In Search of a Spatial Equilibrium in the Developing World*

Douglas Gollin† | Martina Kirchberger§ | David Lagakos‡

University of Oxford | *Trinity College Dublin* | *UCSD and NBER*

February 2017

Abstract

In most developing countries, there is a large gap in income per head between urban and rural areas. One interpretation of this gap is a spatial equilibrium, in which the higher incomes of urban areas are offset by lower non-monetary amenities. In this paper, we draw on new spatial evidence to document how amenities vary across space within a large set of developing countries. We focus on measures of health, housing quality, crime and pollution, which vary substantially across space, and for which high quality data are available. We find that in almost all countries, and for almost all measures, amenities are constant or increasing in population density. In addition, in most countries net internal migration rates are toward denser areas. These finding are hard to reconcile with a spatial equilibrium. Instead, they suggest that developing countries are undergoing a reallocation of workers to denser areas, where living standards are on average higher.

*Email: † douglas.gollin@qeh.ox.ac.uk, § martina.kirchberger@tcd.ie, ‡ lagakos@ucsd.edu. For helpful comments we thank Deborah Balk, Banu Demir-Pakel, Dave Donaldson, Marcel Fafchamps, Darby Jack, Pat Kinney, Horacio Larreguy, Ethan Ligon, Matthew Neidell, Natalia Ramondo, Chris Small, Luke Valin, David Weil, and seminar participants at Bristol, Columbia, the LSE, EDRI, Edinburgh, Sussex, Copenhagen, PSE, Barcelona, Nottingham, TCD, NEUDC and the 2015 CSAE conference. For excellent research assistance we thank Jessie Wang and Abed Mutemi, as well as the Bill and Melinda Gates Foundation for supporting his work on this project. We would like to thank Jeff Geddes and Aaron van Donkelaar for sharing the pollution data and their advice. Kirchberger is grateful for many helpful conversations and suggestions while hosted by the Center for International Earth Science Information Network at Columbia University. This paper is a part of a Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. All potential errors are our own.

1. Introduction

Economists have long recognized that there are large income differences across space within countries. A body of recent evidence has documented that these income gaps are particularly pronounced in the developing world. In particular, urban-rural wage gaps are particular large in developing countries (Ferré, Ferreira, and Lanjouw, 2012; Young, 2014), as, similarly, are gaps in average income between non-agricultural workers and agricultural workers (Gollin, Lagakos, and Waugh, 2014). In an accounting sense, these gaps matter a lot for understanding why developing countries have such low aggregate income, since such a large fraction of workers in developing countries live in rural areas and work in agriculture (Caselli, 2005; Restuccia, Yang, and Zhu, 2008; Vollrath, 2009; McMillan and Rodrik, 2011).

One interpretation of these urban-rural income gaps is a spatial equilibrium, in which the higher incomes of urban areas are offset by lower non-monetary amenities. This idea of a spatial equilibrium, with workers indifferent between all locations, is one of the simplest and most appealing ways economists have used to analyze how economic agents locate through space (Rosen, 1979; Roback, 1982; Glaeser and Gottlieb, 2009). The rationale is simple: if any region offered a better bundle of consumption and amenities than the rest, then agents would move until such arbitrage opportunities were gone. The concept of a spatial equilibrium has proven useful in learning about many different economic phenomena, including the wage and size distribution of U.S. cities (Baum-Snow and Pavan, 2012; Desmet and Rossi-Hansberg, 2013), the dynamics of U.S. manufacturing and service activity through space (Desmet and Rossi-Hansberg, 2014), and the welfare effects of transportation infrastructure improvements (Allen and Arkolakis, 2014).[1]

Almost all papers that assume a spatial equilibrium use a model to infer the amenity values of each location, rather than measuring the amenities directly (Glaeser and Gottlieb, 2009). The reason, presumably, is that amenities are simply too hard to measure. In the context of developing countries, which tend to have much higher incomes in urban areas, a spatial equilibrium implies that urban amenities must be much lower than rural amenities. If this is true, researchers and policy makers should like to know which amenities of urban areas are so much worse than in rural areas. If urban amenities are not lower, then a spatial equilibrium may not be the right interpretation of the significantly higher urban wages in developing countries. Its hard to imagine distinguishing between these two alternatives without direct measures of amenities across space.

In this paper, we attempt to fill this gap by building a new database of amenity values across space in a large set of developing countries. We focus on real amenity measures based on health outcomes, housing quality, crime exposure and pollution exposure, which are among the most

[1] The concept of a spatial equilibria has been applied is numerous other applications as well; see Redding and Rossi-Hansberg (Forthcoming) for an overview of the recent literature.

commonly cited amenities varying across space, and for which high-quality data are available.[2] To construct our database, we combine detailed household survey data, satellite imagery and population density data coming from the Gridded Population of the World Version 4 (GPW). The surveys we employ are the Demographic and Health Surveys (DHS), Living Standards Measurement Surveys (LSMS) and Afrobarometer Surveys. These surveys ask a nationally representative sample of individuals detailed questions about their living situation, including questions about their health, housing quality, experience with crime and access to public goods. The surveys also report the GPS coordinates or location names of each geographic region sampled, which we link to data on population density from the GPW. We also draw on estimates on the spatial distribution of air pollution concentrations for fine particulate matter (PM2.5) and nitrogen dioxide (NO2) derived from a range of satellite instruments. An advantage of our approach is that we do not have to rely on reported categories of what is "urban" and "rural" reflecting administrative boundaries and cut-offs that might not be consistent across space and time. Our linked data allows us to look how our amenities measures vary across population density within each country, using internationally comparable measures of density.

We then use our data to assess how these amenity measures vary across population density in each of our countries. We find that in almost all countries, and for almost all measures, amenities are non-decreasing in population density. Housing quality metrics, such as roof and floor quality, electricity access and extent of indoor plumbing are increasing in population density in virtually every country we consider. Health of children, such as percent consuming a minimum acceptable diet, is also almost everywhere increasing in density. Crime measures are surprisingly constant across density, with most countries having similar rates of reported theft, assault and fear of assault with density. An important caveat is that almost all of our crime data comes from Africa.

For atmospheric pollution, we find that PM2.5 and NO2 levels are unrelated to density in most developing countries. In African countries, the lack of much manufacturing activity means that urban areas tend to have low overall air pollution levels. The locations in Africa with the highest concentrations of PM2.5 are those closest to the Saharan desert, which are exposed to high levels of dust. Excluding sand and dust, we find little correlation between density and PM2.5. In India and China, in contrast, we find strong evidence that pollution levels are increasing in density, and at dangerously high levels in many urban areas. So perhaps air pollution is the compensating differential or urban life in China and India. This doesn't help explain Africa, of course, where urban-rural gaps are on average very high. Another type of air pollution comes from indoor cooking, which is prevalent in many low-income countries. When we look across density, we find that most countries have higher rates of indoor cooking (with traditional cooking sources, like

[2] By real we mean quantities rather than prices. In the following section, we discuss how higher housing prices in urban areas are consistent with a spatial equilibrium only if housing quality or consumption levels (or both) are lower in urban areas, hence our analysis of housing quality.

wood) in rural areas, not urban areas.

Our findings on amenity measures across space do not provide much support for a spatial equilibrium in Africa. With few exceptions, measures of health, housing, crime and pollution are constant or increasing in population density. Of course, searching for something and not finding it does not mean it is not there. We did not, and could not, consider all possible amenities. In fact, by definition there is an "amenity" which is higher in denser areas: population density. Its possible that people could simply dislike living in denser areas.[3] One should view our findings so far as the most carefully look to date at the "prime suspects" of amenities mentioned in the literature, rather than a proof of non-existence of a spatial equilibrium, which, even in principle, we could not produce.

To complement our findings for these four amenities, we take two further steps. First, we look at data on internal migration rates within our countries. In a spatial equilibrium, net migration rates should not be systematically toward denser areas or toward rural areas. What we find is that they are in fact mostly toward denser areas. In particular, individuals living in the most remote areas tend to migrate most often to smaller towns and cities, while those in smaller towns and cities migrate most often to large cities. Second, we look at some additional "secondary" measures of amenities, moving down the list of suspects, but for fewer countries (and less data). Here we find one amenity that does seem to decrease with population density: trust in neighbors, though the magnitudes are not overwhelmingly higher in denser areas. For the other secondary metrics we consider – anxiety, reported lack of food, and reported lack of medicine – we again find worse outcomes in rural areas.

We conclude by offering a simple explanation for the facts we document, which is that living standards in developing countries are higher on average in denser areas, and that individuals migrate to denser areas to improve their living conditions. In other words, our facts are consistent with a more dynamic notion of spatial equilibrium, where individuals are not indifferent across space at any point in time, but migrate to better locations to close spatial gaps in living standards (Topel, 1986). Our facts also point to substantial migration frictions, since net migration rates to denser areas are modest in most countries, in spite of the gaps in income and other metrics. To be sure, we are far from the first to point to substantial frictions to rural-urban migration. Other prominent theories include loss of social network (Morten, 2013; Munshi and Rosenzweig, 2016), lack of skill transferability (Bazzi, Gaduh, Rothenberg, and Wong, 2016), risk of bad migration outcomes (Harris and Todaro, 1970; Bryan, Chowdhury, and Mobarak, 2014) or simply disutility from living in a new place away from one's family.

[3] Ciccone and Hall (1996), for example, argue that this is a simple and "realistic" description of the data. Still, there are reasons to believe people prefer to be in denser areas. In the developing world, Fafchamps and Shilpi (2009), for example, provide evidence that rural workers dislike the isolation that comes with rural life.

Our findings can also be squared with the recent work emphasizing sorting of heterogenous agents across space based on comparative advantage (Lagakos and Waugh, 2013; Herrendorf and Schoellman, 2014; Young, 2014). These theories do not have workers indifferent between all locations, and in fact each worker in general prefers one location to another. To reconcile these theories with the net migration rates to the urban areas, however, one must consider a model where, on net, workers are moving to cities, rather than in a stationary equilibrium where just as many workers move to cities as out of cities. As one step toward guiding these classes of models, we look *within* narrowly defined educational groups, and those without any education.

Our paper builds on recent work measuring living standards is also related to the work of Henderson, Storeygard, and Weil (2012); Henderson, Storeygard, and Deichman (2017), which use satellite data on night lights to construct proxies for income at a fine level of geographic detail. One way our work differs is that it tries to isolate the effect of non-monetary amenities, which have not been studied systematically in the developing world. In emphasizing measurement of amenities, our work parallels that of Albouy (2012), and echoes his conclusion that denser areas do not appear to be such bad places to live once draws on richer spatial data and better measurement is brought in.

This paper is structured as follows. Section 2 provides a simple model with a spatial equilibrium, and shows how we consider housing prices. Section 3 outlines our data and how we link our amenity measures with population density. Section 4 presents our main findings for amenities by population density. Section 5 looks at amenity measures for households by education group. Section 6 explores migration flows plus our secondary amenities metrics. Section 7 concludes.

2. Consumption, Amenities and Housing Prices

In this section we present a simple spatial model to illustrate how a spatial equilibrium works, and how housing prices fit into our analysis. It is important to get this clear early on, as virtually all evidence on housing prices suggest that they increase with population density. We show below that this is not enough to conclude that there is a spatial equilibrium. One needs to look at real quantities of consumption and housing quality as well.

Environment: The economy is population by identical households that each have a utility function $U(c,h,a)$, where c is non-housing consumption, h is housing consumption (or "housing quality"), and a is amenities. The utility function satisfies $\frac{\partial U}{\partial c} \geq 0$, $\frac{\partial U}{\partial h} \geq 0$ and $\frac{\partial U}{\partial a} \geq 0$ for all levels of c, h and a, i.e. utility is everywhere non-decreasing in each input.

The economy is partitioned into J regions. Households are freely mobile and must locate in exactly one region. Regions have three exogenous characteristics: wages, w^j, housing prices, p^j and amenities a^j. Any household locating in region j earns the wage w^j, may consume as much housing services as it wants at price p^j per unit, and enjoys amenities a^j. Households anywhere may consume non-housing consumption at a normalized price of one. The budget constraint of a household locating in region j is $w^j = c + p^j h$. We make no attempt to clear markets, but rather focus on the households' choice of where to locate.

Spatial Equilibrium: A spatial equilibrium implies that utility levels are equated across regions. Let this common utility level be denoted $\bar{u}$. The standard approach in the literature is to use the household's (and/or firms') optimality conditions, plus the common utility value $\bar{u}$, to impute impute amenities, a^j, given data on prices w^j and p^j. In contrast, our approach is to focus on quantities, and to assess whether utility levels are equated across regions. To this end, denote the common quantities of non-housing consumption and housing in j by c^j and h^j. Then, note the following basic, almost definitional, properties of the static spatial equilibrium:

1. For any regions j and k, if $c^j > c^k$, than either $h^j < h^k$ or $a^j < a^k$ or both.

2. For any regions j and k, households do not prefer to migrate from j to k.

Property one says simply that if non-housing consumption is higher in one region than a second region, then either housing consumption or amenities must be higher in the second. No region can have higher quantities of every input into the utility function, or that would violate the assumption that utility is equated across regions. Property two says that when offered the choice, households do not systematically prefer another region to the one they are in. This is heart of the logic of a spatial equilibrium: if households were systematically better off in another region than their current region, they would migrate to improve their well being. In what follows, we draw on new data on quantities of consumption, amenities and net migration flows to test these two basic predictions of a static spatial equilibrium.

The Role of Housing Prices: How do housing prices fit into this? Suppose region j has higher wages than region k, but also higher housing prices. That is, $w_j > w_k$ but $p_j > p_k$. Couldn't this mean there is a spatial equilibrium, even if amenities were the same, i.e. $a_j = a_k$? If there were a spatial equilibrium, then one of two basic patterns could be found in the quantity data. First, households use the higher wages in j to obtain higher consumption, so $c_j > c_k$, though consume lower quality housing, $h_j < h_k$. Second, households get lower consumption, $c_j < c_k$, but higher housing, $h_j > h_k$. If both were higher in k then there would not be a spatial equilibrium. The point is that the higher housing prices in high-wage regions is not sufficient to conclude there is a spatial equilibrium. A spatial equilibrium implies that the high-wage, high-price region have either lower consumption or lower housing. If instead, as we later find, both housing quality and consumption are higher in the high-wage, high-price denser areas, it must be true that some other amenity is worse in the denser area, or there is not a spatial equilibrium.

3. Measuring Amenities Across Space

3.1. Data on Consumption, Housing and Amenities

Until recently, measuring disamenities across space in the developing world was not feasible. Exploiting progress in surveying and mapping technology, we construct a new dataset that spatially links household surveys on amenities, crime, mistrust and reported living standards with satellite-derived measures of pollution and gridded population density. To measure household welfare across space, we use data from Demographic and Health Surveys (DHS), Living Standards Measurement Surveys (LSMS), Afrobarometer and remotely sensed pollution data.[4] The micro surveys are high-quality nationally representative surveys and cover large numbers of households (typically more than 5,000 for DHS and LSMS, and more than 1200–2400 for Afrobarometer) in developing countries. The surveys are designed to use consistent methodologies and definitions across countries. DHS and LSMS focus on variables related to population, health, and nutrition, while Afrobarometer focuses on attitudes towards democracy and governance, including experiences of crimes. We outline the main choices related to using these data here, while referring the reader to the detailed data appendices A - D for further details.

We select countries that satisfy four criteria: (i) the survey was conducted no earlier than 2005, (ii) spatial identifiers of the respondents or clusters were collected and are available, (iii) the country is larger than 50,000 square kilometers, and (iv) classified as a low income country by the World Bank in 2012.[5] For the DHS data, our main data to measure housing and non-housing consumption across space, this leaves us with a sample of 276,051 households across 20 African countries as listed in Tables 1 and 2, covering countries with a combined population of about 770 million people. From these data, we use the following variables: durables owned by households (television, car and mobile/landline telephone), housing conditions (electricity, tap water, flush toilet, constructed floor, finished walls and finished roof), and child health (stunting, wasting, anemia, and minimum acceptable diet).

Data on crime at small levels of geographic detail are difficult to obtain for African countries. Official records are either not stored centrally, or they are not available to researchers. We use data from LSMS and the Afrobarometer Surveys to investigate crime. For our set of Sub-Saharan African countries we have LSMS data for Ethiopia (2013), Malawi (2010), Tanzania (2009-2010), Nigeria (2012) and Uganda (2010-2011). Usually in the context of economic shocks experienced by households, these surveys ask household members about experiences of theft, robbery, personal attacks or assault and violence. Given the relatively low frequency of certain types of crimes, we

[4]To investigate disease we complement DHS data with Malaria Indicator Surveys and Aids Indicator Surveys. These are managed under the umbrella of the DHS Program which encompasses a number of survey types; to avoid repetition and given the similarity in methodology we restrict most of our discussion to DHS.

[5]This implies a GNI per capita (Atlas method, current US$) below $4,126 in 2012.

combine measures of different types of crime and compute the fraction of all households in which at least one member reported at least one crime.[6]

We use three rounds of the Afrobarometer surveys (2005, 2009 and 2011) for Benin, Burkina Faso, Cameroon, Cote D'Ivoire, Ghana, Kenya, Liberia, Madagascar, Malawi, Mali, Mozambique, Niger, Nigeria, Senegal, Sierra Leone, Tanzania, Togo, Uganda, Zambia and Zimbabwe. The Afrobarometers are high quality micro surveys covering between 1200-2400 individuals in each of several African countries. Table A.1 in the appendix shows the number of rounds a country is part of the survey, as well as the number of observations per country. The surveys are designed to use consistent methodologies and definitions across countries. The questionnaire focuses on attitudes towards democracy and governance, and includes questions on crime, safety and trust. An advantage of survey data over administrative data on crime is that the latter are likely to be biased towards areas with police presence. A positive relationship between crime and population density could therefore be due to higher rates of reporting in denser areas. Further, official administrative data on crime are often not stored centrally, or they are unavailable to researchers. Both sources of information share a weakness in that their quality is likely to suffer in areas or times of civil conflict as survey teams might not enter due to safety concerns, and accurate record keeping is less of a priority. This only biases our results if conflicts always occur at a specific density and these areas are excluded from the survey. In that case, it would lead to a downward bias of the rate of crime at that particular density. We use multiple rounds of the Afrobarometer survey spanning six years to reduce the potential bias inherent in a particular round.

To capture experiences of crime, the surveys ask "Over the past year, how often (if ever) have you or anyone in your family had something stolen from your house?", and "Over the past year, how often (if ever) have you or anyone in your family been physically attacked?". To measure perceived safety, the questions are "Over the past year, how often have you or anyone in your family feared crime in your own home?", and "Over the past year, how often, if ever, have you or anyone in your family: Felt unsafe walking in your neighborhood?". The answer to these questions on experienced crime and perceived safety are classified as "never", "just once or twice", "several times", "many times", and "always". I define a dummy variable as equal to one if a respondent's reply is anything more than "never".[7] About one third of respondents reports a theft from their house in the previous year. The highest rates of theft are in Liberia (49%), Uganda (42%) and Senegal (39%) and the lowest rates are in Madagascar (13%), Niger (18%) and Mali (21%). The heterogeneity in physical attacks follows a similar pattern for most countries and the pairwise correlation coefficient at the country level between theft and attack is 0.7 and highly significant. Exceptions include Senegal, where theft is high but attacks are reported infrequently.

[6] Our results are unchanged when investigating different crimes individually.

[7] In the 2011 round the categories slightly changed to "yes, once", "yes, twice" and "yes, three or more times". My results are qualitatively the same when I use the ordered variable as my measure of crime rather the binary variable.

Across the whole sample, more than one third of respondents report that they felt unsafe in their neighborhood at least once in the past year, and that they feared crime in their own home.

To explore whether social cohesion is lower in more densely populated areas we rely on questions about trust towards neighbors and co-ethnics. The questionnaire asks “How much do you trust each of the following types of people: Your neighbors?" and “How much do you trust each of the following types of people: People from your own ethnic group?". Responses are classified into four categories: not at all, just a little, somewhat, a lot; we define a dummy variable mistrust as equal to one if respondents report that they trust neighbors/coethnics not at all, or just a little. About 37 percent of respondents report that they either don't or only little trust their neighbors, and average mistrust towards co-ethnics is 43 percent. There are large differences across countries. For example, in Senegal, Burkina Faso and Mali only between 10–18 percent of respondents report mistrust towards their neighbors, compared to 48–60 percent in Liberia, Sierra Leone and Nigeria.

Another often discussed disamenity of cities are levels of pollution. Individuals, in particular in developing countries, are often exposed to both outdoor as well as indoor pollution (WHO, 2014). Sources of outdoor pollution include vehicles, electricity generation, industry, waste and biomass burning, and re-suspended road dust from unpaved roads; indoor pollution is mainly caused by burning of fuels for cooking. To measure exposure to pollution faced by individuals, ideally we would employ measurements taken on the ground at varying population densities and times of the year, as well as indoor and outdoor readings for different types of households. Unfortunately, data on ambient air pollution from ground measurements in African cities is scarce (Petkova, Jack, Volavka-Close, and Kinney, 2013). We therefore use the most recent satellite-derived estimates of PM2.5 concentrations from van Donkelaar, Martin, Brauer, and Boys (2015), NO2 concentrations from Geddes, Martin, Boys, and van Donkelaar (2015), and refer to DHS data for the source of cooking fuel households use and where they undertake their cooking.[8]

3.2. Population Density Measures

To measure population density, we use data from the Gridded Population of the World Version 4 (GPWv4), which provides population density estimates at a resolution of 30 arc-seconds corresponding to about 1km at the equator (Center for International Earth Science Information Network, 2015). The gridded population data employ a minimal amount of modeling by equally distributing non-spatial population data from censuses among spatial datasets of administrative units (Doxsey-Whitfield, MacManus, Adamo, Pistolesi, Squires, Borkovska, and Baptista, 2015).

One attractive feature of GPWv4 for the purpose of this analysis is that the distribution of population data is transparent and performed without using further auxiliary data. This comes at a

[8]PM2.5 refers to particles that are 2.5 micrometers in diameter or smaller.

cost of a lower resolution that is offered by alternative data sources. For example, one higher resolution dataset is WorldPop, which uses a range of input data and has a resolution of 100m (Linard, Gilbert, Snow, Noor, and Tatem, 2012). For my analysis, one important consideration on input data is that they might introduce circularity in measurement. For example, if nighttime lights data from satellites are used to redistribute populations in order to achieve population densities at finer geographical scale, and we would then use these data to estimate the relationship between population density and electrification, by construction, higher densities will have higher rates of electrification. We rule out this circularity by using population density data that are not modeled using further input data. The maximum dispersion assumption of GPWv4 within spatial administrative units therefore biases us towards finding no relationship between population density and outcome variables.

The resolution of the census data that are used as input varies across countries due to availability of data. Some countries provide their data at the level of the enumeration area, while others only share their second administrative level data. we restrict my analysis to countries for which the input census data have sufficiently high spatial detail, roughly more than 40 regions per country.

3.3. Spatially linking Living Standards and Density

We next combine the different sources of data step by step. To link the individual data from the DHS and Afrobarometer with population density, ideally we would have the GPS location of households. The DHS readily collects GPS coordinates, but these have been re-assigned a GPS location that falls within a specified distance of its actual location to preserve anonymity of survey respondents. Urban DHS clusters are randomly displaced by 0-2km and rural clusters are randomly displaced by 0-5km, with 1 percent of clusters randomly selected to be displaced by up to 10km (Perez-Heydrich, Warren, Burgert, and Emch, 2013).[9] We take into account the random offset of DHS cluster locations when linking DHS GPS data with continuous raster data by taking 5 km buffers around both urban and rural DHS clusters as suggested by Perez-Heydrich, Warren, Burgert, and Emch (2013). Appendix A provides more detail on this procedure and discusses the sampling protocol of the surveys.

Unfortunately, the Afrobarometer did not collect the GPS location, but the location name was recorded. We develop an algorithm that performs a series of exact and fuzzy matches of location names relying on data from a global gazetteer that contains the latitude and longitude of a location. Depending on the survey round, this involves between thirteen and twenty-one steps in which the village name, district name and region name are sequentially matched with the ascii name of locations as well as up to four alternative names listed in the gazetteer. To catch mis-

[9]The displacement is done by selecting a random displacement angle between 1-360 degrees as well as a random distance.

spellings, we perform fuzzy matches based on similar text patterns, using a similarity score of 0.7 and a vectorial decomposition algorithm (3-gram) (Raffo, 2015). Appendix C provides further detail on the matching procedure. Using this algorithm we are able to geo-locate between 85–95% of village names in each round.[10] For each respondent we can then extract the population density value.

Both the pollution data and the population density data are gridded data, making it straightforward to link them. The estimated PM2.5 and NO2 concentrations are available at a resolution of 0.1 decimal degrees (about 10km at the equator) compared to the 30 arc-second resolution of the population data. We construct a fishnet grid of the same resolution of the pollution data (the coarser spatial resolution) and for each pixel compute the average pollution measure as well as the average population density from the GPWv4. PM2.5 is measured in $\mu g/m^3$ while NO2 is measured in ppb (parts per billion).[11] Appendix D contains further details on how we link the pollution data with the population density data, and discusses the joint distributions of these variables in detail for the Nigerian case. With these different pieces of information in hand, we can test whether the data are consistent with a simple static spatial equilibrium model when considering this key set of disamenities.

4. Analysis of Living Standards Data

In this section, we analyze the living standards data in relation to population density. It goes without saying that for many of the countries in our sample, living standards are very low compared to today's rich countries. But we focus in particular on the *patterns* and spatial distribution of living standards. The consistency with which we see the same patterns across a large set of countries suggests that there are underlying forces behind the spatial distribution of well-being.

The measures that we do use have the advantage of being real and directly observable; in other words, they are not based on prices or values, nor do they require imputation. As argued by Young (2014), such measures have the advantage that they admit direct comparison across time and space. There may of course be substantial quality differences that are hidden by the variables that we consider. Thus, a "constructed floor" implies only that the floor is made of a finished material rather than dirt or a simple covering such a tarpaulin or carpet. But within the category of "constructed floors," households could in principle have anything from rough-hewn boards or bamboo slats to poured cement or even marble tile. Similarly having access to tap water could mean a single cold-water tap, or it could mean a plethora of chrome-plated faucets. The data do

[10] Nunn and Wantchekon (2011) manually geo-locate the respondents of the 2005 Afrobarometer round. When we compare their locations with ours, we find that median distance is 10km.

[11] Following Vrijheid, Martinez, Manzanares, Dadvand, Schembari, Rankin, and Nieuwenhuijsen (2011), we use a conversion of 1ppb= 1.88 f $\mu g/m^3$ which assumes ambient pressure of 1 atmosphere and a temperature of 25 degrees celsius.

not allow us to distinguish variation *within* the categories reported in the DHS data.

4.1. Basic Descriptive Statistics of Population Density across Space

We start with some observations related to the distribution of population by density. As one might expect, population density distributions vary considerably across the countries in our sample. Some are heavily urbanized, while others are more rural as illustrated in Figure 1. Quite a few countries display bimodal distributions of population density. For instance, Uganda and Zimbabwe have pronounced bimodal distributions, with both showing spikes in population in what are presumably dense urban areas, to go along with substantial numbers of people living in rural areas. Some countries show relatively narrow distributions of population density; for instance, Malawi has a very narrow support for the density distribution. Others (for example, Cameroon, Ethiopia, and Tanzania, for example) have relatively flat distributions suggesting that even in rural areas, population density varies substantially.

The next point worth noting is that the population density measure provides a different – and we will argue, more useful – way of thinking about the spatial distribution of the population than does the typical urban-rural dichotomy, which is largely based on administrative classifications. Figure 2 shows population density distributions for those DHS households that are classified as urban and rural based on the administrative designation of each cluster.

This figure shows that in many countries there are people living in "urban" locations with quite low population density, while there are at least a few households classified as rural that occupy quite dense locations.[12] The data show that some countries have quite distinct distributions of urban and rural population densities, as one might expect. This is true, for instance, in Uganda and Sierra Leone as well as in Mali, Zambia and Zimbabwe. Other countries have urban and rural population density distributions that overlap considerably; e.g., Ghana, Liberia, and Senegal. The key point to make is that the frequently used urban-rural characterization offers a different view from the density-based measures that we use here. We do not argue that the urban-rural classification is flawed; for some purposes, we may care precisely about the administrative designation, which may dictate access to public services and resources. But for other purposes where location in space matters – for example, agglomeration externalities, market thickness, and transaction costs – the density-based measures are likely to be more informative.

Another advantage of the density measures is that they allow us to differentiate rural locations. Particularly in countries where large fractions of the population remain rural, it is useful to con-

[12]We must exercise some caution here; because of the displacement of DHS clusters in the data, it is conceivable that some of these apparent discrepancies are linked to the displacement procedure, which might move change the measured density in our data. But quantitatively speaking, this does not appear to be a strong enough effect to account for much of the overlap in population density between urban and rural locations.

sider the density heterogeneity within rural areas. Small towns and densely populated rural areas – which are usually in proximity to markets, as noted above – may have quite different characteristics than remote rural areas.[13] We see in the data that the rural population densities vary enormously in many of the DHS countries. In this paper, we will argue that the variation in densities is important and has implications for the types and magnitude of frictions that could perhaps prevent people from moving to areas with higher average living standards. We will return to this issue at the close of the paper. Next we examine the evolution of the elements of the utility function outlined in Section 2 across space. We start with non-housing consumption, then turn to housing quality, and finally, disamenities.

4.2. Living Standards Across Space

We will make three broad claims with respect to the data on housing and non-housing consumption. First, we argue that there is enormous variation within countries in the level of each indicator. Different DHS clusters display strikingly different levels of these indicators. Second, there is a strong spatial pattern with respect to living standards. Specifically, in almost all cases, consumption improves with population density – although at very high levels of density, some indicators indicate a slight decline – though never to the level of the most remote areas. Third, the spatial pattern of variation is typically quite continuous and nearly monotonic with respect to population density – at least until the highest levels of density. Admittedly, to some extent the smoothness reflects the fact that we use local polynomial approximations to smooth the data. But the raw data are nearly as smooth. Our data on amenities, particularly crime and pollution, tell a slightly different story: patterns are less smooth, less continuous, and often imprecisely estimated across densities. However, we still find little evidence that they are systematically worse in more densely populated areas. To investigate other compensating differentials, we examine measures of consumption volatility, stress and trust, which might be higher in cities. We find that there is little systematic relationship between consumption volatility and stress across population density. However, we see that trust in general, and more specifically towards neighbors and relatives, is consistently higher in rural areas. Unless stated otherwise, for presentational purposes we exclude households residing in the top and bottom five percentile of the population density distribution whenever we show data along the whole density distribution.

4.3. Non-housing Consumption

As an example for non-housing consumption consider whether household has a phone. Figure 4 shows a kernel-weighted local polynomial regressions of whether a household has a phone and the

[13] For instance, Gollin and Rogerson (2014) find that rural locations that are *ex ante* identical will take on different characteristics *ex post* based only on transport costs.

log of population density including a 95% confidence interval for Tanzania, Nigeria, Ethiopia and Senegal.[14] Several facts are worth noticing. First, there is a large dispersion in phone ownership – one of our real indicators of living standards – from the least populated areas to the most populated areas, with a support from zero to one. Second, across the whole range of densities, phone ownership is increasing almost monotonically and continuously. The bottom graph shows the same gradient for all countries in our sample. The thick red lines represent Ethiopia, Nigeria, Senegal and Tanzania, and the light grey lines represent the remaining 16 countries in out data. The graph illustrates that the pattern is very similar to what we saw in the upper graph. We now show that these patterns hold across a range of other non-housing consumption measures and countries.

