (No, Once, Twice, 3 or more) been physically attacked?*".[5] For the former, the questions are "*Over the past year, how often, if ever, have you or anyone in your family (Never, Just once or twice, Several times, Many times, Always) (i) felt unsafe walking in your neighborhood?, (ii) feared crime in your own home?*" and "*During election campaigns in this country, how much do you personally fear becoming a victim of political intimidation or violence?" (Not at all, A little bit, Somewhat, A lot).* The average score for feeling unsafe or fearing crime is between Never and Just once or twice, and few people on average have experienced crime over the last year.

2.3 Geographical Data

The geographical data from the Afrobarometer surveys are combined with data on the 2013 road network from BingMap. The exact location of the individual respondent is not recorded, but we do know the coordinates of the enumeration area in which they live.[6] The

[5] This last question was not asked in Tanzania.

[6] This produces measurement error in the measure of individual distance to the nearest paved road. However, the households interviewed in each enumeration area are quite close to each other, so that the gap between the actual (unobserved) distance from the GPS point to the paved road and the measured distance will be only small, and small positive and negative gaps will likely compensate each other in the empirical analysis. This type of measurement error is common in the spatial literature. For instance, in Ghani *et al.* (2016) plant-level data are aggregated at the district level and the distance measure is the shortest straight-line from the district's edge to the Golden Quadrilateral highway network. In our

latitude and longitude coordinates of the communities enable us to place each community on a country map and match in geographical data, especially regarding the road network.

Satellite-data from BingMap, available through ArcGIS and providing a satellite-based representation of the road network as of 2013, was used to construct shapefiles that can be used to calculate the straight-line distance in kilometers between the community and the nearest paved road. Figure 1 shows roads in West Africa in 2013 as replicated from the BingMap in ArcGIS on the left panel, and the location of communities on the right panel. The same figure for the rest of our sample is in Figure A2. Taking the straight-line distance (also called the "as the crow flies" or the great circle distance) has a number of advantages over the use of more sophisticated measures such as the use of real distance or time distance, as no assumptions about the means of transportation owned and used by households need to be made. In addition, Combes and Lafourcade (2005) note that the straight-line distance and alternative measures of distance are highly correlated (with a correlation coefficient of over 0.97). Our respondents here live an average of 12km away from the nearest paved road (see Table 2).

We calculate all three possible instrumental variables at the community level in ArcGIS. We construct the measure of land gradient using the SRTM digital-elevation map[7] and we calculate the straight-line distances between the communities and historical routes or hypothetical lines. Figure 1c shows the location of explorer routes (in red) and hypothetical lines (in blue). We use the GIS shapefile from Nunn and Wantchekon (2011) that provides the location of the explorer routes used during the pre-colonial and early colonial periods (between 1768 and 1894); this information comes from the Century Company (1911).

Jedwab and Moradi (2016) provide the hypothetical lines or "straight lines" connecting two historical settlements (the capital, largest and second-largest cities, the other cities with over 10,000 inhabitants, and ports in each country). More precisely Jedwab and Moradi create an Euclidean Minimum Spanning Tree (EMST) network based on the initial urban network as of 1900 major cities and ports, and define this network, EMST, as *"the network that the colonial powers would have built if they had collaborated to optimally connect the initial cities while minimizing construction costs (using the Euclidean distance between them)"* (page 275).

As shown in the first column of Table 2, the average slope figure is 1.6%; respondents live on average 97km away from the nearest hypothetical line, and 157km from the nearest explorer route. In the context of distance to roads, it might be thought that urban areas

case, the same measurement error appears for the current road network and our instrumental variables (hypothetical lines or explorer routes).

[7]Using Shuttle Radar Topography Mission data, at a resolution or cell size of approximately 90 meters.

are not informative, as they will all be close to a paved road. This turns out not to be the case: comparing the urban and rural samples (columns 2 and 3 of Table 2 respectively), the distance to the nearest paved road is 5.8km on average in urban areas and 14.9km in rural areas. The standard deviations of this distance are substantial in both areas, at 16.6 and 20.1. The distance to hypothetical lines is about twice as large in rural than urban areas, while the distance to explorer routes is lower (140km vs. 187km). The distance between the communities and the nearest historical settlement is calculated in ArcGIS: this is 127km in the whole sample, and 83km and 153km for the urban and rural samples respectively.

3 Empirical Model and OLS Results

3.1 Estimation Equations

To quantify the effect of distance to the nearest road on subjective and objective well-being, we first estimate the following equation:

$$WB_{ijr} = \alpha + \beta distance_{jr} + \gamma X_{ijr} + \delta_r + \varepsilon_{ijr} \tag{1}$$

where WB is either the self-assessed current living conditions or the index of access to basic needs for individual i living in community j in region r, and *distance* is the log of 1 + distance to the nearest paved road in kilometers. We use the log specification to account for our *a priori* expectation of a decreasing marginal effect of distance (so that an extra kilometer matters less at 100km from a road than at 10km from a road). As it turns out, the data prefer this specification (in terms of fit) to a regression where distance enters linearly. In Equation (1), X_{ijr} contains age, age-squared and sex, as well as the community's altitude, longitude and latitude. We control for region dummies, denoted by δ_r. There are on average 13 regions in each of the 24 countries we will analyse here, and 11 communities per region. The errors in Equation (1) above, ε_{ijr}, are clustered at the community level as our measure of distance to the nearest paved road is defined at this level. If access to roads improves well-being, then the estimated value of β will be negative (distance to roads is bad for well-being).

Equation (1) does not control for any of the other explanatory variables that are potentially also caused by distance to the paved road, such as education, labour-market status and urban/rural location: these are thus likely bad controls. We will below in Section 6 consider a variety of these kinds of variables as potential mediators of our main results regarding the effect of road access on living conditions and deprivation. We will show that adding these variables makes no major difference in our results. For example, considering education as a control variable, we will estimate the following equation:

$$WB_{ijr} = \alpha^* + \beta^* distance_{jr} + \gamma^* X_{ijr} + \psi Educ_{ijr} + \delta_r^* + \varepsilon_{ijr}. \quad (2)$$

If, as we suspect, educational opportunities are better closer to roads, and education improves well-being outcomes, then the estimated value of ψ will be positive, and the estimate of β^* in Equation (2) will be less negative than the estimate of β in Equation (1): holding education constant turns off part of the well-being benefit of roads. We will carry out analogous analyses for the mediating effect of labour-market status, health, living in an urban area, other types of public goods, crime, social capital (trust and participation in social or religious groups) and the use of mobile phones and internet. We estimate linear models to be able to compare the size of the road effect when we add the possible mediator and when we modify the sample. The core results are qualitatively unchanged when ordered probit models are estimated in Online Appendix Table O1.

3.2 The Correlation between Road Distance and Well-Being

We first discuss OLS results for the whole sample, as shown in the left part of Panel A in Table 3. As we might think that the effect of roads differs by location, we also estimate the model separately for urban and rural residents in Panels B and C. All of the specifications in this table include over 300 regional dummies as well as the exogenous controls (age, age-squared, female, altitude, longitude and latitude).

The first outcome variable is self-assessed current living conditions (taking values from 1 to 5). In column (1), those living further from a paved road are less satisfied with their current living conditions. The effect size is small here, and is arguably not economically significant: a one standard-deviation rise in the log of distance (1.29, from Table 2) is associated with subjective living conditions that are 0.037 points lower (equivalent to 3% of a standard deviation, using the value of 1.19 from Table 1). Panels B and C of Table 3 distinguish between the living conditions of urban and rural residents. The point estimates here remain negative, and are not significantly different from that in the whole sample in Panel A.

We now switch from looking at subjective living conditions to objective measures of deprivation (which is also called functioning failure). The results appear in column (2) of Table 3. The dependent variable here is the lack of deprivation in four dimensions, each of which is measured from 0 to 4. Our resulting lack of deprivation index ranges from 0 to 16, with higher numbers reflecting better outcomes: 0 refers to someone who is always lacking in all four dimensions, and 16 to someone who never lacks in any dimension. We will also carry out robustness checks on each dimension separately below in Table 7.

The results show that those who live further from roads are more deprived (i.e. lack of deprivation is lower). For the whole sample, one standard-deviation greater log of distance is associated with deprivation that is around 8% of a standard deviation higher. In Panels B and C, for the urban and rural samples, the estimated coefficient is somewhat smaller, but remains significant at all conventional levels.

Appendix Table A3 shows the full results for the whole sample, including the estimated coefficients on all of the control variables. There is a notable U-shape in age in both living conditions and lack of deprivation (as is very often found in the subjective well-being literature), and women have slightly less-good outcomes for both measures.

4 IV Strategy and Results

4.1 The Non-random Location of Roads: IV Strategy

Road placement may well be endogenous, with roads being built in areas with more resources. We thus carry out two-stage least squares estimation including a set of potential instrumental variables. Our first-stage regressions are as follows, where Z_{jr} is the instrumental variable:

$$distance_{ijr} = \phi + \theta Z_{jr} + \pi X_{ijr} + \mu_r + \nu_{ijr}. \tag{3}$$

Building on previous research on the endogenous placement of transport infrastructure, we will consider the standard instrumental variables in this literature (see Redding and Turner, 2015, for a review): routes that were used a long time ago, hypothetical lines between historical settlements, and land characteristics that may affect the cost of road-building. Section 2.4 above described how all three of these are measured in the African data that we use. These three are thought to be good instruments for the following reasons.

First, the location of explorer routes or routes that were used a long time ago could affect the placement of current roads without directly affecting our outcomes. As stated in the review by Redding and Turner (2015), the historical-route instrumental-variable approach relies on the location of old transportation routes as a source of quasi-random variation in the location of current transportation infrastructure. This approach has been used, for example, in Agrawal *et al.* (2017), Baum-Snow *et al.* (2017), Duranton and Turner (2012) and Martincus *et al.* (2017). The value that this instrument takes in our African case will be the distance between the respondent's location and the closest pre-colonial explorer route.

Second, the hypothetical lines connecting current or historical major cities are candidates for the instrumentation of the location of current roads and railways (Atack *et al.*, 2010; Banerjee *et al.*, 2020; Bird and Straub, 2020; Donaldson, 2018; Faber, 2014; Ghani *et al.*, 2016; Jedwab and Moradi, 2016; Michaels, 2008).[8] The straight line connecting historical settlements is a proxy for the most cost-effective way of linking major cities abstracting from natural constraints (e.g. lakes or mountains), and thus provides relevant exogenous variation. Along these lines, Michaels (2008) uses the straight-line distance of the centroid of the county to the nearest city to instrument for the location of highways in the US. Here we will use the network of hypothetical lines from Jedwab and Moradi (2015 and 2016) that relates the urban network as of 1900, consisting of the capital city, the largest and second-largest cities, other cities with above 10,000 inhabitants, and ports:[9] there are on average 3.5 of these historical settlements per country, with a median figure of 2 (see Online Appendix Table O3). These hypothetical lines could have joined major cities and ports at the end of the 19^{th} Century, but were not necessarily built (given natural constraints and the absence of cooperation among colonial empires).

Last, land characteristics such as the slope may determine the construction of roads in a given area as they can influence the associated costs: roads are more likely to be built in flatter than steeper areas.[10] Previous work has used slope measures to instrument the location of physical infrastructure.[11]

The simultaneous-equation model we estimate is:

$$\begin{cases} distance_{ijr} = \phi + \theta Z_{jr} + \pi X_{ijr} + \mu_r + \nu_{ijr} \\ \\ WB_{ijr} = \alpha + \beta \widehat{distance}_{ijr} + \gamma X_{ijr} + \delta_r + \varepsilon_{ijr}. \end{cases} \tag{4}$$

[8]As hypothetical lines may instrument a variety of transportation infrastructure, it is also of interest to estimate a reduced form, regressing well-being directly on the distance to the hypothetical line. Our main results continue to hold in these estimations, which are presented in Table O2.

