Pollution, Crime, and Mistrust across Space: Evidence from Sub-Saharan Africa*

Martina Kirchberger

Columbia University

March 2016

Abstract

Striking differences in living standards between urban and rural areas in low income countries pose a puzzle for growth economists. In a conventional spatial equilibrium model, utility is assumed to be equal across locations; otherwise an individual would have an incentive to move. Whether disamenities compensate individuals for the higher standard of living in urban areas remains largely untested. This paper asks whether the spatial distribution of key disamenities can explain the observed differences in living standards in Sub-Saharan Africa. I construct a new dataset that links geo-located household surveys on crime, mistrust, living conditions, and satellite-derived measures of pollution with gridded population density data. This allows me to view outcomes through the lens of population density in addition to a traditional urban/rural distinction. I reject the possibility that pollution is a key disamenity of high population density areas in the twenty African countries in my sample, but the evidence supports the notion that crime and mistrust are indeed higher in denser areas. However, the magnitudes of these effects are small compared to the differences in living standards, suggesting that these variables do not offset measured differences in income, wages, or living conditions. The results call into question the usefulness of spatial equilibrium concepts for Sub-Saharan Africa.

*I am grateful to Deborah Balk, Banu Demir-Pakel, Stefan Dercon, Dave Donaldson, Shauna Downs, Marcel Fafchamps, James Fenske, Doug Gollin, Christian Helmers, Darby Jack, Pat Kinney, Pramila Krishnan, David Lagakos, Horacio Larreguy, Clare Leaver, Ethan Ligon, Friederike Niepmann, Matthew Neidell, Natalia Ramondo, Tim Schmidt-Eisenlohr, Chris Small, Gerhard Toews, Luke Valin, Eric Verhoogen, David Weil, and participants at various seminars and conferences for helpful comments and thoughtful discussions. I would like to thank Jeff Geddes and Aaron van Donkelaar for sharing the pollution data and their advice. I am grateful to the people at the Center for International Earth Science Information Network at Columbia, in particular Bob Chen, Alex De Sherbinin, Erin Doxsey-Whitfield, Alex Fisher, Marc Levy, and Greg Yetman, for support with the gridded population data, many helpful suggestions, and for hosting me. I am thankful to Abed Mutemi for truly outstanding research assistance, and the Bill and Melinda Gates Foundation for supporting his work on this project. 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 my own. I look forward to receiving your comments via mkirchberger@gmail.com.

1. Introduction

Development economists have long recognized that apparent living standards in urban areas are substantially higher than those of rural areas. These gaps exist in real measures of well-being, and thus do not merely reflect price differences (Young, 2014). The pattern is also present for more and less populated areas within rural and urban locations, and within education groups (Gollin, Kirchberger, and Lagakos, 2015). Much of the economics literature assumes that the allocation of people across space is determined to a first order by a spatial equilibrium, in which utility is equalized across space (Rosen, 1979; Roback, 1982; Glaeser and Gottlieb, 2009).[1] In this view, higher living standards in cities must be accompanied by corresponding disamenities associated with living in urban areas or densely populated areas.[2] If a location provided a better bundle of consumption and amenities for a given set of prices, it would attract in-migration, until various congestion-related disamenities equalize well-being across locations. Previous research has shown that the income advantages in urban areas are compounded by differences in housing quality and health outcomes such as anemia prevalence and the proportion of children who consume a minimum acceptable diet (Gollin, Kirchberger, and Lagakos, 2015).

It remains untested whether differences in living standards in Africa can be rationalized by disamenities of more densely populated areas. This paper tests whether gaps in living standards can be understood by offsetting differences in key intangible disamenities (crime, pollution, and mistrust). These disamenities have hitherto been difficult to measure at a fine geographic scale. However, this set of disamenities is often assumed to be worse in urban areas than in rural areas, and on that basis, offered as a potential disadvantage of urban life.[3] The central finding of the paper is that the magnitude and sign of these disamenities suggests that they are unable to offset the differences in other dimensions of living standards. Somewhat surprisingly, pollution levels are higher in rural areas than in urban areas.[4] Crime and mistrust are slightly higher in denser areas, but the differences are small compared to the differences in living

[1] Examples of recent papers that assume a spatial equilibrium holds within countries include Desmet, Nagy, and Rossi-Hansberg (2015); Henderson, Squires, Storeygard, and Weil (2015); Allen and Arkolakis (2014); Albouy (2009); Harari (2015); Diamond (2015); and Hanlon (2015).

[2] Nordhaus and Tobin (1972) articulated this idea arguing that "the persistent association of higher wages with higher population densities offers one method of estimating the costs of urban life as they are valued by people making residential and occupational decisions", and pricing the "disamenity premium" at 5% of Gross National Product.

[3] Banzhaf and Walsh (2008) find pollution to be an important determinant of locational choice in the United States.

[4] This is based on the level of particulate matter less than 2.5 microns in diameter (PM2.5), a common measure of pollution (World Bank, 2015). As discussed below, this somewhat surprising result differs from the pattern observed in other parts of the world and reflects the high levels of naturally-occurring dust in rural Africa, along with the lack of polluting industry.

standards. This evidence challenges the idea that a spatial equilibrium currently holds in Africa.

My main contribution is to exploit new sources of data and advances in mapping technology to test whether the within-country spatial distribution of amenities is consistent with a simple spatial equilibrium in a large number of countries in Sub-Saharan Africa. In the process, I move beyond traditional urban-rural distinctions by viewing locations through the lens of population density. If cities are defined as the absence of space between people, this is an intuitive way of examining the data (Jacobs, 1961; Glaeser, 2011). Further, I do not have to rely on official classifications of what is urban and rural, which vary both across space and time. For example, linking location data from the Financial Inclusion survey for Bangladesh, India and Tanzania with population density data, I find that the average density of rural respondents is 638 people per square kilometer in India, which is almost as high as the average density for respondents in Tanzania classified as urban (725 people per square kilometer), and less than half the population density of rural respondents in Bangladesh (1,517 people per square kilometer). Classifications also change across time: before 1991 Uganda classified villages with more than 1000 people as urban, after 1991 the cut-off was raised to 2000 people (UN, 2015).[5]

My approach involves constructing a new dataset that spatially links household surveys on crime, mistrust and reported living standards with satellite derived measures of pollution, and gridded population density data. Specifically, I geo-locate the respondents from several rounds of the Afrobarometer surveys and link them with recent estimates on pollution concentrations for particulate matter (PM2.5) and nitrogen dioxide (NO2) distributions derived from satellite observations (van Donkelaar, Martin, Brauer, and Boys, 2015; Geddes, Martin, Boys, and van Donkelaar, 2015). I also use a number of further geo-referenced household surveys: Living Standards Measurement Surveys (LSMS) and Demographic and Health Surveys (DHS). The population density data come from the Gridded Population of the World, Version 4, a recent population density dataset that involves a minimal level of modelling by distributing aspatial population counts from census data across spatial administrative units (Center for International Earth Science Information Network, 2015).[6] This allows me to assign a particular pollution level and population density level to each respondent in the Afrobarometer survey. This new dataset enables me to test whether the data is consistent with a simple static spatial equilibrium, using an intuitive prediction of the Rosen-Roback framework. As geo-referenced data become available for more countries and time periods, my methodology can be readily extended to

[5] Dorélien, Balk, and Todd (2013) find substantial disagreement when comparing urban-rural classifications used in the Demographic and Health Surveys which are based on official classifications with urban extents from the Global Rural-Urban Mapping Project, a satellite-based dataset: about 66 percent of clusters that were classified as rural in the survey-based dataset fell into urban extents as captured by nighttime lights.

[6] For instance, pollution or other disamenities are not used to model the distribution of the population across space.

additional countries and studying dynamics.

I reject the hypothesis that pollution increases with population density in Sub-Saharan Africa. On the contrary, natural sources of PM2.5 measures are substantial, and densely populated areas tend to have lower PM2.5 concentrations compared to sparsely populated areas. For example, the average PM2.5 concentration for residents of the lowest quartile of population density in Nigeria is about 1.5 times the level of the most densely populated quartile (34μg per m^3 versus 24 μg per m^3). Average NO2 levels in the whole sample are low on average with 0.15 μg per m^3 and there is no relationship with population density.[7] Household-level data from the DHS suggest that indoor pollution is also lower in more densely populated areas for two reasons: first, households living in denser areas are more likely to use liquid fuels and electricity; and second, cooking is more likely to take place outside the home in denser areas if households use solid fuels. This implies that there is no evidence that higher pollution in dense areas compensates households for the lower living standards in rural areas.