To compactly present data for our whole set of countries, Figure 5 shows the proportion of households with a phone or a television at the top and bottom quartile of the population density distribution, using household weights provided by the DHS.[15] The solid line represents the 45 degree line, indicating equality between households located in the upper and lower quartile of the population density distribution. Almost all countries are located in the lower triangle, indicating that ownership of these durables is consistently higher for households located in the upper quartile of the population density distribution. Appendix F shows sub-plots of kernel-weighted local polynomial regressions for each country for all measures discussed in this section.

Food represents an important part of non-housing consumption of households. The DHS data do not contain questions on detailed food consumption of survey households which we could use to compute caloric intake of household members. It might be that prices are significantly higher in cities so that despite higher incomes, households are worse off. The DHS carefully collects high-quality information on child health outcomes, which we consider to be measures of *realized consumption*. In the absence of consumption aggregates we use these child health outcomes to investigate whether there is evidence that differences in price levels are the mechanism by which households across population density space are made indifferent. One caveat of using child health measures is that household members living in urban areas might be more informed about health problems of children, thus affecting the propensity to report a health problem. Mindful of this, we only selected outcomes which are objectively measured; in other words, they are not dependent on reporting of caretakers and thereby possibly capturing a combination of differences in information as well as outcomes across space.

Figure 6 shows the average incidence of anemia, stunting and wasting, and the proportion of children who do not consume a minimum acceptable diet, which is defined by a minimum meal frequency and dietary diversity (WHO, 2015). In the vast majority of measures and countries,

[14]The graphs exclude the top and bottom five percentile of the distribution.
[15]We include all households to compute the quartiles.

households living in the lower quartile fare on average either as well as, or worse than households in the upper quartile of the population density distribution.

4.4. Housing

In this section, we ask how housing quality varies across population density. We focus on five measuring quality, in particular the percentage of households that have: (1) a finished roof, (2) a constructed floor (as opposed to dirt), (3) finished walls, (4) electricity, (5) a flush toilet, and (6) tap water. Together these measures provide a fairly comprehensive view of housing quality for households in the developing world.

Figures 7–9 plot the fraction of households having each of these characteristics in the top and bottom quartiles of the density distribution. Figure 7 plots the results for finished roofs and constructed floors, which are mostly the result of private investments. In most countries, three-quarters or more of residents in the densest regions have constructed floors and finished roofs, compared to half or fewer of residents of the least dense regions. Only Benin and Ethiopia are near parity, and differences are perhaps starkest for Zimbabwe, for whom virtually all households in the densest areas have finished roofs and floors, while less than one half do in the least dense areas. In Liberia, more than eighty percent of households have finished walls in the densest quartile, compared to less than one in five households in the lowest density quartile.

Figure 8 and 9 show the results for whether a household has a finished floor, electricity, flush toilets and tap water, where the latter three have relatively more of a public-investment component. Differences between the densest and least dense areas are even starker here. In no single country do the least dense areas have more than half of households with access to electricity or tap water, or more than one quarter of households having flush toilets. In the densest areas, flush toilets are also quite uncommon, though still substantially higher than in the least dense areas in every country. Electricity and tap water have more variance, but are again much more common in the densest areas. We conclude that by these six measures, housing quality is unambiguously higher in denser areas in these developing countries.[16]

4.5. Disamenities

In 1900, life expectancy at birth for a white male living in a city in the United States was approximately 48 years compared to 54 years in the countryside (Haines, 2001). Some urban disamenities, such as pollution and particular types of crimes, even today, are higher in larger, than

[16]Some measures of housing quality may already be incorporated into measured rural-urban income differences. For example, surveys reporting average consumption expenditures at the household level may deflate housing expenditure by regional quality-adjusted price indices. In practice, it is an open question how well quality adjustments are done in these surveys.

in smaller cities in the United States; however, this gap has decreased substantially over the past decades (Kahn, 2010). Kahn (2010)'s results show higher rates of murder and crime between 1994 and 2002 in counties located in metropolitan areas with higher population. Albouy (2012) finds lower levels of property crime and higher levels of violent crime rates in larger metro areas, but no relationship of both measures with population density. Clean water technologies have further contributed to the mortality gap to disappear (Cutler and Miller, 2005). Characterizations of slums in developing countries often resemble cities at the turn of the century: lacking clean water and sewerage systems, polluted, congested, and disease-ridden. Our analysis so far points towards the opposite of the mortality penalty experienced in the earlier stages of industrial cities: for most countries, higher densities seem to offer better sources of water and higher likelihoods of access to sanitation systems. Regarding pollution and crime, due to a lack of data, little is known about the distribution of these two disamenities across the rural-urban continuum. This section relies on the most recently available data and discusses their role as possible compensating differentials.

4.5.1. Crime

Figure 10 shows differences in experienced crime and fear of crime across space. We show both of these categories of variables, as fear of crime might matter at least as much as experiences of crime for location choices. Both figures illustrate that most countries are located close to the 45 degree line. Property crime appears to be slightly higher in denser areas, but the differences for most countries are fairly small when comparing them with observed differences in living standards, for example. One limitation of the theft variable is that it does not consider livestock theft, a type of crime common in rural areas. It is therefore likely that the difference is even smaller when taking into account livestock theft. The results are similar for fear of crime and perceived feeling of safety in the neighborhood, where most countries cluster around the 45 degree line.

This finding resonates with evidence from Madagascar and South Africa suggesting that crime is higher in less densely populated areas (Fafchamps and Moser, 2003; Demombynes and Ozler, 2002). It is against the conventional wisdom considering crime as a main disamenity of cities by raising the frequency of interactions between individuals. Explanations for the inverse relationship include a lack of police presence in rural areas, higher levels of organized crime and higher alcohol consumption due to a lack of other entertainment activities (Fafchamps and Moser, 2003). Fafchamps and Moser (2003) instrument for police presence with amenities and find that the negative relationship between crime and population density is not driven by the bias in policing. We have data on a woman's reported alcohol consumption of her partner for Burkina Faso, Cameroon, Ghana, Ivory Coast, Kenya, Malawi and Zambia; apart from Cameroon, alcohol consumption appears to be either unrelated to population density (Burkina Faso, Ghana, Ivory Coast, Malawi) or higher in cities (Kenya and Zambia). Further investigating this relationship and underlying reasons is a topic beyond this paper. Figure A.5 in the appendix shows the fractions of LSMS households having experienced a crime in the last twelve months. While patterns of crime across space are different in each of these five countries, one common theme is a lack of evidence that rates of crime are systematically increasing in density, or a lack of evidence that the densest areas have the highest rates of theft; visually, a constant rate of crime seems like it would just about fit within the confidence intervals.

4.5.2. Pollution

Banzhaf and Walsh (2008) find pollution to be an important determinant of locational choice in the United States. Exposure to pollutants significantly affects health, human capital and productivity (Adhvaryu, Kala, and Nyshadham, 2014; Currie and Walker, 2011; Currie, Hanushek, Kahn, Neidell, and Rivkin, 2009; Graff Zivin and Neidell, 2012).[17] This section examines satellite-derived estimates of PM2.5 and NO2 distributions across space and types of fuels used by households, including the location of cooking activities.

Figure 11 shows the average PM2.5 concentrations for the highest and lowest population density quartiles as well as for PM2.5 concentrations with dust and sea salt removed, as measured as micrograms per cubic meter ($\mu g/m^3$). Figure 12 shows the corresponding graphs for the entire density support in each country. Given the high concentration of dust and sea salt in the air in Africa, we separately examine total PM2.5 concentrations and PM2.5 concentrations with dust and sea salt removed.[18] There is no evidence suggesting that antropogenic sources of PM2.5 are more or less harmful for health than natural sources, but it is possible that individuals perceive antropogenic sources as more hazardous to their health. Taking logs removes uninhabited pixels.

The World Health Organization recommends mean annual exposures of 10 $\mu g/m^3$ or less for PM2.5 and 40 $\mu g/m^3$ or less for NO2, at the same time, highlighting that there are no levels of pollution exposure that have been proven not to negatively affect health, referred to sometimes as a "no-threshold model" (Geddes, Martin, Boys, and van Donkelaar, 2015; WHO, 2006). Further, the consequences of a particular pollutant mix remain unclear. These guidelines should be applied with caution when examining satellite-derived estimates, as they refer to point measurements of ground stations (Geddes, Martin, Boys, and van Donkelaar, 2015); nevertheless, in the absence of more conclusive evidence, they give an indication of the current recommended thresholds.

There are several observations worth noting when examining the quartile graph and PM2.5 along

[17] For a comprehensive surveys of the literature on pollution and individual welfare see Graff Zivin and Neidell (2013); on environmental amenities and city growth see Kahn and Walsh (2015).

[18] About half of the population-weighted ten-year mean of PM2.5 concentration in Eastern Sub-Saharan Africa is estimated to be due to dust and sea salt; in Western Sub-Saharan Africa the proportion amounts to about three quarters (van Donkelaar, Martin, Brauer, and Boys, 2015).

the whole population density distribution: first, there is a strong difference between the two PM2.5 measures, particularly in West African Sub-Saharan countries, such as Senegal, Burkina Faso, Mali and Sierra Leone. Second, PM2.5 levels are often above the WHO recommended thresholds at all levels of population density, such as in Ghana, Mali, Liberia, Nigeria and Sierra Leone. Third, non-anthropogenic drivers of PM2.5 are substantial, so that removing dust and sea salt leads to PM2.5 concentrations below the WHO recommended limit of 10 $\mu g/m^3$ in several countries, including Burkina Faso, Ghana, Mali, Senegal and Sierra Leone. Fourth, density gradients are either relatively flat, for example for Kenya and Nigeria, or downward sloping such as for Ivory Coast or Malawi. The only country with a sharp rise at high densities is Mozambique, which, even with the increase from 3 to 7 $\mu g/m^3$, remains below the WHO recommended threshold. Figures 13 and 14 illustrate the relationship between population density and NO2 concentrations. We find that on average NO2 levels are very low across the African countries in our sample. The WHO-recommended threshold of 40 $\mu g/m^3$ it is far above the concentration inferred from satellite data. The data also does not suggest clear patterns with population density. In some countries there is a positive relationship between NO2 and population density, such as in Kenya, Mozambique and Uganda; others exhibit a negative relationship, such as Benin, Liberia and Sierra Leone. Still, all of the countries have very low levels.

What do these gradients look like in other parts of the world, known to suffer from high levels of urban pollution, such as China? The global coverage of the PM2.5 map as well as the population density map allows us to look at further countries. Figure 15 shows the relationship between pollution and population density in China, India, and the United States. The left-hand side axis denotes levels for India and China, while the level in the U.S. is illustrated on the right-hand side axis. All three countries show visible density gradients. In China, PM2.5 levels for the top population density decile amount to 66 $\mu g/m^3$, more than six times the WHO recommended threshold; the lowest population density decile has a level of 13 $\mu g/m^3$. In India, the top decile has a level of 41 $\mu g/m^3$, still four times the WHO recommended threshold, compared to 6 $\mu g/m^3$ in the lowest decile. The bottom figure shows NO2 distribution for these other countries. The levels are much lower, but there are gradients again for China, India and the United States.

We do not make a claim here that pollution does not matter in African cities. Our satellite derived pollution estimates do not capture pollution exposure in at least these dimensions: they are annual series and therefore average over temporarily high values. Second, at a 10km resolution they are spatially rather coarse, ignoring local effects such as proximity to roads which have been demonstrated to matter significantly. For example, Kinney, Gichuru, Volavka-Close, Ngo, Ndiba, Law, Gachanja, Gaita, Chillrud, and Sclar (2011) find average PM2.5 concentrations at four traffic sites between 7.30am and 6.30pm in Nairobi to amount to between 58.1 and 98.1 $\mu g/m^3$; the maximum multi-annual average PM2.5 concentration for Kenya in our sample is 13.9 $\mu g/m^3$, and this pixel is at Lake Turkana, the world's largest desert lake (Avery, 2012). Finally, satellite derived measures reflect the column of pollution as observed from space, rather than the concentration experienced on the ground. Nevertheless, as the graphs from India, China, and the US illustrate, the two datasets still capture meaningful variation in concentrations levels across space. What emerges, rather, is that in Africa, cities are not large enough and concentration of industries is not significant enough to create large clouds of pollution around cities while background non anthropogenic pollution is high, to produce similar gradients as observed in other parts of the world.

As a proxy for indoor air quality we examine the main material used for cooking as reported in the DHS. The World Health Organization estimates that over 4 million people suffer from pre-mature deaths due to illnesses attributable to cooking with solid fuels, such as wood and charcoal (WHO, 2014). Indoor air pollution is also estimated to contribute significantly to outdoor air pollution-related deaths. Figure 16 shows the proportion of households using solid fuels for cooking across population density. One potential advantage of rural areas is that there might be more space to accommodate outdoor cooking, thereby somewhat mitigating the negative effect of using solid fuels. Therefore we also show the interaction effect: out of the households using solid fuels, what is the proportion cooking indoors. Figures 17 and 18 show the gradients along the whole population density support. The figure shows that in the bottom quartile of the density distribution, almost all households use solid fuels for cooking. In some countries, the proportion decreases when moving to higher levels of density, such as in Cameroon, Kenya or Nigeria. The data suggests that if anything, households who use solid sources of cooking fuel are less likely to cook inside in more densely populated areas. One possible explanation is that rooms are smaller in denser areas, so that cooking is done outside.

4.5.3. Infectious Diseases

Over the course of history, outbreaks of cholera and the plague have shaped public health policies in cities now located in high income countries. Certain aspects of modern day cities still make them more prone to a faster transmission of infectious diseases, including the higher frequency of interactions between humans, mobility of residents, and traditionally rural pathogens adopting to urban conditions. At the same time, the incidence of infectious diseases might be lower in cities: for example, wealthier, more educated and health-aware individuals sort into cities; urban residents might be able to access health care more easily containing the spread of diseases; and the urban environment is potentially less attractive to disease vectors (Alirol, Getaz, Stoll, Chappuis, and Loutan, 2011).

Unfortunately, (geo-referenced) data on infectious diseases in Africa are often non-existent due to a lack of of systems to observe these diseases. One exception are malaria test samples takes as part of the Demographic and Health Surveys, Malaria Indicator Surveys and Aids Indicator Surveys. We combine data for our set of countries from any of these surveys. Figure A.21 shows malaria incidence of children aged 6-59 months for the 12 countries for which we have data. Disease incidence is decreasing with population density consistently in most of the countries. This is in line with Tatem, Gething, Smith, and Hay (2013) who document a negative relationship between malaria and urbanization on a global scale. We can not draw any conclusion whether this relationship is due to sorting, better access to health care or specificities to the disease vector, or whether this holds for any other disease.

5. Gradients by Education Group

One explanation for the patterns we observe is that workers select into different regions and occupations according to comparative advantage. This type of sorting has recently been emphasized by e.g. Lagakos and Waugh (2013), Young (2014), Bryan and Morten (2015) and others as a way of explaining regional income differences in the developing world. To address this issue further, we compare how living standards measures vary by density within specific educational groups. Educational attainment is perhaps the simplest measure of skills or ability that is availability at the individual level in the data, though of course there are components of ability not well captured by educational attainment.

For simplicity, we divide households into two education groups: those whose head finished primary school, and those whose head did not finish primary school. To get started, Figure 19 plots one metric by educational group, namely electricity access. The top graph shows the proportion of households who have electricity by the highest education of the household head. The histograms show that the different educational groups are represented at various population density levels. Households with more educated household heads are experiencing better access to electricity at almost all levels of population density. Still, the urban-rural gradients documented earlier persist even for different levels of education.

To study these slopes by education group for all metrics, we estimate the following linear projection for households i in country c:

$$x_{ic} = \theta_0 + \theta_1 P_{ic} + \theta_2 E_{ic} + \theta_3 (P_{ic} * E_{ic}) + \epsilon_{ic}$$

where x_{ic} is a measure of consumption, P_{ic} is the log of population density and E_{ic} is a dummy variable that is equal to one if the household head has completed primary education or more.

Figure 20, Panel (a), shows the linear gradients for households with household heads who have less than complete primary education, and panel (b) shows gradients for households with household heads who have complete primary education or more. Each dot represents a slope estimate for one country. The y-axis indicates the size of the coefficient and the grey horizontal bars show the median slope coefficient. The figures show that in virtually all countries and metrics, these urban-rural gradients persist *within* populations at similar educational levels. That is, almost without exception, the relationship between population density and housing consumption is positive for the two main education groups.

6. Mobility Across Space and Secondary Amenities Measures

The descriptive statistics of outcomes across population densities in section 4 suggest there is no easily observable dimension of consumption that does not markedly improve with population density. The only variable that seems to be somewhat lower is trust. In this section we provide further evidence on migration flows and our secondary amenities metrics.

6.1. Net Migration Rates

We now explore whether there is evidence in the data that individuals migrate to close these gaps in living standards. For a subset of countries, we have information on how long an individual has resided in the current location, as well as the type of the previous location (countryside, town, city). When the most recent DHS survey did not provide this information, we searched through earlier surveys, or complemented DHS surveys with data from AIDS Indicator Surveys or Malaria Indicator surveys.[19] We label an individual as a mover if the individual has moved within the last five years, which is the case for about a fourth of the sample.

We start by examining the proportion of respondents who have moved within the last 5 years, as shown in the top panel of table 6. The first column shows the average for the whole sample, and the following four columns show averages for the different quartiles. The first fact worth noting is that the fraction of individuals who moved in the past 5 years varies substantially across countries: more than a third of individuals moved in the past five years in Tanzania and Zambia; slightly less than a third of individuals moved in Ghana, Kenya, Liberia, Nigeria and Uganda, compared to between ten and fifteen percent in DRC, Ethiopia and Madagascar. Migration rates are a multiple of what Chauvin, Glaeser, Ma, and Tobio (2016) find for the U.S., Brazil, China and India, where according to 2010 figures, 13.8%, 7.1%, 12.8% and 2.0%, respectively, migrated in the last five years.

There are significant differences across countries where mobility is concentrated. For most countries, the upper population density quartiles host the largest fraction of individuals who moved within the past five years. In other countries the proportions are more equally distributed such as

[19] Table 2 shows the exact survey taken for the migration variables.

DRC, Madagascar, and Mali.

Ideally, we would know the location an individual migrated from. Unfortunately the DHS data does not contain this information. However, we know if an individual moved from the capital or a large city, town, or the countryside. This allows us to further understand migration flows in these countries. The middle panel shows the proportion of individuals at the various population density quartiles who migrated from the countryside and the bottom panel shows the same numbers for individuals who migrated from a city or town. There are again large differences across countries. For example, in the DRC where about 15% of individuals moved in the past 5 years, 10% of individuals migrated from a countryside location, while only 4% did so from a city or town. Among people living in the third and fourth quartile of the population density distribution, 30 percent have moved from a countryside location. In Kenya, 29% of the sample moved within the last 5 years; 18% migrated from the countryside, and 11% did do from a city or town. Among the top two quartiles, 22 percent of indivdiuals migrated from cities and towns. The general pattern is that individuals from the countryside move to the middle densities, while individuals migrating from cities or towns move to higher densities.

6.2. Secondary Amenities Measures

6.2.1. Loss of network, social insurance, increased stress and commuting

By bringing together large numbers of people cities are more anonymous places. One possible downside is that networks are weaker, and an important consideration of individuals when making migration choices is their access to insurance provided by these networks. Busy, buzzing city life might also be more stressful, lonely and isolating. If Individuals have sufficiently strong preferences for the social integration, possibly smoother consumption (at least in terms of smaller deviations from the individuals surrounding them), this could be part of the explanation for the stark differences in living standards that we observe.

The Afrobarometer asks a number of questions that allow us to further investigate these possible explanations related to shortages of necessities, stress and trust. To examine consumption variability we look at whether and how often a household has gone without enough food or medicines in the past year. Representing crucial consumption goods these variables are possibly good proxies for insurance provided by community members. We define a household as lacking food or medicines if they state that they lacked the good either several times, many times or always. Finally, the Afrobarometer asks “In the past month, how much of the time: Have you been so worried or anxious that you have felt tired, worn out, or exhausted?”. If respondents reply with “many times” or “always” we take this as a measure of anxiety.

There are also a number of questions related to trust towards people in general, and neighbors and relatives, among others, more specifically. The questionnaire asks "Generally speaking, would you say that most people can be trusted or that you must be very careful in dealing with people?". We create a dummy variable that is equal to one if the individual responds with "most people can be trusted" and zero otherwise. The questions related to specific individuals are framed slightly differently: whether respondents trust their neighbors or relatives on a scale from 0 to 3 (not at all, just a little, somewhat, a lot); in line with our framing of lower levels of trust as a downside of city life, we define a dummy variable equal to one if respondents say they trust their neighbors not at all or just a little.

The left panel of Figure 21 shows the results for social insurance and anxiety. The data do not support social insurance as an explanation which is somewhat surprising. The probability that an individual reports having lacked food or medicines many times of the last year is in most contexts higher in rural areas than in urban areas. One explanation is that rural communities face covariant shocks causing the lack of insurance, but especially in the case of medicine, we would expect a significant idiosyncratic component. To further explore this question, we also used data from the Financial Inclusion Insights program which asks a representative sample of individuals in Uganda, Kenya, Nigeria and Tanzania. The survey asks whether an individual could get extra money from relatives in an emergency. In all four countries the proportion answering "yes" is higher in the most densely populated quartile (31%) compared to the least densely populated quartile (25%). Finally, anxiety levels appear largely to be higher in the countryside rather than in cities, weakening the argument that people prefer the countryside due to its higher level of tranquility and peace.

Figure 22 shows trust towards people in general, relatives, neighbors and coethnics. Several observations are worth pointing out: first, all these are higher in the lowest population density quartile compared to the highest population density quartile. Second, there appear to be very low levels of trust in general among the population: most trusting are individuals in Madagascar where 32% thinks that people can in general be trusted. Other countries have much lower levels of trust; for example, 90% of respondents in Tanzania and Kenya report that one must be very careful with people. Observing the same pattern in trust towards co-ethnics strengthens the argument that the observed patterns in expressed trust towards neighbors are not simply capturing a positive correlation between trust towards coethnics and higher propensity to be located next to coethnics in the sparsely populated ares. [20] We do not argue that these results settle the question

[20] We are not the first to find that mistrust is higher in denser areas in Africa. Nunn and Wantchekon (2011) use the 2005 Afrobarometer survey, demonstrating that a higher exposure to the slave trade reduced levels of trust. They control for urban location as classified by the survey, but the coefficient is not reported in the main paper. We geo-locate two further rounds of the same survey and link the data with spatial population density data. Replicating their results, we find that the coefficient on the urban dummy is negative in all their models and highly significant. The patterns found in the two papers are therefore similar in that urban location is associated with higher levels of mistrust.

of compensating differentials, nor do they confirm or reject the existence of a spatial equilibrium. There are many issues with these kinds of self-reported data, both within the same cultural contexts and even more so across different contexts. Further many of these quartile averages are rather close to the 45 degree line. Nevertheless, we think these results are novel and interesting and potentially point towards avenues for further research.

A final "prime suspect" for an urban disamenity is congestion. We do not have any comparable measure for the time spent to get to one's job across different density quartiles. However, we have a consistent measure of the time it takes to fetch water. This can be viewed as the lower bound to a daily commute that at least one household member has to undertake. We find that in the lowest density quartile, a household spends about 30 mins on average to the source of drinking water and back, compared to 12 mins for the average household in the highest density quartile.

7. Conclusion

This paper examined whether gaps in livings standards in Africa are consistent with a spatial equilibrium such that higher living standards in denser areas are associated with higher disamenties. Guided by a simple framework, we build a new dataset on housing consumption, non-housing consumption and amenities across space. Newly derived satellite data for pollution, geo-coded household surveys, our matching algorithm, and gridded population density data make this possible. Our dataset allows us to depart from administrative classifications of what is 'rural' and 'urban' to shed light on the distribution of these key groups of variables – non-housing consumption, housing consumption, and amenities – across the entire population density distribution.

We find that non-housing consumption, housing consumption, and most amenity measures are non-decreasing in population density. Many variables exhibit a continuously increasing relationship with population density. While can not prove that there isn't a spatial equilibrium, based on our results it would have to contain a strong component of intangibles in order to compensate individuals for the lower living standards in a broad range of observable characteristics. Rather, considering our evidence on migration, the findings are consistent with a more dynamic notion of spatial equilibrium in which individuals are not indifferent across space at any point in time, but migrate to better locations to close spatial gaps in living standards.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2014): "Management and Shocks to Worker Productivity: Evidence from Air Pollution Exposure in an Indian Garment Factory," mimeo.

Albouy, D. (2012): "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," Unpublished Working Paper, University of Michigan.

Alirol, E., L. Getaz, B. Stoll, F. Chappuis, and L. Loutan (2011): "Urbanisation and infectious diseases in a globalised world," *The Lancet infectious diseases*, 11(2), 131–141.

Allen, T., and C. Arkolakis (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129(3), 1085–1140.

Avery, S. (2012): "Lake Turkana & the Lower Omo: hydrological impacts of major dam and irrigation developments," mimeo, African Studies Centre, the University of Oxford.

Banzhaf, H. S., and R. P. Walsh (2008): "Do People Vote With Their Feet? An Empirical Test of Tiebout's Mechanism," *The American Economic Review*, pp. 843–863.

Baum-Snow, N., and R. Pavan (2012): "Understanding the City Size Wage Gap," *Review of Economic Studies*, 79(1), 88–127.

Bazzi, S., A. Gaduh, A. Rothenberg, and M. Wong (2016): "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 106(9), 2658–98.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014): "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 82(5), 1671–1748.

Caselli, F. (2005): "Accounting for Cross-Country Income Differences," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf., 679-741. Elsevier.

Center for International Earth Science Information Network (2015): "Gridded Population of the World, Version 4 (GPWv4), Preliminary Release 2 (2010)," Accessed 23 September 2015; available: http://www.ciesin.columbia.edu/data/gpw-v4.

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (2016): "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States," *Journal of Urban Economics*.

Ciccone, A., and R. E. Hall (1996): "Productivity and the Density of Economic Activity," *American Economic Review*, 86(1), 54–70.

Currie, J., E. A. Hanushek, M. E. Kahn, M. Neidell, and S. G. Rivkin (2009): "Does Pollution Increase School Absences?," *Review of Economics and Statistics*, 91(4), 682–694.

Currie, J., and R. Walker (2011): "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal. Applied Economics*, 3(1), 65.

Cutler, D., and G. Miller (2005): "The role of public health improvements in health advances: the twentieth-century United States," *Demography*, 42(1), 1–22.

Demombynes, G., and B. Ozler (2002): "Crime and local inequality in South Africa," *World Bank Policy Research Working Paper*, (2925).

Desmet, K., and E. Rossi-Hansberg (2013): "Urban Accounting and Welfare," *American Economic Review*, 103(6), 2296–2327.

——— (2014): "Spatial Development," *American Economic Review*, 104(4), 1211–1243.

DHS (2013): "DHS Sampling and Houshold Listing Manual," Discussion paper, ICF International.

Doxsey-Whitfield, E., K. MacManus, S. B. Adamo, L. Pistolesi, J. Squires, O. Borkovska, and S. R. Baptista (2015): "Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4," *Papers in Applied Geography*, 1(3), 226–234.

Fafchamps, M., and C. Moser (2003): "Crime, Isolation and Law Enforcement," *Journal of African Economies*, 12(4), 625–671.

Fafchamps, M., and F. Shilpi (2009): "Isolation and subjective welfare: evidence from South Asia," *Economic Development and Cultural Change*, 57(4), 641–683.

Ferré, C., F. H. Ferreira, and P. Lanjouw (2012): "Is There a Metropolitan Bias? The relationship between poverty and city size in a selection of developing countries," *The World Bank Economic Review*, p. lhs007.

Geddes, J., R. Martin, B. Boys, and A. van Donkelaar (2015): "Long-Term Trends Worldwide in Ambient NO2 Concentrations Inferred from Satellite Observations.," *Environmental health perspectives*.

Glaeser, E. L., and J. D. Gottlieb (2009): "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 47(4), 983–1028.

Gollin, D., D. Lagakos, and M. E. Waugh (2014): "The Agricultural Productivity Gap," *Quarterly Journal of Economics*, 129(2), 939–993.

GOLLIN, D., AND R. ROGERSON (2014): "Productivity, Transport Costs and Subsistence Agriculture," *Journal of Development Economics*, 107, 38–48.

GRAFF ZIVIN, J., AND M. NEIDELL (2012): "The Impact of Pollution on Worker Productivity," *American Economic Review*, 102(7), 3652–73.

——— (2013): "Environment, Health, and Human Capital," *Journal of Economic Literature*, 51(3), 689–730.

HAINES, M. R. (2001): "The Urban Mortality Transition in the United States, 1800-1940," Working Paper 134, National Bureau of Economic Research.

HARRIS, J. R., AND M. P. TODARO (1970): "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 60(1).

HENDERSON, J. V., A. STOREYGARD, AND U. DEICHMAN (2017): "Has Climate Change Driven Urbanization in Africa?," *Journal of Development Economics*, 124, 60–82.

HENDERSON, V., A. STOREYGARD, AND D. WEIL (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, (2).

HERRENDORF, B., AND T. SCHOELLMAN (2014): "Wages, Human Capital, and Structural Transformation," mimeo.

KAHN, M. E. (2010): "New evidence on trends in the cost of urban agglomeration," in *Agglomeration Economics*, pp. 339–354. University of Chicago Press.

KAHN, M. E., AND R. WALSH (2015): "Chapter 7 - Cities and the Environment," in *Handbook of Regional and Urban Economics*, ed. by J. V. H. Gilles Duranton, and W. C. Strange, vol. 5 of *Handbook of Regional and Urban Economics*, pp. 405 – 465. Elsevier.

KINNEY, P. L., M. G. GICHURU, N. VOLAVKA-CLOSE, N. NGO, P. K. NDIBA, A. LAW, A. GACHANJA, S. M. GAITA, S. N. CHILLRUD, AND E. SCLAR (2011): "Traffic impacts on PM 2.5 air quality in Nairobi, Kenya," *Environmental science & policy*, 14(4), 369–378.

LAGAKOS, D., AND M. E. WAUGH (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 103(2), 948–980.

LINARD, C., M. GILBERT, R. W. SNOW, A. M. NOOR, AND A. J. TATEM (2012): "Population distribution, settlement patterns and accessibility across Africa in 2010," *PLoS One*, 7(2), e31743.

MCMILLAN, M. S., AND D. RODRIK (2011): "Globalization, Structural Change and Productivity Growth," NBER Working Paper No. 17143.

MORTEN, M. (2013): “Temporary Migration and Endogenous Risk Sharing in Village India,” Unpublished Working Paper, Stanford University.

MUNSHI, K., AND M. ROSENZWEIG (2016): “Networks and misallocation: Insurance, migration, and the rural-urban wage gap,” *The American Economic Review*, 106(1), 46–98.

NUNN, N., AND L. WANTCHEKON (2011): “The Slave Trade and the Origins of Mistrust in Africa,” *American Economic Review*, 101(7), 3221–52.

PEREZ-HEYDRICH, C., J. WARREN, C. BURGERT, AND M. EMCH (2013): “Guidelines on the use of DHS GPS data,” Discussion paper.