[9]Some examples are Ouagadougou and Bobo-Dioulasso in Burkina Faso, Kumasi in Ghana, Luanda in Angola and Mombasa in Kenya.

[10]It is worth underlining that all of these three instruments are time-invariant, so there is no particular advantage to analysing panel as opposed to cross-section data.

[11]Duflo and Pande (2007) instrument dam construction using river gradient across Indian districts. Dinkelman (2011) uses land gradient as an instrument for the locations chosen to benefit from an electrification project in South Africa. Land gradient has also been identified as a determinant of cellular-phone coverage in Malawi by Batzilis *et al.* (2016).

For the instrumental variable to be valid, it has to be correlated with the current distance to the nearest paved road and not correlated with the error term of the second-stage equation ε. The first correlation is discussed when presenting the results from the first-stage estimations in Section 4.2. The second requirement is the exclusion restriction: the instrument should have no effect on the outcome other than through the first-stage channel.

We will below address two major points regarding the exclusion restriction. First, the exclusion restriction is conditional on all covariates, and does not assume the unconditional orthogonality of the instrument and the outcome. This distinction is emphasised in Agrawal *et al.* (2017) and Duranton and Turner (2012). It is therefore important to control for exogenous variables, in particular community-level variables measured at the same level as distance to the road: all of our regressions will control for latitude, longitude and altitude, as exogenous proxy variables for variations in climate and agricultural outcomes across communities within administrative regions. One concern with using land slope as an instrument for road location is that, in a rural setting, it may affect agricultural outcomes, as suggested in Dinkelman (2011). In our case, the direct impact of gradient on farm productivity would appear in living conditions and material deprivation, our two outcome variables. With respect to the two other candidate instruments, distances to explorer routes and hypothetical lines, these will likely affect current town size and urbanisation that will in turn affect our outcome variables. This concern will be taken into account in the mediation analysis of Section 6.

The second major point is that some of the locations that are close to a hypothetical line were already very close to the capital city or a major port, for example, in 1900, while others were far away from both. We should not consider those who were already close to an urban area as being treated by the hypothetical-line instrument: communities that were in the suburbs of Dakar in 1900 are still in the suburbs of Dakar in 2000. The instrument will thus provide less exogenous variation the closer the community is to a historical city or port. We will deal with this issue in Section 4.4 by progressively dropping communities that are within a certain number of kilometers from a historical city or port, where the number of kilometers considered will vary between 1 and 20. This strategy is similar to that used in Bird and Straub (2020) and Ghani *et al.* (2016). As we will see below, this sample restriction will make only little difference to the econometric results.

4.2 The Non-random Location of Roads: First-stage Results

The first-stage results appear in Table 4. This table has three columns and six panels, referring to two specifications for each of our three potential instruments: distance to hypothetical lines, slope, and distance to explorer routes. The instruments appear first as levels and then in logs. The first column refers to the entire sample, and the second and third to the urban and rural samples. All equations include the instrument, regional dummies (so that we estimate the effect of the instrument within an administrative region), and the exogenous demographic and community characteristics (sex, a quadratic in age, altitude, longitude and latitude). Online Appendix Table O4 shows the full results, including the estimated coefficients on the control variables.

The results in Table 4 show that distance to the hypothetical line is the only instrument with a convincing first-stage F-statistic (usually considered to be greater than 10: see Staiger and Stock, 1997). In column (1) of Table 4, the log of 1 + the distance to the nearest hypothetical line in Panel B attracts an F-statistic of over 100, and the alternative linear specification (with distance divided by 100 to make the coefficients easier to read) an F-statistic of about 20. The estimated coefficient in the log specification implies an elasticity of current road distance to the distance to historical hypothetical lines that is around 0.25. The other potential instruments, slope and distance to explorer routes, have F-statistics of under 10.

We show the separate first-stage results for the urban and rural samples in columns (2) and (3) of Table 4. These again explore the six different potential instrument specifications. The resulting F-statistics are in line with those in column (1), with only the log of 1 + distance to the nearest hypothetical line being consistently satisfactory. As discussed in Section 2, there remains substantial variation in the distance to paved roads even in the areas that enumerators describe as urban.

4.3 The Effects of Road Distance on Well-being: Second-stage Results

We first discuss the second-stage results for the whole sample, before progressively dropping communities that are close to historical cities or ports, as discussed in Section 4.1 above.

The first outcome variable is self-assessed current living conditions (taking values from 1 to 5). The results for the whole sample are shown in Panel A in Table 5, and the separate estimations for urban and rural residents in Panels B and C. As in the OLS specifications, all of the IV specifications include regional dummies and the exogenous controls (age,

age-squared, female, altitude, longitude and latitude).

While in the naïve regressions with uninstrumented distance in column (1) of Table 3, those living further from a paved road were less satisfied with their current living conditions, the IV results in Table 5 suggest that there is no effect of road distance on current living conditions.

Panels B and C of Table 5 distinguish between the living conditions of urban and rural residents. In the urban sample, a greater distance to a road leads to worse living conditions, but only at the ten percent level, while the analogous coefficient in the rural sample is positive and insignificant.

The second outcome variable is the lack of material deprivation in column (2). The instrumented distance coefficients here are all negative (although that in the rural sample is not significant), as was the case in the OLS results in Table 3. A greater distance to roads therefore increases deprivation (i.e. it reduces the lack of deprivation).

4.4 Instrument Validity and Distance from Historical Settlements

Section 4.1 raised a major concern about the validity of distance to the hypothetical lines (connecting historical settlements) as an instrument. The instrument will not be valid for locations that were already very close to a city or major port in 1900 and will still be so today, and the location of the city or port in 1900 cannot be considered as exogenous (with settlements being more likely where conditions were more favourable).

We deal with this issue by re-estimating the regressions in Tables 3, 4 and 5, progressively dropping locations that are within a certain number of kilometers from historical settlements. The value of this exclusion distance varies from 1 to 20 kilometers.

Figure 2 depicts the F-statistic from the first stage in each of these 20 new regressions. The F-statistic falls as we progressively exclude locations that are further from the 1900 historical settlements, but the instrument remains sufficiently strong until we reach a figure of 16 kilometers (the horizontal line shows the standard value of 10 below which the instrument is considered not to be strong). This fall is consistent with our first-stage results in Table 4, where the F-statistic was larger for urban than rural locations (with the former being more likely to be close to historical settlements). The high F-statistic in Table 4 then partly reflects the mechanical correlation for communities that were already close to 1900 historical settlements.

The results for the 20 corresponding second-stage regressions are plotted in Figure 3. Those on the right refer to the effect of the distance to a road on the lack of deprivation. The point estimates turn out to be remarkably stable across all of the regressions, although the smaller sample sizes render the estimated coefficients insignificant at distances of over 16 kilometers. The effect of the distance to a current road on material deprivation is thus not driven by those who live close to historical settlements.

The left-hand side of Figure 3 shows the corresponding results for the effect of road distance on subjective living conditions. All of the estimated coefficients here are insignificant (as in Panel A of Table 5). The living conditions that individuals report in Africa do not vary with their distance from a road.

In what follows, and in particular for all of the robustness tests, we will restrict the sample to locations that are situated over 10km from a historical settlement.[12] The descriptive statistics for this restricted sample appear in Appendix Table A4 (which shows that the means and standard deviations of the outcome variables are very similar to those in the whole sample). The benchmark first- and second-stage results (as well as the OLS counterparts) appear in Table 6. The F-statistic in column (1) is satisfactory. As suggested in Figure 3, the results for deprivation are similar to those for the whole sample in Table 5. The effect size for deprivation is such that moving from 10km to 20km from a paved road is estimated to reduce access to basic needs by one half of a point (12% of a standard deviation). The instrumented coefficient for road distance in the living-conditions regression continues to be insignificant.

The instrumented coefficient for lack of deprivation in column (5) of Table 6 is more negative than the naïve version in column (4). The difference between the OLS and IV estimates is designed to reflect the endogenous location of roads. The estimated difference here is consistent with roads being built closer to individuals who are more deprived: were roads to be randomly allocated, the deprivation gap by distance would be even larger.

Our key result is then that the OLS correlation between roads and deprivation continues to hold in an IV analysis: roads have served to improve the objective living conditions of the deprived in Africa. However, we find no such effect for subjective living conditions. There are a number of explanations of this contrast. The first is that the four basic needs we have considered here are relatively unimportant in individuals' evaluation of their living conditions (the correlation coefficient between the two is only 0.22), although this may seem unlikely. The second is that basic needs are important, but that roads also bring problems in other domains of life that offset the positive effect on basic needs (for example via worse health, pollution and crime): we will consider the role of mediators (that is, other variables that change as a result of roads) in Section 6.

[12]With this restriction, we will not be able to carry out separate analyses for the urban and rural samples, as the F-statistic in the first stage for the former is unsatisfactory.

Last, exposure to roads, and to new goods and richer people, may change the benchmark that individuals use to evaluate their lives. We do not have direct information on the extent to which individuals compare to others, nor to whom they compare. However, we can make some progress by considering a third well-being Afrobarometer question, asking about *relative* living conditions, as opposed to the absolute living conditions that we analysed above. For example, this relative question in Mozambique is "*In general, how do you rate your living conditions compared to those of other Mozambicans?*", with answers on a five-point scale from "much worse" to "much better". The regression results using this relative living conditions variable are almost identical to those for absolute living conditions (see Appendix Table A5), suggesting that the answers to the absolute question already contain a substantial relative component. Those living closer to roads may be objectively better off, but not feel better off as their reference group has changed.

We considered potential heterogeneity in the effect of roads on well-being, based on institutions and ethnic fractionalisation using data at the country level. Splitting our sample of countries into two groups by their level of fractionalisation (according to ethnicity, language, or religion)[13] or institutional quality (transparency, property rights, or public administration),[14] produces no sharp differences in the size of the estimated coefficients.

5 Robustness Checks

We now turn to a number of robustness tests. These are all carried out on the sample of locations over 10km from a historical settlement, the baseline results for which appeared above in Table 6. Our tests are split up into two groups: the first refers to the measures of well-being and road distance, and the second to the estimation sample.

5.1 Robustness Checks 1: Measures

Measures of outcome variables We above took a counting approach to material deprivation, adding up the unweighted sum of the scores in each of the four dimensions. This relies on two assumptions: that each dimension is equally important, and that the scores for each dimension are cardinal (in the sense that moving from "Never" (0) to "Just once or twice" (1) represents the same rise in deprivation as moving from "Many times"

[13]From Alesina *et al.* (2003), and the updated version in the Historical Index of Ethnic Fractionalization (HIEF) by Drazanova (2020).

[14]A set of four indices from Freedom House (legal, political, economic, and the total score) and a set of three indices from the World Bank database "Country Policy and Institutional Assessment (CPIA)".

(3) to "Always" (4)). We relax both of these assumptions in Table 7, first by considering each of our four dimensions separately, and then looking at a dummy variable for no or low deprivation (and thus high well-being), defined as the individual reporting having had difficulty in satisfying their basic needs in the dimension under consideration over the past year never or just once or twice.

The effect of road distance on the four separate dimensions of the deprivation index, considered cardinally (i.e. from 0 to 4), appears in the left-hand part of Table 7. The results show that distance is harmful for three of the four dimensions: the exception is cooking fuel. The difference between the naïve and instrumented regressions for these three dimensions is analogous to that for the overall index in Table 6.