I next show that property crime and violent crime, fear, and lack of perceived safety are more prevalent in more densely populated locations. Mistrust towards neighbors and co-ethnics is significantly higher in more densely populated locations. To gauge whether these disamenities are likely to provide a satisfactory explanation for the large gap in living standards, I construct a simple index of a household's assets and housing quality and show that a doubling of population density is associated with an increase in the index of 13% of a standard deviation. For the same increase in population, the probability of reporting a violent crime increases by 2% of a standard deviation or 5% of a standard deviation for measures of mistrust. The heterogeneity in welfare associated with population density is thus 8 times the heterogeneity in crime associated with population density, and 5 times the heterogeneity in mistrust associated with population density. One possibility is that I fail to measure a key disamenity. The questionnaire asks individuals how they rate their own living conditions, a measure which is arguably an all-encompassing measure of the quality of life across different locations. I find that self-reported living conditions are significantly higher in denser areas too. Although it is difficult to know how to interpret these subjective measures of well-being, if taken at face value, the data suggests that a spatial equilibrium does not currently hold in Africa.

To my knowledge, this is the first paper that tests whether key intangible disamenities can explain differences in living standards across space in Africa. These gaps in living standards have been documented both in consumption between urban and rural locations, and between agri-

[7] For both pollution measures I find significant gradients in other developing countries, such as India and China: PM2.5 concentration rises until a population density of 400 people per square kilometer and then remains largely constant. The level at which it flattens out is substantially different as well, 60 μg per m^3 in China versus 40 μg m^3 in India.

cultural and non-agricultural productivity (Young, 2014; Gollin, Lagakos, and Waugh, 2014). Why these large differences across sectors and space exist in low income countries remains an open question. Gollin et al. (2014) speculate that amenities of rural areas and the lower cost of living are possibly keeping workers in sectors with lower productivity. Alternative explanations for differences in living standards across locations include mobility costs (Morten and Oliveira, 2014), selection and sorting (Young, 2014; Lagakos and Waugh, 2013), the presence of insurance networks (Munshi and Rosenzweig, 2015) and fear of negative outcomes (Harris and Todaro, 1970). The presence of disamenities is a further possible explanation; given the large differences in living standards, it is likely that a combination of factors is driving these.

Viewing data through the lens of population density is in line with the fairly recent movement away from a dichotomous urban-rural distinction towards thinking about allocations in a continuous space (Desmet and Rossi-Hansberg, 2014; Henderson, Storeygard, and Deichman, 2014; Desmet and Henderson, 2015; Henderson, Storeygard, and Weil, 2012). In Gollin, Kirchberger, and Lagakos (2015), using data from Demographic and Health Surveys we show that the large differences in living standards are continuously and monotonically increasing over population density space, and this pattern holds within education groups. Desmet et al. (2015) evaluate the welfare effects of migration restrictions assuming that a spatial equilibrium holds within countries. Amenities play a key role in their model, and their parameter characterizing the relationship between amenities and population comes from estimates relating amenities in metropolitan statistical areas in the United States to population. My paper differs in that, rather than assuming a spatial equilibrium holds within countries, I test this assumption for African countries by directly measuring and documenting the evolution of key disamenities across population density space.

I am 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. I geo-locate two further rounds of the same survey and link the data with spatial population density data. Replicating their results, I 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.[8] The weak evidence for classical disamentities such as crime and

[8]The most common generalized trust question asks individuals whether they think that "Most people can be trusted" or whether he or she thinks that one "has to be very careful". Overall, generalized levels of trust in Africa are lower than what similar data suggest for Europe and the United States (Algan and Cahuc, 2014). The average in the 18 countries of 2005 Afrobarometer survey is 16% of individuals reporting that most people can be trusted, compared to 40% in the United States, and above 68% in Norway. Within Africa and Europe regional differences are substantial too, both within countries as well as across countries. The lowest level of generalized trust in my

pollution adds another dimension in which the urbanization process in Africa differs from other parts of the world, such as the absence of the commonly observed link between urbanization and industrialization (Jedwab, 2013).

My use of data on self-reported quality of living conditions as a summary measure of overall living standards contributes the literature on the use of subjective well-being data as alternative measures of consumption and income (Deaton and Stone, 2013; Stevenson and Wolfers, 2013) and the role of non-material consumption goods such as social connections and happiness determining location choices (Glaeser, Gottlieb, and Ziv, 2015; Barnhardt, Field, and Pande, 2015).[9] This paper differs from these existing studies in that my geo-referenced data allow me to look at these measures across space within a large range of developing countries. In this sense, I present a spatial version of the Easterlin Paradox. Further, the interpretation of my results is slightly different due to the phrasing of the question in the Afrobarometer, which is less likely to reflect happiness or life satisfaction in general but more specifically a comprehensive self-assessment of living conditions.

Determining whether large observed gaps in living standards are due to inefficiencies matters for policy. If there are large gaps in outcomes between urban and rural areas that remain difficult to explain by looking at the distribution of consumption and amenities across locations, there might be substantial efficiency gains from alleviating factors restricting internal mobility. On the other hand, if the observed outcomes appear efficient, intervention must be justified by concerns other than efficiency. My findings also imply that, while cities share certain features across different contexts, the type of amenities and disamenities denser areas deliver vary substantially not only across levels of development but also across contexts. The significant pollution-population density gradients present in China and India are not present in most Sub-Saharan African countries. This emphasizes that models aimed at explaining location choices of individuals ought to be tailored to the particular context and process of urbanization they try to characterize. For example, my results suggest that modeling pollution as an endogenous disamenity as in Hanlon (2015) is very insightful for the Asian context where there is a strong relationship between pollution and urbanisation, but is less likely to capture location choices across space in most of Africa where this relationship is largely absent.

While it is still possible that a spatial equilibrium holds under certain conditions, I find little evidence that the first order explanations for a spatial equilibrium fully rationalize the observed differences in living standards. It is possible that I fail to measure a relevant disamenity.

sample is in Malawi with 6.9% compared to 27% in Benin.

[9] Glaeser et al. (2015) argue that their measure of happiness, which comes from a question on life satisfaction in most of their surveys, should be seen rather as an input into the utility function than utility itself. Individuals in the United States get compensated for lower levels of happiness by higher wages.

However, it is not clear what disamenity this is, and its effect would have to be large.

The paper is structured as follows. Section 2 sketches a conceptual framework that guides the empirical analysis; Section 3 discusses the data and how I link them with population density data. I discuss my main results in section 4 and provide a conclusion in Section 5.

2. Conceptual Framework

The workhorse model of location choice across space is the Rosen-Roback model, developed by Rosen (1979) and Roback (1982). The aim of this section is to fix ideas that guide the empirical analysis; I mainly rely on the treatment presented in Glaeser (2008). The Rosen-Roback model's goal is to understand the key drivers affecting consumers' and firms' location choices. The most basic form of this static spatial equilibrium model has three main components: (i) consumers maximize utility from amenities, a tradable good, and non-tradable housing; (ii) firms maximize profits by choosing the optimal amount of labor and capital; and (iii) builders maximize profit by chosing the optimal amount of height and land. This yields a spatial equilibrium condition for consumers, firms, and builders.

I use the consumption side of the model to illustrate the basic intuition behind this framework. Assume that there is a large number of locations individuals can choose from, and that these are indexed by population density d. At each location, consumers receive utility from consuming amenities as captured by an index θ_d (for example, fresh air or safety), a tradable good c_d and non-tradable housing h_d. They have a Cobb-Douglas utility function

$$U_d = \theta_d c_d^{1-\alpha_d} h_d^{\alpha_d}$$

where $0 < \alpha_d < 1$, $\partial U_d/\partial c_d \geq 0$, $\partial U_d/\partial h_d \geq 0$ and $\partial U_d/\partial \theta_d \geq 0$ for all levels of c, h and θ. Consumers earn wages w_d and pay rent p_d per unit of housing. The price of the tradable consumption good c_d is normalized to one so that the budget constraint is

$$c_d = w_d - p_d h_d.$$

The key equilibrium condition is that utility $\bar{U}$ is equalized across space. If there was a location that provided a better bundle of every input into the utility function, an individual would move. A testable proposition follows from the simple model for any regions j and k:

Proposition 1 (Compensating Differentials.) *If $c_j > c_k$ and $h_j > h_k$ then it has to be that at least for one element of θ, it is the case that $\theta_j < \theta_k$.*[10]

[10]A stronger version of the proposition is that for utilities to be equal, the higher non-housing and housing

To illustrate the intuition behind Proposition 1, let us fix $\bar{U} = 20$, $\alpha = 0.5$, $h = 1$, and proxy c with an asset and housing quality index, counting the number of durables a household has and housing quality indicators using data from Gollin, Kirchberger, and Lagakos (2015) for 20 Sub-Saharan African countries.[11] 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 $\bar{U}$, α, and h, this allows me to back out how the value of amenities evolves across space.

Figure 1 shows this relationship between consumption, amenities, and utility. The x-axis shows different locations d as measured by deciles of the log of population density. Consumption and amenities are measured on the left hand side axis, utility on the right hand side axis. The Rosen-Roback model implies that households located at very low population density levels have low levels of consumption but are compensated by higher levels of amenities, for example better air quality and lower levels of crime. For a given increase in living standards across population density space, amenities have to decrease to ensure equality of utilities across space. If a location provided better consumption and better amenities for a given set of wages and prices, individuals would move to the location until arbitrage opportunities have been exploited. If I do not find any amenity that decreases with population density, this is suggestive that something prevents individuals from exploiting these arbitrage opportunities.