PETKOVA, E. P., D. W. JACK, N. H. VOLAVKA-CLOSE, AND P. L. KINNEY (2013): “Particulate matter pollution in African cities,” *Air Quality, Atmosphere & Health*, 6(3), 603–614.

RAFFO, J. (2015): “MATCHIT: Stata module to match two datasets based on similar text patterns,” Statistical Software Components, Boston College Department of Economics.

REDDING, S., AND E. ROSSI-HANSBERG (Forthcoming): “Spatial Development,” *Annual Review of Economics*.

RESTUCCIA, D., D. T. YANG, AND X. ZHU (2008): “Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis,” *Journal of Monetary Economics*, 55, 234–250.

ROBACK, J. (1982): “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90(6), 1257–1278.

ROSEN, S. (1979): “Wage-Based Indexes of Urban Quality of Life,” in *Current Issues in Urban Economics*, ed. by P. Mieszkowski, and M. Straszheim. Johns Hopkins University Press.

TATEM, A. J., P. W. GETHING, D. L. SMITH, AND S. I. HAY (2013): “Urbanization and the global malaria recession,” *Malaria journal*, 12(1), 133.

TOPEL, R. H. (1986): “Local Labor Markets,” *Journal of Political Economy*, 94(3), S111–S143.

VAN DONKELAAR, A., R. MARTIN, M. BRAUER, AND B. BOYS (2015): “Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter.,” *Environmental Health Perspectives*, 123(2), 135–144.

VOLLRATH, D. (2009): “How Important are Dual Economy Effects for Aggregate Productivity?,” *Journal of Development Economics*, 88(2), 325–334.

VRIJHEID, M., D. MARTINEZ, S. MANZANARES, P. DADVAND, A. SCHEMBARI, J. RANKIN, AND M. NIEUWENHUIJSEN (2011): "Ambient air pollution and risk of congenital anomalies: a systematic review and meta-analysis," *Environmental Health Perspectives*, 119(5), 598–606.

WHO (2006): *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization.

——— (2014): *Indoor air quality guidelines: household fuel combustion*. World Health Organization.

WHO (2015): "Nutrition Landscape Information System (NLiS): Children 6-23 months of age who receive a minimum acceptable diet," Accessed 1 september 2015; available: here.

YOUNG, A. (2014): "Inequality, the Urban-Rural Gap and Migration," *Quarterly Journal of Economics*, 129(2), 939–993.

Tables and Figures

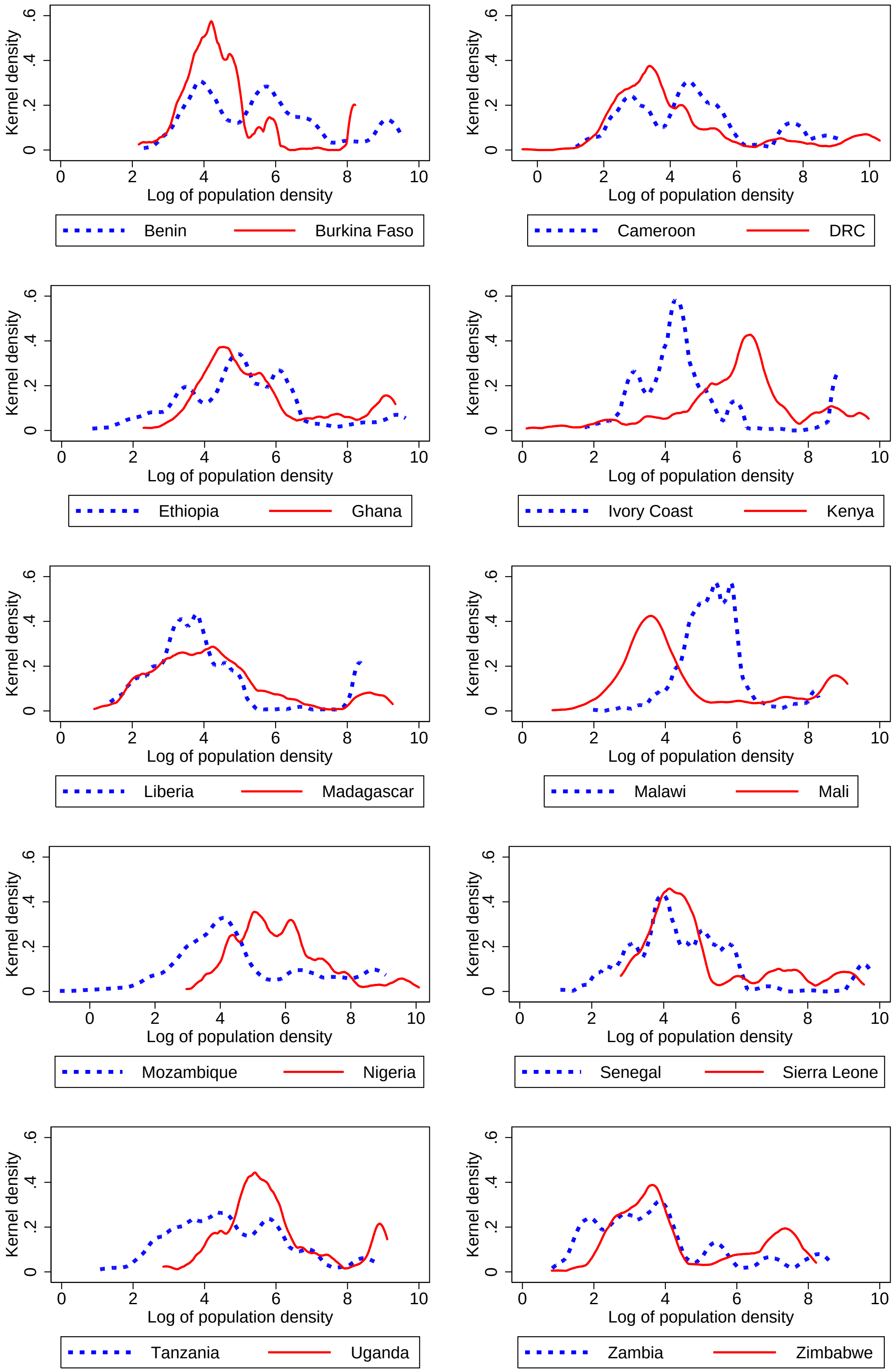
Table 1: Sample of households

Country	Households	Population
Benin	17,332	10,050,702
BurkinaFaso	13,617	16,460,141
Cameroon	14,189	21,699,631
DRC	16,344	65,705,093
Ethiopia	16,037	91,728,849
Ghana	11,574	25,366,462
IvoryCoast	9,394	19,839,750
Kenya	9,033	43,178,141
Liberia	9,333	4,190,435
Madagascar	17,578	22,293,914
Malawi	24,210	15,906,483
Mali	10,105	14,853,572
Mozambique	13,899	25,203,395
Nigeria	38,170	168,800,000
Senegal	7,780	13,726,021
SierraLeone	12,629	5,978,727
Tanzania	9,282	47,783,107
Uganda	8,939	36,345,860
Zambia	7,164	14,075,099
Zimbabwe	9,442	13,724,317
Total	276,051	769,082,846

Table 2: Surveys

	household	malaria	migration	crime (Afrobarom.)	crime (LSMS)
Benin	Benin 2011-12 Standard DHS	Benin 2011-12 Standard DHS		x	
BurkinaFaso	Burkina Faso 2010 Standard DHS	Burkina Faso 2010 Standard DHS			
Cameroon	Cameroon 2011 Standard DHS	Cameroon 2011 Standard DHS			
DRC	Congo Democratic Republic 2013-14 Standard DHS	Congo Democratic Republic 2013-14 Standard DHS	Congo Democratic Republic 2007 Standard DHS		
Ethiopia	Ethiopia 2011 Standard DHS	Ethiopia 2011 Standard DHS	Ethiopia 2005 Standard DHS		2013/14 Ethiopian Socioeconomic Survey
Ghana	Ghana 2008 Standard DHS		Ghana 2008 Standard DHS	x	
IvoryCoast	Cote d'Ivoire 2011-12 Standard DHS	Cote d'Ivoire 2011-12 (14) Standard DHS			
Kenya	Kenya 2008-09 Standard DHS		Kenya 2008-09 Standard DHS	x	
Liberia	Liberia 2013 Standard DHS	Liberia 2011 MIS	Liberia 2007 Standard DHS		
Madagascar	Madagascar 2008-09 Standard DHS	Madagascar 2013 MIS DHS-VI	Madagascar 2008-09 Standard DHS	x	
Malawi	Malawi 2010 Standard DHS	Malawi 2012 MIS	Malawi 2010 Standard DHS	x	LSMS 2004/05
Mali	Mali 2012-13 Standard DHS	Mali 2012-13 Standard DHS	Mali 2006 Standard DHS	x	
Mozambique	Mozambique 2011 Standard DHS	Mozambique 2011 Standard DHS	Mozambique 2009 Standard AIS	x	
Nigeria	Nigeria 2013 Standard DHS	Nigeria 2010 MIS	Nigeria 2008 Standard DHS	x	NGHS, Panel Wave 2, 2012-2013; Post-harvest household questionnaire
Senegal	Senegal 2010-11 Standard DHS	Senegal 2010-11 Standard DHS	Senegal 2005 Standard DHS	x	
SierraLeone	Sierra Leone 2013 Standard DHS		Sierra Leone 2008 Standard DHS		
Tanzania	Tanzania 2010 Standard DHS	Tanzania 2011-12 Standard AIS	Tanzania 2007-08 Standard AIS	x	Tanzania NPS 2008
Uganda	Uganda 2011 Standard DHS	Uganda 2009 MIS	Uganda 2011 Standard AIS	x	Uganda NPS 2009/10
Zambia	Zambia 2007 Standard DHS		Zambia 2007 Standard DHS	x	
Zimbabwe	Zimbabwe 2010-11 Standard DHS		Zimbabwe 2010-11 Standard DHS		

Figure 1: Distribution of population density



Notes: For expositional simplicity we exclude 84 observations with negative log of population density.

Figure 2: Urban/Rural classifications

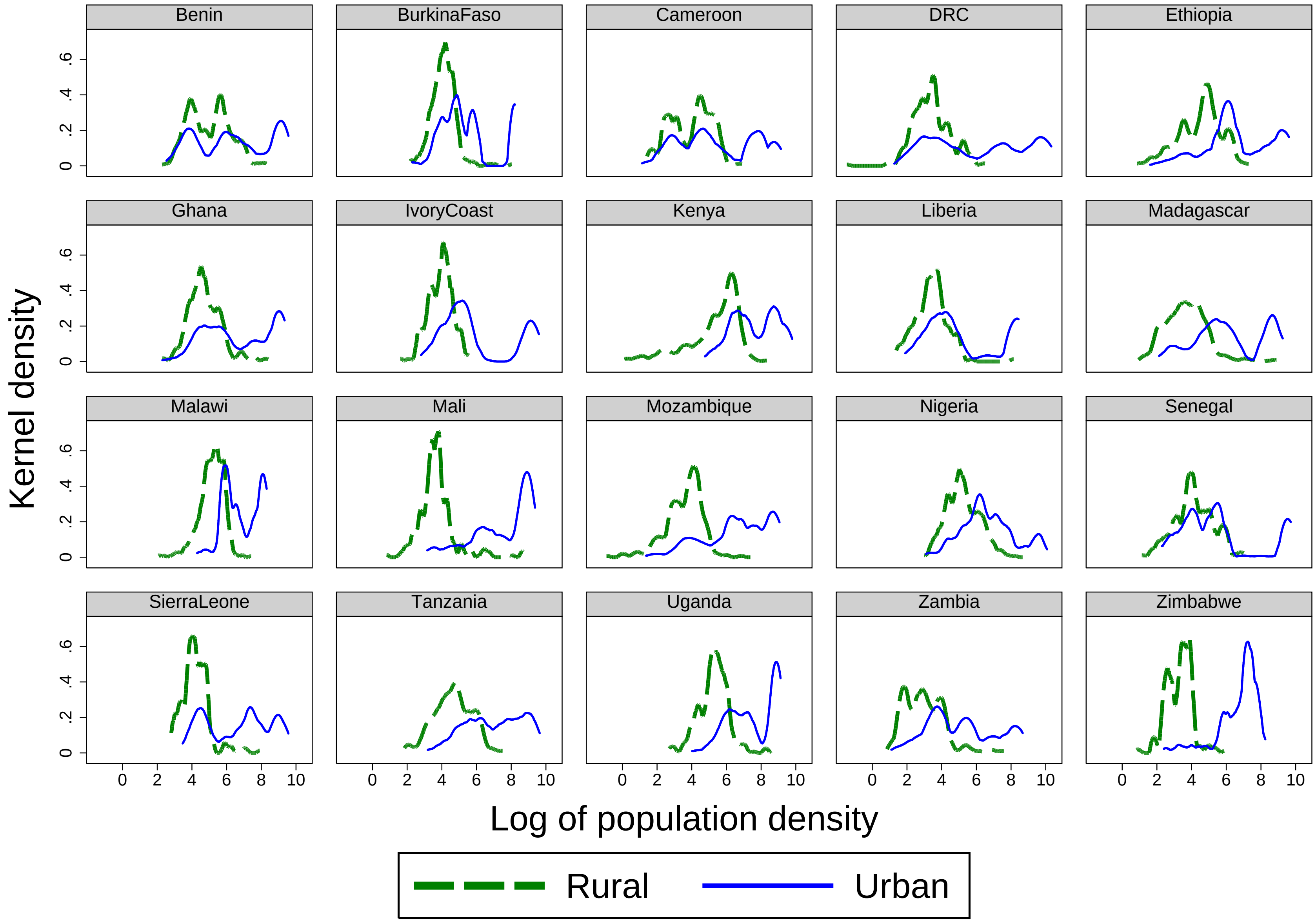
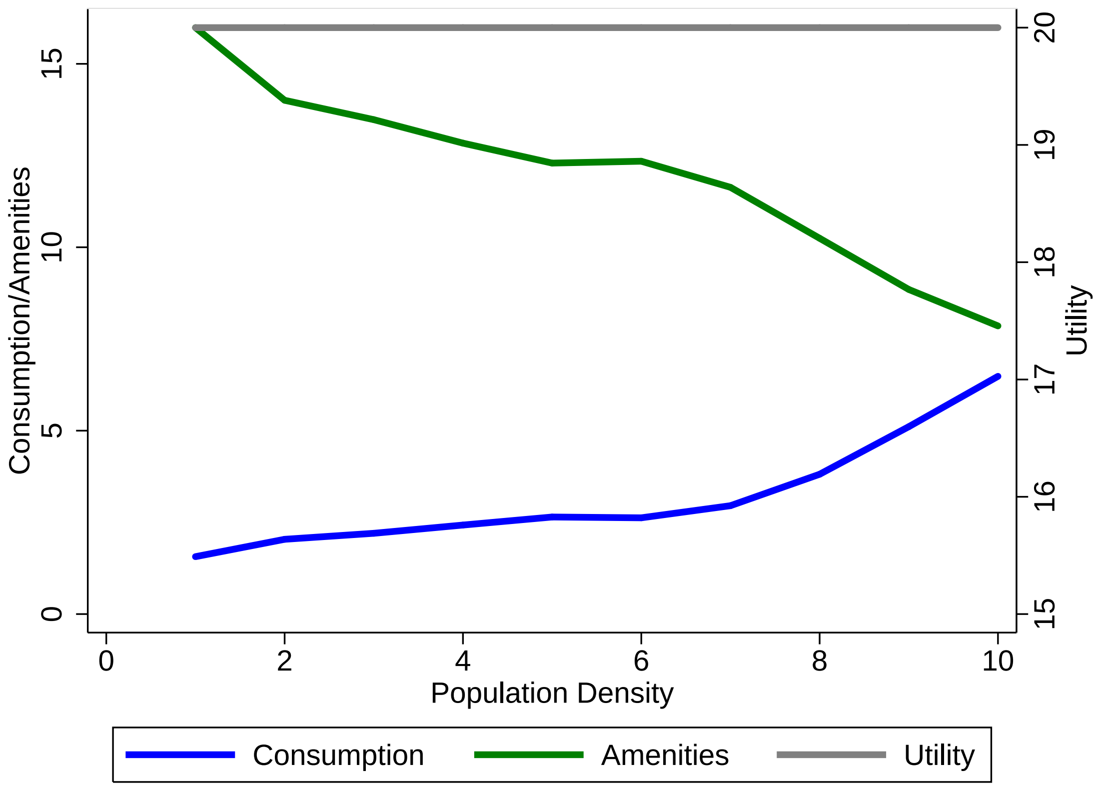


Figure 3: Amenities, consumption and utility across population density space



Notes: The figure shows the relationship between consumption, amenities, and utility as predicted in a standard Rosen-Roback model. Consumption c is proxied with an asset and housing quality index, counting the number of durables a household has and housing quality indicators using data from **?** for 20 Sub-Saharan African countries. The index simply counts a household's assets (tv, car, phone, motorcycle) and measures of basic housing quality (electricity, tap water, flush toilet, finished wall, finished roof). It therefore ranges from zero to ten; the mean over the entire sample is 3.1. I then divide individuals into population density deciles and compute the average of the index for the different deciles across the entire sample. Having fixed $\bar{U} = 20$, $\alpha = 0.5$, and $h = 1$, this allows me to back out how the value of amenities evolves across space. For a given increase in living standards across population density space, amenities have to decrease to ensure equality of utilities across space.