The right-hand side of Table 7 then drops the cardinality assumption, and considers dummy variables for no or low deprivation, to see whether the effect of roads is concentrated at one or other of the extremes of the scale. The estimation here is linear (the same qualitative results are found in probit regressions). The estimated coefficients for distance to roads in these binary regressions are all similar to those using all of the values of the dependent variable in the left-hand side of Table 7.[15] As such, roads shift some people from being more deprived (2 to 4) to less deprived (0 or 1), as well as probably shifting some individuals within these categories.[16]

For completeness, Panel E of Table 7 presents analogous results for subjective living conditions. The left-hand side shows the baseline results from the cardinal regression in Table 6. We then create a dummy variable for good or very good living conditions, the results of which appear on the right-hand side of Panel E. As was the case for the dimensions of deprivation, the binary results are entirely consistent with the cardinal results.

Measures of distance to the nearest paved road Table 8 checks whether the results from the instrumented regressions are robust to changes in the distance measure. Panel A reproduces the benchmark estimates from Table 6. We then create three dummy variables for road distance as communities that are located more than 5km, 10km and 20km from a paved road: the results appear in Panels B to D. Distance is treated as a level (rather than a log) in row E and as a quadratic in row F. Last, the measure in row G is one of road proximity: the inverse of one plus the distance in km. The results in Table 8 show that our

[15]The estimated coefficient on medical care is insignificant in the IV regressions: the point estimate is actually more negative under IV, but the standard error is much higher.

[16]The Afrobarometer also includes a question allowing us to create a dummy variable for whether the dwelling has a solid roof. The 2SLS analysis of this variable, using the same approach as in the right-hand side of Table 7, reveals a negative significant relationship between road distance and solid roofs. These results appear in Online Appendix Table O5.

conclusions are not sensitive to the measure of distance: roads reduce material deprivation but almost never have an effect on reported living conditions.[17]

5.2 Robustness Checks 2: Samples

We now change the estimation sample in the two following ways.

1. Dropping observations that are "too close" to the nearest paved road.

2. Removing one country from the sample at a time in a Jackknife analysis.

The results for Checks 1 and 2 appear in Table 9. The first row of this table includes the benchmark results, for ease of comparison. We first drop observations that are less than 5 and then 10 kilometers from the current paved road. We do so to make sure that our results are not only driven by local observations that are close to roads, rather than being more global. This is a particular concern in our case as we use log distance, where the variation is higher for smaller distance values. In rows 2 and 3, the results for more-distant observations are of the same sign and significance as the benchmark results: our findings then seem to be global rather than local.[18] Another reason to drop observations close to roads is that these areas may attract migrants who have different characteristics (richer, more entrepreneurial etc.), which would bias our results: these migrants are not treated by roads, but rather attracted by them. That the results in rows 2 and 3 change only little may reveal that migration is not behind our results; we will consider migration further in Section 7.

The results from the Jackknife analysis appear in Appendix Table A6. Again the first line refers to the benchmark results. The remaining 24 lines list the estimated coefficients on the distance to roads in the well-being regressions when the country in the first column is dropped. In no case is the jackknife coefficient significantly different from that in the main analysis.

[17]The results for lack of deprivation in Panel G are consistent with the others, but are not significant. This may reflect that the inverse of distance approaches its asymptotic value faster than does log, so that identification is more concentrated on small distance values.

[18]We may also worry about distance values that are "too large": top trimming the 1% (over 98km) or 5% (over 51km) of the observations with the highest values for distance to the nearest paved road makes very little difference to the results.

6 Potential Mediators

Roads may affect many outcomes, which themselves help determine well-being: for example, health, criminality, and public goods.[19]

We evaluate the role of these different mediators in the effect of road distance on well-being by adding a sequence of additional control variables to the instrumented benchmark Equation (4): these are urban/rural location, education, labour-market variables, health, public goods, the use of computers and mobile phones, social capital and crime.[20] The estimated road-distance coefficients once we control for each of these different mediators in turn appear in Table 10 (where the first row shows the benchmark values) and are illustrated in Figure 4 (along with their respective 95-percent confidence intervals). Table 10 has two columns, the first for living conditions and the second for lack of deprivation.

None of the specifications in column (1) produces a significant effect of road distance on living conditions. In terms of the mediation for lack of deprivation (see Figure 4b and column (2) of Table 10), the road-distance coefficient is notably closer to zero when we control for the provision of public goods in row F (although not significantly different from the benchmark coefficient in row A). Public-good provision is thus better closer to roads, and helps reduce material deprivation: were people closer to and further from paved roads to have the same level of public goods, there would be no difference in their deprivation. The last row in Table 10 introduces all of the mediators at the same time (except for experience of crime, for which one question is missing for Tanzania). There continues to be no effect of road distance on living conditions, and the estimated coefficient for lack of deprivation is also insignificant once all of the mediating variables are held constant. However the instrumental variable is weaker when we control for all of the mediators at the same time (see Table O6).

We have not been able here to directly investigate all of the potential mediating variables. Additional candidates in this respect include deforestation, pollution, health outcomes, family break-up, worse community life, unwanted pregnancies or the observation of richer people (including tourists and traders) closer to roads. All of these may affect subjective living conditions more than the lack of deprivation, and help to explain why roads

[19]Our estimation model is that hypothetical line lead to roads, which then produce hospitals (for example), and hospitals produce well-being. Hospitals thus mediate the effect of (instrumented) roads on well-being. Considering hospitals as a problem for the exclusion restriction implies on the contrary that hospitals are located close to hypothetical lines, even when a paved road was not built. For locations that are not close to a historical settlement, this does not seem likely.

[20]The first-stage results for all of these mediation analyses appear in Table O6. All of the F-statistics are satisfactory (except for the last row where the F-statistic is equal to 9).

have no effect on the former but a significant effect on the latter. The data constraints do not allow us to investigate these. As we have regional dummies, we require (for example) data on air quality by Enumeration Area within region for the early 2010s. In addition, this information would need to be available in a harmonised way across the 24 countries in our analysis. Further work when more-detailed data becomes available would be useful in this respect.

7 The Role of Migration

Our analysis above has corrected for the endogenous placement of roads via an instrumental variable. However, we have not yet addressed the question of the endogenous placement of individuals. An infrastructure project may not benefit natives, but rather attract richer individuals: in the latter case, the social benefit of the project is over-estimated. If individuals of a certain type systematically move towards (or away from) exogenously-placed roads, part of what we think is the causal effect of roads may instead reflect composition. If those who move to live closer to the current road network are less deprived, we will over-estimate the effect of roads on deprivation.

Unfortunately, the Afrobarometer data does not include information about migration status. We thus turn to Demographic and Health Survey (DHS) data. Of the 24 countries that appeared in our Afrobarometer analysis above, there are 15 that appear in the DHS data and for which we have geo-coded location information and migration status. We take the latest available geo-coded wave for each country, which gives us observations spread out between 2003 and 2017.

We try to replicate as far as possible our Afrobarometer analysis on DHS data. We calculate distance to observed roads in the same way as above, and continue to use distance to hypothetical lines as an instrument. We do not have the same outcome measures, and take the wealth index constructed by the DHS on the basis of objective information on housing characteristics and durable-good ownership. We standardise this index to have a mean of zero and a standard deviation of one. The other controls are as in our analysis above, except that we now have information on migration status. The exact question in the DHS is as follows: *"How long have you been living continuously in (name of current city, town or village of residence)?"* The answers were either always, a number of years, or a special code for those who do not live there but are visitors from outside.[21] Those

[21] We do not include visitors in our analysis, as we observe neither their permanent place of residence (and its distance to roads) nor their wealth index.

who report a number of years are migrants, picking up both international and internal migration, and at all respondent ages (including migration as a child). The question also identifies as migrants those who left their place of birth and subsequently returned to it.

The initial sample size is 276, 346 for these 15 countries. Appendix Table A7 lists the countries, sample size and year in the DHS surveys that we use. Dropping the 5, 624 visitors reduces this figure to 270, 722. We will apply the same method as in Table 6 to the DHS data. We drop the 33, 968 observations on individuals who are in locations that are 10km or closer to historical settlements, producing a final DHS analysis sample of 236, 754 respondents. All equations include dummy variables for the 160 administrative regions in these 15 countries. The descriptive statistics for this DHS sample appear in Appendix Table A8 and the location of the respondents in Appendix Figure A3.

Almost half of our DHS sample have changed location at least once in their life. Figure 5 plots the percentage of migrants by distance to the nearest current paved road: as can be seen, the distribution is fairly uniform, with a slightly higher percentage who are either close to a road or very far from one (Jacoby and Minten, 2009, note that migrants may decide to live in remoter areas in order to profit from lower land prices and greater agricultural-land availability). Migration has been frequent, but seemingly not related to major roads.

The first-stage results appear in Appendix Table A9. As was the case for the Afrobarometer, the distance to the hypothetical line is the only instrument that has an F-statistic of over 10. The ensuing second-stage results from the benchmark regressions on DHS data appear in Panel A of Table 11. These show a consistent negative effect of road distance on material well-being, which is larger in the IV results in column (2) (analogous to our findings for the lack of deprivation in Afrobarometer data in column (5) of Table 6).

We then carry out separate analyses by migration status in Panels B and C in order to see whether our main results continue to hold in the sample of non-migrants (in the same spirit as the work on the returns to education, program effectiveness and migration reviewed in Strauss and Thomas, 1995). The F-statistics in the first stages for these two groups continue to be satisfactory. The estimated road coefficient for non-migrants in Panel B is negative and significant, so that distance to roads has a sizeable negative effect on the material well-being of non-migrants. The analogous coefficient for migrants in Panel C is also negative and significant, and of the same size. This suggests that migration may not be behind our core Afrobarometer results, as the effect of roads on well-being is strikingly similar for both migrants and non-migrants.[22]

[22]This migration status refers to the respondent. It is possible that non-migrant respondents have migrant parents, who migrated because of the presence of roads. In this case some non-migrants are not treated by the road, but rather selected (due to the migration of their parents). We thus have two

8 Conclusion

This paper has matched road-network information to nationally-representative household survey data from 24 African countries in the early 2010s to analyse the relationship between the distance to paved roads and both living conditions and the lack of material deprivation (regarding food, water, cooking fuel and medical care). The naïve correlations between distance to roads and both of the well-being outcomes are negative (although that for living conditions is very small): those who live further from roads are less well-off. However, roads are not located exogenously (and neither are people). We instrument road location by the hypothetical lines drawn between major cities and ports at the end of the 19^{th} Century. The instrumented effect of road distance on the lack of deprivation continues to be negative; however, that on self-assessed living conditions becomes zero. This conclusion continues to hold when we exclude areas that are "too close" to historical settlements. We also provide some supporting evidence that these estimated road coefficients do not entirely reflect the migration of certain kinds of individuals towards areas with roads.

Roads therefore have contrasting effects in Africa. While they provide access to basic needs, and likely reduce inequality in this dimension, they have no effect on self-assessed living conditions. The positive consequences for access to water, food and medical care seem to be counterbalanced by other (unmeasured) aspects of life that matter to individuals. We hope that future data will be able to match road location to both subjective measures of well-being and variables such as health, family break-down, environment and relative standing. These will help our understanding of the role of roads in determining the quality of life.

types of non-migrants: those with migrant parents and those with non-migrant parents. If non-migrant respondents with migrant parents "behave" in the same way as migrant individuals do in Table 11, in the sense of having the same value of the estimated coefficient, then the true coefficient for non-migrant respondents with non-migrant parents (this is the group that is treated by roads) is even more negative than that in Panel B. We can thus consider the coefficient of -0.535 as a lower-bound in absolute terms of the true effect in the case of the intergenerational transmission of traits.

References

Afrobarometer (2015), Data codebook for a Round 5 Afrobarometer survey in 34 African countries. Prepared by: Chunho Park Michigan State University July 2015. Report (107 pages).

Afrobarometer Data, Burundi, Benin, Burkina Faso, Botswana, Cameroon, Cote d'Ivoire, Ghana, Guinea, Kenya, Liberia, Mali, Malawi, Mozambique, Namibia, Niger, Nigeria, Senegal, Sierra Leone, Sudan, Tanzania, Togo, Uganda, Zambia and Zimbabwe, Round 5, 2011-2013, available at http://www.afrobarometer.org.