A short discussion of migration costs is warranted at this point. The simple model assumes that individuals are fully mobile and mobility costs are equal to zero. Morten and Oliveira (2014) relax this assumption. With non-zero migration costs, individuals will move to a location as long as the utility from the new location is higher than the utility in the old location plus the cost of moving. In the absence of other types of frictions, this implies that the gap in living standards would be explained by migration costs. Another simplification of the present model is that it does not explicitly model differences *within* city boundaries, as explicitly modeled, among others, by Ahlfeldt, Redding, Sturm, and Wolf (2015). This is at odds with one of the most striking features of rapidly growing cities: the web of extreme wealth and poverty side by side. Empirically, my approach is far less restrictive as I take into account differences in amenities from the densest areas moving to the most remote areas in the country. The goal of this example is to illustrate the intuition behind the model rather than a specific functional form for amenities across space. A prediction of the simple spatial equilibrium model is that at least one element of the amenity vector has to deteriorate with population density space. The

consumption in region j has to be *exactly* offset by lower levels of amenities in region j. In the results section I come back to what preferences would have to be in order to rationalize the data.

[11] 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.

rest of this paper uses new sources of data on key disamenities across locations in Africa to test this prediction.

3. Data

Until recently, measuring disamenities across space was not feasible. Exploiting progress in surveying and mapping technology, I construct a new dataset that spatially links household surveys on crime, mistrust and reported living standards with satellite-derived measures of pollution and gridded population density. Specifically, I geo-locate the respondents of several rounds of the Afrobarometer surveys and link them with recent estimates of pollution concentrations for particulate matter less than 2.5 microns in diameter (PM2.5) and nitrogen dioxide (NO2) distributions derived from satellite data (van Donkelaar et al., 2015; Geddes et al., 2015). I also use a number of further geo-referenced household surveys: Living Standards Measurement Surveys (LSMS) and Demographic and Health Surveys (DHS). The population density data come from the Gridded Population of the World v4 (Center for International Earth Science Information Network, 2015). I select countries that satisfy four criteria: (i) the survey or dataset is from 2005 or more recent[12], (ii) spatial identifiers are available; these could be location names, latitude and longitude of a household or survey cluster, or gridded geo-spatial data; (iii) the country is larger than 50,000 square kilometers[13], and (iv) classified as a low income country by the World Bank in 2012.[14] This section describes the various sources of data I combine in more detail.

3.1. 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.[15] To measure exposure to pollution faced by individuals, ideally I would employ measurements taken on the ground at varying population densities and times of the year, as well as measure pollutants at indoor and outdoor locations for different types of households. Unfortunately, data on ambient air pollu-

[12] This is to ensure that I combine data from similar time periods.

[13] This restriction leads me to exclude small countries such as Burundi with one major city or island states such as Cape Verde.

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

[15] A growing literature shows how pollution 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). 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).

tion from ground measurements in African cities is scarce (Petkova, Jack, Volavka-Close, and Kinney, 2013). I use the most recently available satellite-derived estimates of concentrations of particulate matter equal to 2.5 micrometers or less in diameter (PM2.5) from van Donkelaar et al. (2015) and nitrogen dioxide concentrations (NO2) from Geddes et al. (2015). These use measures of aerosol optical depth and tropospheric vertical nitrogen dioxide column density to approximate the distribution of pollutants in the atmosphere as observed from satellites. PM2.5 is harmful to health as particles of this small size are able to enter deeply into the respiratory tract.

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 et al., 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 et al., 2015); nevertheless, in the absence of more conclusive evidence, they give an indication of the current recommended thresholds.

As my population density data is approximately from 2010, I take pollution data that are closest in time. Specifically, I use the tri-annual mean (2009-2011) for both datasets. To capture indoor pollution, I rely on Demographic and Health Surveys which collect data on the source of cooking fuel households use. Solid sources of cooking fuel include coal, lignite, charcoal, wood, straw and animal dung, compared to liquid sources which include electricity, liquified petroleum gas, natural gas, biogas, and kerosene. The DHS questionnaire also asks where households undertake their cooking. I define a dummy for indoor cooking as equal to one if the respondent states that food is cooked in the house or in a separate building.

3.2. Crime and Mistrust

To measure violent and property crime at small levels of geographic detail, I 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. These countries form my base sample for which I will present all other results.

The Afrobarometers are high quality micro surveys covering between 1200-2400 individuals in each of several African countries. Table 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 my 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. I 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".[16] 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. 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 I 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

[16]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.

into four categories: not at all, just a little, somewhat, a lot; I 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.

3.3. Population Density Measures

To measure population density, I 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 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 I then use these data to estimate the relationship between population density and electrification, by construction, higher densities will have higher rates of electrification. I 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 me 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. I restrict my analysis to countries for which the input census data have sufficiently high spatial detail, roughly more than 40 regions per country.

3.4. Spatially linking Pollution, Crime, Mistrust and Density

I next combine the different sources of data step by step. 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. I 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).[17]

Figure 2 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 (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. There is no evidence that the health effects of non-anthropogenic sources of PM2.5 are different from anthropogenic sources, so for the remainder of the paper I present most results for the basic PM2.5 estimate that contains sea salt and dust, and show the full set of results for PM2.5 after removing anthropogenic sources in the Appendix.

The Nigerian example illustrates further that looking separately at PM2.5 and NO2 is instructive.[18] The pairwise correlation between PM2.5 and NO2 is 0.05. Across my whole set of African countries, the correlation of these two measures ranges from 0.64 in Cameroon to -0.47 in Senegal.

[17] Following Vrijheid, Martinez, Manzanares, Dadvand, Schembari, Rankin, and Nieuwenhuijsen (2011), I 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.

[18] This is in line with what Geddes et al. (2015) find when they inspect population weighted average PM2.5 and NO2 levels and trends.

To link the individual data from the DHS and Afrobarometer with population density, ideally I would have the GPS location of households. Unfortunately, the Afrobarometer did not collect the GPS location, but the location name was recorded. I develop an algorithm that performs a series of exact and fuzzy matches of location names relying on data from a global gazeteer 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-spellings, I 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 A provides further detail on the matching procedure. Using this algorithm I am able to geo-locate between 85–95% of village names in each round.[19] For each respondent I can then extract the population density value as well as the pollution exposure.

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).[20] I 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 et al. (2013). Figure A.1 illustrates this procedure and discusses the sampling protocol of the surveys. With these different pieces of information in hand, I can test whether the data are consistent with a simple static spatial equilibrium model when considering this key set of disamenities.

4. Disamenities across Space

As Young (2014) and Gollin, Kirchberger, and Lagakos (2015) show, real measures of living standards are consistently better in denser areas suggesting that prices are not the equilibrating mechanism. If we take a spatial equilibrium model seriously, then this means that for utilities to be equal across space, key disamenities have to be higher in denser areas. To do justice to the richness of the data, I present my results in four steps that highlight different features of the data: I start by showing the spatial evolution across the entire density spectrum for several selected variables. I will not impose any restrictions on functional form here and estimate

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

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

the relationship between disamenities and population density non-parametrically. Second, I present summary graphs for the entire set of countries in my sample, depicting outcomes for the highest compared to the lowest population density quartile. Using this comparable metric allows me to compactly present differences in disamenities across space for many countries. Third, using parametric restrictions on the relationship I estimate population density gradients for each of the variables and countries. Finally, I pool the data and formally test if there is a positive relationship between population density and disamenties. For the first two parts I discuss each group of disamenities separately, and then combine them in the last two parts.

4.1. Pollution

I first examine satellite-derived estimates of PM2.5 and NO2 distributions across space and then move to types of fuels used by households, including the location of cooking activities. Both pollutants are measured as micrograms per cubic meter ($\mu g/m^3$). Figure 3 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.[21] Taking the log of population density removes uninhabited pixels.[22] The figure confirms what the visual inspection of figure 2 suggested: there is no evidence that pollution levels are higher in denser areas such as cities. If anything, there is a negative relationship between PM2.5 levels and population density. When sea salt and dust are removed, the relationship is flat. One interpretation of this relationship is that cities in Nigeria are located in areas where PM2.5 levels are lower. Whatever pollution is being emitted by denser areas in Nigeria, it does not dominate the high non-anthropogenic levels, leading to a positive relationship between population density and pollution. Nitrogen dioxide concentrations show an increase first at low levels of density, and then decrease until they increase again at higher levels of density (but the confidence interval becomes large at the highest level of density). The average level of NO2 in Nigeria is 0.49 $\mu g/m^3$, which is low compared to the WHO recommended threshold of 40 $\mu g/m^3$.

Figure 4 shows PM2.5 level for all African countries in my sample. Several countries show a negative correlation between PM2.5 and population density, such as Benin, Ghana and Ivory Coast. Density gradients are sometimes flat, for example in Kenya, Malawi, Tanzania and Zambia; only a handful of countries show a positive gradient, such as Liberia, Cameroon and

[21] 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 et al., 2015).

[22] I remove the top and bottom five percentile of the population density distribution as very few observations fall into this bin and patterns are potentially misleading.