Figure 4: Phone ownership

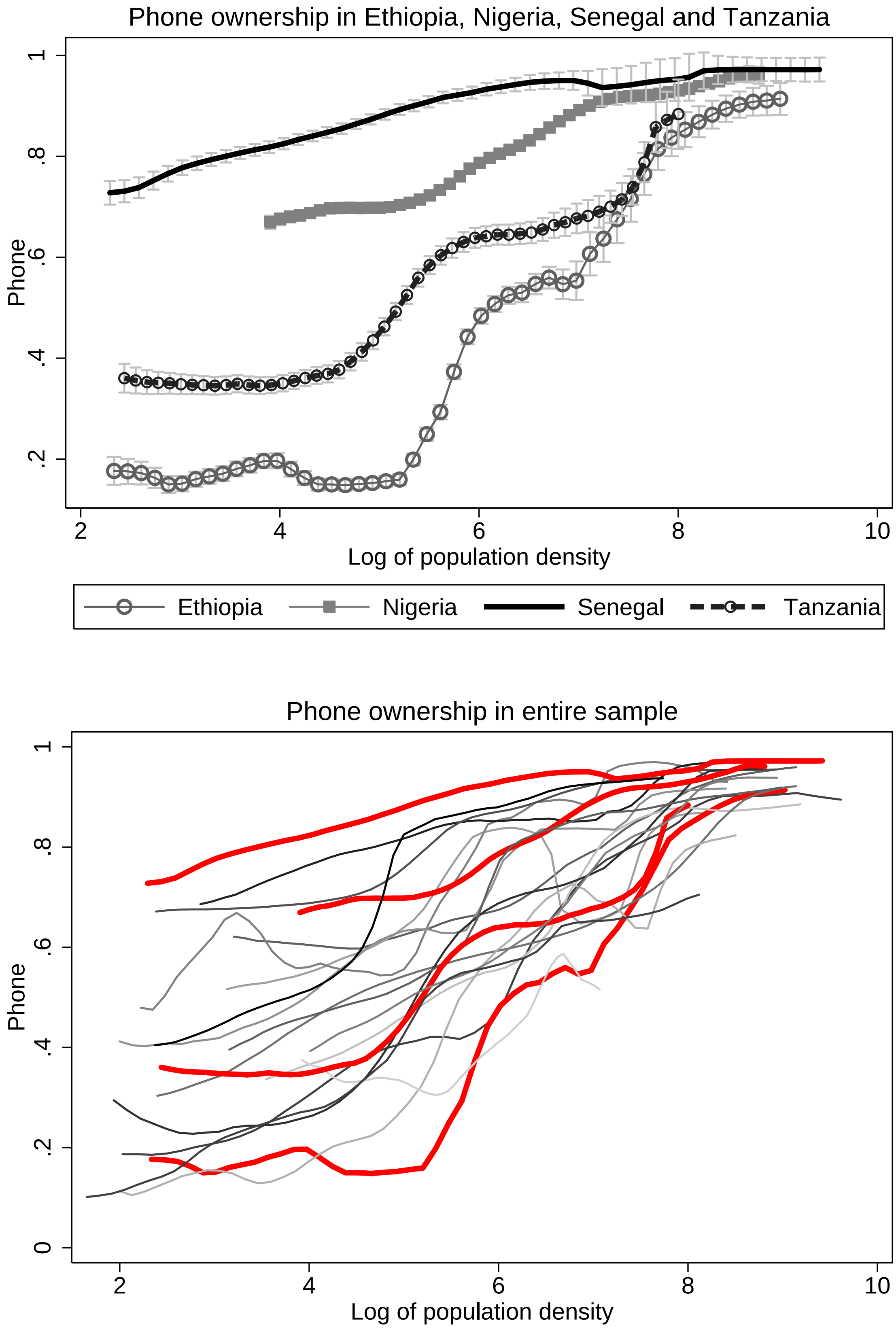


Figure 5: Durables ownership across space

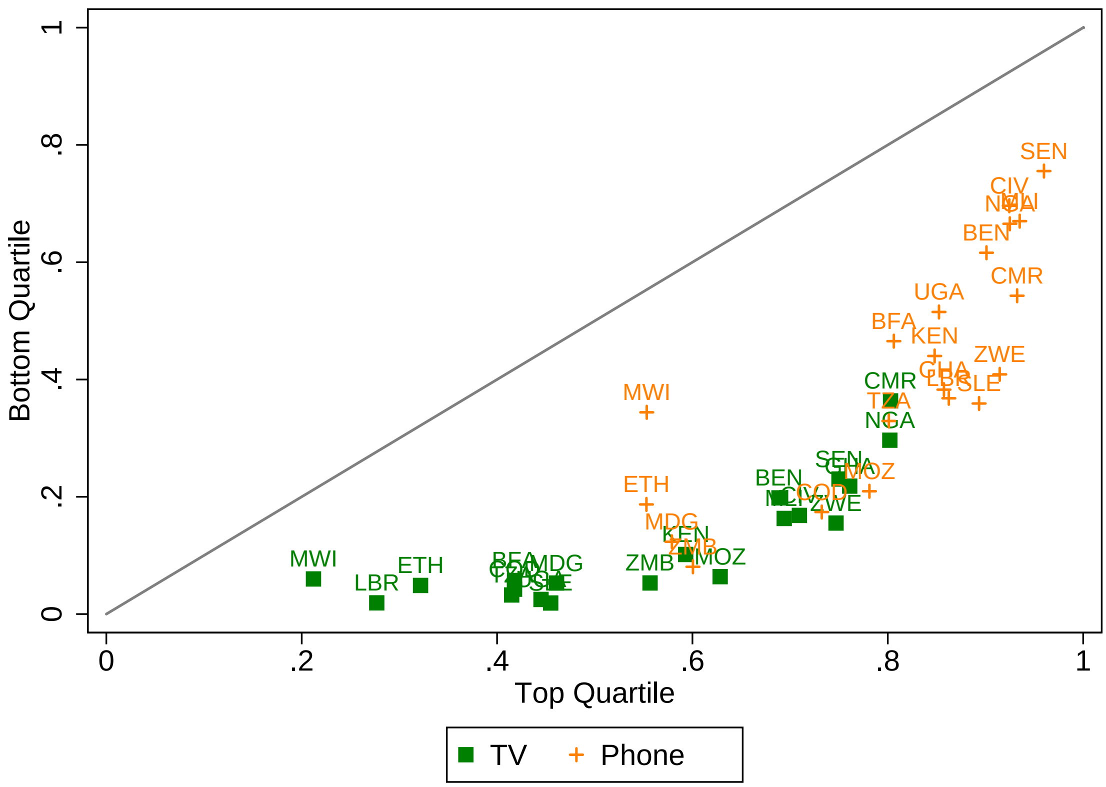


Figure 6: Child Health

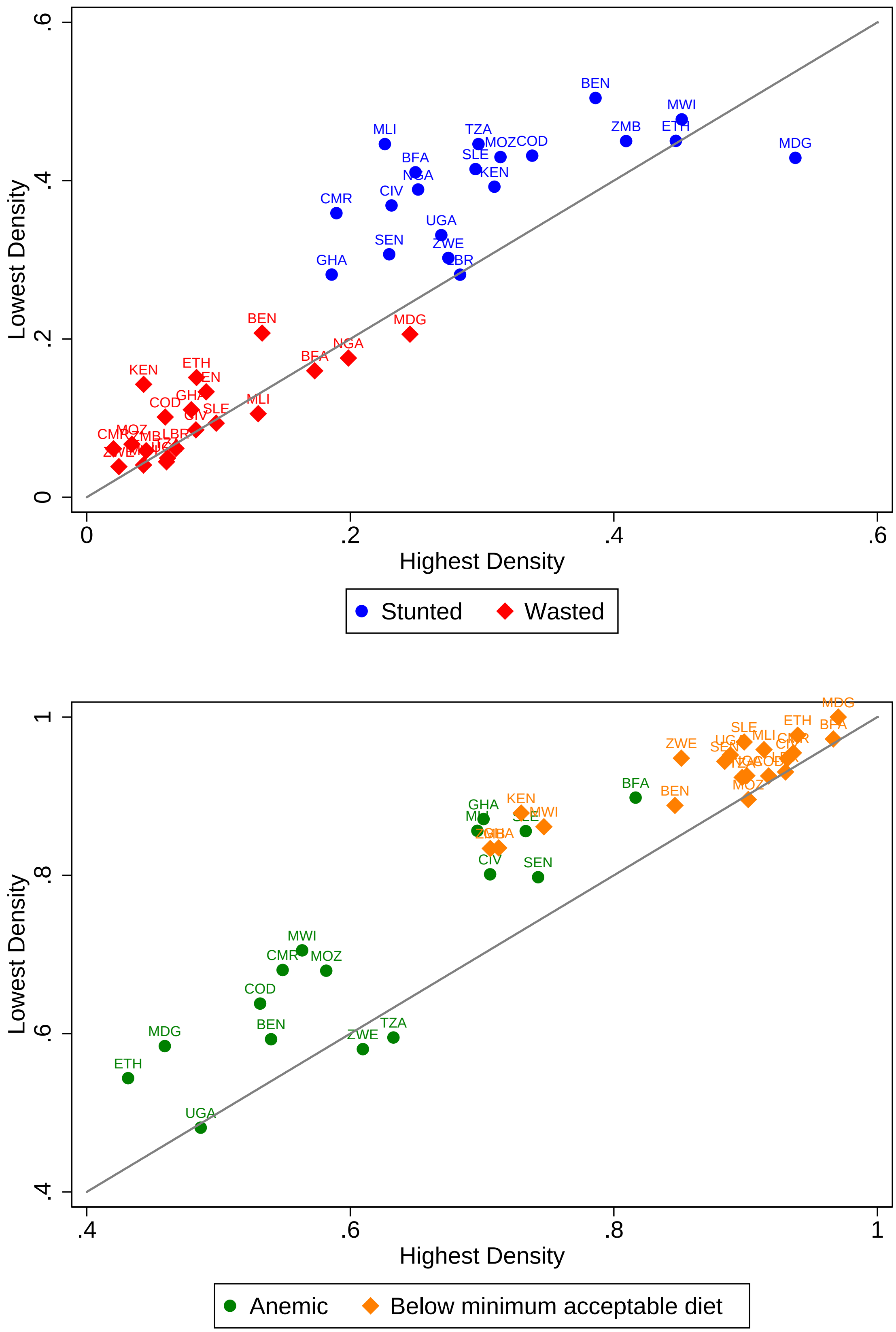


Figure 7: Housing amenities: Finished wall and roof

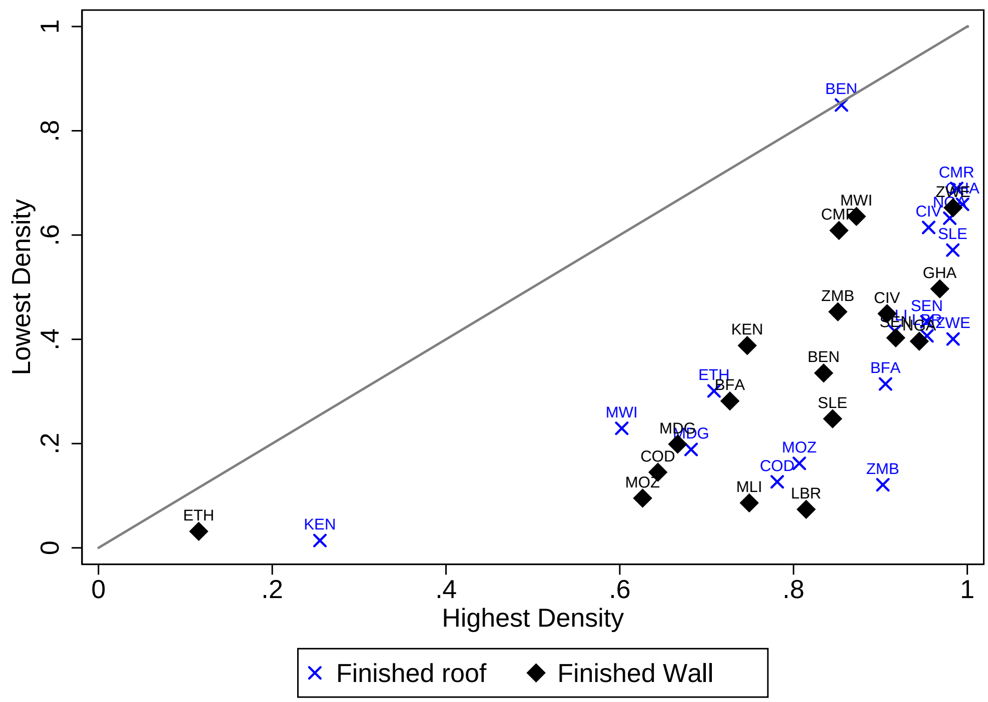


Figure 8: Housing amenities: Finished floor and electricity

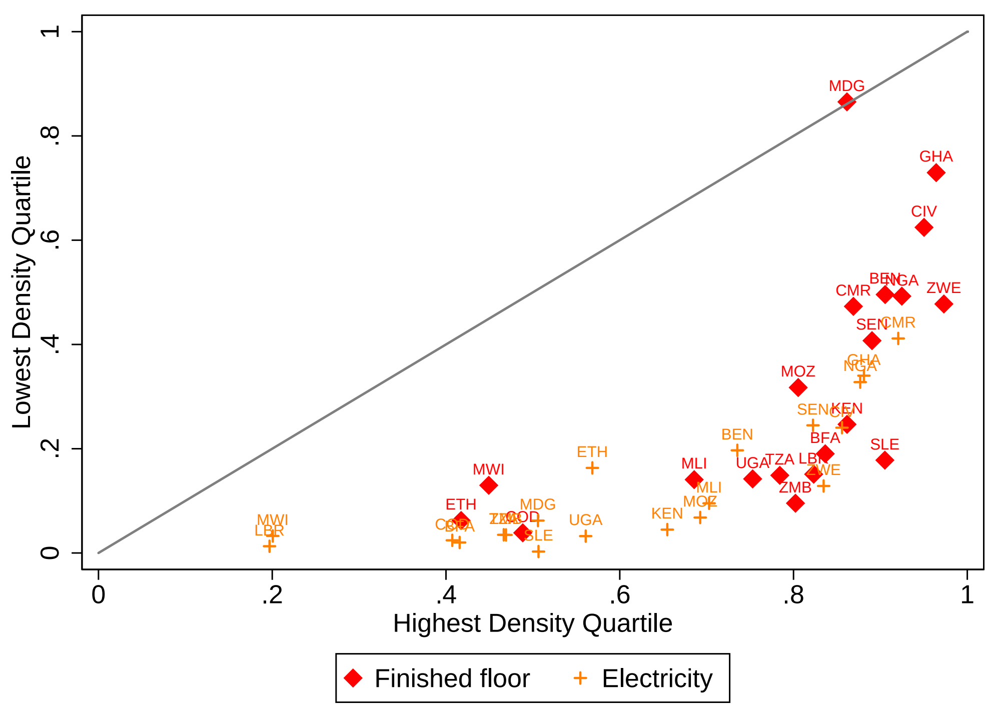


Figure 9: Housing amenities: Tap water and flush toilet

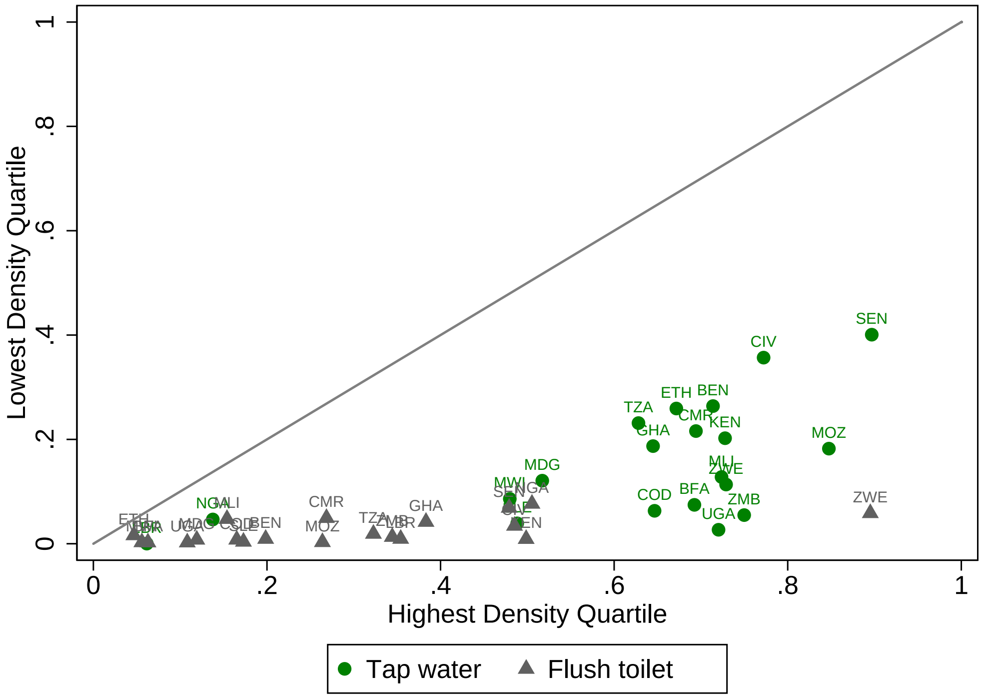
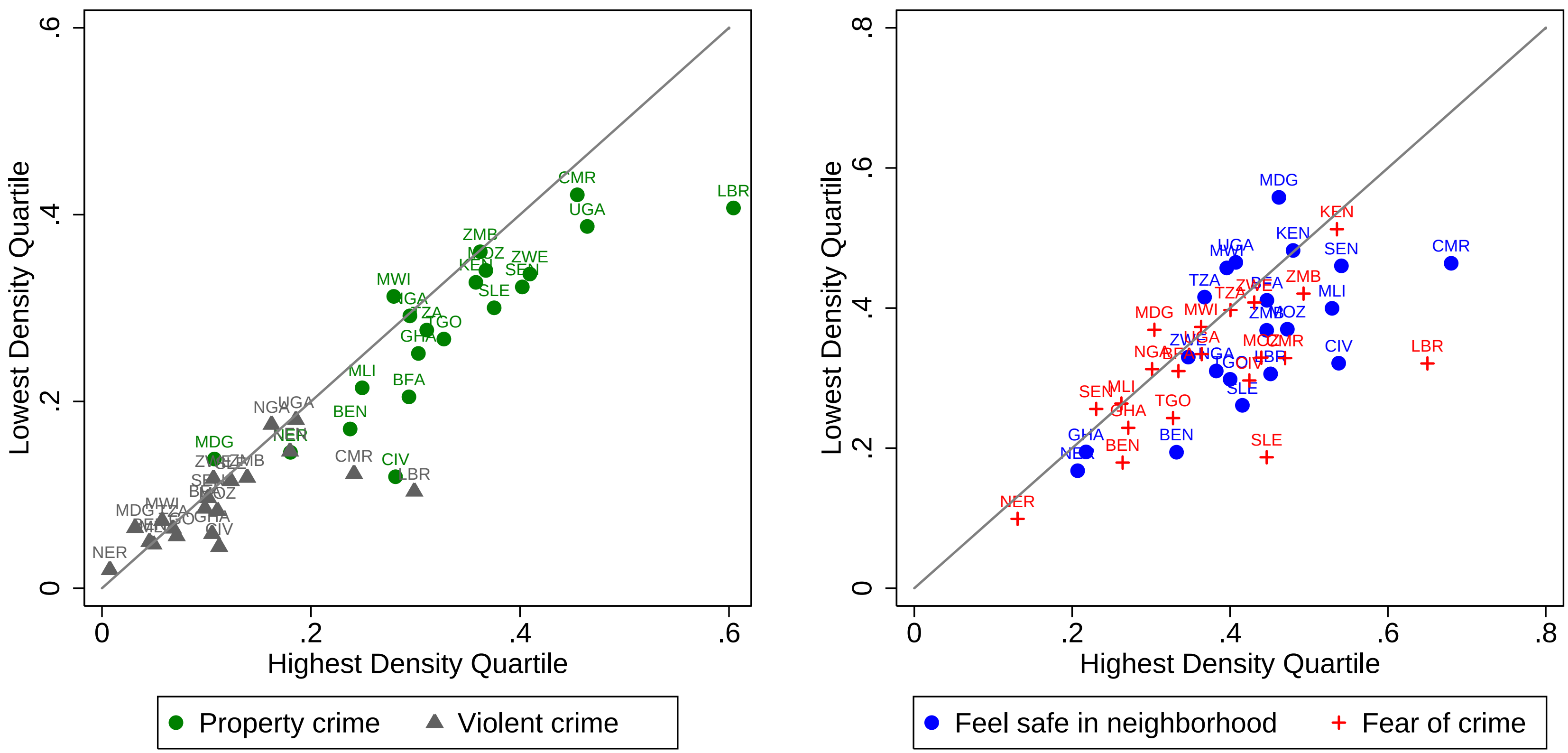


Figure 10: Crime - Afrobarometer



Notes: The figure compares the proportion of individuals who report experiences of crime (shown in the figure on the left) or fear thereof (shown in the figure on the right) in different population density quartiles. The x-axis shows the highest density quartile, and the y-axis shows the lowest density quartile. The closer a country aligns to the 45 degree line, the more similar the prevalence of crime or fear of crime is.

Figure 11: Pollution: PM2.5

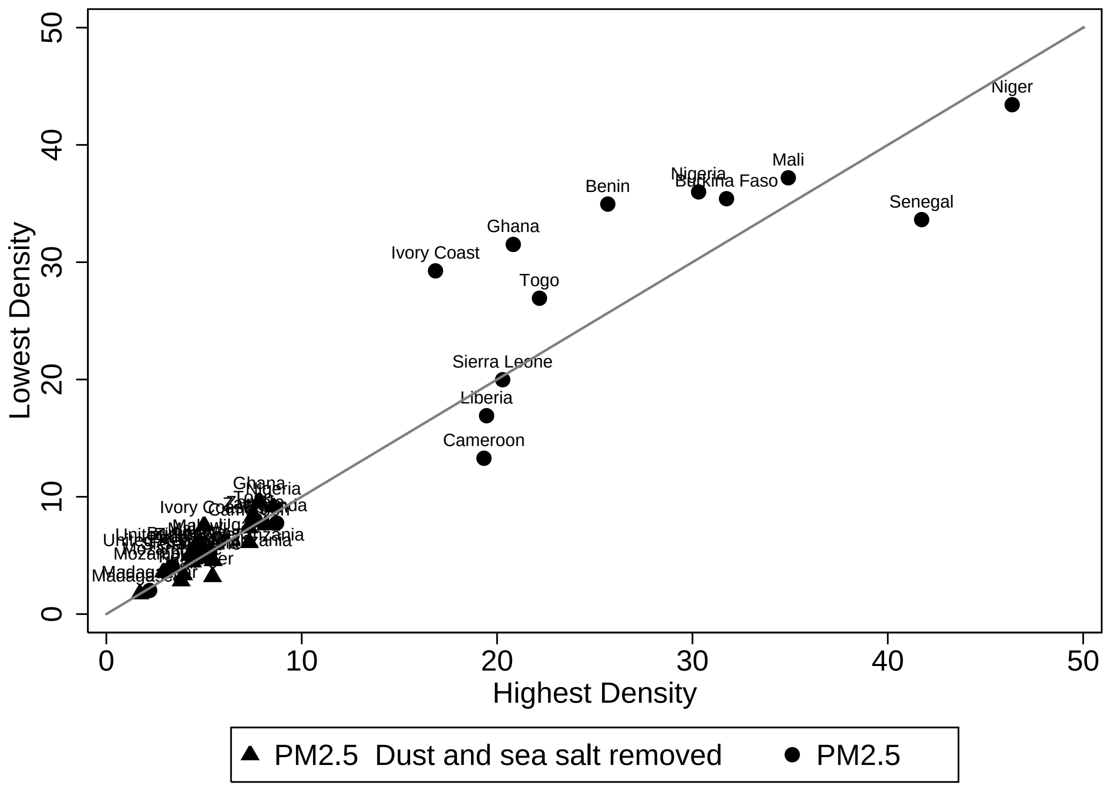
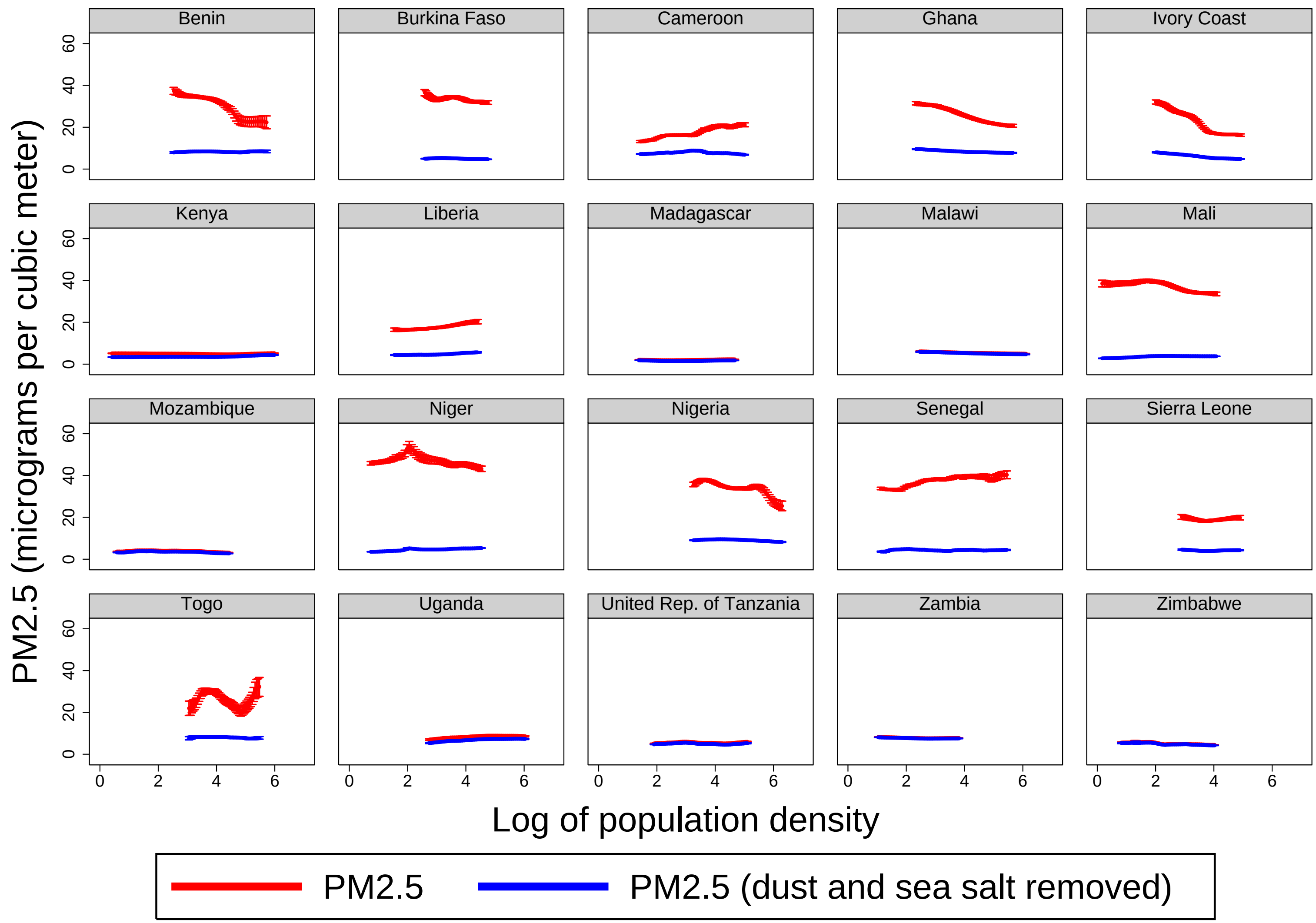


Figure 12: PM2.5 concentration



Notes: The figure shows a kernel-weighted local polynomial regression of the level of PM2.5 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

Figure 13: Pollution: NO2

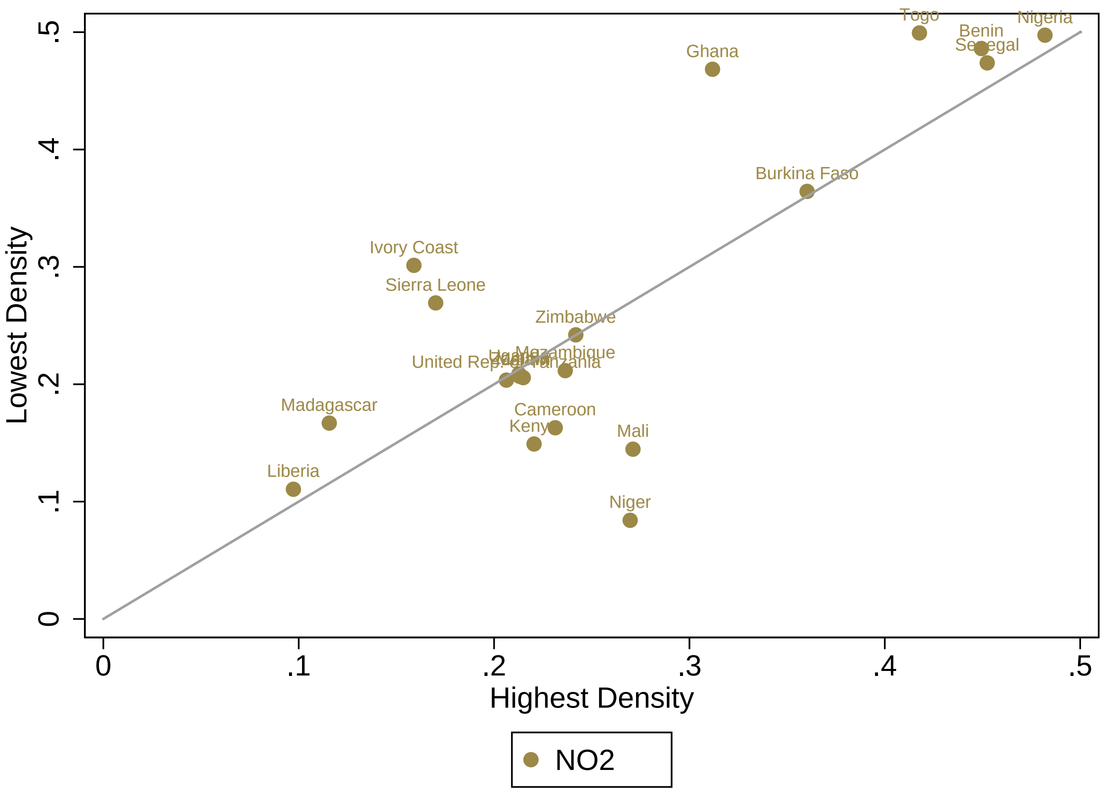
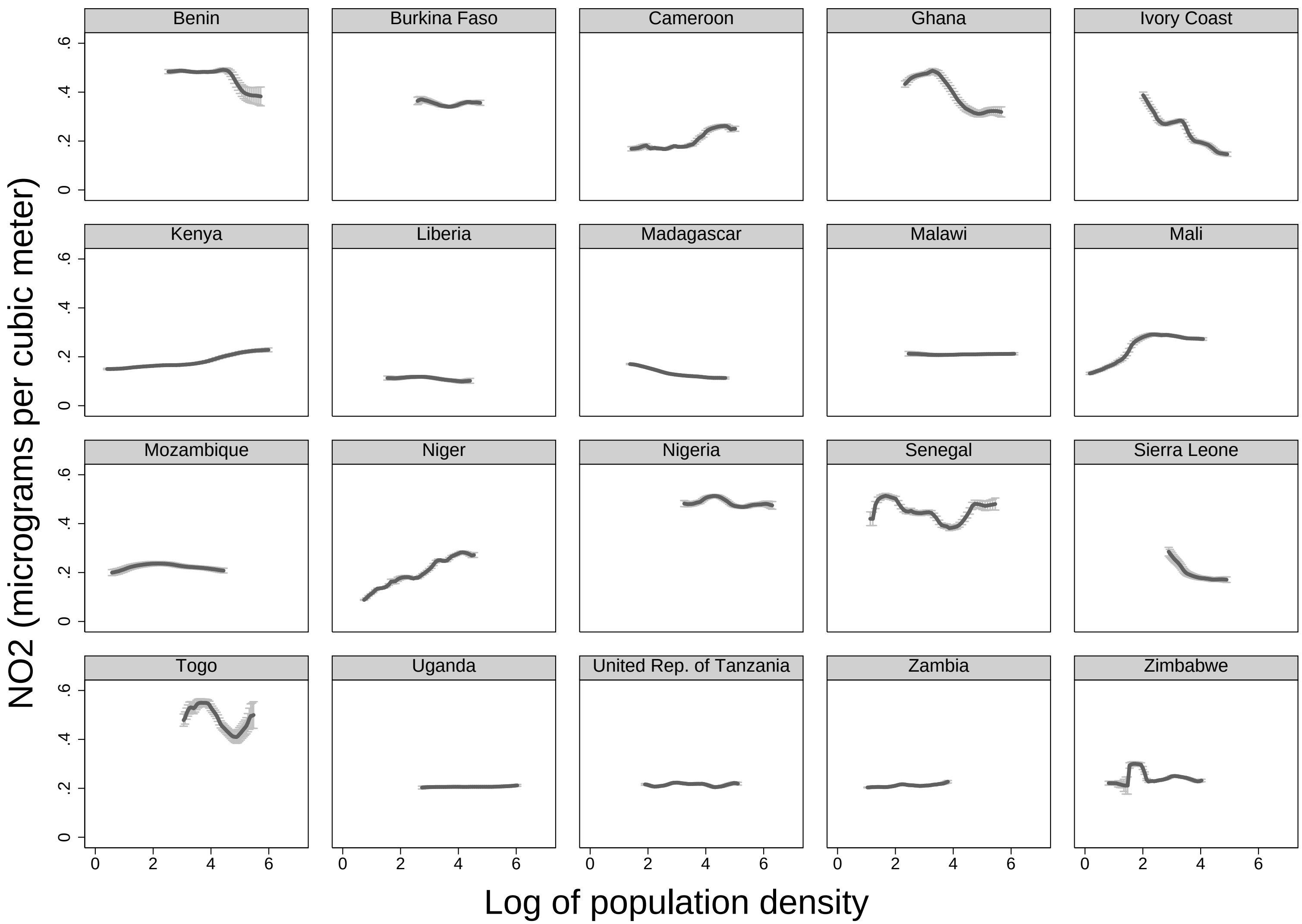
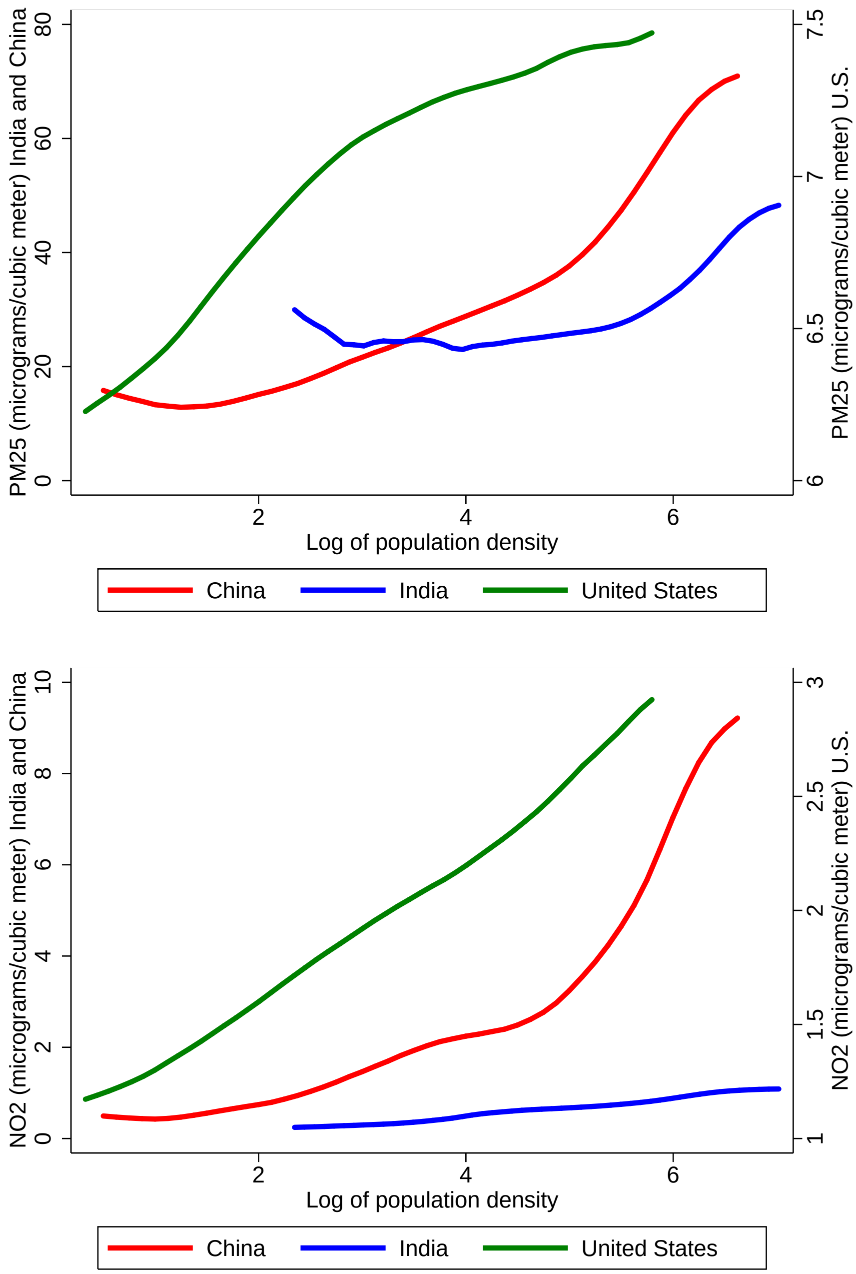


Figure 14: Nitrogen dioxide concentration



Notes: The figure shows a kernel-weighted local polynomial regression of the level of NO2 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

Figure 15: Pollution-Population Density gradients in other countries



Notes: The figures shows a kernel-weighted local polynomial regression of pollution on the log of population density for China, India and the United States. The top panel shows PM2.5, and the bottom panel shows NO2, both measured in micrograms per cubic meter. All three countries show visible density gradients. In China, PM2.5 levels for the top population density decile amount to 66 $\mu g/m^3$, more than six times the WHO recommended threshold; the lowest population density decile has a level of 13 $\mu g/m^3$. In India, the top decile has a level of 41 $\mu g/m^3$, still four times the WHO recommended threshold, compared to 6 $\mu g/m^3$ in the lowest decile. The bottom figure shows NO2 distribution for these other countries. The levels are much lower, but there are gradients again for China, India and the United States.