Aggarwal, S. (2018), "Do rural roads create pathways out of poverty? Evidence from India", *Journal of Development Economics*, 133, 375-395.

Agrawal, A., Galasso, E., and Oettl, A. (2017), "Roads and Innovation", *Review of Economics and Statistics*, 99, 417-434.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003), "Fractionalization.", *Journal of Economic Growth*, 8, 155-94.

Asher, S., Garg, T. and Novosad, P. (2020). "The Ecological Impact of Transportation Infrastructure", *Economic Journal*, 130, 1173-1199.

Asher, S. and Novosad, P. (2020), "Rural Roads and Local Economic Development", *American Economic Review*, 110, 797-823.

Atack, J., Bateman, F., Haines, M. and Margo, R.A. (2010), "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60", *Social Science History*, 34, 171-197.

Banerjee, A., Duflo, E. and Qian, N. (2020), "On the Road: Access to Transportation Infrastructure and Economic Growth in China", *Journal of Development Economics*, 145, 102442.

Batzilis, D., Dinkelman, T., Oster, E., Thornton, R. and Zanera, D. (2016), "New Cellular Networks in Malawi: Correlates of Service Rollout and Network Performance", in Edwards S., Johnson S. and Weil D., (Eds.), *African Successes: Modernization and Development*, University of Chicago Press, 215-246.

Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A. and Zhang, Q. (2017), "Roads, Railroads and Decentralization of Chinese Cities", *Review of Economics and Statistics*, 99, 435-448.

BenYishay, A., Rotberg, R., Wells, J., Lv, Z., Goodman, S., Kovacevic, L. and Runfola, D. (2017), "Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality," AidData. Available online at *http : //geo.aiddata.org*.

Bird, J. and Straub, S. (2020), "The Brasilia Experiment: The heterogeneous Impact of Road Access on Spatial Development in Brazil", *World Development*, 127, 104739.

Bucheli, J.R., Bohara, A.K. and Villa, K. (2018), "Paths To Development? Rural Roads and Multidimensional Poverty in the Hills And Plains Of Nepal," *Journal of International Development*, 30, 430-456.

Buys, P., Deichmann, U. and Wheeler, D. (2010), "Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa," *Journal of African Economies*, 19, 399-432.

Cagé, J. and Rueda, V. (2016), "The Long-Term Effects of the Printing Press in Sub-Saharan Africa", *American Economic Journal: Applied Economics*, 8, 69-99.

Clark, A.E. and D'Ambrosio, C. (2019), "Living Conditions and Well-Being: Evidence from African Countries", *South African Journal of Economics*, 87, 91-109.

Century Company. 1911, "The Century Atlas: Africa [map]" Buffalo, NY: Matthews-Northrup (470 English statute miles to 1 inch).

Combes, P.-P. and Lafourcade, M. (2005), "Transport costs: measures, determinants, and regional policy implications for France", *Journal of Economic Geography*, 5, 319-349.

Damania, R., Russ, J., Wheeler D. and Barra, A.F. (2018), "The Road to Growth: Measuring the Tradeoffs Between Economic Growth and Ecological Destruction", *World Development*, 101, 351-376.

De Janvry, A. and Sadoulet, E. (2015), Development Economics: Theory and Practice. Routledge, 2015.

Dercon, S., Gilligan, D.O., Hoddinott, J. and Woldehanna, T. (2009), "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages", *American Journal of Agricultural Economics*, 91, 1007-1021.

Dillon, A., Sharma, M. and Zhang, X. (2011), "Estimating the Impact of Rural Investments in Nepal", *Food Policy*, 36, 250-258.

Dinkelman, T. (2011), "The Effects of Rural Electrification on Employment: New Evidence from South Africa", *American Economic Review,* 101, 3078-3108.

Djemaï E. (2018), "Roads and the spread of HIV in Africa", *Journal of Health Economics*, 60, 118-141.

Donaldson, D. (2018), “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure”, *American Economic Review*, 108, 899-934.

Drazanova, L. (2020), “Introducing the Historical Index of Ethnic Fractionalization (HIEF) Dataset: Accounting for Longitudinal Changes in Ethnic Diversity”, *Journal of Open Humanities Data*, 6, 6.

Duflo, E. and Pande, R. (2007), “Dams”, *Quarterly Journal of Economics*, 122, 601-646.

Dumas, C. and Játiva, X. (2020), “Better roads, better off? : Evidence on improving roads in Tanzania”, Université de Fribourg, Working Papers SES, 518.

Duranton, G. and Turner, M.A. (2012), “Urban Growth and Transportation”, *Review of Economic Studies*, 79, 1407-1440.

Faber, B. (2014), “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System”, *Review of Economic Studies*, 81, 1046-1070.

Ghani, E., Goswami, A and Kerr, W. (2016), “Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing”, *Economic Journal*, 126, 317-357.

Gibbons, S., Overman, H., Lyytikainen, T. and Sanchis-Guarner, R. (2019), “New Road Infrastructure: The Effects on Firms”, *Journal of Urban Economics*, 110, 35-50.

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

Jacoby, H.G. (2000), “Access to Markets and the Benefits of Rural Roads”, *Economic Journal*, 110, 713-737.

Jacoby, H.G. and Minten, B. (2009), “On Measuring the Benefits of Lower Transport Costs”, *Journal of Development Economics*, 89, 28-38.

Jedwab, R. and Moradi, A. (2016), “The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa”, *Review of Economics and Statistics*, 98, 268-284.

Jedwab, R. and Moradi, A. (2015), “Replication data for: The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa”, doi:10.7910/DVN/29908, Harvard Dataverse, V1.

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

Lipscomb, M., Mobarak, A.M. and Barham, T. (2013), "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil", *American Economic Journal: Applied Economics*, 5, 200-231.

Martincus, C.V., Carballo, J. and Cusolito, A. (2017), "Roads, exports and employment: Evidence from a developing country", *Journal of Development Economics*, 125, 21-39.

Michaels, G. (2008), "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System", *Review of Economics and Statistics*, 90, 683-701.

Nunn, N. and Wantchekon, L. (2011), "The Slave Trade and the Origins of Mistrust in Africa", *American Economic Review*, 101, 3221-3252.

Pradhan, M. and Ravallion, M. (2000), "Measuring Poverty Using Qualitative Perceptions of Consumption Adequacy", *Review of Economics and Statistics*, 82, 462-471.

Ravallion, M. (2014), "Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty", In A.E. Clark and C. Senik (Eds.), *Happiness and Economic Growth: Lessons from Developing Countries*. Oxford: Oxford University Press.

Redding, S.J. and Turner, M.A. (2015), "Transportation Costs and the Spatial Organization of Economic Activity", in Cheshire P. and Mills E.S. (Eds.), *Handbook of Regional and Urban Economics*, Volume 5B, 1339-1398.

Riley-Powell A.R., Lee G.O., Naik N.S., Kelly E. Jensen, K.E., O'Neal, C., Salmon-Mulanovich, G., Hartinger, S.M., Bausch, D.G. and Paz-Soldan, V.A. (2018), "The Impact of Road Construction on Subjective Well-Being in Communities in Madre de Dios, Peru", *International Journal of Environmental Research and Public Health*, 15, 1271.

Staiger D. and Stock J.H. (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, 557-586.

Straub, S., Vellutini, C. and Warlters, M. (2008), "Infrastructures and Economic Growth in East Asia", Policy Research Working Paper No. 4589. World Bank.

Strauss, J and Thomas, D. (1995), "Human Resources: Empirical Modeling of Household and Family Decisions", in J. Behrman and T.N. Srinivasan (Eds.), *Handbook of Development Economics*, Volume III, 1883-2023.

Wang, Y. and Wu, B. (2015), "Railways and the Local Economy: Evidence from Qingzang Railway", *Economic Development and Cultural Change*, 63, 551-588.

Tables and Figures

Table 1: Sample Descriptive Statistics

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Current living conditions (LC)	38,249	2.63	1.19	1	5
Current LC, good or very good	38,249	0.30	0.46	0	1
Lack of depriv., over 4 items	37,985	11.4	3.7	0	16
Age	38,027	36.4	14.1	18	105
Female	38,395	0.50	0.50	0	1
Urban	38,395	0.36	0.48	0	1
Primary education	38,290	0.32	0.47	0	1
Secondary education	38,290	0.34	0.47	0	1
Higher education	38,290	0.12	0.32	0	1
Paid work	38,216	0.31	0.46	0	1
Visited Public Health Centre	38,124	0.85	0.36	0	1
Electricity	38,395	0.54	0.50	0	1
Piped water	38,331	0.47	0.50	0	1
Sewage	38,108	0.21	0.41	0	1
Mobile	38,371	0.92	0.27	0	1
Post Office	38,283	0.18	0.38	0	1
School	38,267	0.88	0.32	0	1
Police Station	38,203	0.34	0.47	0	1
Clinic	38,196	0.59	0.49	0	1
Market	38,291	0.66	0.47	0	1
Use computer	37,753	0.47	1.11	0	4
Use internet	37,529	0.47	1.12	0	4
Use mobile, call	38,050	2.31	1.48	0	4
Use mobile, text message	37,996	1.51	1.53	0	4
Use mobile, money	37,592	0.33	0.80	0	4
Religious group, inactive	38,157	0.17	0.37	0	1
Religious group, active	38,157	0.29	0.45	0	1
Religious group, leader	38,157	0.07	0.25	0	1
Association, inactive	38,049	0.14	0.34	0	1
Association, active	38,049	0.20	0.40	0	1
Association, leader	38,049	0.06	0.24	0	1
Trust general	37,618	0.20	0.40	0	1
Trust relatives	38,240	2.44	0.87	0	3
Trust neighbours	38,260	1.82	1.01	0	3
Trust others	38,148	1.35	1.02	0	3
Feeling unsafe	38,238	0.79	1.18	0	4
Fearing crime	38,242	0.67	1.14	0	4
Experience stolen	38,358	0.53	0.93	0	3
Experience attacked	35,929	0.13	0.48	0	3
Fearing election	37,726	1.04	1.18	0	3

Notes: Unweighted statistics. Sample: Whole.

Table 2: Geographical Descriptive Statistics

	(1) Whole sample N = 38,395		(2) Urban sample N = 13,968		(3) Rural sample N = 24,427	
	Mean	SD	Mean	SD	Mean	SD
Distance to paved road (km)	11.6	19.4	5.8	16.6	14.9	20.1
Log distance to paved road	1.69	1.29	1.04	1.07	2.07	1.25
Distance to hypothetical line (km)	97.1	110.4	63.1	101.7	116.6	110.4
Log distance to hypothetical line	3.73	1.58	2.90	1.73	4.20	1.27
Distance to explorer route (km)	156.8	168.9	186.7	193.0	139.7	150.7
Log distance to explorer route	4.44	1.28	4.62	1.29	4.34	1.26
Slope (%)	1.63	2.60	1.41	2.34	1.75	2.72
Log slope	0.72	0.62	0.66	0.58	0.76	0.63
Distance to historical settlement (km)	127.4	122.8	82.7	114.5	152.9	120.1

Notes: Unweighted statistics. Sample: Whole. All of the differences between the urban and rural means are significant at the one percent level.