Senegal. There are large differences in levels. Some countries have PM2.5 levels that are above the WHO recommended thresholds of 10 $\mu g/m^3$ at all levels of population density, such as in Ghana, Mali, Liberia, Niger, Senegal, and Sierra Leone; others, including Kenya, Uganda, Zambia and Zimbabwe are below the threshold for all levels of density. Finally, there is also a strong difference between the two PM2.5 measures, particularly in West African Sub-Saharan countries, such as Benin, Burkina Faso, Mali, Nigeria, Senegal, and Sierra Leone.

Figure 5 shows the same figures for NO2 concentrations. Average NO2 levels are very low across the African countries in my sample: 0.26 $\mu g/m^3$ compared to the WHO recommended threshold of 40 $\mu g/m^3$. Similar to the relationship between PM2.5 levels and population density, 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, Cameroon, and Niger; others exhibit a negative relationship, such as Benin, Ivory Coast, and Sierra Leone. Still, all of the countries have very low levels. Figures B.1 and B.2 in the appendix show PM2.5 and NO2 levels rescaled for each country. Overall the findings suggest that there are no clear patterns between pollution levels and population density.

What do these gradients look like in other parts of the world? For example, what do the data show for countries known to suffer from high levels of urban pollution, such as China? The global coverage of the PM2.5, NO2 and population density data allows me to look at further countries. Figure B.3 in the Appendix shows the correlation between population density and pollution for China, India, and the United States. All three countries show density gradients. In China, PM2.5 levels for the top population density decile amount to 67 $\mu g/m^3$, more than six times the WHO recommended threshold; the lowest population density decile has a level of 14 $\mu g/m^3$. 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.

I do not make a claim here that pollution does not matter in African cities. The satellite-derived pollution estimates do not capture pollution exposure in several dimensions: they are annual series and therefore average out 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).

Nevertheless, as the graphs from India, China, the US and the UK illustrate, the series still capture meaningful variation in concentrations levels across space. What emerges, rather, is that in Africa, cities are not large enough and their 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 I 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 pollution-related morbidity and mortality. Figure 6 shows the proportion of households using solid fuels for cooking across population density. In several countries, such as Liberia, Malawi, Mali, Sierra Leone and Uganda, solid materials are used by almost all households irrespective of location. For other countries, including Cameroon, Kenya or Nigeria, there are substantial gradients: households are more likely to use liquid fuels in more densely populated areas. 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. The data suggests quite the opposite.

Figure 7 shows the probability that a household cooks inside. The red line restricts the sample to households whose main source of cooking fuel is solid. For countries in which solid fuels dominate across the whole population density spectrum, the grey line is not visible. Overall, there are no clear patterns that hold for all countries across density space, as well as between households using solid fuels compared to the whole sample; if anything, households using solid fuel for cooking are almost in all countries less likely to cook inside in *denser* areas compared to less dense areas. One possible explanation is that rooms are smaller in denser areas, so that cooking is done outside. I now turn to Afrobarometer data to examine differences in crime and mistrust across population density space.

4.2. Crime

Figure 8 shows differences in experienced crime and fear of crime across space. I compare the highest density with the lowest density areas. To do this, I divide the dataset into population density quartiles and compute the average level of the variable for each quartile in a country, using the within country sample weights provided by the survey.[23] The closer they are to the 45 degree line, the more similar outcomes are across population density space. If countries

[23] An alternative would be to compute quartiles for each country separately. I prefer to define quartiles over the entire sample in order to compare outcomes for similar population densities.

exhibit a significant positive relationship between population density and crime, we would expect these averages to cluster in the lower triangle of the figure.

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. One weakness of the 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.

The weak association between crime and density is against the conventional wisdom considering crime as a main disamenity of cities. Denser areas are thought to have higher levels of crime by raising the frequency of interactions between individuals, the potential gain from committing a crime (such as theft) due to higher welfare, and the lower likelihood of being caught due to higher anonymity. However, my finding resonates with evidence from Madagascar and South Africa suggesting that crime in Africa might be even higher in less densely populated areas (Fafchamps and Moser, 2003; Demombynes and Ozler, 2002). Madagascar also in my data exhibits this pattern: in the lowest population density quartile 13% of respondents report theft compared to 9% in the highest population density decile. 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.

4.3. Mistrust

By bringing together large numbers of people cities are more anonymous places. One possible downside is that networks are weaker. Busy, buzzing city life might also be more stressful, lonely and isolating compared to life in rural areas. If individuals have sufficiently strong preferences for living in an area with strong social cohesion this could be part of the explanation for the stark differences in living standards that we observe.

Figure 9 shows differences in trust towards relatives, neighbors and co-ethnics for the highest and lowest population density quartile. Almost all quartile-level means are located in the upper triangle, suggesting that higher levels of trust in the lowest population density quartile compared to the highest population density quartile. 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. This evidence suggests that trust appears to be a potential compensating differential that keeps people from moving to more densely populated areas.

A seminal paper on trust is Nunn and Wantchekon (2011) who use the 2005 Afrobarometer survey and demonstrate that a higher exposure to the slave trade reduced levels of trust. They control for urban location as defined by the survey, but the coefficient is not reported in the main paper. I replicate their results with a focus on the coefficient on the urban dummy. I find that the coefficient is always negative and highly significant when including their full set of controls (Table 2 in their paper), in the IV specification (Tables 5 and 6 in their paper), as well as in the regressions controlling for individual level distance to the coast (Tables 7 and 8 in their paper).[24] The patterns found in the two papers are therefore similar in that urban location appears to be associated with less trust.

4.4. Parametric tests

The results so far do not suggest that disamenities are substantially higher in more densely populated areas. A simple spatial equilibrium would predict a relationship as illustrated by Figure 1 in which dissamenities are higher in more densely populated locations. To formally test whether there is a positive gradient of disamenities with respect to population density, I estimate 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. The number of observations varies as certain questions are not asked in some countries in a particular round.

I start by estimating these regressions for each of the countries separately. Figure 10 presents the coefficients of these regressions where each of the blue dots represents a coefficient. 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. For the different measures of crime, the slope coefficient is overall small and a median coefficient close to zero; a substantial number of countries has a negative relationship between crime and population density, as indicated by the dots below the horizontal line. The size of the coefficients, however, tends to be small, for example, considering the average levels of theft of 31 percent. The lower panel indicates that for both types of pollution, countries are rather

[24] For example, taking column (1) of table 2, I find that urban residents are 3.5% less likely to report trusting their relatives "somewhat", or "a lot".

equally distributed above and below the zero line, again, with coefficients that are small in magnitude. The only variable that shows a clear pattern is trust in the bottom right panel; for all measures of trust there is a negative relationship.

To test whether the observed pattern is statistically significant, I next pool all countries but include country fixed effects ξ_c to allow for different levels of disamenities across countries. As before, fixed effects for the different survey rounds. Each column in table 2 presents a different regression where the disamenity tested is listed as the column header. The stars denote signficance levels with p-values corrected for multiple hypothesis testing using a correction proposed by Benjamini-Hochberg (1995). I find that two measures of crime are significantly different from zero and positive; for the remaining measures of crime, the sign and size of the coefficients of the is similar. I find that pollution is not significantly higher in more densely populated areas. Levels of PM2.5 exposure are significantly lower in denser areas, and NO2 is not significantly related to population density. It is possible that the functional form is incorrect so I have also estimated this equation using the log of the two pollution measures; the results remain the same. Unsurprisingly considering the evidence from Figure 10, mistrust towards neighbors and co-ethnics is significantly higher in denser areas.

Is it likely that compensating differentials are sufficiently high to account for the differences in observed living standards? I approach this question from three angles: first, I compare the size of the correlation of disamenities and population density with the change in assets for the same change in population density. Second, I examine the gap in consumption levels that would have to be explained by differences in disamenities. Third, I use self-reported living standards as a comprehensive assessment of living conditions that encompasses disamenities I did not consider.

To compare the magnitudes of disamenities with changes in living standards ideally I would use data on consumption across population density space. Unfortunately, the Afrobarometer survey does not collect such data. However, the 2011 round of the survey has data on variables representing real measures of consumption: whether the respondent owns a television, radio, or phone; if the source of water for household use is inside the house or compound and if the toilet or latrine is located inside the house or the compound. I compute a simple linear index of these assets and housing characteristics which ranges from zero to one. The relationship with assets and population density is highly significant and suggests that doubling population density is associated with an increase in assets by 13% of a standard deviation. This magnitude is large compared to the relationship of crime and mistrust with population density: a doubling of population density is associated with a 1.7% or 5% of a standard deviation for the experience of theft or higher levels of mistrust. The heterogeneity in assets associated with population density is thus 8 times the heterogeneity in crime associated with population density. This suggests that while there is evidence for compensating differentials, it is unlikely that they compensate for the significantly lower living standards.

I next use data from National Panel Survey in Tanzania to compare the magnitude to changes in consumption as one moves across population density space. I find that a doubling of population density is associated with a 12% higher per adult equivalent annual real expenditure on food, furnishing and recreation. To obtain the appropriate crime-population density gradient I obtain the density gradients for Tanzania as shown in figure 10. The coefficient on theft is 0.01 and the coefficients on the other crime variables and on trust are are very close to zero and insignificant.[25] This indicates that, for a spatial equilibrium to hold, households would have to be willing to give up 12% of their annual real expenditure for a decrease in the probability of theft of 0.01%.