Figure 16: Solid fuels and indoor pollution

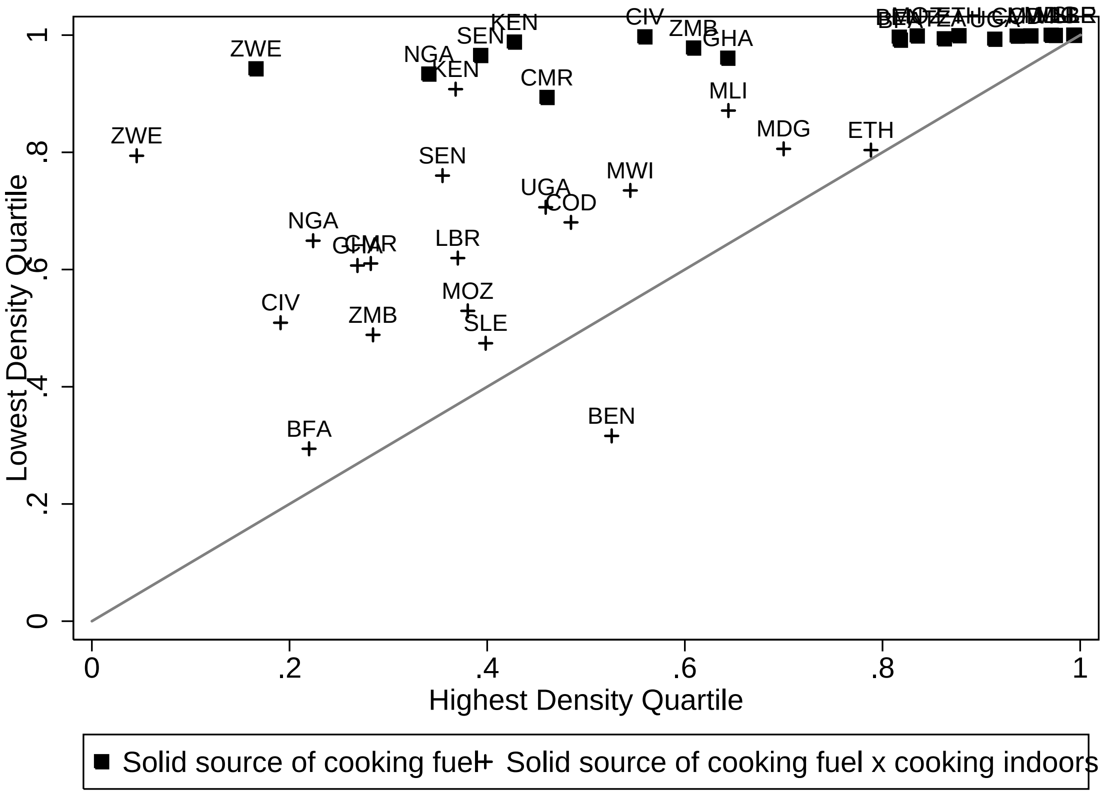


Figure 17: Solid type of fuel for cooking

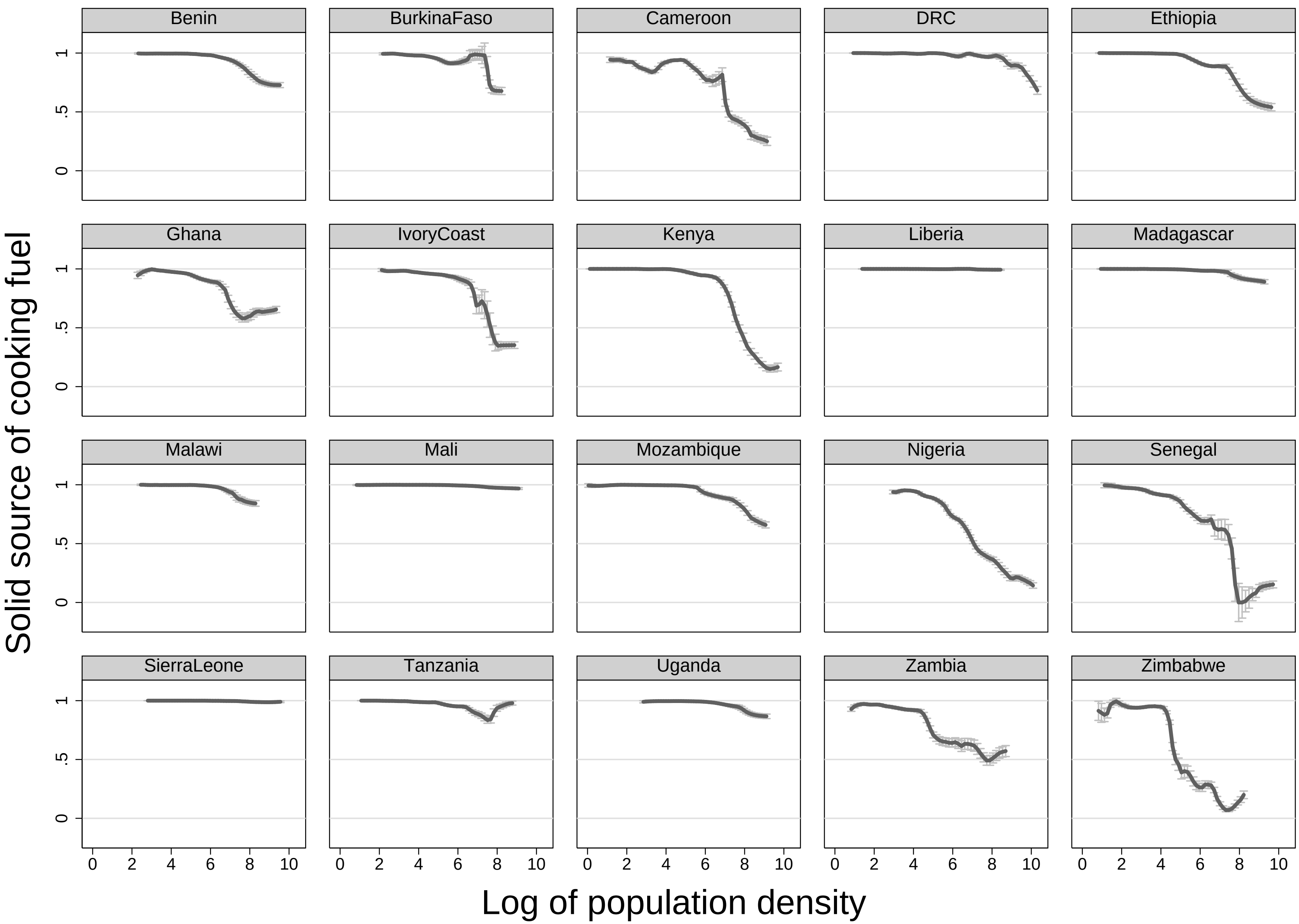


Figure 18: Food cooked indoors

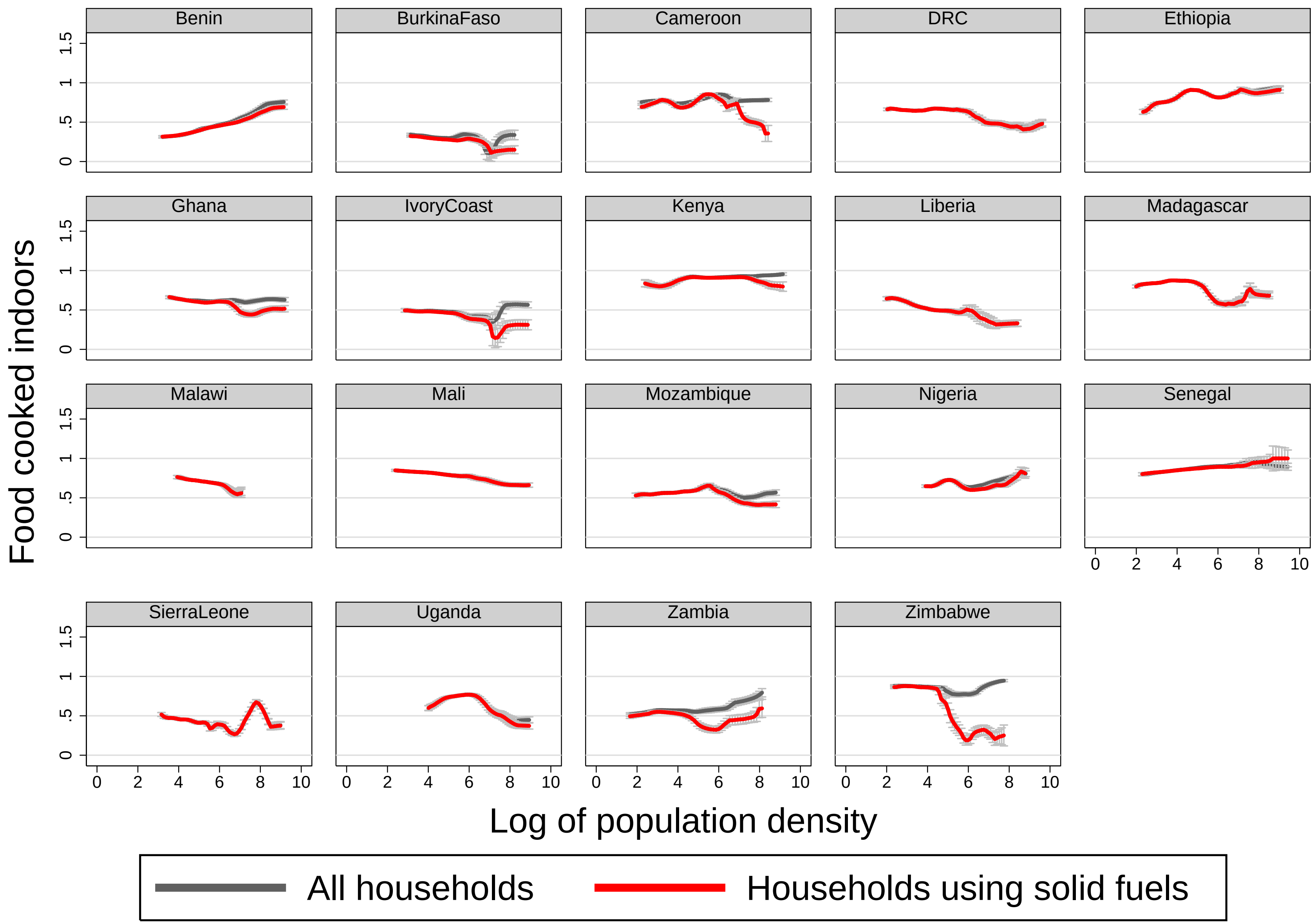


Figure 19: Electricity by highest education of household head in Nigeria

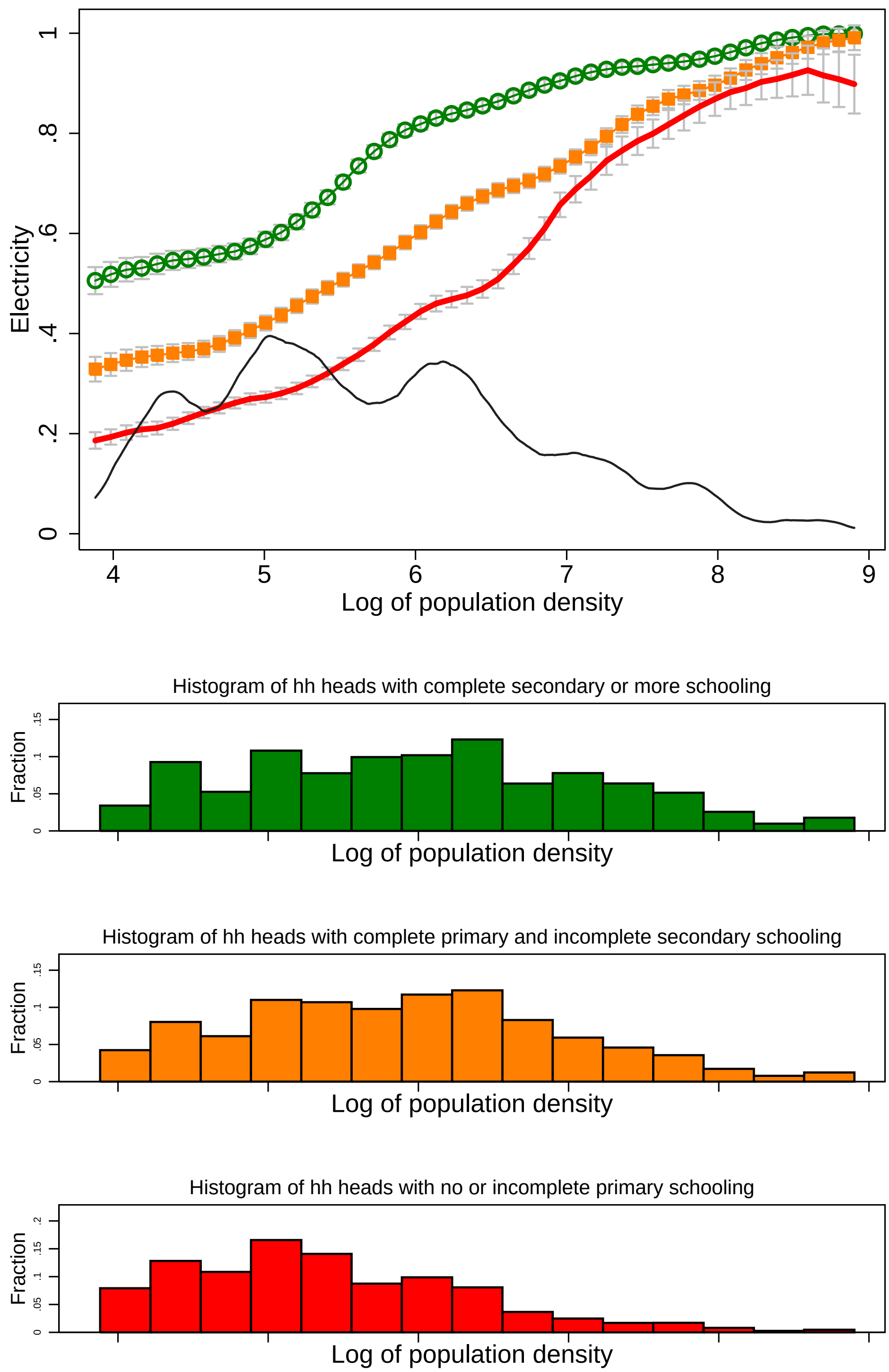
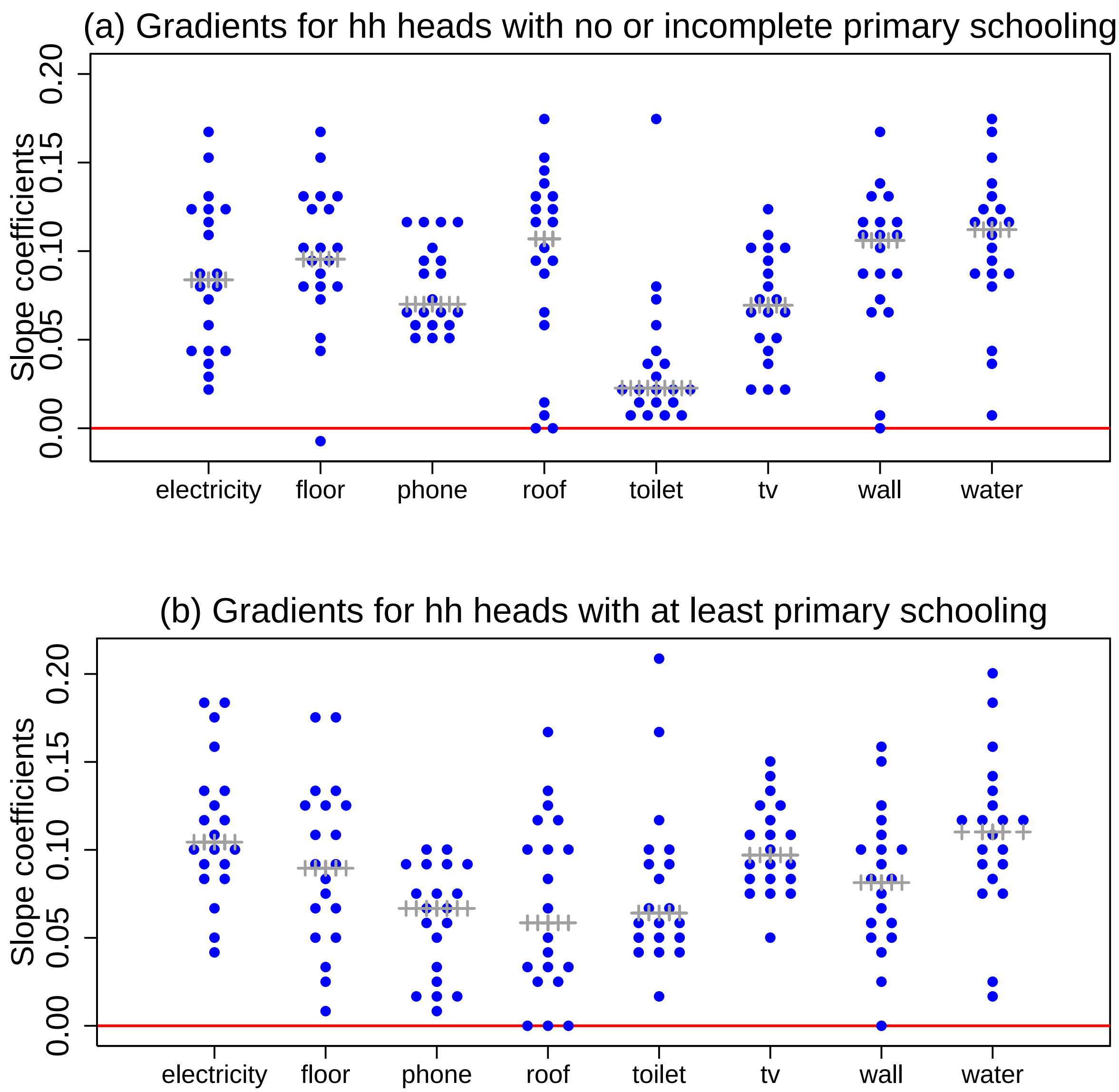


Figure 20: Slopes by education across the sample



Notes: Each dot in the Figure represents the coefficient estimate of a linear projection for household i in country c:

$$x_{ic} = \theta_0 + \theta_1 P_{ic} + \theta_2 E_{ic} + \theta_3 (P_{ic} * E_{ic}) + \epsilon_{ic}$$

where x_{ic} is a measure of consumption, P_{ic} is the log of population density and E_{ic} is a dummy variable that is equal to one if the household head has completed primary education or more. Panel (a) shows the linear gradients for households with household heads who have less than complete primary education, and panel (b) shows gradients for households with household heads who have complete primary education or more. The y-axis indicates the size of the coefficient and the grey horizontal bars show the median slope coefficient.

Table 3: Average Differences in Amenities by Density

	Differences From Q1			
	Q2	Q3	Q4	Regional Std. Dev
phone	0.072	0.200	0.426	0.467
	14	18	20	
tv	0.028	0.164	0.463	0.413
	15	18	20	
car	0.005	0.019	0.086	0.182
	6	15	20	
motorcycle	-0.010	-0.007	0.002	0.264
	8	13	14	
electricity	0.034	0.188	0.512	0.412
	16	17	20	
watert	0.020	0.211	0.505	0.436
	12	18	20	
floorc	0.045	0.193	0.480	0.456
	13	17	20	
toiletf	0.014	0.090	0.284	0.283
	14	17	20	
timewater	-4.058	-7.762	-16.426	35.011
	11	16	18	
solid	-0.006	-0.074	-0.271	0.243
	10	13	20	
rooff	0.092	0.261	0.478	0.445
	16	18	17	
wallf	0.063	0.180	0.482	0.461
	14	17	18	
inside	0.019	0.026	0.014	0.441
	14	16	16	

Note: The first three columns report the average differences from the second, third and fourth density quartiles relative to the first (least dense) quartile across our set of 20 countries. The fourth column reports the average standard deviation across regions across our set of 20 countries. Numbers in parentheses below each average difference are the number of countries with a difference that is statistically significant at the one-percent level.

Table 4: Average Differences in Pollution and Crime by Density

	Differences From Q1			
	Q2	Q3	Q4	Regional Std. Dev
no2	0.006	-0.007	0.002	0.085
	11	13	12	
pm25	-0.569	-1.301	-1.090	6.633
	10	11	11	
theft	0.021	0.023	0.047	0.453
	6	4	6	
attack	-0.004	0.003	0.022	0.289
	2	2	5	
fear	0.011	0.024	0.051	0.463
	4	5	8	
anxiety	0.011	-0.005	-0.047	0.464
	1	2	3	
unsafe	0.014	0.015	0.078	0.478
	4	5	8	
tv	0.021	0.102	0.305	0.435
	5	12	19	
radio	0.017	0.044	0.099	0.456
	3	7	13	
toilet	0.035	0.073	0.260	0.433
	7	8	18	
phone	0.017	0.078	0.219	0.446
	5	8	17	
water	0.031	0.076	0.320	0.437
	5	9	19	
trust neighbor	-0.022	-0.074	-0.154	0.457
	4	8	13	
trust any	0.007	0.003	-0.073	0.388
	0	3	11	
trust ownethnic	-0.029	-0.081	-0.134	0.467
	2	3	5	
swb bin	-0.022	-0.024	0.005	0.444
	2	2	7	

Note: The first three columns report the average differences from the second, third and fourth density quartiles relative to the first (least dense) quartile across our set of 20 countries. The fourth column reports the average standard deviation across regions across our set of 20 countries. Numbers in parentheses below each average difference are the number of countries with a difference that is statistically significant at the one-percent level.

Table 5: Average Differences in Child Health by Density

	Differences From Q1			
	Q2	Q3	Q4	Regional Std. Dev
anemic	-0.003	-0.041	-0.080	0.453
	1	6	11	
stunted	0.010	-0.017	-0.106	0.479
	4	7	15	
wasted	-0.005	-0.015	-0.022	0.282
	4	7	8	
notmad	-0.013	-0.027	-0.053	0.276
	3	3	9	

Note: The first three columns report the average differences from the second, third and fourth density quartiles relative to the first (least dense) quartile across our set of 20 countries. The fourth column reports the average standard deviation across regions across our set of 20 countries. Numbers in parentheses below each average difference are the number of countries with a difference that is statistically significant at the one-percent level.

Figure 21: Afrobarometer: Social insurance and anxiety

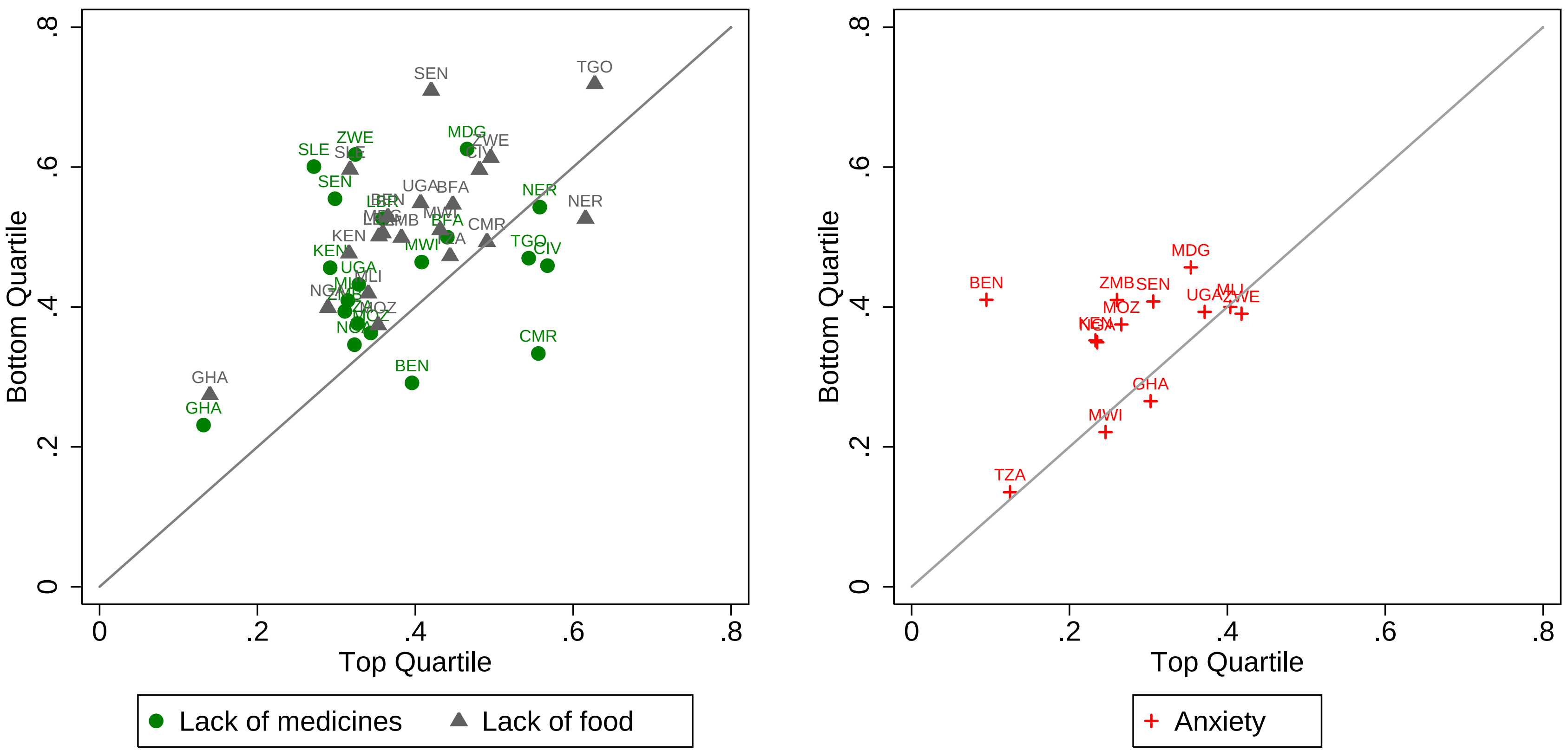


Figure 22: Afrobarometer: Trust

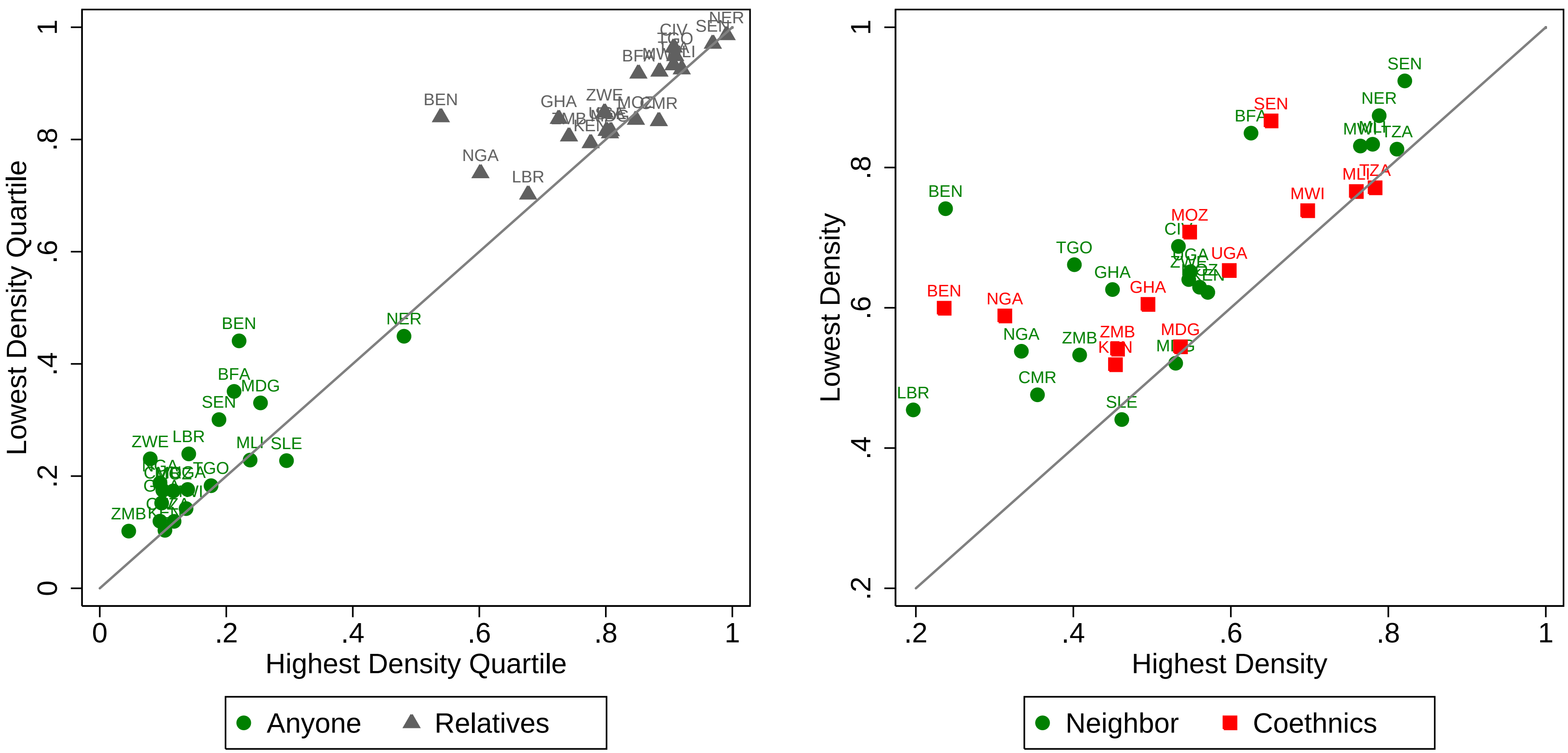


Table 6: Mobility

	Average	Quartile 1	Quartile 2	Quartile 3	Quartile 4
		Migrated in last 5 years			
DRC	15%	12%	22%	26%	13%
Ethiopia	11%	10%	9%	10%	17%
Ghana	30%	23%	25%	29%	38%
Kenya	29%	17%	19%	26%	34%
Liberia	29%	23%	26%	35%	38%
Madagascar	15%	13%	13%	16%	23%
Malawi	26%	21%	24%	22%	41%
Mali	16%	13%	14%	12%	23%
Nigeria	29%	20%	20%	23%	39%
Senegal	22%	19%	20%	25%	24%
SierraLeone	25%	15%	17%	25%	45%
Tanzania	36%	30%	30%	46%	49%
Uganda	28%	23%	24%	21%	40%
Zambia	36%	32%	33%	46%	43%
			Migrated from the countryside		
DRC	10%	9%	14%	23%	7%
Ethiopia	8%	9%	7%	7%	8%
Ghana	6%	9%	8%	6%	3%
Kenya	18%	14%	14%	18%	19%
Liberia	4%	4%	7%	1%	2%
Madagascar	9%	9%	9%	8%	9%
Malawi	17%	17%	19%	16%	17%
Mali	9%	8%	8%	6%	11%
Nigeria	10%	14%	12%	11%	7%
Senegal	9%	10%	12%	11%	5%
SierraLeone	8%	9%	8%	12%	8%
Zambia	15%	20%	17%	7%	4%
			Migrated from the city or town		
DRC	4%	3%	7%	3%	6%
Ethiopia	3%	2%	2%	2%	8%
Ghana	23%	13%	17%	23%	34%
Kenya	11%	3%	5%	8%	15%
Liberia	24%	18%	19%	34%	34%
Madagascar	6%	3%	4%	7%	14%
Malawi	9%	3%	5%	6%	24%
Mali	7%	5%	6%	5%	11%
Nigeria	18%	6%	8%	12%	31%
Senegal	11%	6%	7%	12%	17%
SierraLeone	16%	5%	9%	12%	36%
Zambia	21%	11%	16%	38%	39%

Appendix

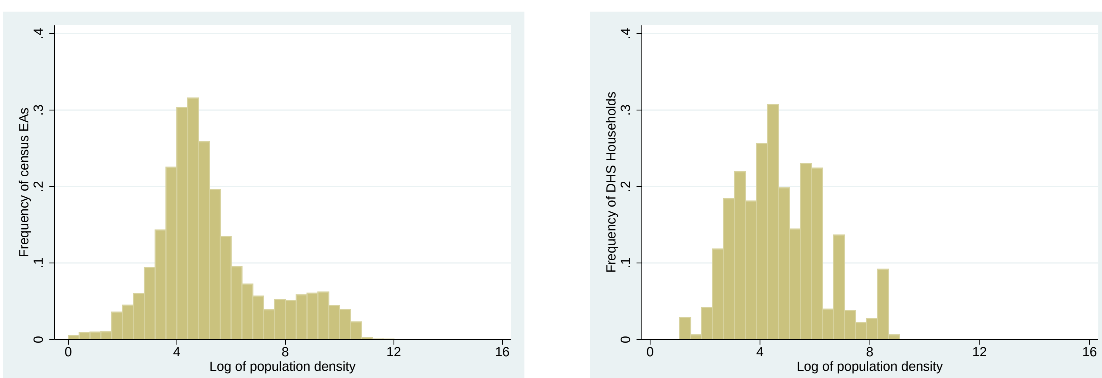
A. DHS data

1.1. DHS sample

While the DHS aim to make survey instruments and samples comparable across countries, the exact sampling differs according to the particular survey.[21] The target population of most DHS surveys are women aged 15-49 and children under the age of five living in residential households with the most common sampling following a two-stage cluster sampling procedure (DHS, 2013). If a recent census is available, the sampling frame of the census is used to define primary sampling units which are usually enumeration areas. Alternative sample frames include lists of electoral zones, estimated structures per pixel derived from high-resolution satellite imagery or lists of administrative units. Clusters will then be stratified depending on the number of domains that are desired for the particular survey, where a typical stratification is first at the geographical level and then at rural/urban clusters. In the first stage, from each of the strata a random sample of enumeration areas is selected inversely proportional to size. Unless a reliable listing of households exists, households will be listed for each of the selected primary sampling units. In the second stage, households are selected with equal probability.

For Tanzania we can compare the population density distribution of the DHS clusters with those of the overall population from the census data where we weight the population density of enumeration areas by the population. As is evident from Figure A.1, the DHS appears to capture a representative sample of the population with respect to density.

Figure A.1: Distribution of population and DHS clusters in Tanzania



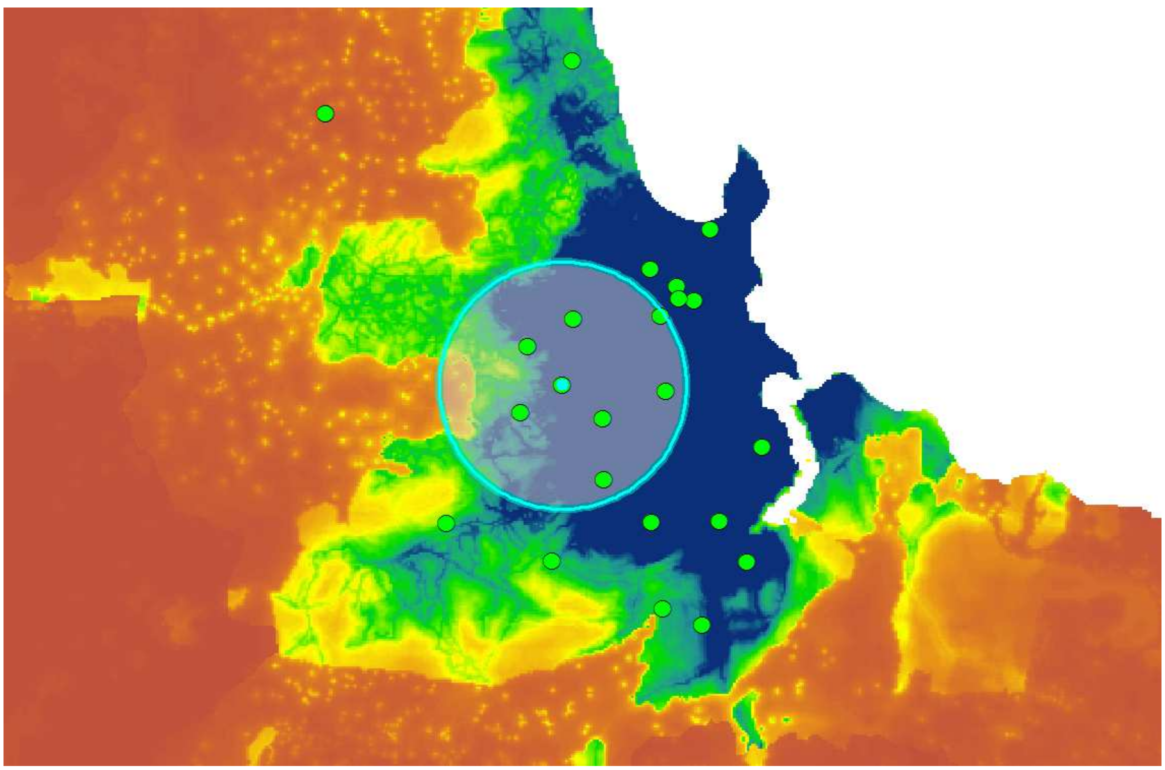
[21]For further information see: http://www.dhsprogram.com.

1.2. Linking DHS data with population density data

To link the DHS data with population density we draw a buffer of 5km around each cluster and extract the average population density around each cluster. We perform these calculations in WGS1984, since the different areas of the pixel sizes when moving away from the equator has been taken into account when constructing the population density grid, which is defined as the population count divided by the area. Many urban DHS clusters are in proximity closer than 5 km so that buffer polygons around clusters are overlapping. We therefore compute our zonal statistics using the Spatial Analyst Supplemental Tools in ArcGIS, a supplemental toolbox that allows computing zonal statistics for overlapping polygons. All computations were performed in ArcGIS 10.4.

1.3. Spatial linking of DHS

Figure A.2: DHS clusters in Dar Es Salaam



Notes: Figure shows a 5km circle around a DHS cluster in Dar Es Salaam; the gridded data come from WorldPop.