Table 3: The Correlation between Roads and Self-assessed Current Living Conditions and Lack of Deprivation

Outcomes	Living Conditions (1)	Lack of Deprivation (2)
A. Whole sample		
Log distance to paved road	-0.029***	-0.238***
	(0.006)	(0.024)
N	37764	37498
Adjusted R^2	0.144	0.170
B. Urban sample		
Log distance to paved road	-0.027**	-0.170***
	(0.013)	(0.043)
N	13754	13671
Adjusted R^2	0.129	0.158
C. Rural sample		
Log distance to paved road	-0.015*	-0.156***
	(0.008)	(0.030)
N	24010	23827
Adjusted R^2	0.157	0.171

Notes: These are OLS regressions. Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Whole. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: First-Stage OLS Regressions

Dependent variable: Log distance to the nearest paved road

	Whole	Urban	Rural
	(1)	(2)	(3)
A. Dist to hypothetical lines	0.264***	0.390***	0.133**
(divided by 100)	(0.060)	(0.125)	(0.062)
N	37902	13822	24080
Adjusted R^2	0.374	0.475	0.334
F (excluded instruments)	19.52***	9.65***	4.68**
B. Log Dist to hypothetical lines	0.253***	0.239***	0.162***
	(0.025)	(0.033)	(0.035)
N	37902	13822	24080
Adjusted R^2	0.393	0.498	0.341
F (excluded instruments)	103.10***	53.95***	21.94***
C. Slope (%)	0.030***	0.029**	0.023**
	(0.010)	(0.012)	(0.011)
N	37397	13403	23994
Adjusted R^2	0.367	0.476	0.330
F (excluded instruments)	9.61***	5.95**	4.34**
D. Log slope	0.061	0.058	0.040
	(0.049)	(0.065)	(0.058)
N	37397	13403	23994
Adjusted R^2	0.365	0.475	0.329
F (excluded instruments)	1.55	0.78	0.47
E. Dist to explorer routes	-0.128*	-0.034	-0.111
(divided by 100)	(0.071)	(0.126)	(0.086)
N	37902	13822	24080
Adjusted R^2	0.367	0.463	0.333
F (excluded instruments)	3.26*	0.07	1.65
F. Log Dist to explorer routes	-0.051	-0.045	-0.049
	(0.033)	(0.069)	(0.035)
N	37902	13822	24080
Adjusted R^2	0.366	0.463	0.332
F (excluded instruments)	2.32	0.42	1.98

Notes: Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Whole. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: The Effect of Roads on Self-assessed Current Living Conditions and Lack of Deprivation

Outcomes	Living Conditions (1)	Lack of Deprivation (2)
A. Whole sample		
Log distance to paved road	-0.044	-0.531***
	(0.029)	(0.128)
N	37764	37498
Adjusted R^2	0.144	0.163
B. Urban sample		
Log distance to paved road	-0.080*	-0.400**
	(0.047)	(0.185)
N	13754	13671
Adjusted R^2	0.128	0.156
C. Rural sample		
Log distance to paved road	0.057	-0.158
	(0.079)	(0.271)
N	24010	23827
Adjusted R^2	0.153	0.171

Notes: These are 2SLS regressions. The log of 1 plus the distance to the nearest paved road is instrumented by the log of 1 plus the distance to the nearest hypothetical line. Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Whole. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: The Effect of Roads for Communities over 10km from Historical Settlements

Dependent variable	Dist. to Road	Living Conditions		Lack of Deprivation	
Model	OLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)
Log distance to hyp. line	0.137***				
	(0.031)				
Log distance to paved road		-0.026***	0.028	-0.222***	-0.701**
		(0.007)	(0.073)	(0.026)	(0.317)
N	32882	32769	32769	32545	32545
Adjusted R^2	0.348	0.149	0.147	0.171	0.153
F (excluded instruments)	20.09***				

Notes: In columns 3 and 5, the log of 1 plus the distance to the nearest paved road is instrumented by the log of 1 plus the distance to the nearest hypothetical line. Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Robustness Checks: The Dependent Variables

Model	Categorical variable		Binary variable	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
Dimensions of basic needs				
A. Having enough food				
Log distance to paved road	-0.055***	-0.238**	-0.019***	-0.096**
	(0.008)	(0.101)	(0.003)	(0.039)
N	32825	32825	32825	32825
Adjusted R^2	0.126	0.103	0.109	0.082
B. Having enough water				
Log distance to paved road	-0.087***	-0.320**	-0.027***	-0.085**
	(0.011)	(0.132)	(0.004)	(0.043)
N	32826	32826	32826	32826
Adjusted R^2	0.130	0.100	0.115	0.100
C. Having enough medical care				
Log distance to paved road	-0.079***	-0.188**	-0.025***	-0.039
	(0.009)	(0.093)	(0.003)	(0.032)
N	32732	32732	32732	32732
Adjusted R^2	0.145	0.138	0.122	0.121
D. Having enough cooking fuel				
Log distance to paved road	-0.002	0.036	0.001	0.013
	(0.008)	(0.085)	(0.003)	(0.029)
N	32708	32708	32708	32708
Adjusted R^2	0.141	0.140	0.125	0.124
Dummy variable for Living Conditions				
E. Having good or very good living conditions				
Log distance to paved road	-0.026***	0.028	-0.008***	-0.001
	(0.007)	(0.073)	(0.003)	(0.027)
N	32769	32769	32769	32769
Adjusted R^2	0.149	0.147	0.121	0.120

Notes: In columns 2 and 4, the log of 1 plus the distance to the nearest paved road is instrumented by the log of 1 plus the distance to the nearest hypothetical lines. Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies are included in each model. Columns 1 and 2 use the ordinal measures of whether each basic need is satisfied (values from 0 to 4) and living conditions (1 to 5); columns 3 and 4 use the dummy variables for reporting having had difficulty in satisfying their basic needs in each dimension over the past year never or just once or twice, and for having good or very good living conditions. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Robustness Checks: Measures of Distance to the Nearest Paved Road

Outcome	Living Conditions (1)	Lack of Deprivation (2)
A. Log distance to road	0.028	-0.701**
(Benchmark)	(0.073)	(0.317)
B. Dist to road > 5km	0.110	-2.755*
	(0.294)	(1.627)
C. Dist to road > 10km	0.065	-1.632**
	(0.169)	(0.738)
D. Dist to road > 20km	0.060	-1.505**
	(0.154)	(0.642)
E. Distance to road	0.369	-3.640**
(divided by 100)	(0.387)	(1.740)
F. Distance to road	0.942*	-4.580*
(divided by 100)	(0.539)	(2.723)
Squared distance to road	-0.005	0.007
(divided by 100)	(0.003)	(0.022)
G. Inverse of (1+ Dist. to road)	1.894	11.431
	(2.447)	(13.725)

Notes: Instrument: Distance to the nearest hypothetical lines (the log of 1 plus, or levels, or in quadratic, or 1 over 1+distance to HL). Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Communities over 10km from historical settlements.

Table 9: Robustness Checks regarding the Sample

Outcome	Living Conditions (1)	Lack of Deprivation (2)
Benchmark	0.028	-0.701**
	(0.073)	(0.317)
N	32769	32545
Adjusted R^2	0.147	0.153
Removing obs. less than 5 km from road	0.093	-0.916**
	(0.092)	(0.376)
N	16588	16473
Adjusted R^2	0.139	0.154
Removing obs. less than 10 km from road	0.155	-1.047**
	(0.133)	(0.517)
N	12405	12319
Adjusted R^2	0.132	0.150

Notes: These are the instrumented estimates for the log of 1 plus distance to the nearest paved road in linear regressions. Standard errors clustered at the community level appear in brackets. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Mediation Analysis

Outcome	Living Conditions (1)	Lack of Deprivation (2)
Additional controls:		
A. Basic specification only	0.028	-0.701**
	(0.073)	(0.317)
B. Urban location	0.107	-0.516
	(0.110)	(0.418)
C. Education	0.068	-0.545*
	(0.079)	(0.305)
D. Paid work	0.045	-0.678**
	(0.075)	(0.315)
E. Visited Public Health Centre	0.032	-0.699**
	(0.074)	(0.317)
F. Public goods	0.097	-0.344
	(0.100)	(0.335)
G. Mobile phone: Calls	0.060	-0.575*
	(0.078)	(0.308)
H. Mobile phone: SMS	0.060	-0.612*
	(0.078)	(0.314)
I. Mobile phone: Money	0.040	-0.690**
	(0.077)	(0.330)
J. Computer	0.052	-0.653**
	(0.075)	(0.310)
K. Internet	0.056	-0.649**
	(0.075)	(0.307)
L. Association	0.030	-0.720**
	(0.072)	(0.317)
M. Trust general	0.034	-0.698**
	(0.076)	(0.327)
N. Trust w.r.t. relatives, neighb. and others	0.022	-0.682**
	(0.074)	(0.321)
O. Feelings of security and crime	0.027	-0.712**
	(0.074)	(0.314)
P. Experiences of crime	0.002	-0.725**
	(0.074)	(0.313)
Q. Social capital (L, M, N, O, P)	0.012	-0.674**
	(0.077)	(0.322)
R. All Mediators (except P)	0.194	-0.179
	(0.147)	(0.411)

Notes: These are the instrumented estimates for the log of 1 plus distance to the nearest paved road in linear regressions. Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. The models in rows P and Q do not include Tanzania, for which the experience of being attacked variable is missing. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

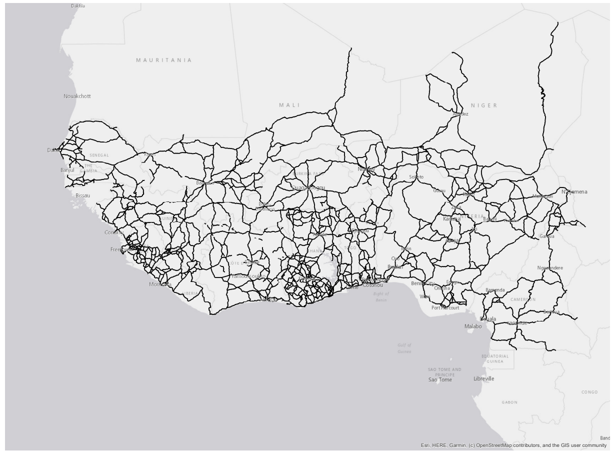
Table 11: The Effect of Roads on the Wealth Index: DHS Data

Dependent Variable: Standardised Wealth Index

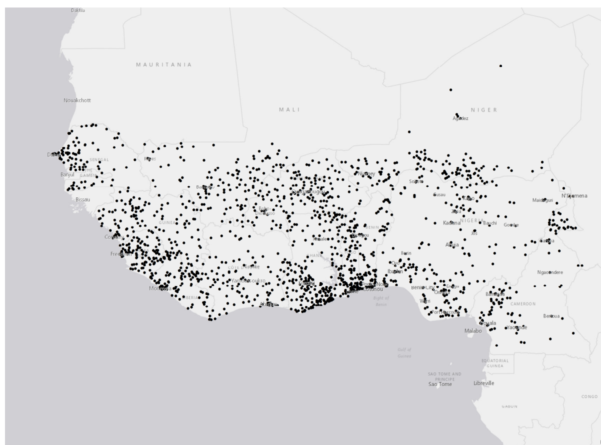
	(1) OLS	(2) 2SLS
A. Whole sample		
Log distance to paved road	-0.231***	-0.544***
	(0.012)	(0.122)
N	234962	234962
Adjusted R^2	0.176	0.062
B. Non-migrant sample		
Log distance to paved road	-0.160***	-0.535***
	(0.009)	(0.121)
N	119161	119161
Adjusted R^2	0.266	0.025
C. Migrant sample		
Log distance to paved road	-0.281***	-0.483***
	(0.016)	(0.143)
N	115440	115440
Adjusted R^2	0.169	0.134