It is possible that there is an unobserved dimension of quality of life that I do not capture in my vector of amenities, and households have a strong preference for this amenity. The Afrobarometer survey has a question on self-reported living conditions. More specifically, the questionnaire asks: "In general how would you describe your own present living condition?" The respondent's options are: very bad, fairly bad, neither good nor bad, fairly good and very good.[26] Using the distinction raised by Deaton and Stone (2013), the Afrobarometer measure qualifies as an evaluative rather than a hedonic measure, as individuals are required to contemplate. It is likely to represent a comprehensive measure of living standards, encompassing dimensions of disamenities I did not consider so far. I find that reported living conditions are significantly higher in denser areas. 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" I take this as a measure of anxiety. Again, I find that anxiety is generally lower in denser areas.

A limitation of this paper is that I can not prove or disprove the existence of a spatial equilibrium in Africa. My estimates suggest that it is unlikely that disamenities fully rationalize the observed differences in living standards. However, it is possible that I fail to measure a relevant disamenity. I argue that self-reported living conditions are likely to contain dimensions of disamenities which I could not measure. The fact that I still find on average higher self-reported living conditions in denser areas suggests that disamenities such as crime and mistrust explain

[25] If I don't truncate the sample by the top and bottom five percentile there is also no gradient on theft.

[26] The question just before asks how households would describe the economic situation of their country. Thus, although the question is phrased as an absolute measure, there is reason to believe that the prior question might have introduced a reference point. It is at the beginning of the questionnaire, right after asking a respondent about their age, language, and whether the respondent is the head of the household.

part of the differences, but not all of them. However, I can not rule out that there are certain disamenities households fail to report when asked about their overall living conditions, which in turn fully compensate them for the lower living standards in rural areas. It is not clear what they are and they would have to be large, but this is possible. Second, the focus of this paper is on testing the static relationship between disamenities and population density space. Unfortunately, the Afrobarometer data does not contain information on past migration choices of households to explore.

5. Conclusion

This paper aims to make progress on our understanding of the persistence of lower living standards in rural compared to urban areas in Africa. These have been found in real measures of consumption, within densely and sparsely populated areas, and within education groups; they are compounded by better housing quality and health outcomes in denser areas (Young, 2014; Gollin, Kirchberger, and Lagakos, 2015). A common assumption in the literature is the existence of a spatial equilibrium in which utilities are equalized across locations (Rosen, 1979; Roback, 1982; Glaeser and Gottlieb, 2009). One possible explanation for the large differences in living standards are unobservable dimensions of rural life that compensate individuals for the lower living standards in rural areas. Whether disamenities can plausibly explain differences in living standards in Africa remains largely untested.

The paper's main contribution is to exploit new sources of data and advances in mapping technology to fill this gap by testing whether the spatial distribution of amenities is consistent with a simple spatial equilibrium in a large number of countries in Sub-Saharan Africa. In the process, I move beyond traditional urban-rural distinctions by viewing locations through the lens of population density. My approach involves constructing a new dataset that spatially links household surveys on crime, mistrust and reported living standards with satellite derived measures of pollution, and gridded population density data. The central finding of the paper is that the magnitude and sign of these key disamenities suggests that they are unable to offset the differences in other dimensions of living standards. Somewhat surprisingly, pollution levels are lower in rural areas than in urban areas. Crime and mistrust are slightly higher in denser areas, but the differences are small compared to the differences in living standards. Self-reported living standards are better on average in denser areas. Taken together, this evidence challenges the idea that a spatial equilibrium currently holds in Africa.

Determining whether large observed gaps in living standards are due to inefficiencies matters for policy implications. If there are large gaps in outcomes between urban and rural areas that remain difficult to explain by looking at the distribution of consumption and amenities across locations, there might be substantial efficiency gains from alleviating factors restricting internal mobility. On the other hand, if the observed outcomes appear efficient, intervention requires to be justified by concerns other than efficiency. My findings also imply that while cities share certain features across different contexts, the type of amenities and disamenities denser areas deliver vary substantially not only across levels of development but also across contexts. This emphasizes that models aimed at explaining location choices of individuals ought to be tailored to the particular context and process of urbanization they try to characterize.

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. Working paper.

Ahlfeldt, G. M., S. Redding, D. M. Sturm, and N. Wolf (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, 1–55.

Albouy, D. (2009). The unequal geographic burden of federal taxation. *Journal of Political Economy 117*(4), 635–667.

Algan, Y. and P. Cahuc (2014). *Trust, Growth, and Well-Being: New Evidence and Policy Implications*, Volume 2, pp. 49–120. Elsevier.

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics 129*(3), 1085–1140.

Avery, S. (2012). Lake Turkana & the Lower Omo: Hydrological impacts of major dam and irrigation developments. Working paper, 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 98*(3), 843–863.

Barnhardt, S., E. Field, and R. Pande (2015). Moving to Opportunity or Isolation? Network Effects of a Randomized Housing Lottery in Urban India. Working Paper 21419, National Bureau of Economic Research.

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.

Currie, J., E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin (2009). Does pollution increase school absences? *The 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.

Deaton, A. and A. A. Stone (2013). Two happiness puzzles. *The American economic review 103*(3), 591.

Demombynes, G. and B. Ozler (2002). Crime and local inequality in south africa. World Bank Policy Research Working Paper 2925, The World Bank.

Desmet, K. and J. V. Henderson (2015). The Geography of Development Within Countries. In J. V. H. Gilles Duranton and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, pp. 1457 – 1517. Elsevier.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2015). The geography of development: Evaluating migration restrictions and coastal flooding. Working Paper 21087, National Bureau of Economic Research.

Desmet, K. and E. Rossi-Hansberg (2014). Spatial development. *American Economic Review 104*(4), 1211–43.

Diamond, R. (2015). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. Working paper.

Dorélien, A., D. Balk, and M. Todd (2013). What Is urban? Comparing a satellite view with the Demographic and Health Surveys. *Population and Development Review 39*(3), 413–439.

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.

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*. Forthcoming.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes US richer, smarter, greener, healthier and happier*. Pan Macmillan.

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. Oxford University Press.

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.

Glaeser, E. L., J. D. Gottlieb, and O. Ziv (2015). Unhappy cities. *Journal of Labor Economics*. Forthcoming.

Gollin, D., D. Lagakos, and M. E. Waugh (2014). The agricultural productivity gap. *The Quarterly Journal of Economics 129*(2), 939–993.

Gollin, Kirchberger, and Lagakos (2015). Living Standards Across Space in the Developing World: Evidence from Geo-referenced Micro Data. Working Paper.

Graff Zivin, J. and M. Neidell (2012). The impact of pollution on worker productivity. *American Economic Review 102*(7), 3652–73.

Graff Zivin, J. and M. Neidell (2013). Environment, health, and human capital. *Journal of Economic Literature 51*(3), 689–730.

Hanlon, W. W. (2015). Endogenous City Disamenities: Lessons from Industrial Pollution in 19th Century Britain. Working paper.

Harari, M. (2015). Cities in Bad Shape: Urban Geometry in India. Working paper.

Harris, J. R. and M. P. Todaro (1970). Migration, unemployment and development: A two-sector analysis. *American Economic Review 60*(1), 126–142.

Henderson, J. V., A. Storeygard, and U. Deichman (2014). 50 Years of Urbanization in Africa: Examining the Role of Climate Change. World Bank Policy Research Working Paper 6925, The World Bank.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Henderson, V., T. Squires, A. Storeygard, and D. Weil (2015). The Global Spatial Distribution of Economic Activity: Nature, History and the Role of Trade. Working paper.

ICF International (2012). *Demographic and Health Survey Sampling and Household Listing Manual*. MEASURE DHS, Calverton, Maryland, U.S.A.: ICF International.

Jacobs, J. (1961). *The death and life of great American cities*. Vintage.

Jedwab, R. (2013). Urbanization without Structural Transformation: Evidence from Consumption Cities in Africa. Working paper.

Kahn, M. E. and R. Walsh (2015). Cities and the Environment. In J. V. H. Gilles Duranton and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, 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. *The 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.

Morten, M. and J. Oliveira (2014). Migration, roads and labor market integration: Evidence from a planned capital city. Working paper.

Munshi, K. M. and M. Rosenzweig (2015). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. Working paper.

Nordhaus, W. D. and J. Tobin (1972). Is growth obsolete? In *Economic Research: Retrospect and Prospect, Volume 5, Economic Growth*, pp. 1–80. Nber.

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*. DHS Spatial Analysis Reports No. 8. Calverton, Maryland, USA: ICF International.

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, April). MATCHIT: Stata module to match two datasets based on similar text patterns. Statistical Software Components, Boston College Department of Economics.

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 P. Mieszkowski and M. Straszheim (Eds.), *Current Issues in Urban Economics*. Johns Hopkins University Press.