1.4. DHS Variables

The main variables to measure durables are whether the household has a television (hv208=1), mobile or landline (hv221=1 or hv243a=1).[22]. For housing consumption, we examine whether the household has electricity (hv206=1), tapped water (hv201<21), a constructed floor (hv213>13), a flush toilet (hv205<20), finished walls (hv214>=30) and a finished roof (hv215>=30). For the highest level of education completed by the household head we use variable hv109.

WHO (2015) which defines a minimum acceptable diet as follows:

"The composite indicator of a minimum acceptable diet is calculated from: (i) the proportion of breastfed children aged 6-23 months who had at least the minimum dietary diversity and the minimum meal frequency during the previous day and (ii) the proportion of non-breastfed children aged 6-23 months who received at least two milk feedings and had at least the minimum dietary diversity not including milk feeds and the minimum meal frequency during the previous day.

Dietary diversity is present when the diet contained four or more of the following food groups:

- grains, roots and tubers;
- legumes and nuts;
- dairy products (milk, yogurt, cheese);
- flesh foods (meat, fish, poultry, liver or other organs);
- eggs;
- vitamin A-rich fruits and vegetables; and
- other fruits and vegetables.

The minimum daily meal frequency is defined as:

- twice for breastfed infants aged 6-8 months;
- three times for breastfed children aged 9-23 months;
- four times for non-breastfed children aged 6-23 months."

[22]For Liberia there is no data on whether the household has a landline

To compute a child's minimum dietary diversity from the DHS we use data on breastfeeding (m4), age of the child (hw1), number of times the child ate (m39), the type of food groups the child consumed (v414a-v414j), consumption of powdered/tinned/fresh milk and infant formula (469e and 469f). If the survey is from Phase 5 there is no information on milk feedings for children who are not breastfed anymore; in this case we only calculate the minimum acceptable diet for children who are breastfed.

Migration status is determined using the years lived in the current location (v104) and type of place of previous residence which is classified into capital, large city; city; town; countryside; and abroad (v105).

Our data for malaria incidence comes from a combination of DHS, MIS and AIS. In a subset of countries, blood samples were collected for children aged 6-59 months as part of the DHS data collection in households that were selected for the men's questionnaire (every 1 out of 8 households). If the data is not available as part of the DHS, we use data from the most recent geo-referenced AIS survey or MIS survey during which blood samples from children aged 0-59 or 6-59 months were taken depending on the specific survey. Table 2 shows the exact survey used for each country. Malaria tests were administered via rapid diagnostic testing and blood smear microscopy. [23] We construct a dummy variable that is equal to one if the malaria rapid test for a child was positive (hml25).

B. LSMS data

[TO DO][24]

C. Afrobarometer

3.1. Geo-locating Afrobarometer respondents

Afrobarometer surveys collect data on attitudes towards democracy and governance, as well as a range of other measures of the quality of life.[25] The Afrobarometer surveys do not coordinates of respondents, but record the village, district and region names. The 2011 round provides four different administrative names. We use a matching algorithm that matches village names and other provided administrative names to locations as listed in gazetteers; specifically, we follow Nunn and Wantchekon (2011) and use the geonames gazeteer available on www.geonames.com. This website provides a list of locations where each location is assigned an id along with several

[23]Note that in Madagascar areas without malaria have been excluded from the survey.
[24]For further information see http://iresearch.worldbank.org/lsms.
[25]For further information see http://www.afrobarometer.org.

names: the geographical name of the point in utf8 and plain ascii characters; alternative names, the associated latitude and longitude coordinate. There is also auxiliary information such as the modification date of each entry, administrative codes, elevation, and feature classes. If a name is associated with several entries we keep the most recent entry.

Our matching algorithm uses a mixture of exact matches and fuzzy matches in multiple stages (depending on the survey round, between thirteen and twenty-one). Whenever a location name is identified, we assign it the latitude and longitude and remove it from the dataset that is fed into the next stage. In essence, matching is achieved in the following way: first, we perform a series of exact matches based on the village name from Afrobarometer with the asciiname listed in the gazetteer; if there are no exact matches with the village name and the asciiname, I search through the next four alternative names listed in the gazeteer for the specific location. In this first stage we find almost forty percent of locations. We then use the most precise administrative classification. For example, if the data set has information on the village name, district and region, this would be the district. We perform the exact same series of matches on the district name, using again the asciiname as well as four possible alternative names listed in the gazetteer. In rounds three and four of the survey in which we have only district and region names in addition to the village names, this step finds 49–52 percent of the locations.

Third, we match on the region name which finds another 4–6 percent of the sample. Finally, to catch any remaining misspellings we perform a fuzzy match based on similar text patterns between the village name and the asciiname using a command developed by Raffo (2015). We use a similarity score of above 0.70 and a vectorial decomposition algorithm (3-gram). This finds another 1–3 percent of locations. In total, we are able to match between 92 and 95 percent of individuals in each round.

In addition to random checks of the identified locations we use the 2005 data to check the consistency between our algorithm and Nunn and Wantchekon (2011)'s location data. For the subset of locations for which they provide geo-locations, we find that the median distance between their location and my location is 12.46 km. Further, considering the population density data vary largely at the district and region level, we expect the difference to be even smaller when looking at the resulting population densities. Indeed, the correlation coefficient between the population density from their and our data is 0.65 with a p-value of 0.000.

We use variables related to fear of crime in own home (q9a), theft (q9b), physical attack (q9c); trust in general (q83), towards relatives (q84a) and towards neighbors (q84b); frequency of lack of food (q8a) and medicine (q8c); and anxiety (q96b).

Table A.1: Afrobarometer Sample

	Individuals	Round 3	Round 4	Round 5
Benin	3550	x	x	x
Burkina Faso	2112		x	x
Cameroon	944			x
Cote D'Ivoire	1168			x
Ghana	4037	x	x	x
Kenya	4573	x	x	x
Liberia	2204		x	x
Madagascar	3880	x	x	x
Malawi	4768	x	x	x
Mali	3667	x	x	x
Mozambique	4565	x	x	x
Niger	1151			x
Nigeria	6836	x	x	x
Senegal	2400		x	x
Sierra Leone	1039			x
Tanzania	4505	x	x	x
Togo	800			x
Uganda	7192	x	x	x
Zambia	3600	x	x	x
Zimbabwe	3016		x	x

Notes: Column (2) shows the number of individuals in my sample for each of the countries; columns (3)–(5) indicate when a country was added to the Afrobarometer sample. Round 3 took place in 2005, round 4 in 2008, and round 5 in 2011.

D. Pollution data

Our pollution data for PM2.5 and NO2 concentrations come from van Donkelaar, Martin, Brauer, and Boys (2015) and Geddes, Martin, Boys, and van Donkelaar (2015), respectively. As the date for the Gridded Population of the World v4 (GPWv4) data is approximately 2010, we take the pollution measures that are closest in time: the tri-annual mean (2009-2011) both for PM2.5 series; for NO2 we have the exact year 2010. The estimated PM2.5 and NO2 concentrations are available at a resolution of 0.1 decimal degrees (about 10km at the equator). We construct a fishnet of the same resolution and for each pixel compute the average pollution measure as well as the average population density from the GPWv4. PM2.5 is measured in $\mu g/m^3$ while NO2 is measured in ppb (parts per billion). Following Vrijheid, Martinez, Manzanares, Dadvand, Schembari, Rankin, and Nieuwenhuijsen (2011), we use a conversion of 1ppb= 1.88 f $\mu g/m^3$ which assumes ambient pressure of 1 atmosphere and a temperature of 25 degrees celsius.

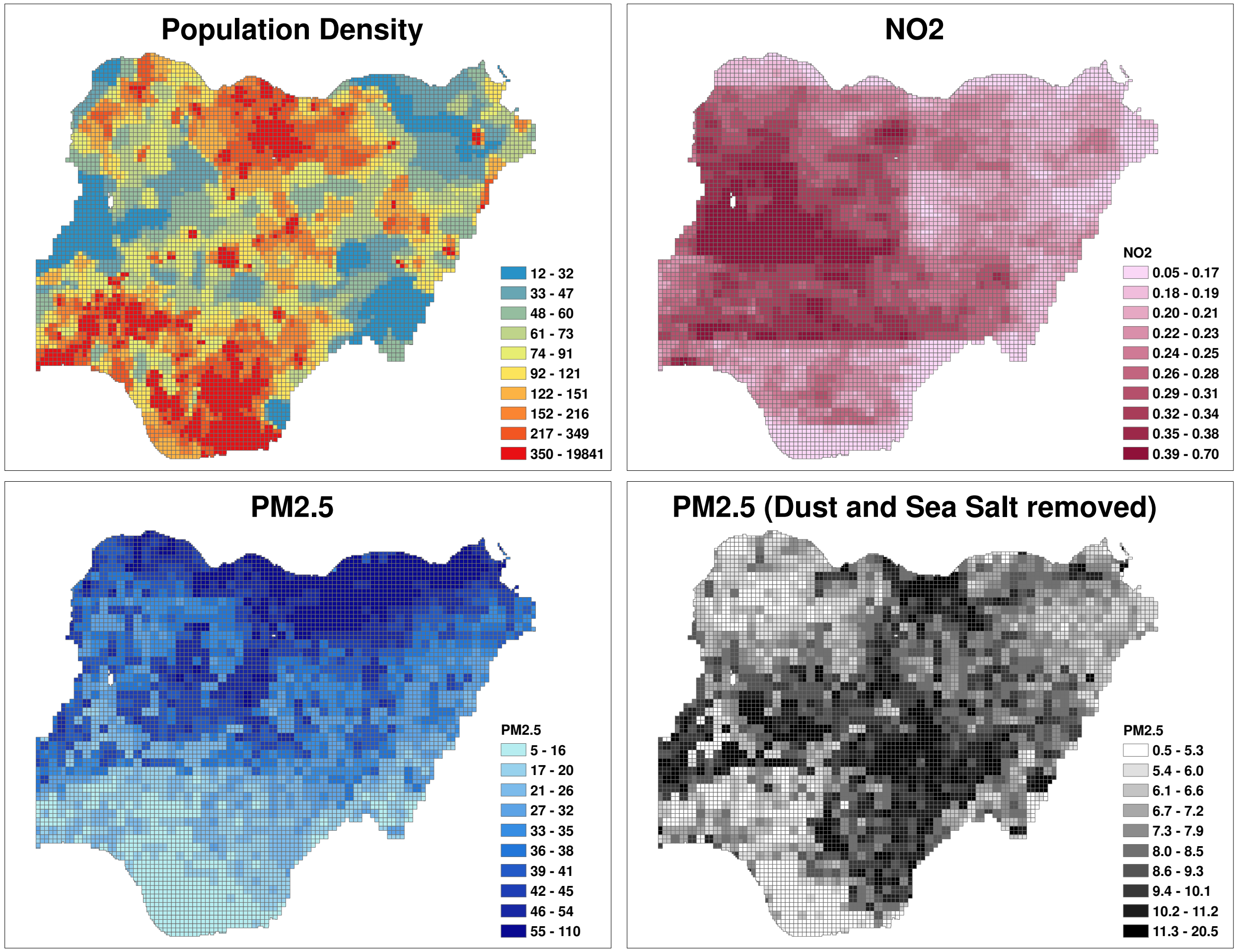
Figure A.3 illustrates this procedure and shows the distributions of pollutants and population density across space in Nigeria. The top left graph shows the distribution of population density, the top right graph shows the NO2 distribution, and the two bottom graphs show PM2.5, where the graph on the right removes sea salt and dust. Warmer colors denote higher values, and the bins are formed by dividing the data into deciles. Population density in the North is highest around Kano; in the center around Abuja; in the South West close to Lagos and Ibadan; and in the South East between Benin City, Port Hartcourt and Enugu.

Moving to the pollution measures, several observations are worth highlighting: first, at least visually, population density does not appear to be strongly correlated with either of the pollution measures. Nitrogen dioxide levels are very low, with a maximum of 0.7 ppb (1.316 $\mu g/m^3$), far below the WHO recommended thresholds of 40 $\mu g/m^3$. Values are higher over Lagos, Ibadan, Abuja, Kanduna and Kano, but not over cities in the South East in the Delta, and there are high levels in the center towards the West of the country where few people live. PM2.5 levels appear to be mainly driven by dust of the Sahara when inspecting the bottom left graph. Removing sea salt and dust produces quite a different distribution, with higher levels in the center, and over some cities.

The Nigerian example illustrates further that looking separately at these two indicators for pollution is instructive.[26] The pairwise correlation between PM2.5 and NO2 is -0.0085 with a p-value of 0.4633. Across our whole set of African countries, the correlation of these two measures ranges from 0.65 (Cameroon) to -0.47 (Senegal).

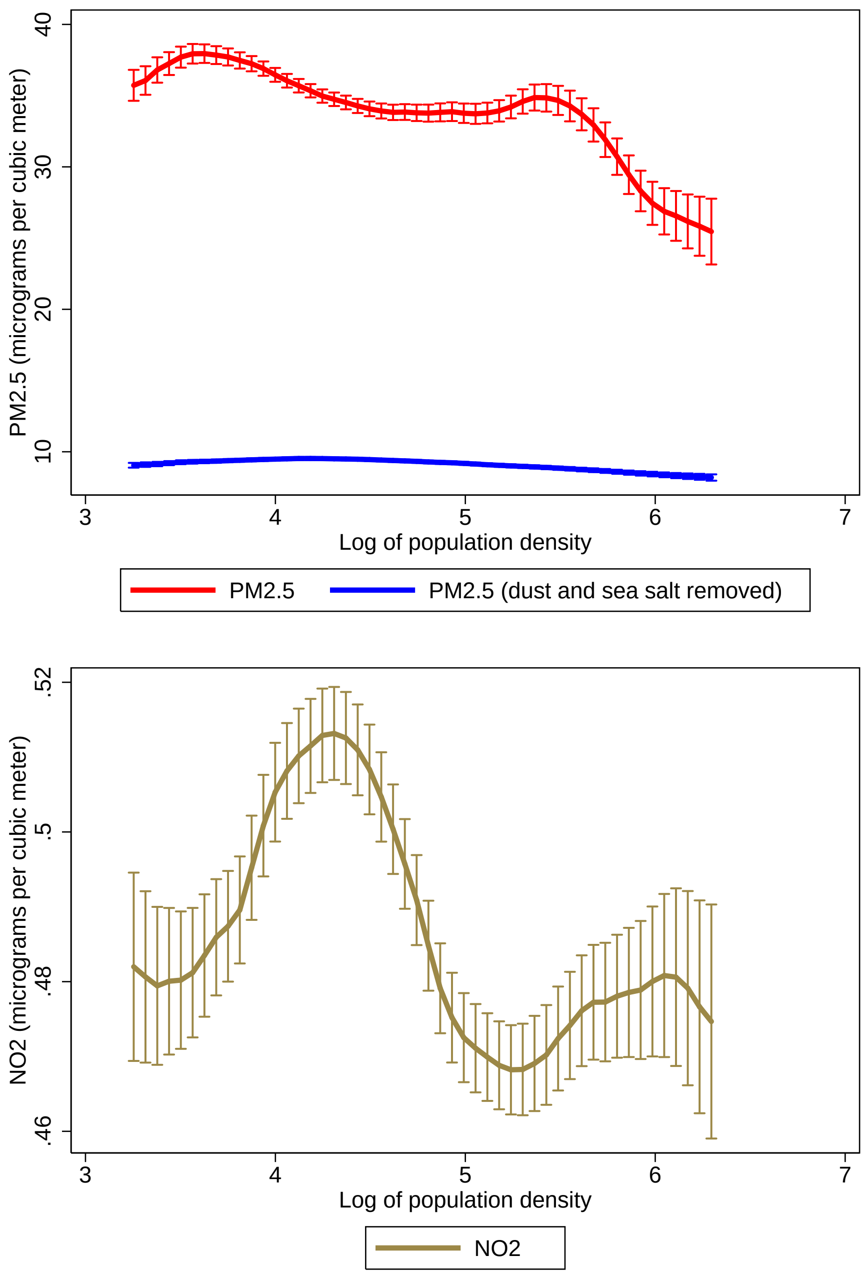
[26]This is in line with what Geddes, Martin, Boys, and van Donkelaar (2015) find when they inspect population weighted average PM2.5 and NO2 levels and trends.

Figure A.3: Pollution in Nigeria



Notes: The top left graph shows the distribution of population density, the top right graph shows the NO2 distribution, and the two bottom graphs show PM2.5, where the graph on the right removes sea salt and dust. Warmer (darker) colors denote higher values, and the bins are formed by dividing the data into deciles. Population density in the North is highest around Kano; in the center around Abuja; in the South West close to Lagos and Ibadan; and in the South East between Benin City, Port Hartcourt and Enugu. At least visually, population density does not appear to be strongly correlated with either of the pollution measures. Nitrogen dioxide levels are very low, with a maximum of 0.7 ppb (1.316 $\mu g/m^3$), far below the WHO recommended thresholds of 40 $\mu g/m^3$. Values are higher over Lagos, Ibadan, Abuja, Kanduna and Kano, but not over cities in the South East in the Delta, and there are high levels in the center towards the West of the country where few people live. PM2.5 levels appear to be mainly driven by dust from the Sahara when inspecting the bottom left graph. Removing sea salt and dust produces quite a different distribution, with higher levels in the center, and over some cities.

Figure A.4: Pollution and population density in Nigeria



Notes: The figure shows a kernel-weighted local polynomial regression of the level of pollution on the log of population density in Nigeria using data from the entire country, and plotting 95 percent confidence intervals. The top panel shows the results for PM2.5 and the bottom panel shows NO2 levels across population density space. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

E. Further tables

Figure A.5: Crime - LSMS

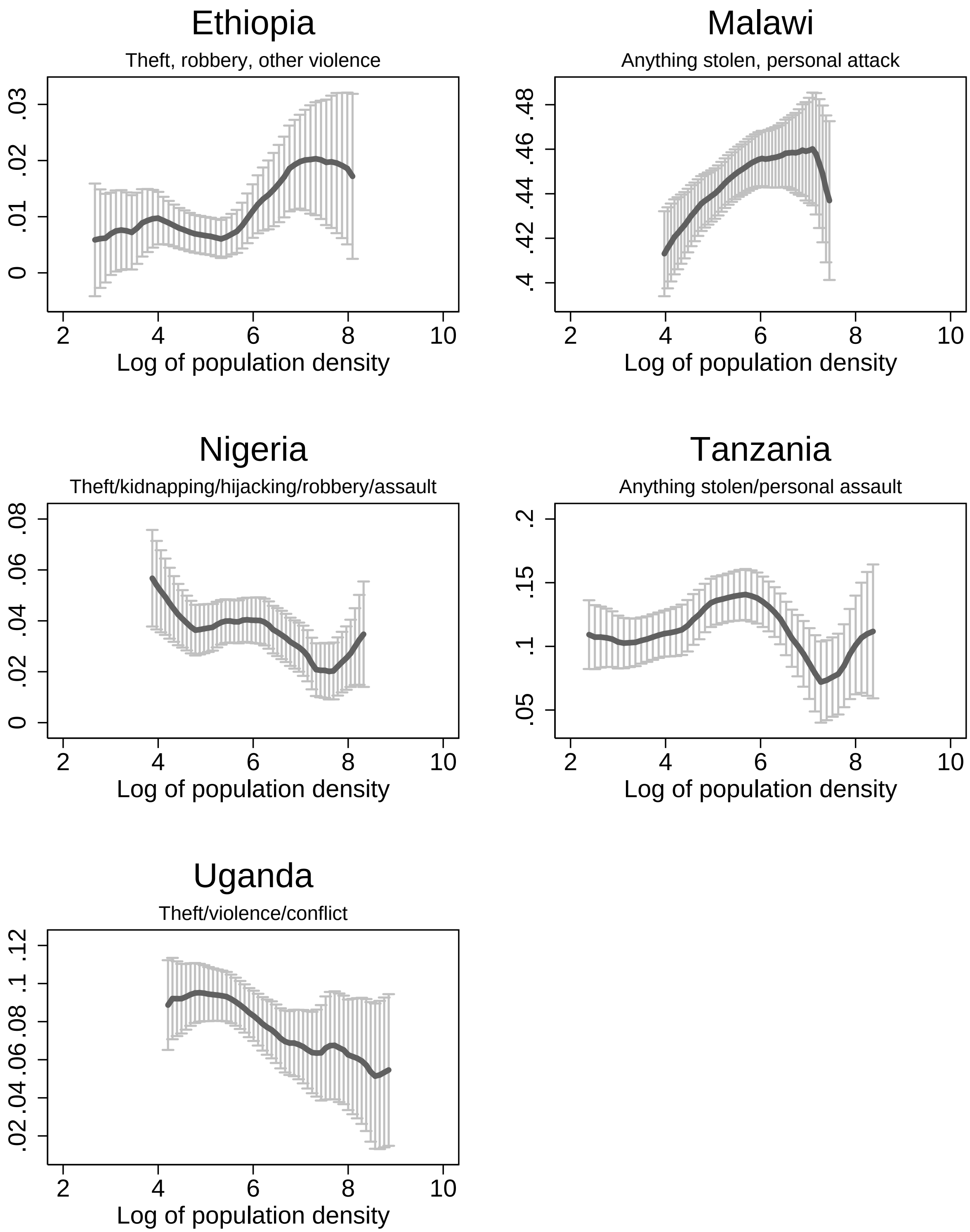
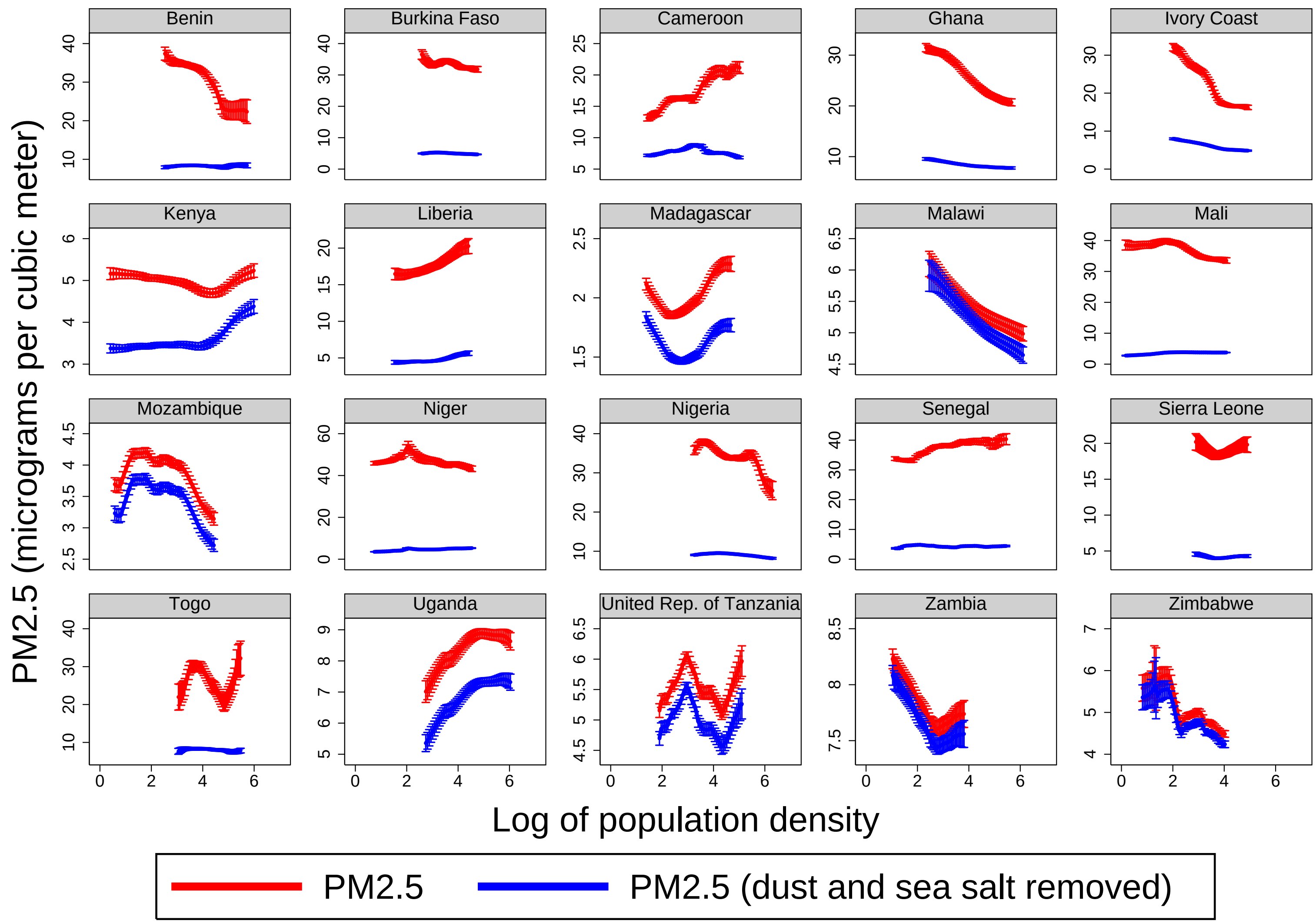
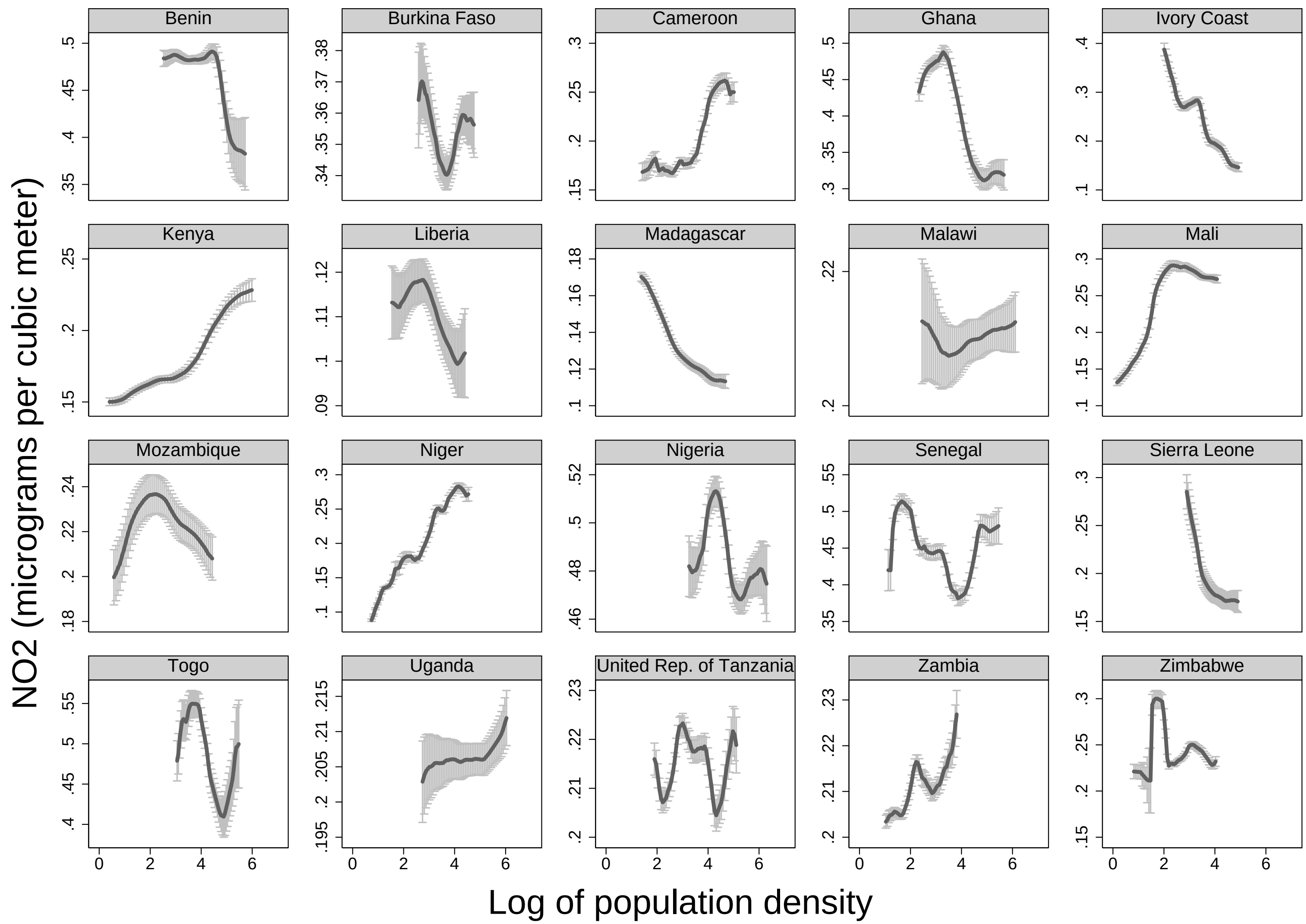


Figure A.6: PM2.5 concentration across Africa- different scales



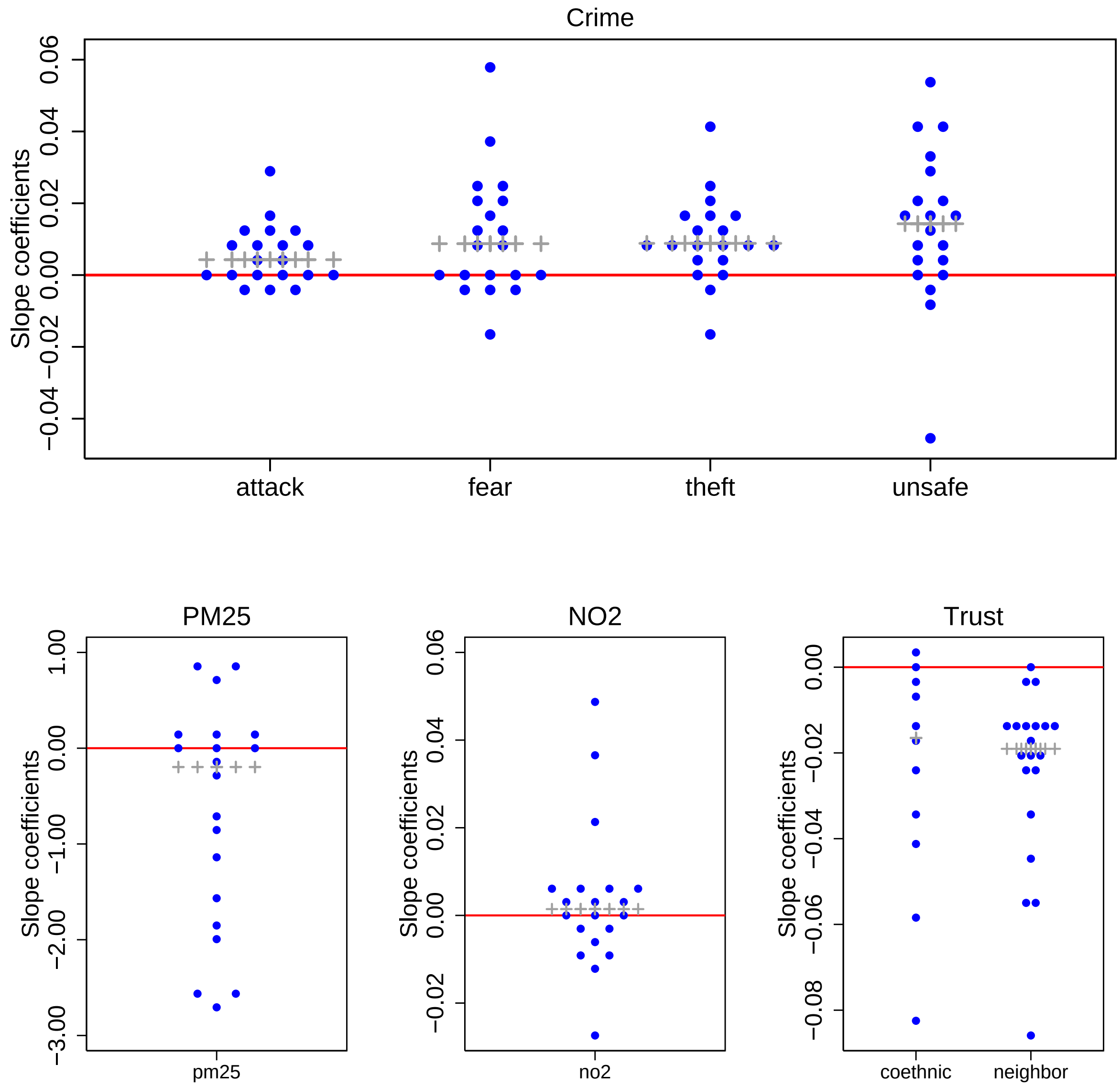
Notes: The figure shows a kernel-weighted local polynomial regression of the level of PM2.5 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

Figure A.7: Nitrogen dioxide concentration across Africa - different scales



Notes: The figure shows a kernel-weighted local polynomial regression of the level of NO2 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

Figure A.8: Slope Coefficients



Notes: The figure presents the coefficients of estimating the following equation for individual i at location j at time t

$$\theta_{ijt} = \alpha + \gamma_t + \beta \ln d_{ij} + \varepsilon_{ijt}$$

where θ captures different measures for disamenities: crime, pollution and mistrust; and d represents population density. All models include survey round fixed effects γ_t. I estimate these regressions for each of the countries separately, so that every dot represents a coefficient β for a specific country. The upper panel shows the four variables on crime, and the lower panel show the results for pollution and trust. The red line indicates zero, and the grey crosses show the median coefficient.