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

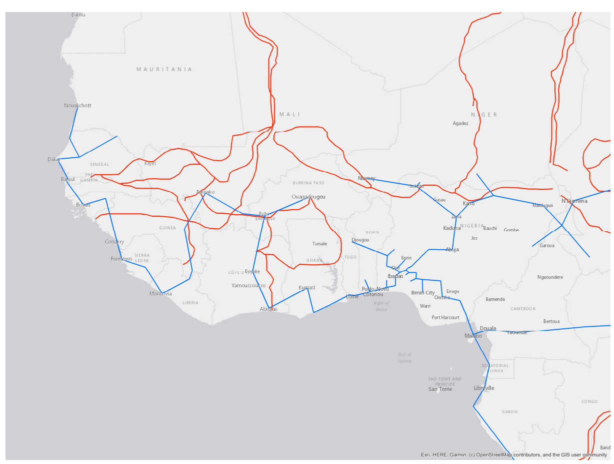
Figure 1: Geographic Data, West Africa



(a) Location of Current Paved Roads



(b) Location of the Enumeration Areas



(c) Instruments - Location of the Hypothetical Lines (blue) and Explorer Routes (red)

Figure 2: F-statistic from the First-stage Regressions, by Minimum Distance to the Nearest Historical Settlement

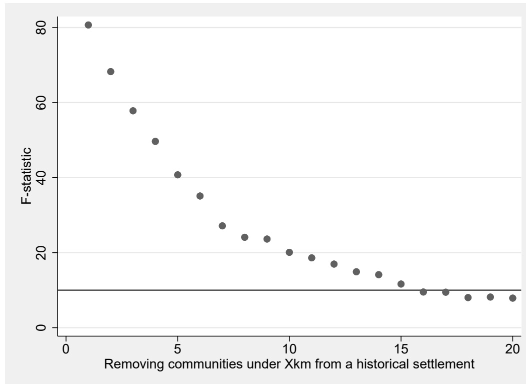
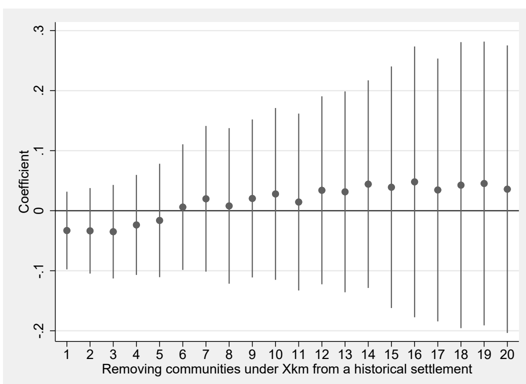
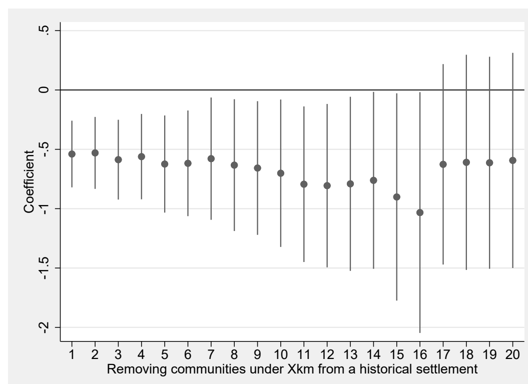


Figure 3: Effects of Roads, by Minimum Distance to the Nearest Historical Settlement

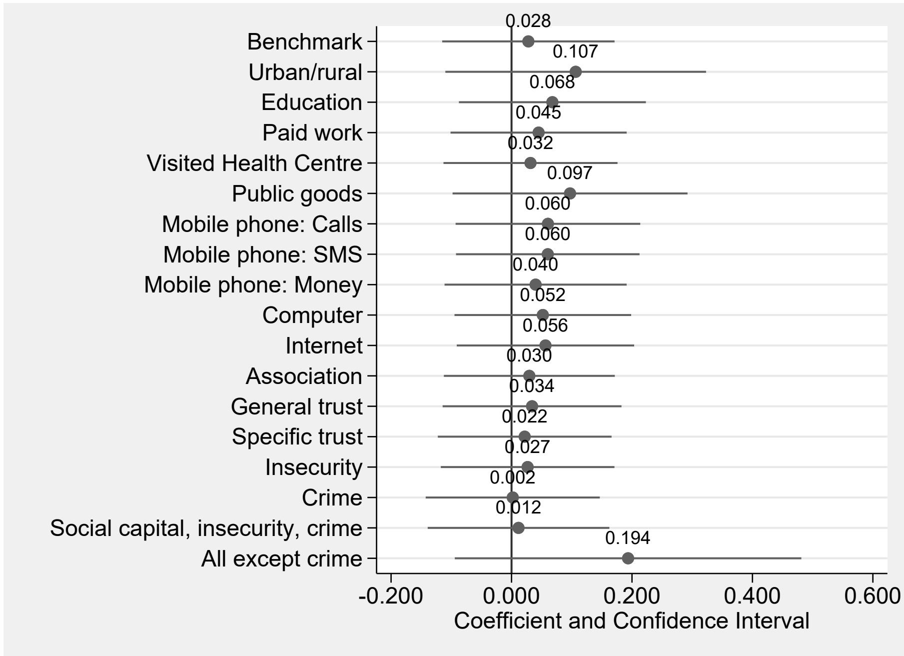


(a) Effects on Subjective Living Conditions

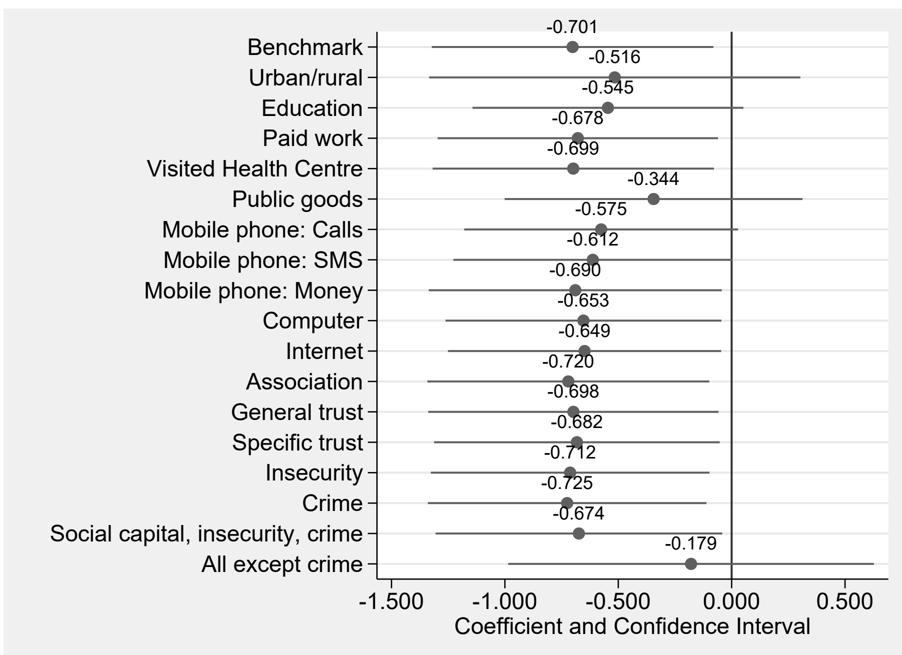


(b) Effects on the Lack of Deprivation

Figure 4: Mediation in the Effects of Roads on Well-being

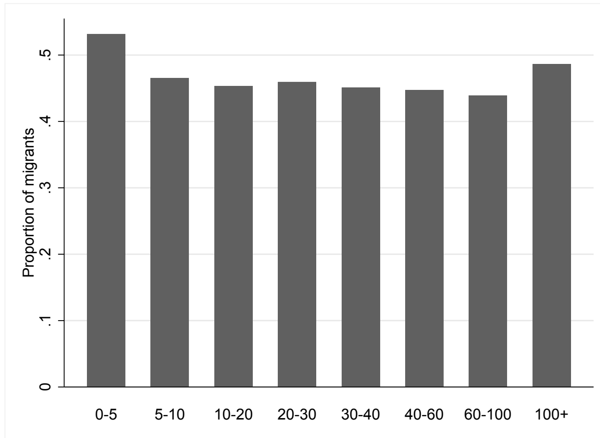


(a) Effects on Subjective Living Conditions



(b) Effects on Lack of Deprivation

Figure 5: Proportions of Ever Migrants, by Distance to the Current Paved Road



Appendix

Table A1: List of Countries, Sample Size and Means of Key Variables

Countries	N	Living Conditions	Lack of Deprivation	Distance to Road	Log Distance to Road
Benin	1,200	2.58	11.8	3.5	1.05
Burkina Faso	1,200	2.90	10.3	14.1	1.84
Burundi	1,200	2.59	9.8	8.4	1.67
Botswana	1,200	2.39	12.5	14.5	1.65
Cameroon	1,200	2.93	10.0	11.6	1.52
Cote d'Ivoire	1,200	2.62	10.2	6.8	1.33
Ghana	2,400	2.52	14.0	2.5	0.83
Guinea	1,200	2.62	10.0	7.6	1.58
Kenya	2,399	1.96	12.1	6.8	1.24
Liberia	1,199	3.07	10.8	5.7	1.22
Mali	1,200	2.70	11.8	11.5	1.69
Malawi	2,408	2.51	11.5	9.9	1.88
Mozambique	2,400	2.86	12.0	24.1	2.32
Namibia	1,200	3.04	12.9	14.3	1.62
Niger	1,199	3.09	9.7	11.7	1.66
Nigeria	2,400	2.91	11.5	11.7	1.74
Senegal	1,200	2.73	10.1	4.8	1.11
Sierra Leone	1,190	2.99	11.1	5.8	1.30
Sudan	1,199	2.60	11.5	19.6	1.63
Tanzania	2,400	2.13	11.5	17.3	2.22
Togo	1,201	2.30	9.7	4.1	1.12
Uganda	2,400	2.33	11.2	12.3	2.08
Zambia	1,200	3.21	11.9	24.4	2.52
Zimbabwe	2,400	2.76	11.4	16.5	2.54

Notes: Unweighted statistics. Sample: Whole.

Table A2: Sample Descriptive Statistics

	N	Mean	SD	Min	Max
Having enough food	38,323	2.84	1.23	0	4
Having enough water	38,325	2.76	1.39	0	4
Having enough medical care	38,207	2.69	1.31	0	4
Having enough cooking fuel	38,184	3.08	1.21	0	4
Having enough food	38,323	0.60	0.49	0	1
Having enough water	38,325	0.59	0.49	0	1
Having enough medical care	38,207	0.56	0.50	0	1
Having enough cooking fuel	38,184	0.69	0.46	0	1

Notes: Unweighted statistics. Sample: Whole.

Table A3: Distance to Roads and Well-being - All Controls

	Living Conditions		Lack of Deprivation	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
Log distance to paved road	-0.029***	-0.044	-0.238***	-0.531***
	(0.006)	(0.029)	(0.024)	(0.128)
Age	-0.029***	-0.029***	-0.047***	-0.045***
	(0.002)	(0.002)	(0.007)	(0.007)
Age-squared/100	0.026***	0.026***	0.032***	0.031***
	(0.002)	(0.002)	(0.007)	(0.007)
Female	-0.030***	-0.030***	-0.066**	-0.065**
	(0.011)	(0.011)	(0.030)	(0.029)
Latitude	0.034***	0.034***	-0.148***	-0.133**
	(0.011)	(0.011)	(0.051)	(0.053)
Longitude	-0.044***	-0.044***	-0.025	-0.034
	(0.013)	(0.013)	(0.052)	(0.056)
Altitude	-0.000	-0.000	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	3.094***	3.108***	14.833***	15.113***
	(0.189)	(0.188)	(0.732)	(0.765)
N	37764	37764	37498	37498
Adjusted R^2	0.144	0.144	0.170	0.163

Notes: Standard errors clustered at the community level appear in parentheses. Regional dummies are included but the coefficients are not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Whole.