Stevenson, B. and J. Wolfers (2013). Subjective well-being and income: Is there any evidence of satiation? *The American Economic Review 103*(3), 598–604.

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.

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.

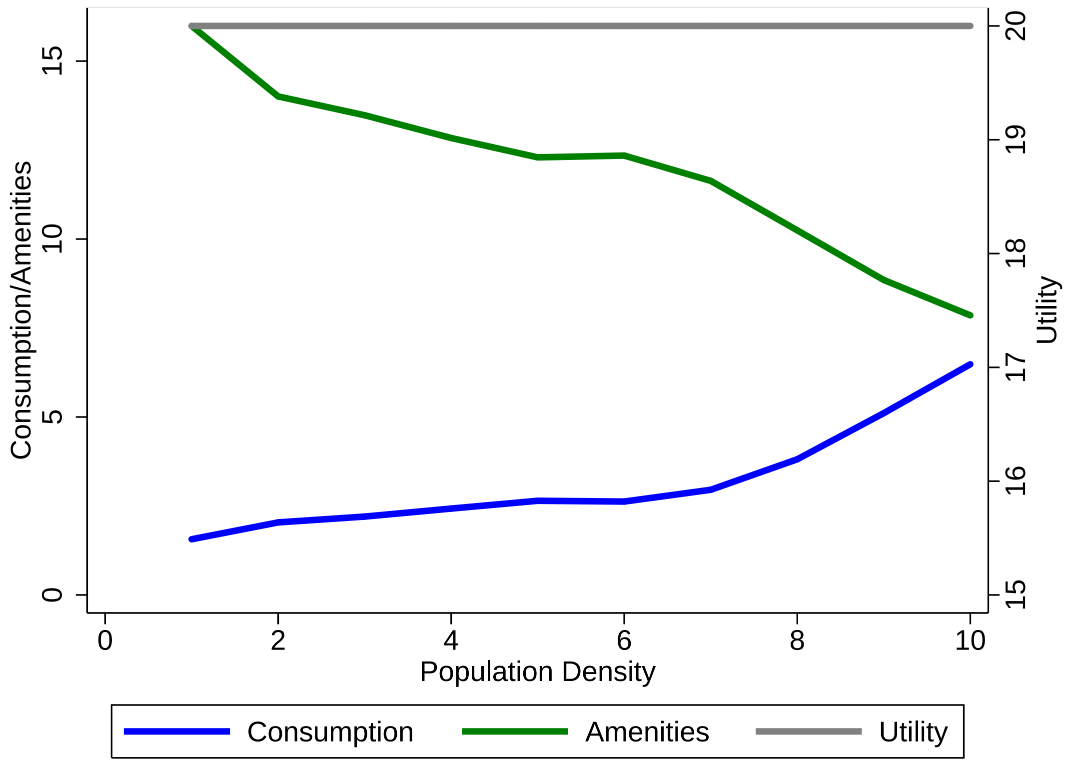
WHO (2014). *Indoor air quality guidelines: household fuel combustion*. World Health Organization.

World Bank (2015). *The Little Green Data*. World Bank Publications.

Young, A. (2014). Inequality, the urban-rural gap and migration. *Quarterly Journal of Economics 129*(2), 939–993.

Tables and Figures

Figure 1: Amenities, consumption and utility across population density space



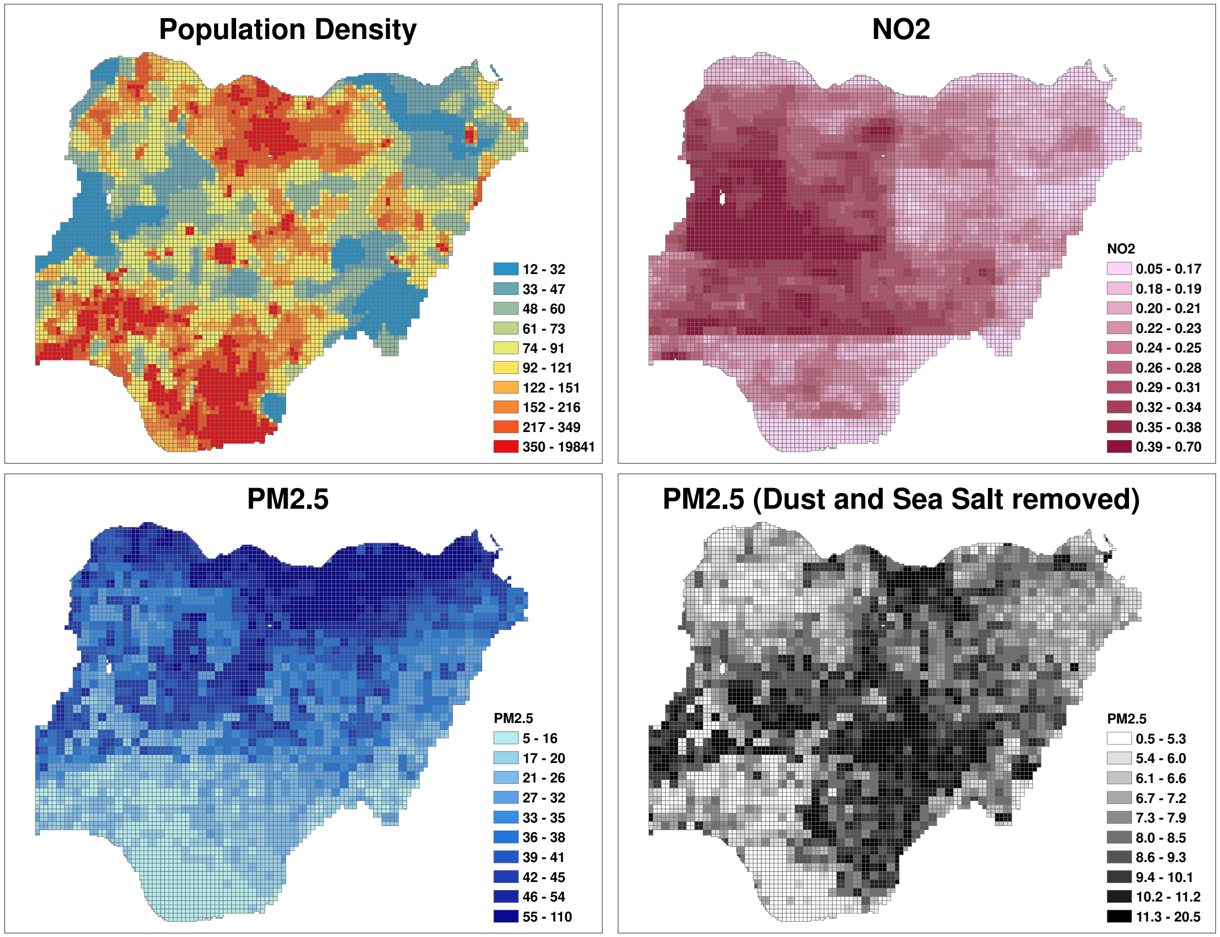
Notes: The figure shows the relationship between consumption, amenities, and utility as predicted in a standard Rosen-Roback model. Consumption c is proxied with an asset and housing quality index, counting the number of durables a household has and housing quality indicators using data from Gollin, Kirchberger, and Lagakos (2015) 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.