F. Sub-graphs for all dimensions of living standards by country

Figure A.9: Phone

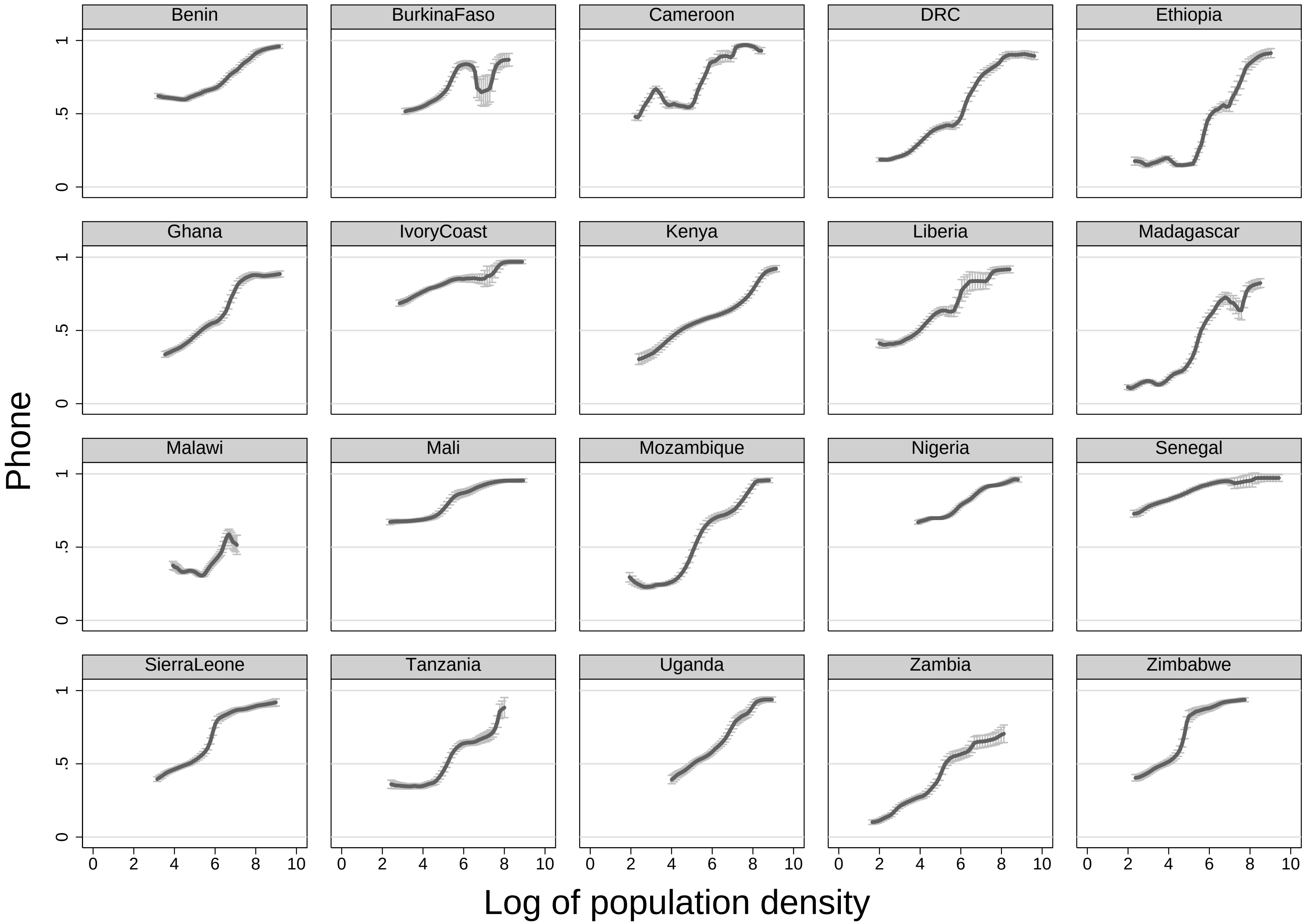


Figure A.10: TV

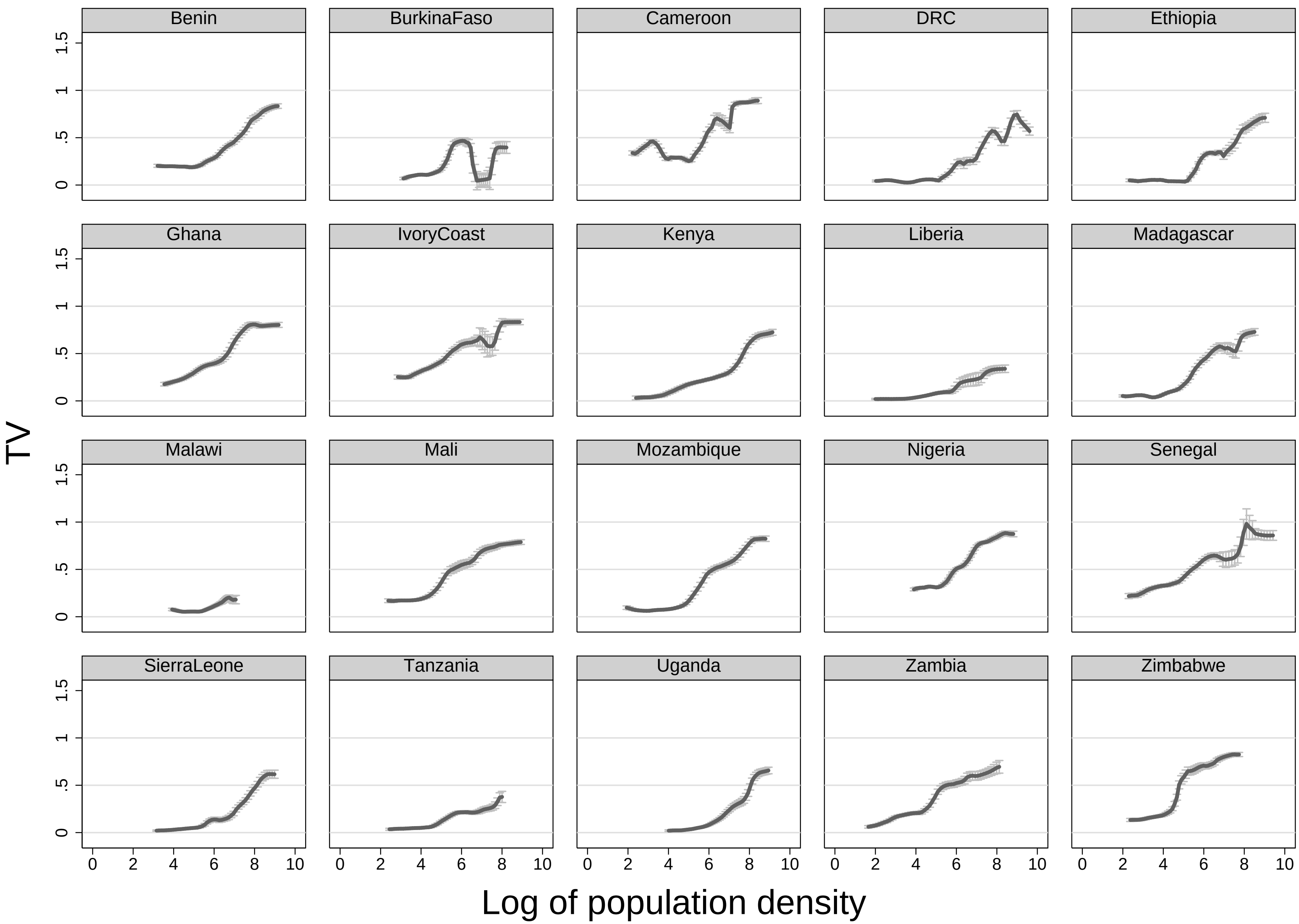


Figure A.11: Stunted (Height-for-age z-score<-2)

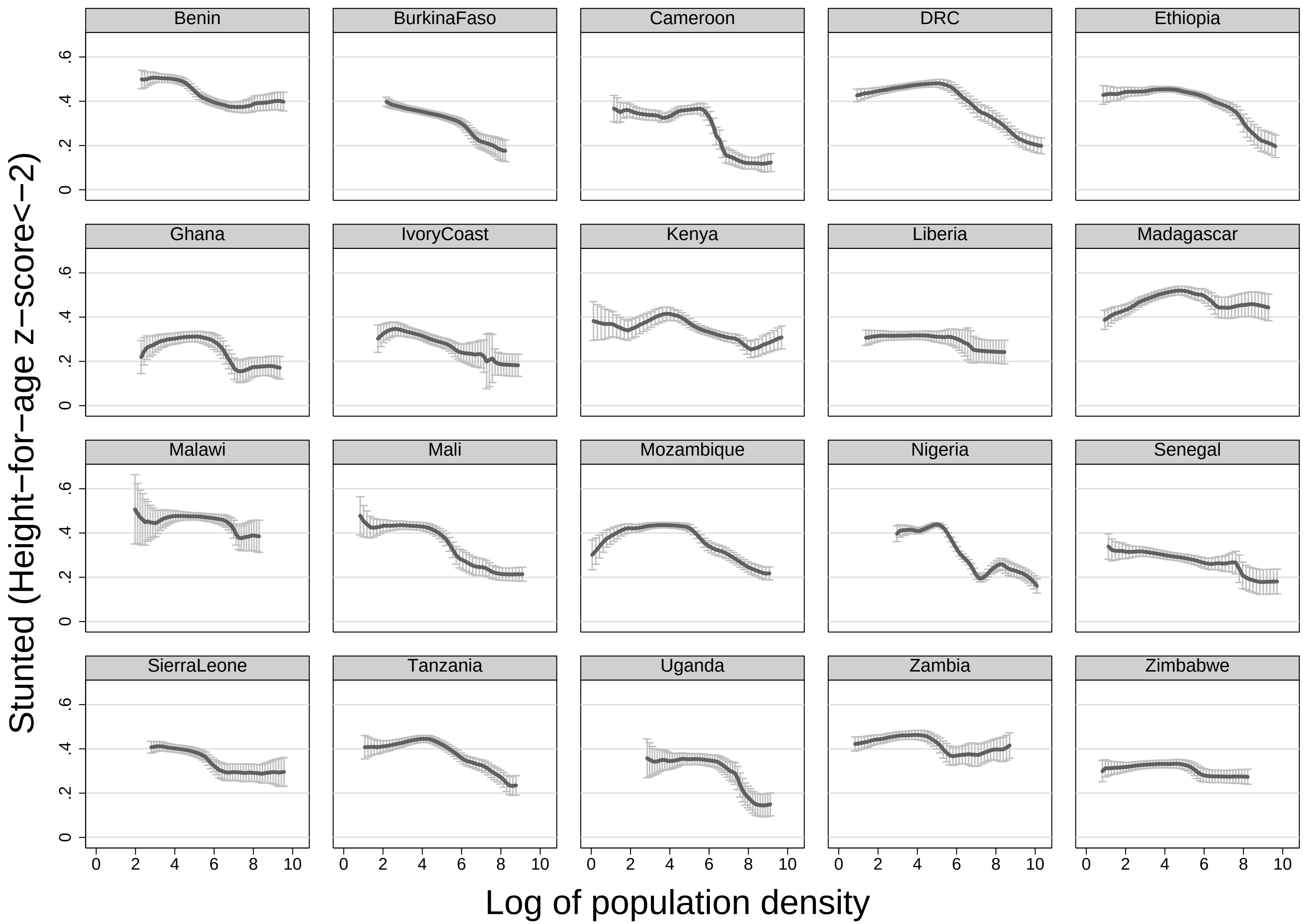


Figure A.12: Wasted (Weight-for-height z-score<-2)

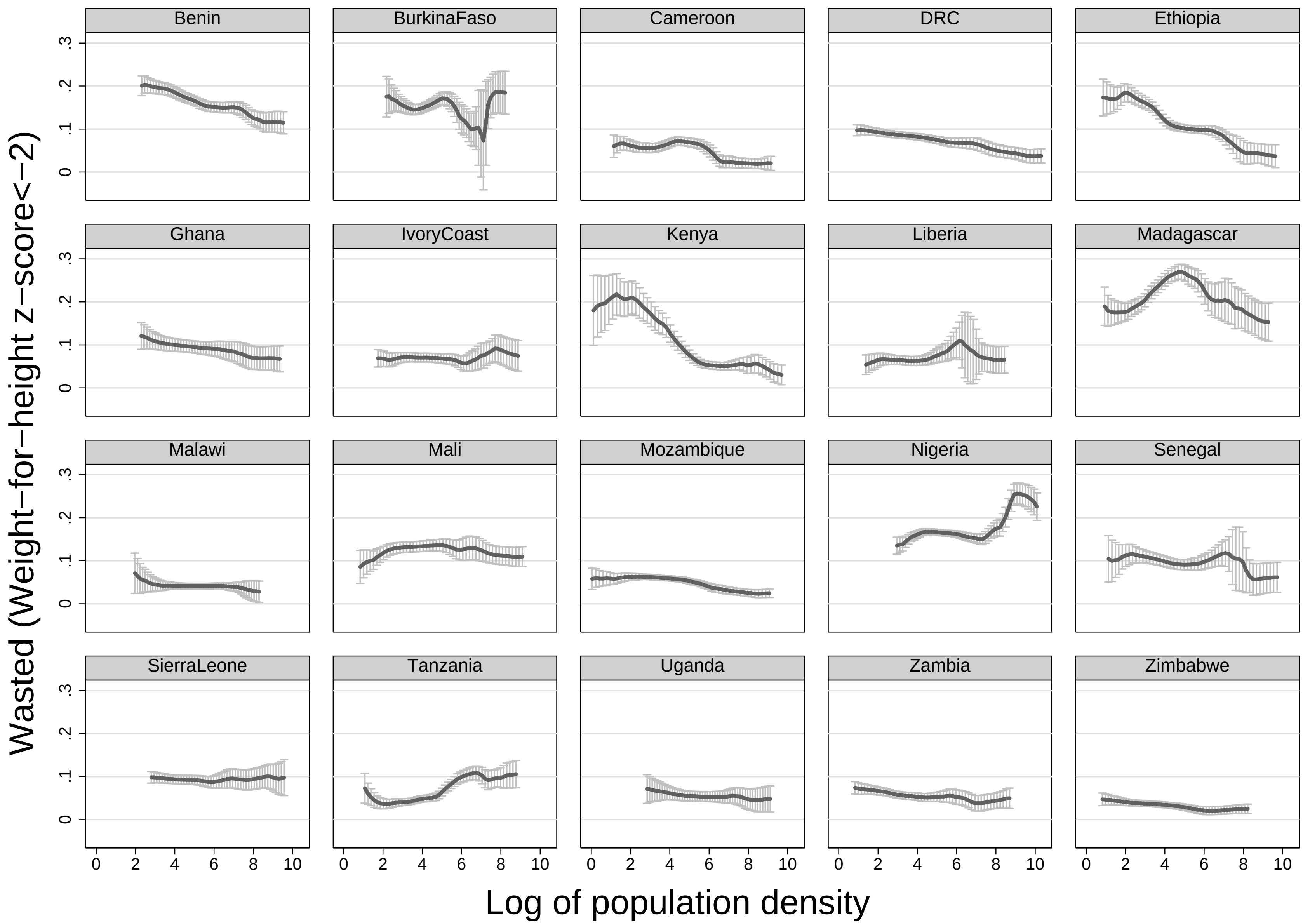


Figure A.13: Anemic (hgb<11 mg/l)