Table A4: Restricted Sample Descriptive Statistics

Variables	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Current living conditions (LC)	33,220	2.62	1.19	1	5
Lack of depriv., over 4 items	32,998	11.29	3.74	0	16
Distance to road, km	33,341	13.2	20.3	0	248.1
Log distance to road	33,341	1.86	1.29	0	5.52
Distance to hypothetical line, km	33,341	111.5	111.7	0.02	733.6
Log distance to hypothetical line	33,341	4.13	1.26	0.02	6.60
Age	32,999	36.7	14.2	18	105
Female	33,341	0.50	0.50	0	1
Urban	33,341	0.29	0.45	0	1
Primary education	33,261	0.34	0.47	0	1
Secondary education	33,261	0.33	0.47	0	1
Higher education	33,261	0.09	0.29	0	1
Paid work	33,207	0.30	0.46	0	1
Visited Public Health Centre	33,097	0.86	0.35	0	1
Electricity	33,341	0.50	0.50	0	1
Piped water	33,285	0.43	0.49	0	1
Sewage	33,125	0.18	0.38	0	1
Mobile	33,317	0.91	0.29	0	1
Post Office	33,245	0.17	0.37	0	1
School	33,221	0.88	0.33	0	1
Police Station	33,197	0.31	0.46	0	1
Clinic	33,165	0.57	0.50	0	1
Market	33,237	0.64	0.48	0	1
Use computer	32,735	0.37	.99	0	4
Use internet	32,519	0.37	1.01	0	4
Use mobile, call	33,039	2.20	1.49	0	4
Use mobile, text message	32,997	1.41	1.50	0	4
Use mobile, money	32,616	0.31	.76	0	4
Religious group, inactive	33,159	0.17	0.37	0	1
Religious group, active	33,159	0.29	0.45	0	1
Religious group, leader	33,159	0.07	0.25	0	1
Association, inactive	33,069	0.14	0.35	0	1
Association, active	33,069	0.21	0.40	0	1
Association, leader	33,069	0.06	0.25	0	1
Trust general	32,673	0.21	0.40	0	1
Trust relatives	33,203	2.45	0.87	0	3
Trust neighbours	33,234	1.86	1.01	0	3
Trust others	33,143	1.38	1.02	0	3
Feeling unsafe	33,216	0.75	1.16	0	4
Fearing crime	33,223	0.64	1.11	0	4
Experience stolen	33,311	0.52	0.92	0	3
Experience attacked	31,144	0.13	0.47	0	3
Fearing election	32,777	1.04	1.18	0	3

Notes: Unweighted statistics. Sample: Communities over 10km from historical settlements.

Table A5: Distance to Roads and Relative Living Conditions

	OLS (1)	2SLS (2)
Panel A. Whole sample		
Log distance to paved road	-0.027***	-0.038
	(0.006)	(0.027)
N	36692	36692
Adjusted R^2	0.130	0.130
Panel B. Far from historical settlements		
Log distance to paved road	-0.026***	-0.016
	(0.006)	(0.062)
N	31877	31877
Adjusted R^2	0.128	0.128

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Whole in Panel A, Communities over 10km from historical settlements in Panel B.

Table A6: Jackknife Analysis

	Living Conditions		Lack of Deprivation	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
All	-0.026***	0.028	-0.222***	-0.701**
	(0.007)	(0.073)	(0.026)	(0.317)
Burundi	-0.028***	0.013	-0.220***	-0.774**
	(0.007)	(0.073)	(0.026)	(0.326)
Benin	-0.027***	0.028	-0.221***	-0.723**
	(0.007)	(0.073)	(0.026)	(0.319)
Burkina Faso	-0.027***	0.025	-0.232***	-0.745**
	(0.007)	(0.074)	(0.027)	(0.325)
Botswana	-0.024***	0.046	-0.220***	-0.824**
	(0.007)	(0.089)	(0.027)	(0.396)
Cameroon	-0.028***	0.025	-0.228***	-0.726**
	(0.007)	(0.078)	(0.027)	(0.341)
Cote d'Ivoire	-0.028***	0.023	-0.226***	-0.678**
	(0.007)	(0.072)	(0.027)	(0.314)
Ghana	-0.025***	0.001	-0.219***	-0.660**
	(0.007)	(0.067)	(0.027)	(0.301)
Guinea	-0.025***	0.039	-0.222***	-0.723**
	(0.007)	(0.069)	(0.027)	(0.304)
Kenya	-0.027***	0.023	-0.228***	-0.769**
	(0.007)	(0.074)	(0.027)	(0.326)
Liberia	-0.025***	0.042	-0.222***	-0.736**
	(0.007)	(0.078)	(0.026)	(0.341)
Mali	-0.030***	0.040	-0.224***	-0.643**
	(0.007)	(0.071)	(0.027)	(0.302)
Malawi	-0.028***	0.017	-0.216***	-0.864***
	(0.007)	(0.070)	(0.027)	(0.312)
Mozambique	-0.031***	0.062	-0.225***	-0.681*
	(0.007)	(0.083)	(0.027)	(0.354)
Namibia	-0.026***	0.049	-0.218***	-0.742**
	(0.007)	(0.080)	(0.026)	(0.350)
Niger	-0.025***	0.063	-0.227***	-0.632**
	(0.007)	(0.078)	(0.027)	(0.322)
Nigeria	-0.027***	0.026	-0.218***	-0.654**
	(0.007)	(0.070)	(0.026)	(0.298)
Senegal	-0.026***	0.023	-0.209***	-0.673**
	(0.007)	(0.071)	(0.026)	(0.305)
Sierra Leone	-0.026***	0.045	-0.232***	-0.807**
	(0.007)	(0.081)	(0.026)	(0.354)
Sudan	-0.026***	0.049	-0.222***	-0.682**
	(0.007)	(0.077)	(0.026)	(0.328)
Tanzania	-0.025***	0.000	-0.230***	-0.727**
	(0.007)	(0.072)	(0.027)	(0.319)
Togo	-0.026***	0.042	-0.218***	-0.565*
	(0.007)	(0.080)	(0.026)	(0.322)
Uganda	-0.025***	0.035	-0.201***	-0.410
	(0.007)	(0.072)	(0.027)	(0.311)
Zambia	-0.023***	0.019	-0.220***	-0.630**
	(0.007)	(0.069)	(0.027)	(0.297)
Zimbabwe	-0.028***	0.097	-0.229***	-0.809**
	(0.007)	(0.078)	(0.026)	(0.346)

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Sample Size, DHS

Country	Year	Obs.	Restricted Obs.
Benin	2017	23320	19138
Burkina Faso	2003	15894	14098
Cameroon	2004	15188	12115
Ghana	2008	9396	8201
Guinea	2005	10959	9953
Kenya	2009	11649	10335
Liberia	2007	12889	10674
Mali	2006	18550	15780
Malawi	2010	29756	27516
Namibia	2007	13128	11492
Nigeria	2008	48153	43065
Senegal	2005	17846	15834
Tanzania	2015	16280	14555
Uganda	2006	9725	8020
Zimbabwe	2015	17989	15978
Total		270722	236754

Table A8: Descriptive Statistics, DHS

Variable	N	Mean	Std. Dev.	Min	Max
Migrant	236,388	.49	.49	0	1
Std. wealth index	236,754	0	1	-6.25	11.56
Wealth index (raw, continuous)	236,754	-11340	260297	-1638620	2996530
Log distance to paved road	236,754	1.95	1.22	.00003	5.31
Log distance to hyp. line	236,754	4.08	1.21	.002	6.69
Dist. to hyp. line (/100)	236,754	1.03	1.02	.00002	8.07
Log distance to explorer route	236,754	4.57	1.38	.012	6.83
Dist. to expl. route (/100)	236,754	1.92	2.07	.0001	9.24
Log slope	232,831	.63	.60	0	4.14
Slope (%)	232,831	1.48	2.54	0	62.02
Age	236,754	29.2	10.3	15	64
Female	236,754	.70	.46	0	1
Latitude	236,754	2.00	11.5	-28.6	20.18
Longitude	236,754	11.48	17.5	-17.5	41.8
Altitude	234,962	573.5	498.2	1	2951

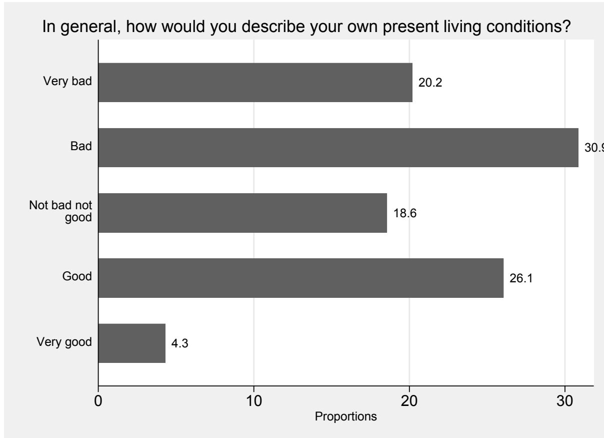
Notes: Unweighted statistics. Sample: Communities over 10km from historical settlements (permanent residents).

Table A9: First-Stage OLS Regressions (DHS)

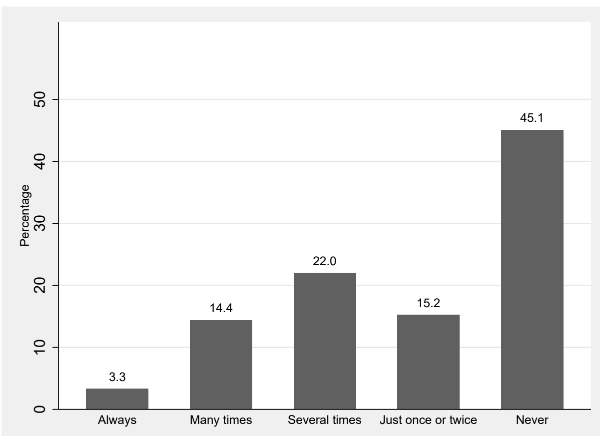
Dependent variable: Log distance to the nearest paved road	(1)
A. Distance to hypothetical lines	0.253***
(divided by 100)	(0.038)
N	234962
Adjusted R^2	0.218
F (excluded instruments)	45.136
B. Log Distance to hypothetical lines	0.149***
	(0.021)
N	234962
Adjusted R^2	0.218
F (excluded instruments)	52.297
C. Slope (%)	0.014**
	(0.006)
N	232831
Adjusted R^2	0.202
F (excluded instruments)	5.320
C. Slope (log)	0.012
	(0.033)
N	232831
Adjusted R^2	0.201
F (excluded instruments)	0.146
E. Distance to explorer routes	0.033
(divided by 100)	(0.040)
N	234962
Adjusted R^2	0.209
F (excluded instruments)	0.686
F. Log Distance to explorer routes	0.054**
	(0.023)
N	234962
Adjusted R^2	0.210
F (excluded instruments)	5.524

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

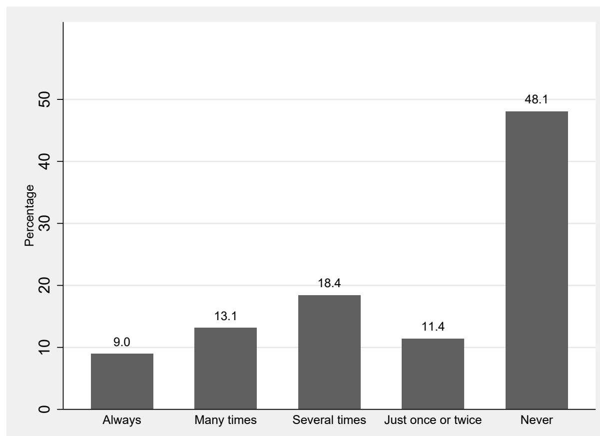
Figure A1: Distribution of the Categorical Variables