Table 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

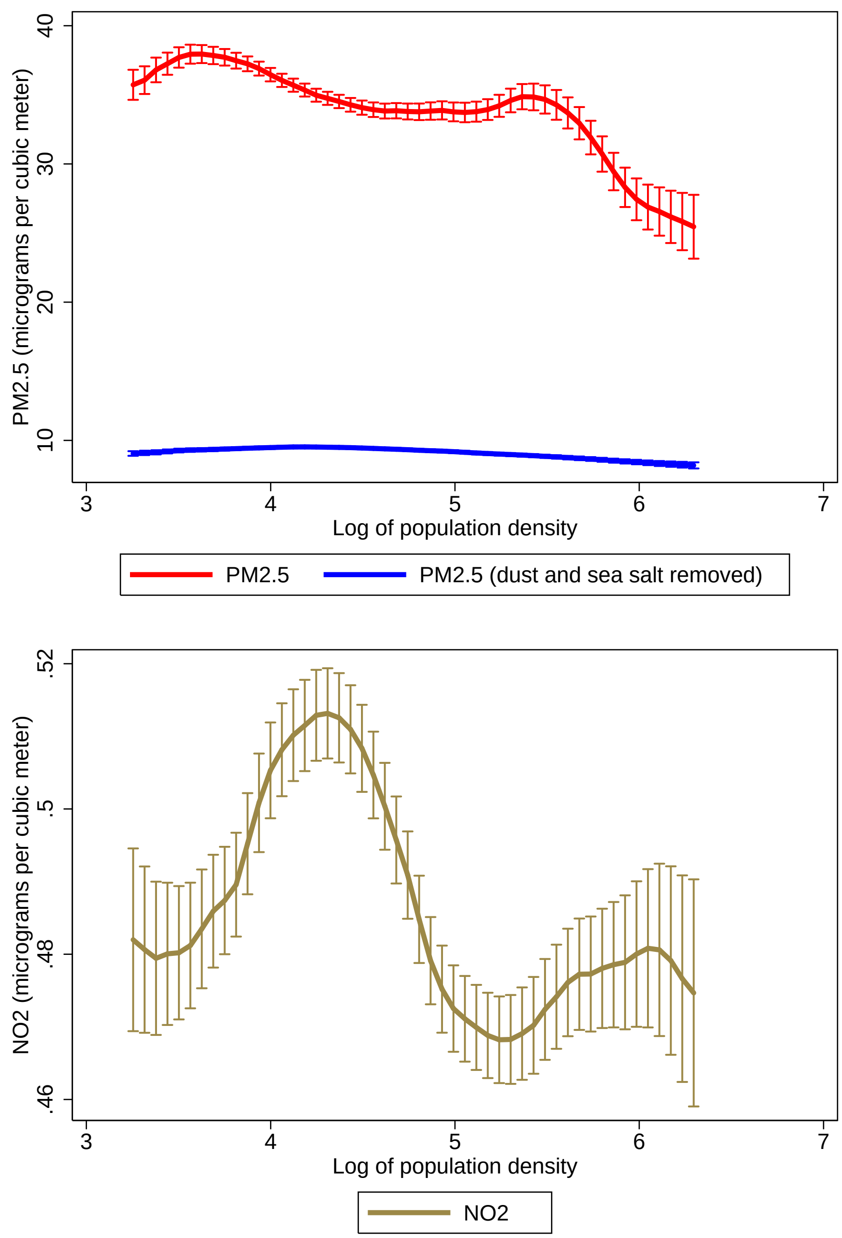
Notes: Column (2) shows the number of individuals in my sample for each of the countries; columns (3)–(5) indicate when a country was added to the Afrobarometer sample. Round 3 took place in 2005, round 4 in 2008, and round 5 in 2011.

Figure 2: Pollution in Nigeria



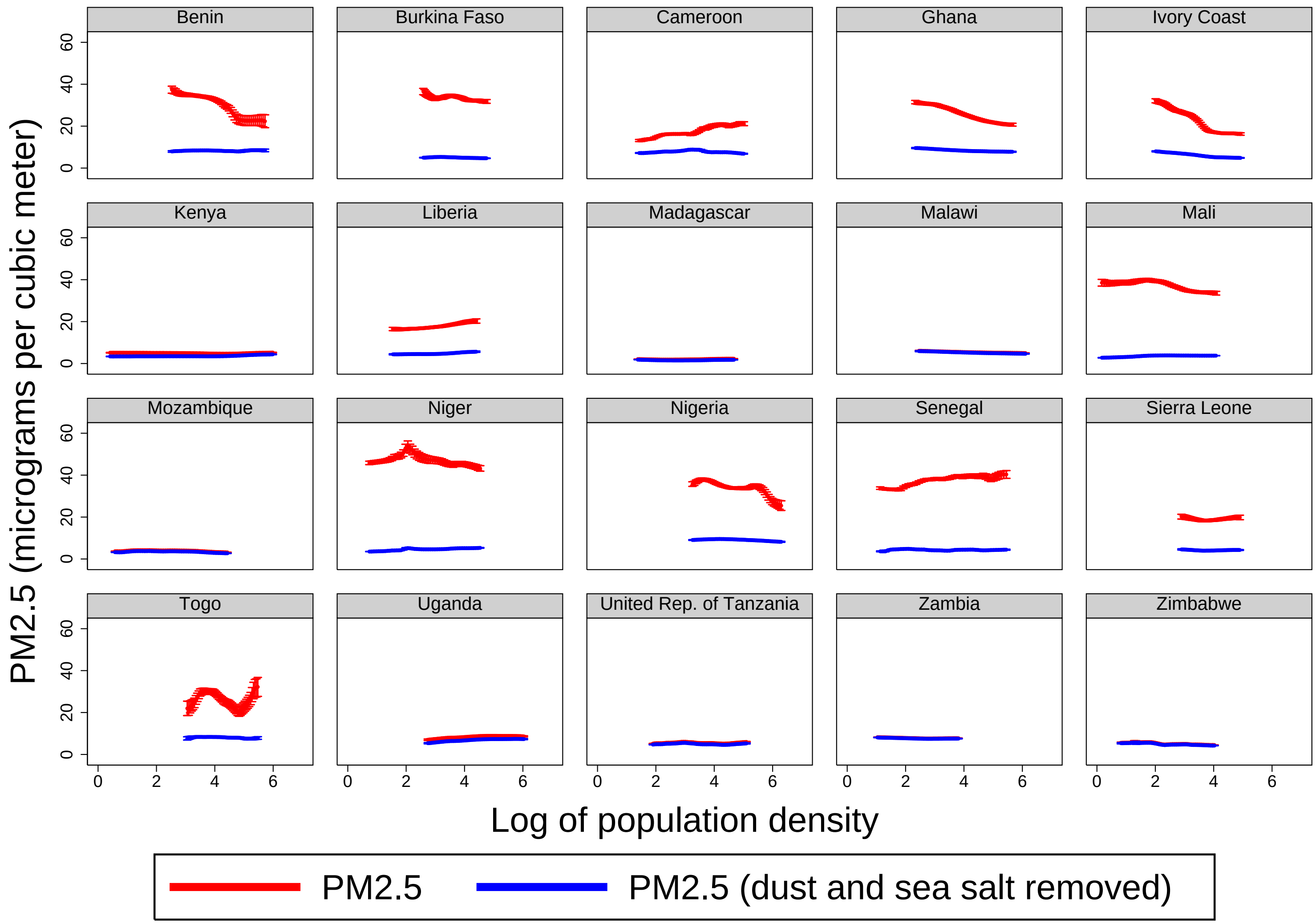
Notes: The top left graph shows the distribution of population density, the top right graph shows the NO2 distribution, and the two bottom graphs show PM2.5, where the graph on the right removes sea salt and dust. Warmer (darker) colors denote higher values, and the bins are formed by dividing the data 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 3: Pollution and population density in Nigeria



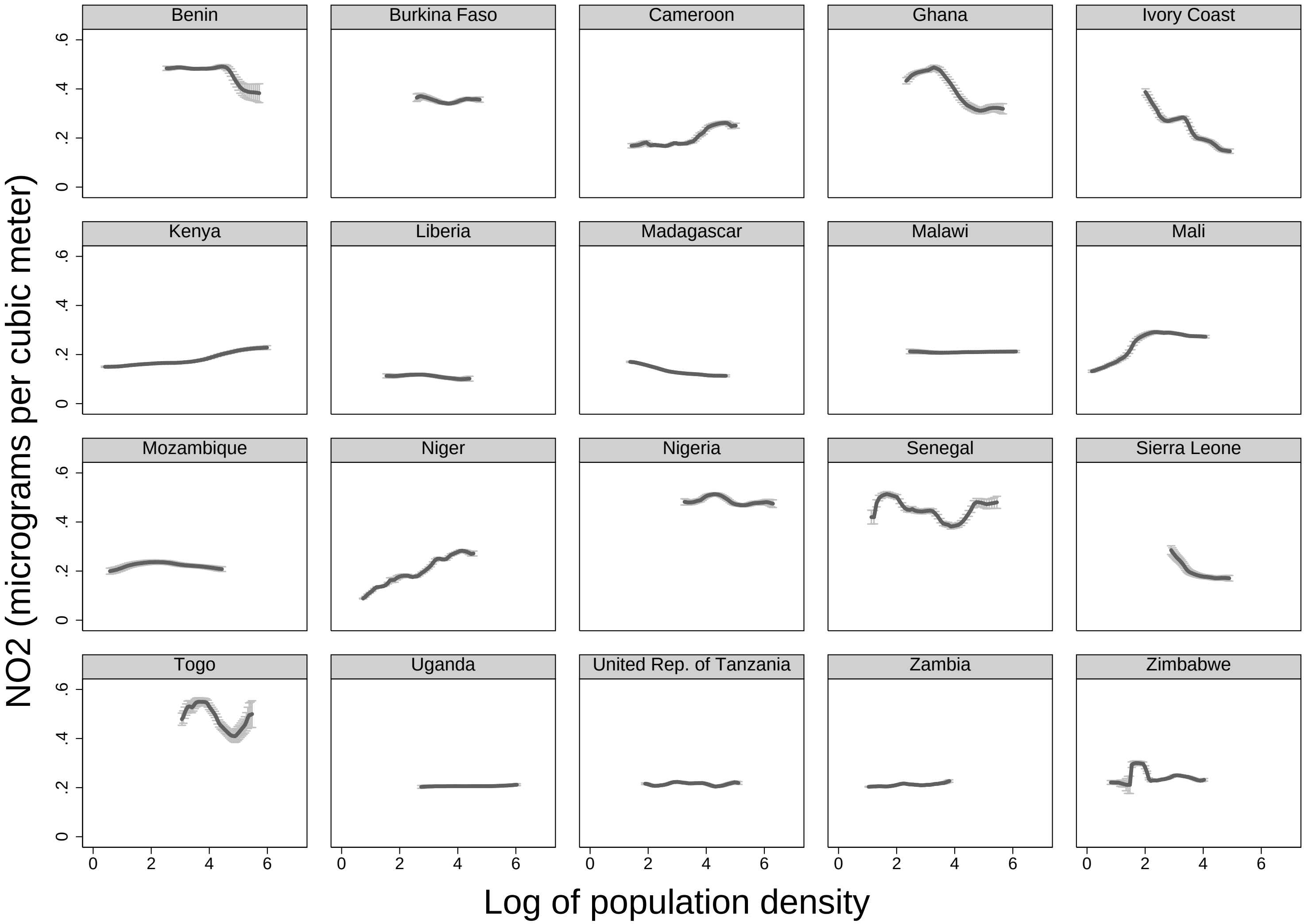
Notes: The figure shows a kernel-weighted local polynomial regression of the level of pollution on the log of population density in Nigeria using data from the entire country, and plotting 95 percent confidence intervals. The top panel shows the results for PM2.5 and the bottom panel shows NO2 levels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