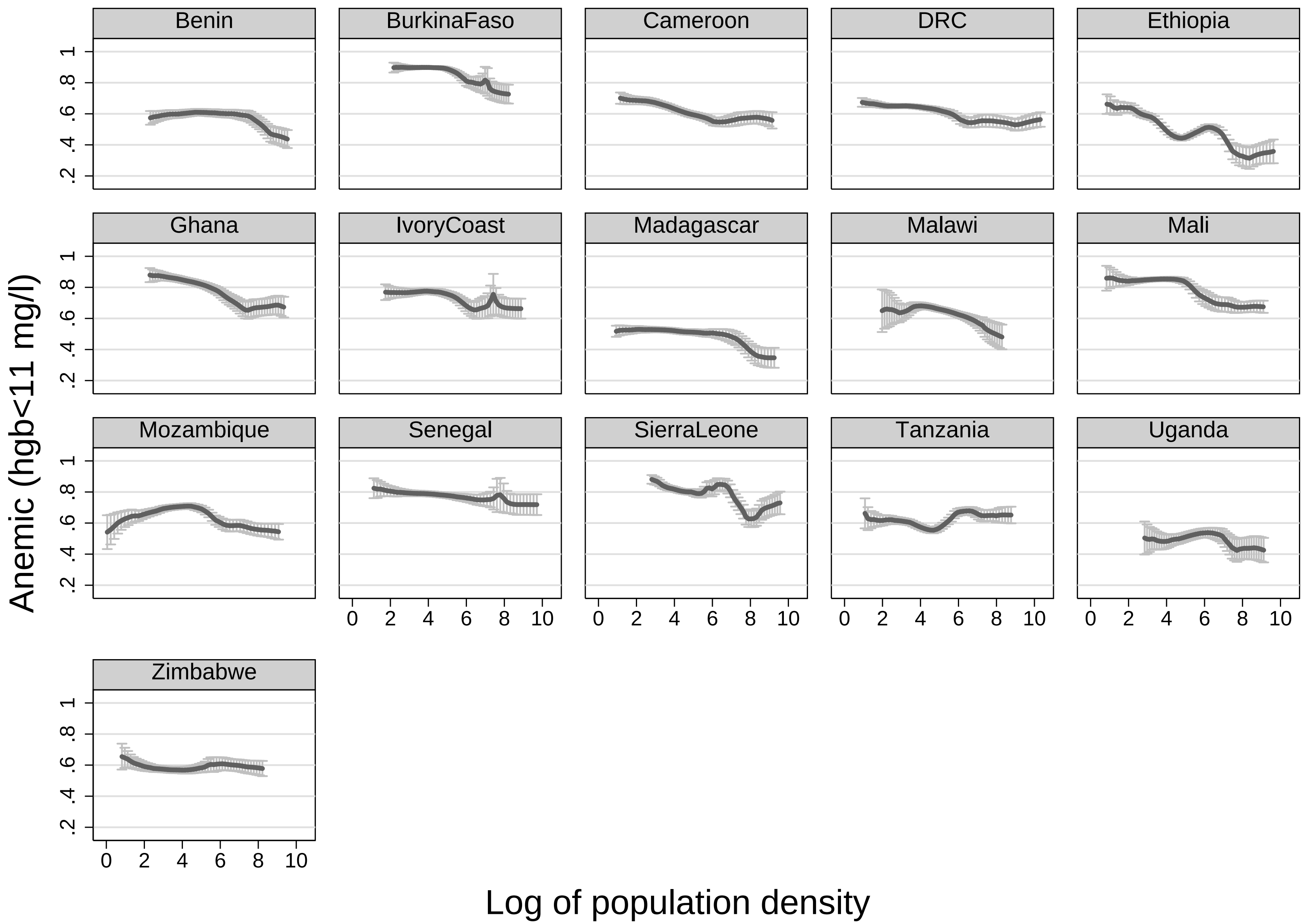


Figure A.14: Below minimum acceptable diet

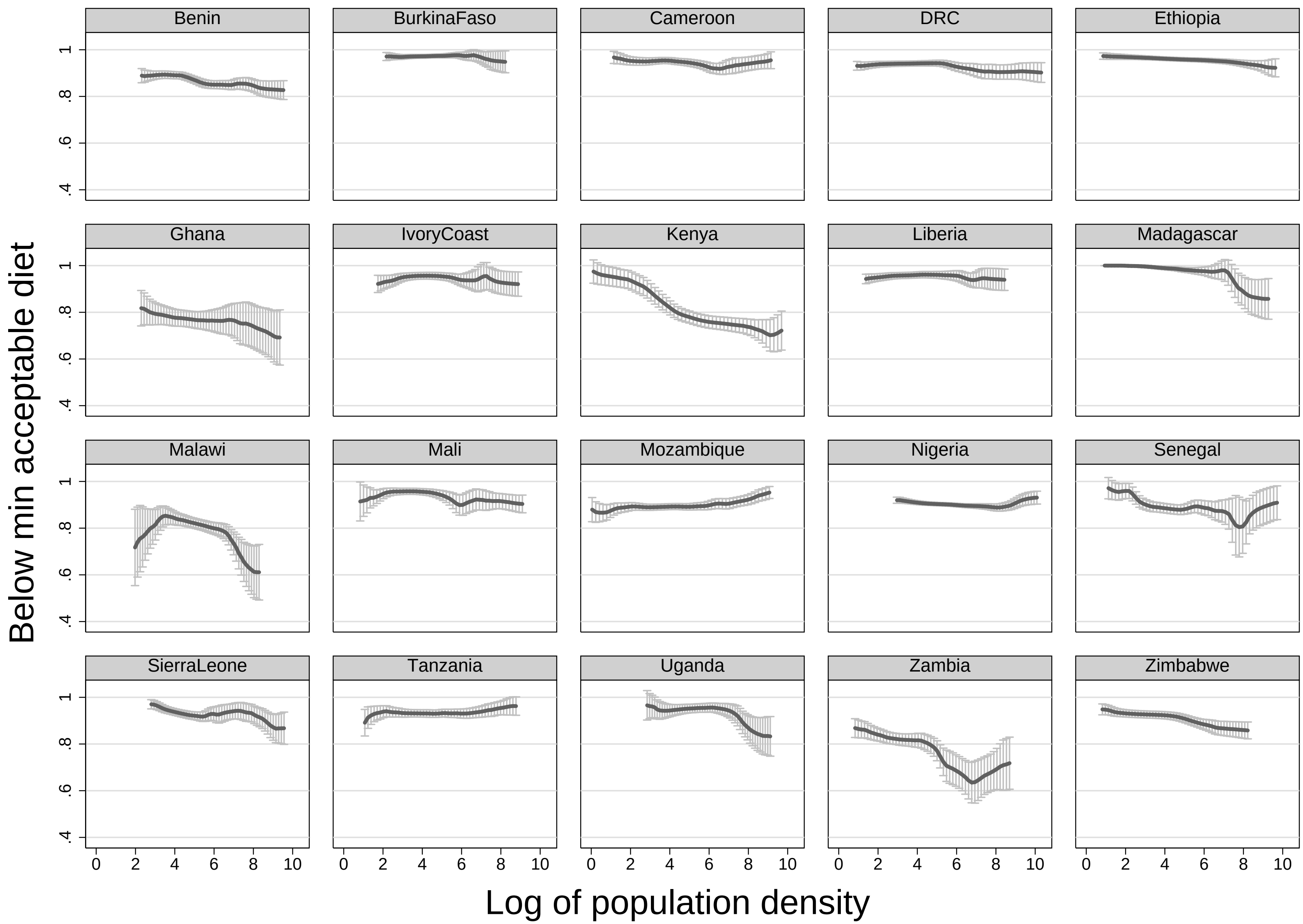


Figure A.15: Finished Roof

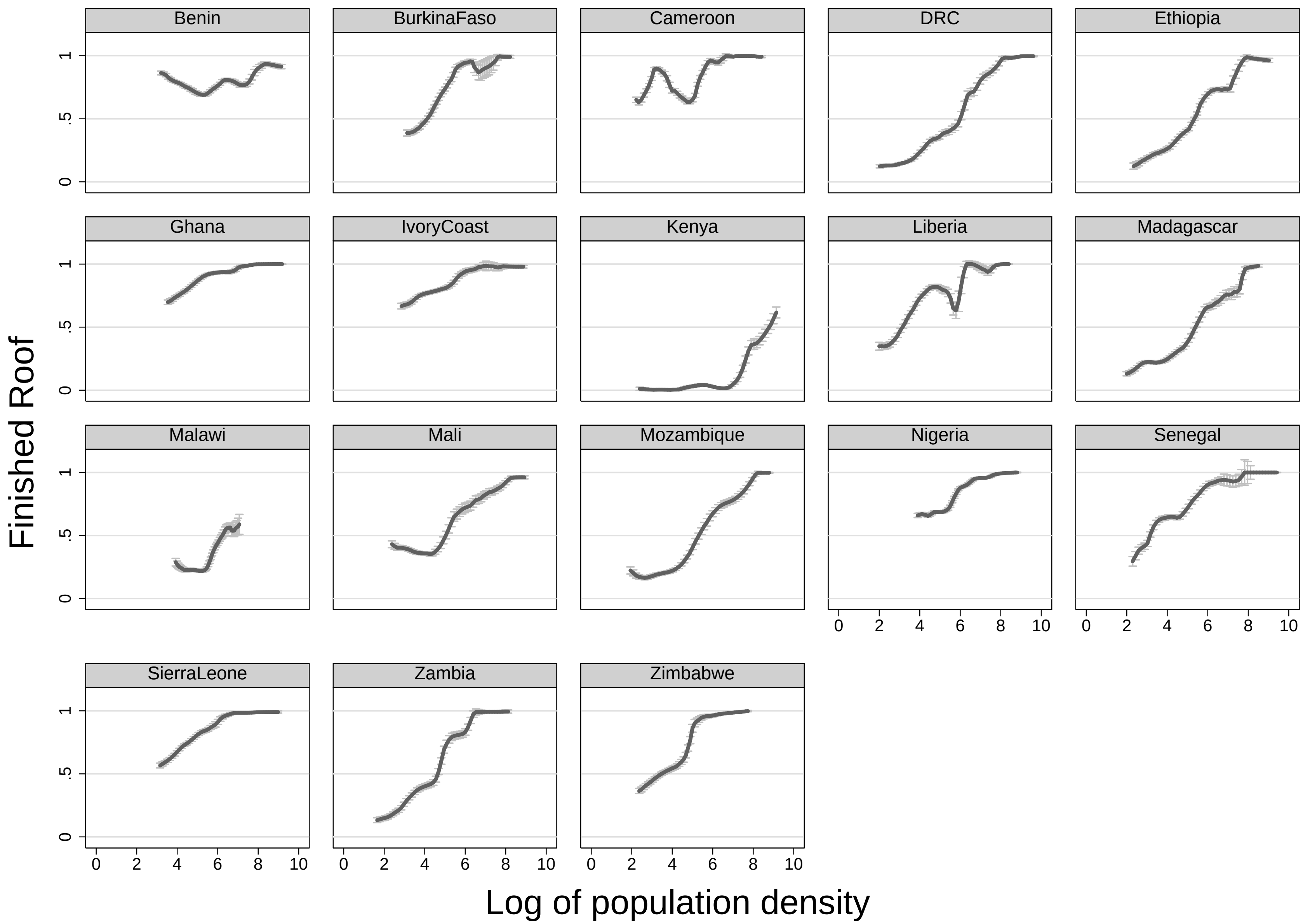


Figure A.16: Constructed Floor

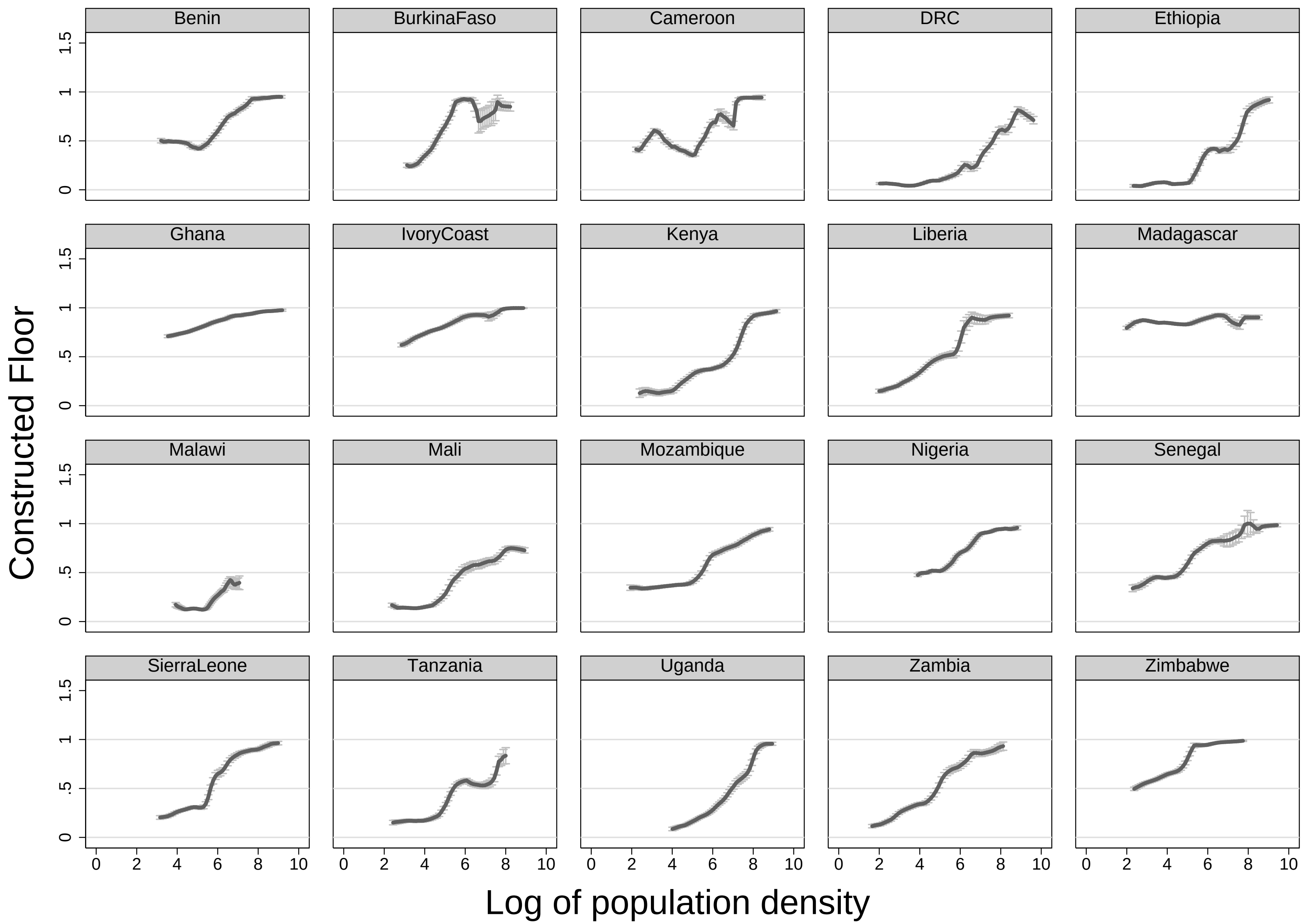


Figure A.17: Finished Walls

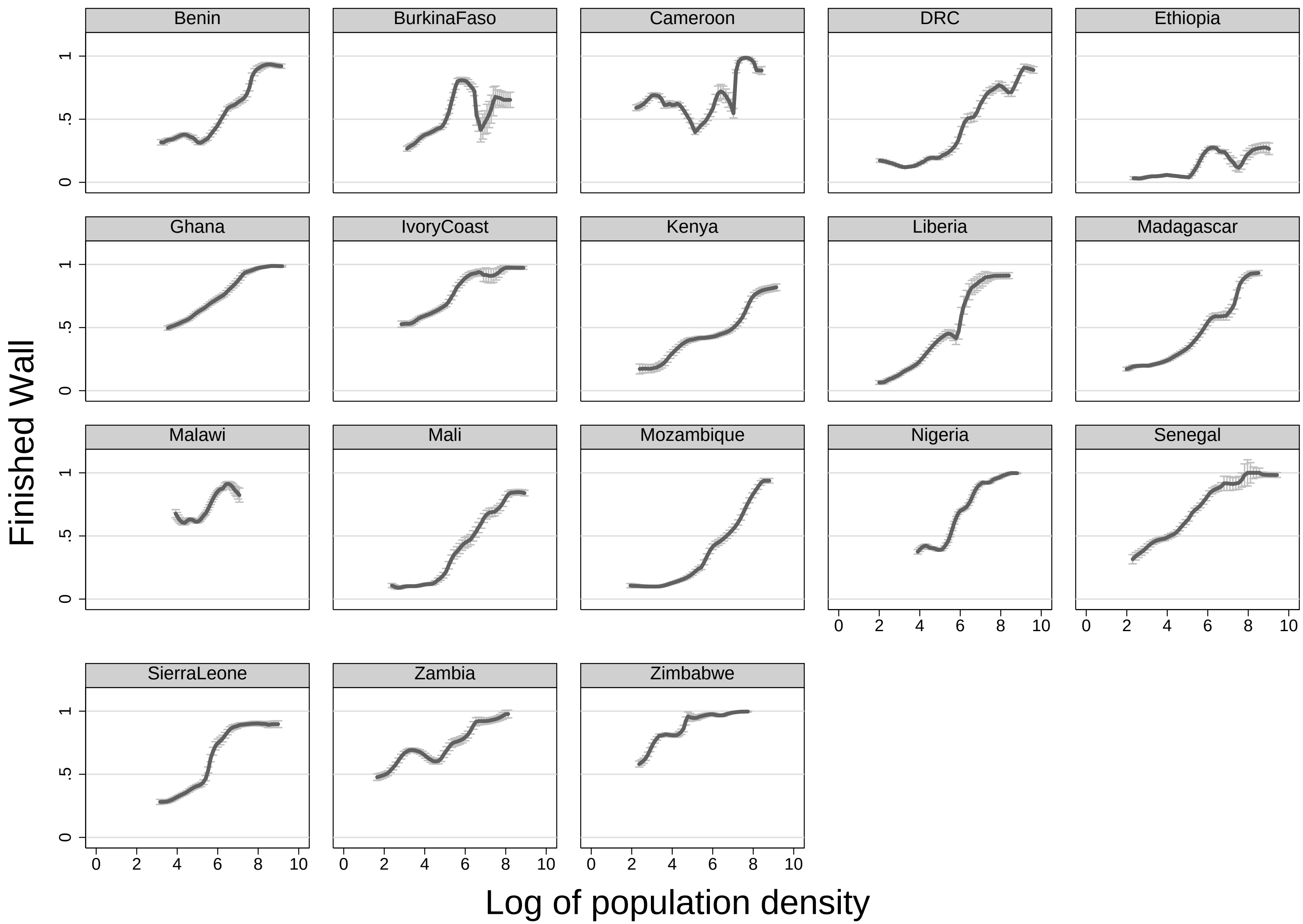


Figure A.18: Water from Tap

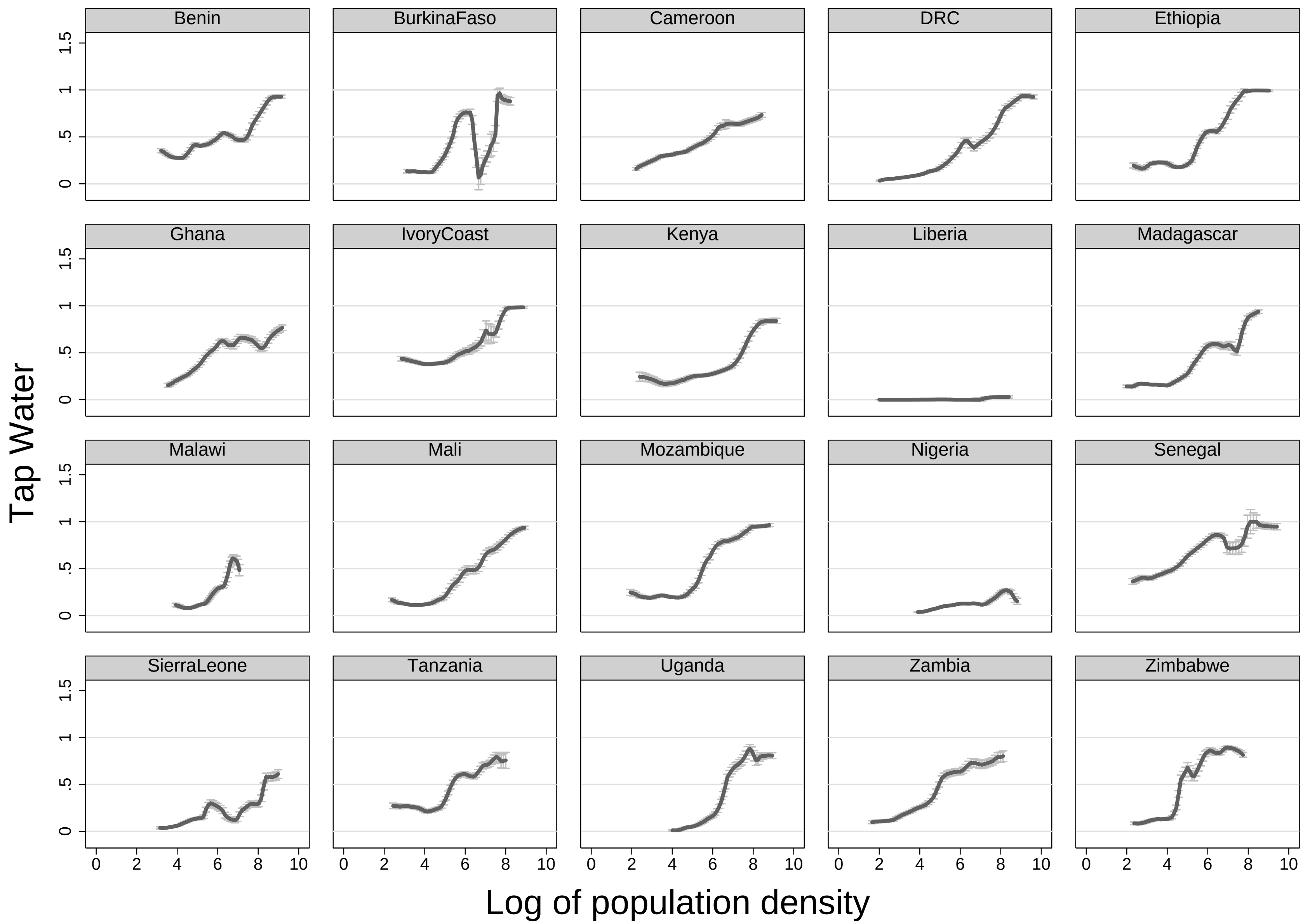


Figure A.19: Flush Toilet

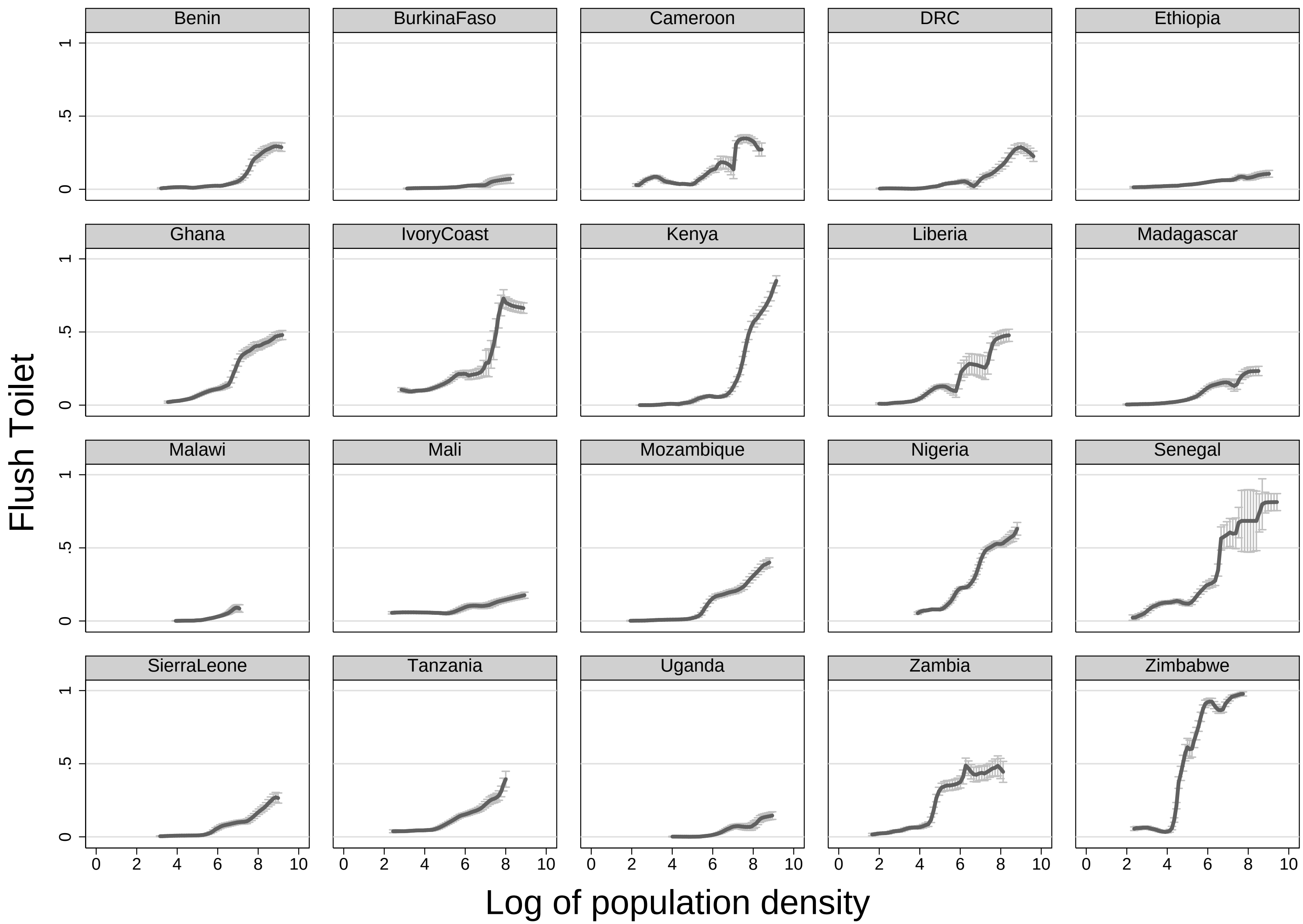


Figure A.20: Electricity

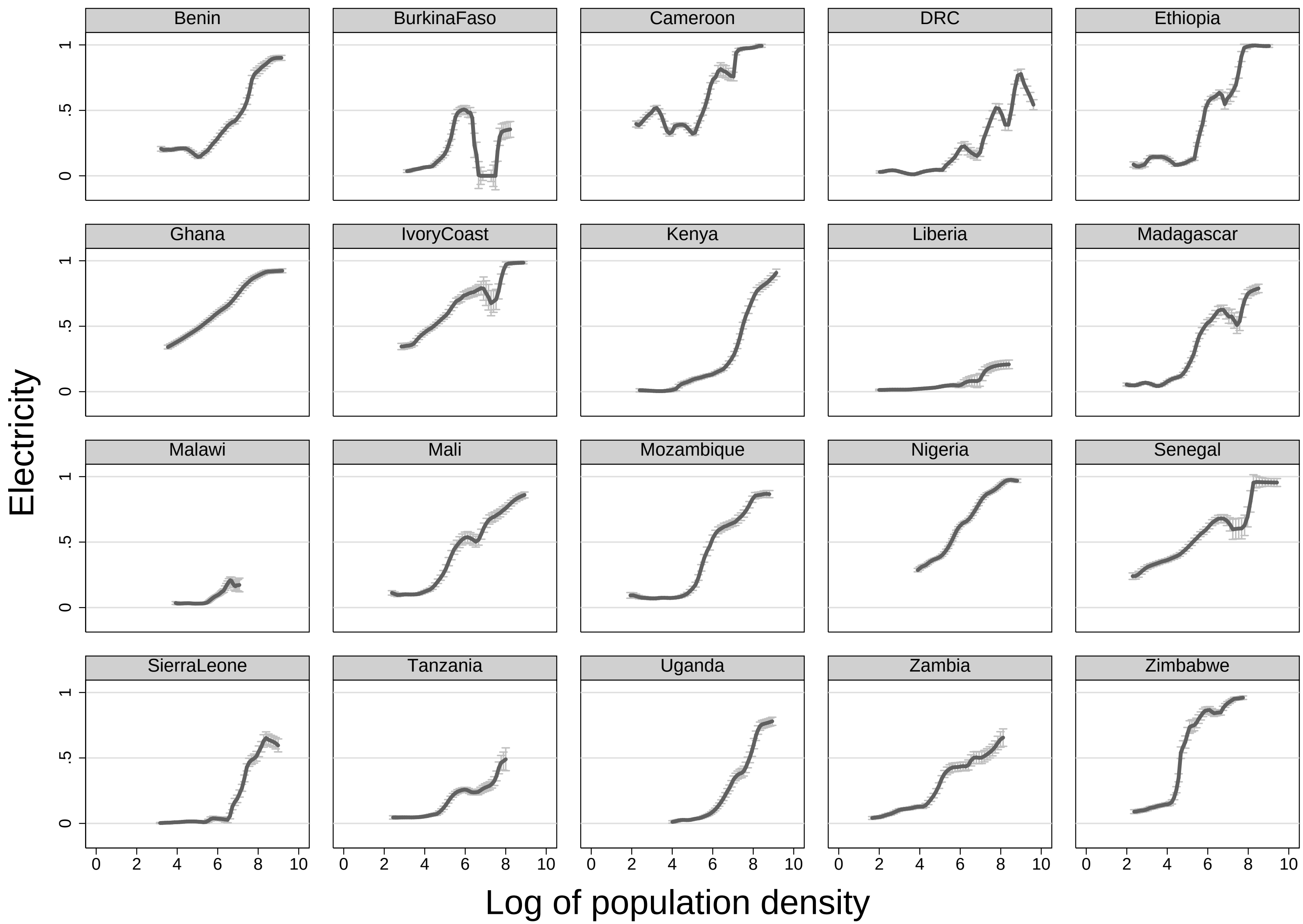


Figure A.21: Malaria of children aged 6-59 months

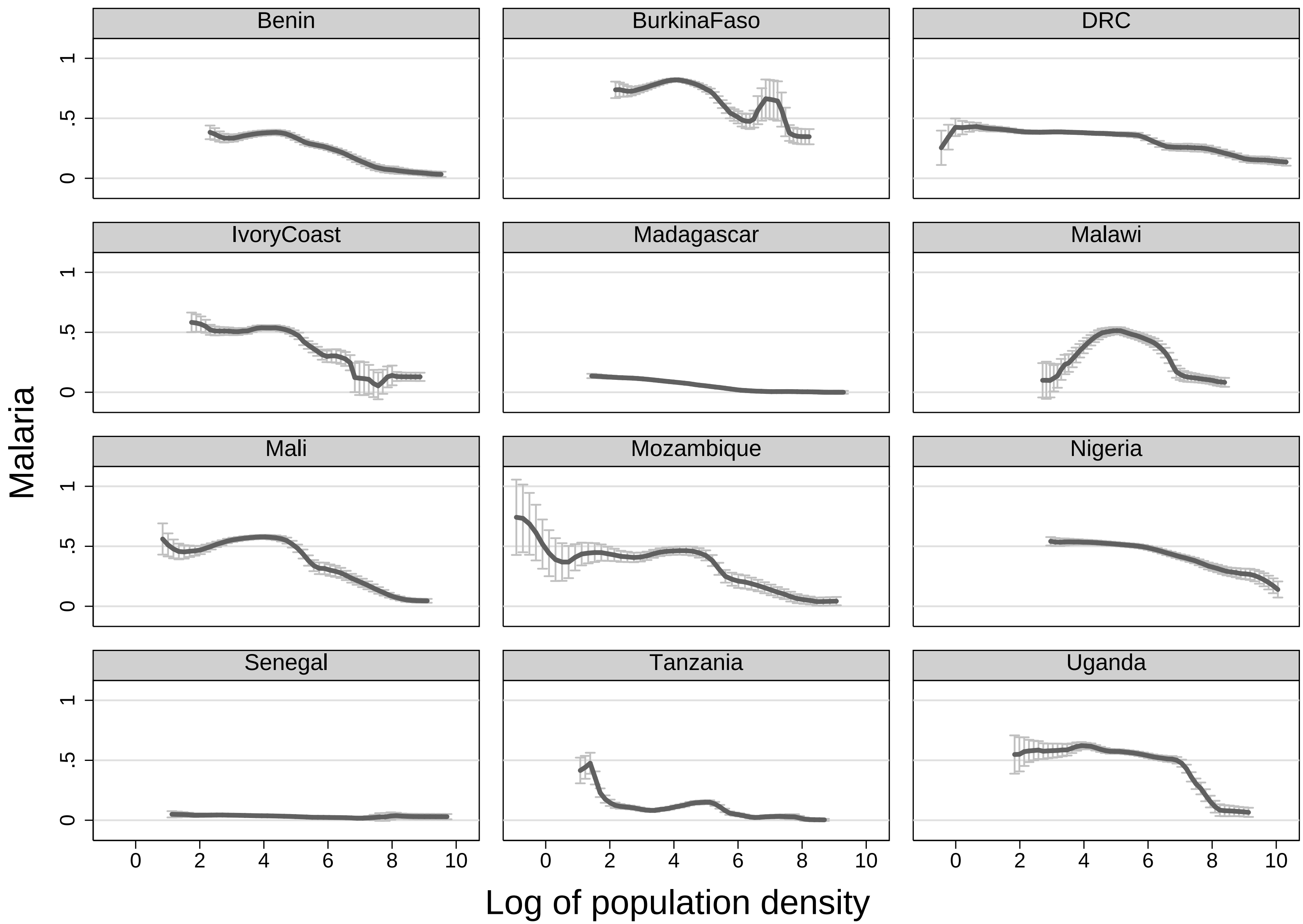


Figure A.22: Years lived in current location

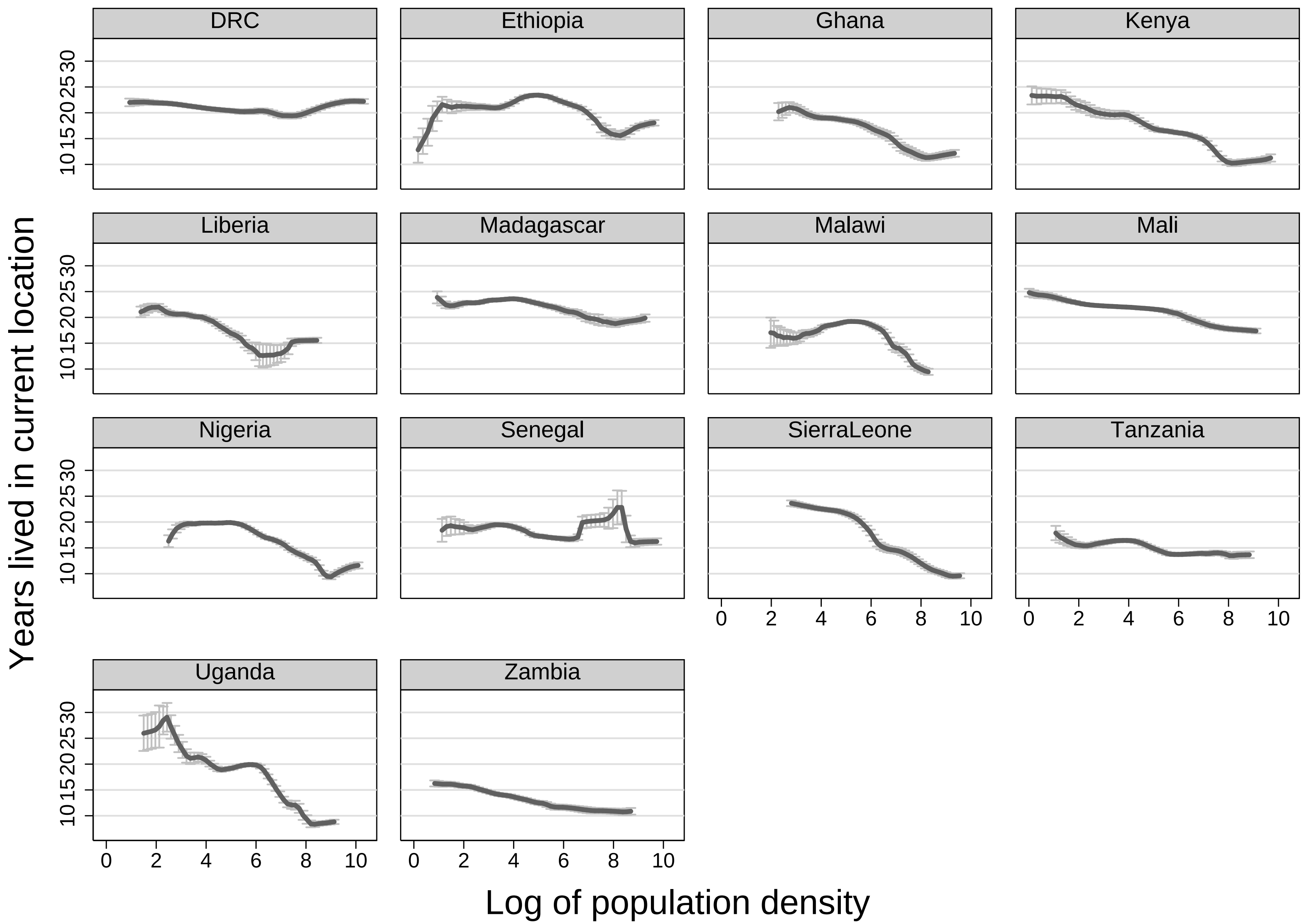


Figure A.23: Migrated in last 5 years from anywhere

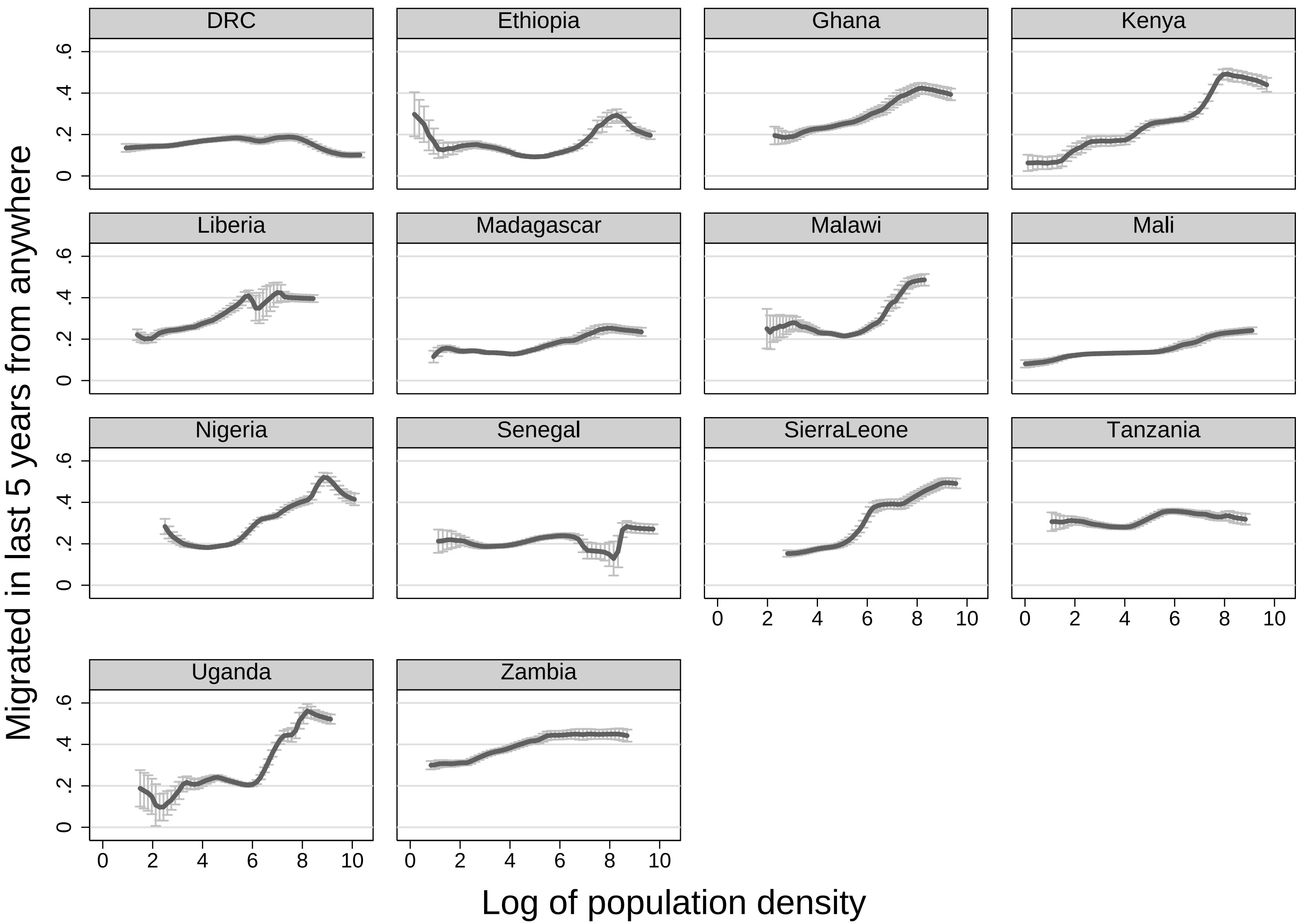


Figure A.24: Migrated in last 5 years from countryside

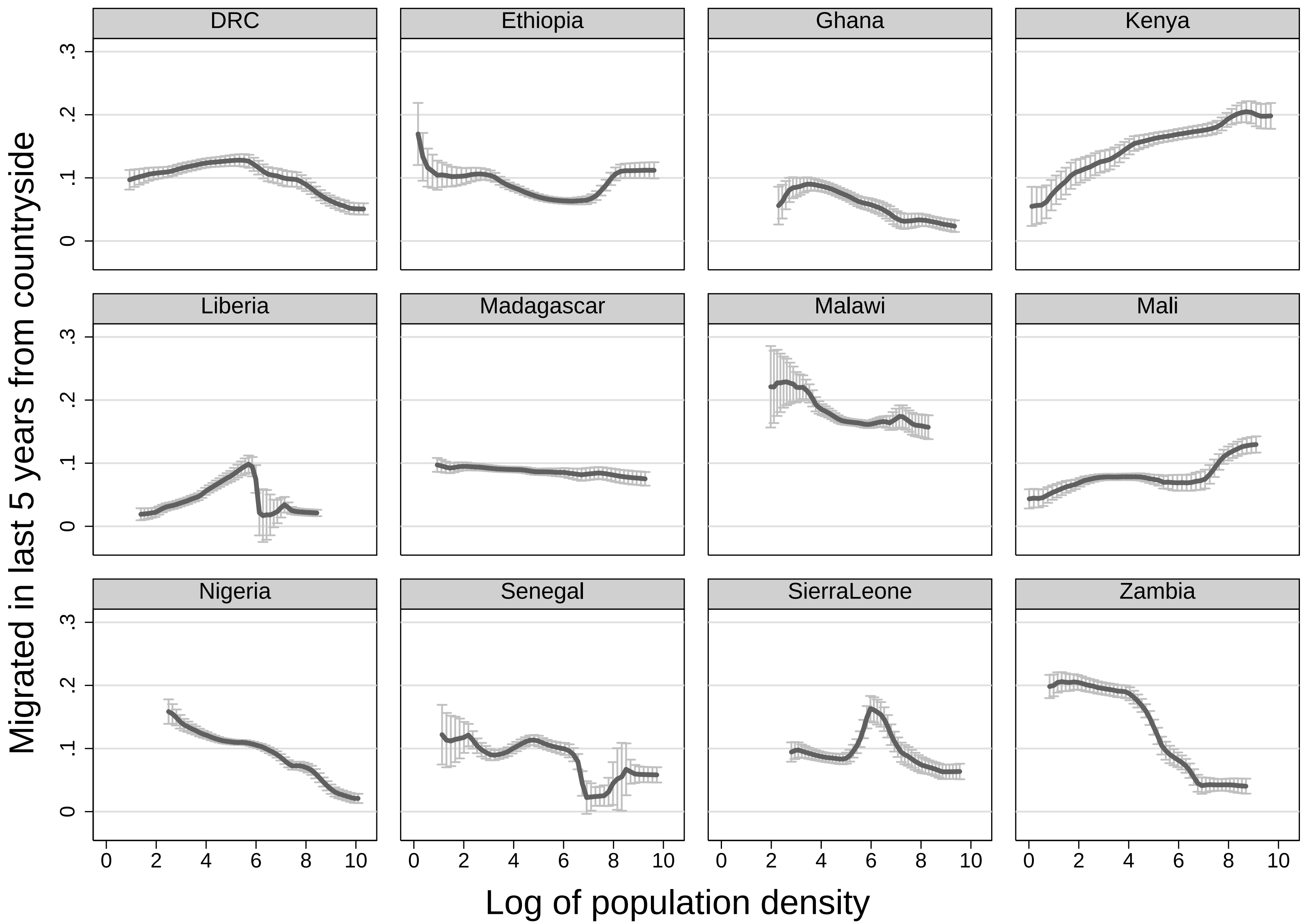


Figure A.25: Migrated in last 5 years from town/city/capital