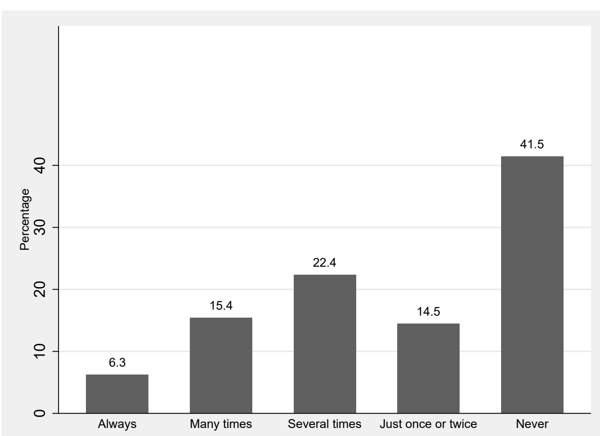
(a) Subjective well-being



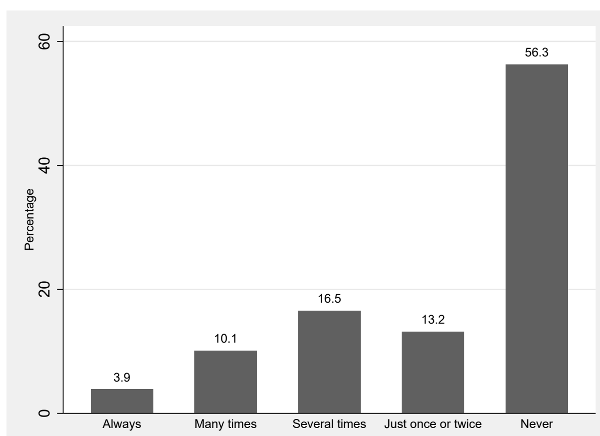
(b) Gone without enough food



(c) Gone without enough water

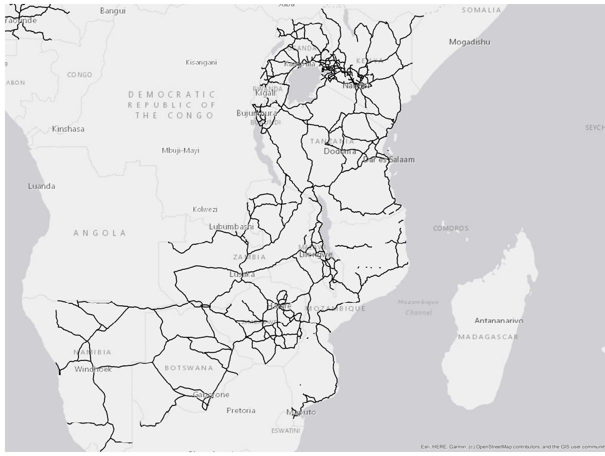


(d) Gone without enough medical care

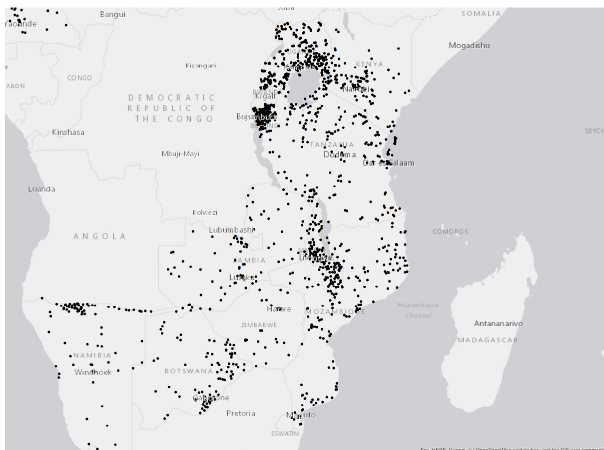


(e) Gone without enough cooking fuel

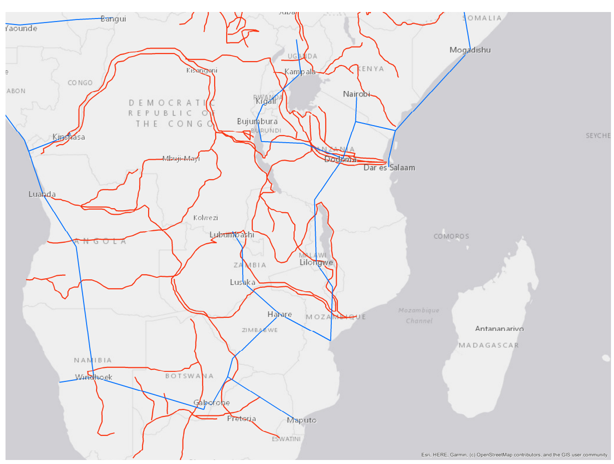
Figure A2: Geographic data, East and Southern Africa



(a) Location of the current paved roads

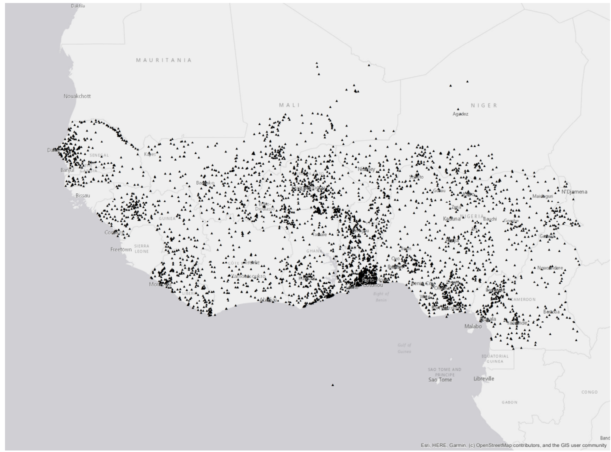


(b) Location of the enumeration areas

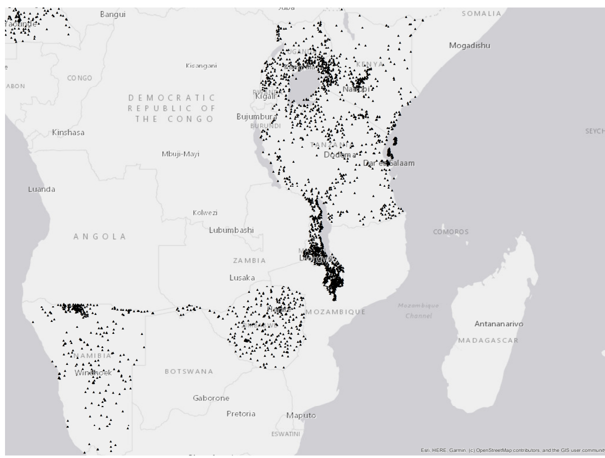


(c) Instruments - Location of the hypothetical lines (blue) and explorer routes (red)

Figure A3: Location of the EA in the DHS



(a) West Africa



(b) South-East

Online Appendix - Not for publication

Table O1: The Effect of Roads on Self-assessed Current Living Conditions and Lack of Deprivation - Ordered Probit Estimations

Outcomes	Living Conditions		Lack of Deprivation	
Instrumented	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
A. Whole sample				
Log distance to paved road	-0.027***	-0.042	-0.073***	-0.161***
	(0.006)	(0.028)	(0.007)	(0.038)
N	37764	37764	37498	37498
B. Urban sample				
Log distance to paved road	-0.025**	-0.074	-0.052***	-0.124**
	(0.012)	(0.045)	(0.013)	(0.059)
N	13754	13754	13671	13671
C. Rural sample				
Log distance to paved road	-0.013*	0.048	-0.049***	-0.048
	(0.008)	(0.077)	(0.009)	(0.082)
N	24010	24010	23827	23827
D. Communities over 10 km from historical settlements				
Log distance to paved road	-0.025***	0.022	-0.069***	-0.217**
	(0.006)	(0.070)	(0.008)	(0.092)
N	32769	32769	32545	32545

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Whole.

Table O2: Distance to Hypothetical Lines and Well-being

	Living Conditions (1)	Lack of Deprivation (2)
A. Whole sample		
Log distance to hypothetical lines	-0.011	-0.134***
	(0.008)	(0.031)
N	37764	37498
Adjusted R^2	0.144	0.167
B. Far from historical settlements		
Log distance to hypothetical lines	0.004	-0.096**
	(0.010)	(0.040)
N	32769	32545
Adjusted R^2	0.149	0.168

Notes: Standard errors clustered at the community level appear in parentheses. Linear models. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample: Whole in Panel A, Communities over 10km from historical settlements in Panel B.

Table O3: The Number of Historical Settlements per Country

Country	Number
Angola	3
Benin	4
Botswana	5
Burkina Faso	2
Burundi	2
Cameroon	3
Cote d'Ivoire	2
Ghana	4
Guinea	3
Liberia	2
Malawi	2
Mali	2
Mozambique	2
Namibia	2
Niger	2
Nigeria	35
Senegal	3
Sierra Leone	2
Sudan	10
Tanzania, United Rep of	6
Togo	2
Uganda	2
Zambia	2
Zimbabwe	2

Table O4: First-Stage OLS Regression - All Controls

Dependent variable: Log distance to paved road	(1)	(2)	(3)	(4)	(5)	(6)
Log distance to hypothetical line	0.253***					
	(0.025)					
Distance to hypothetical line/100		0.264***				
		(0.060)				
Log distance to explorer route			-0.051			
			(0.033)			
Distance to explorer route/100				-0.128*		
				(0.071)		
Log slope					0.061	
					(0.049)	
Slope						0.030***
						(0.010)
Age	0.006***	0.006***	0.005**	0.005**	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age-squared/100	-0.005**	-0.005**	-0.004*	-0.004*	-0.004**	-0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female	0.002	0.003	0.003	0.003	0.003	0.003
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Latitude	0.029	0.035	0.046	0.032	0.044	0.047
	(0.033)	(0.034)	(0.035)	(0.036)	(0.035)	(0.034)
Longitude	-0.034	-0.046	-0.026	-0.013	-0.036	-0.039
	(0.053)	(0.055)	(0.054)	(0.051)	(0.055)	(0.055)
Altitude	-0.000	-0.000	-0.000*	-0.000*	-0.000**	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: Standard errors clustered at the community level appear in parentheses. Regional dummies are included but the coefficients are not shown. Sample: Whole. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O5: Robustness Checks - using Solid Roof as the Outcome Variable

	(1)	(2)
	OLS	2SLS
Log distance to paved road	-0.036***	-0.183***
	(0.004)	(0.059)
N	32727	32727
Adjusted R^2	0.350	0.250

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table O6: First-Stage Regressions Controlling for the Possible Mediators

Dependent variable: Log distance to the nearest paved road

	(1) Coef.	(2) SE.	(3) F-stat
Additional controls			
A. Basic specification	0.137***	(0.031)	20.09
B. Urban location	0.099***	(0.030)	11.23
C. Education	0.133***	(0.031)	18.79
D. Paid work	0.136***	(0.031)	19.73
E. Visited Public Health Centre	0.136***	(0.031)	19.75
F. Public goods	0.111***	(0.030)	13.63
G. Mobile phone: Calls	0.132***	(0.031)	18.69
H. Mobile phone: SMS	0.133***	(0.031)	18.74
I. Mobile phone: Money	0.133***	(0.031)	18.89
J. Computer	0.135***	(0.031)	19.36
K. Internet	0.136***	(0.031)	19.53
L. Association	0.137***	(0.031)	20.05
M. Trust general	0.135***	(0.031)	19.06
N. Trust w.r.t. relatives, neighb. and other	0.136***	(0.031)	19.59
O. Feelings of security and crime	0.136***	(0.031)	19.74
P. Experiences of crime	0.142***	(0.032)	19.95
Q. Social capital (L, M, N, O, P)	0.138***	(0.032)	18.35
R. All mediators (except P)	0.085***	(0.030)	8.03

Notes: Standard errors clustered at the community level appear in parentheses. Controls: Age, age-squared, female, altitude, longitude, latitude and regional dummies. Sample: Communities over 10km from historical settlements. Rows P and Q do not include Tanzania, for which the experience of being attacked variable is missing. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.