Figure 4: PM2.5 concentration



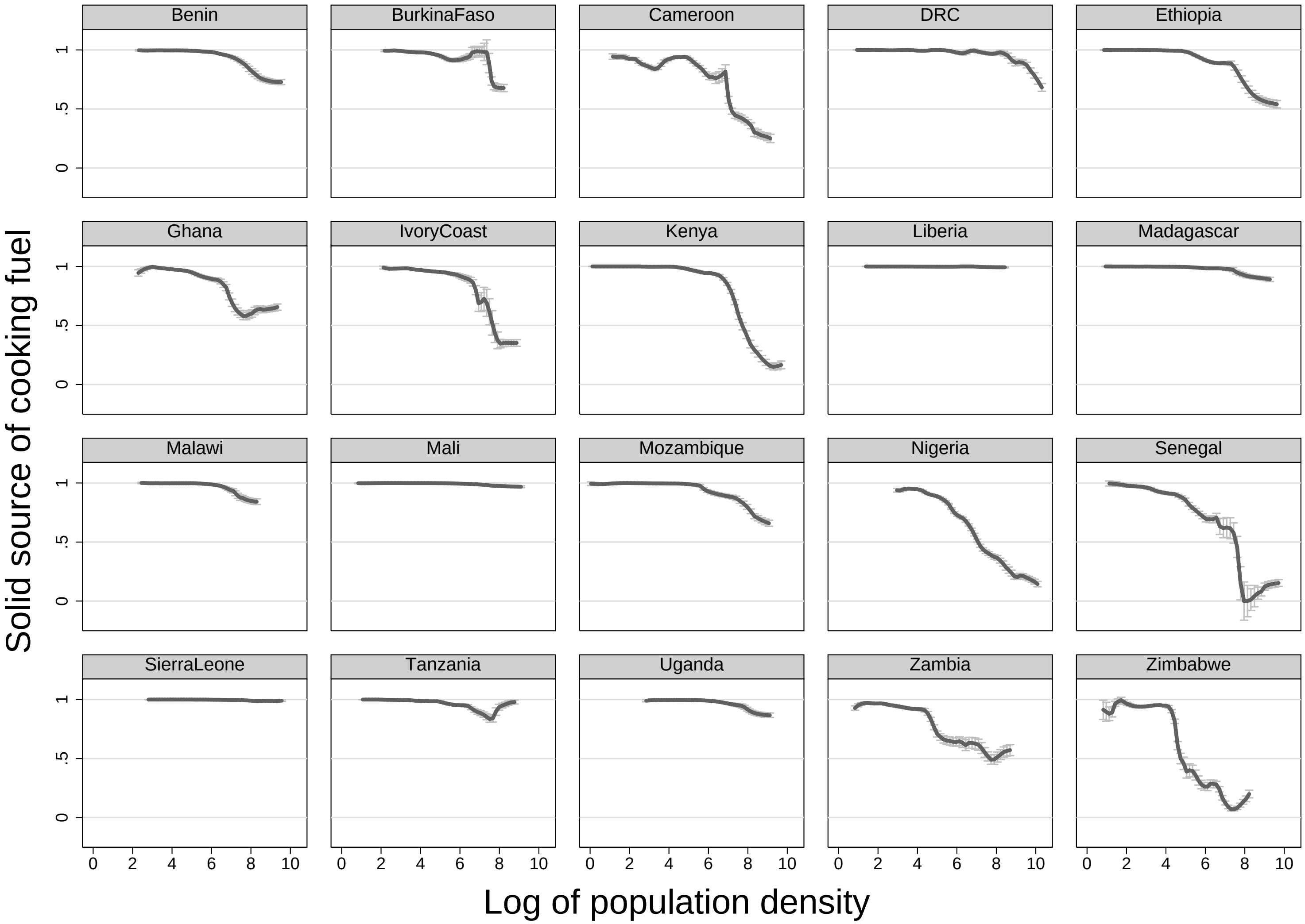
Notes: The figure shows a kernel-weighted local polynomial regression of the level of PM2.5 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density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

Figure 5: Nitrogen dioxide concentration across Africa



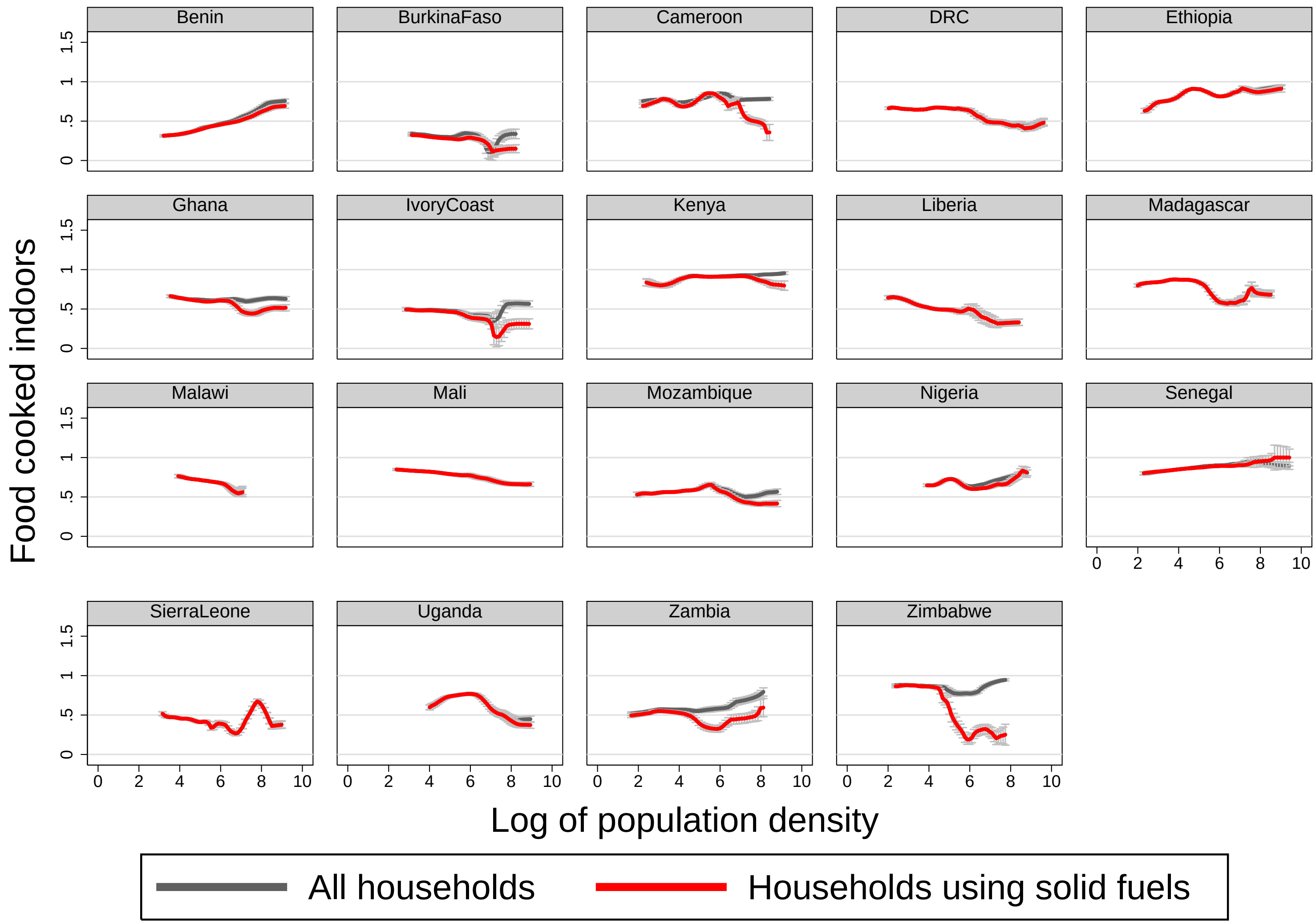
Notes: The figure shows a kernel-weighted local polynomial regression of the level of NO2 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density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

Figure 6: Solid type of fuel for cooking



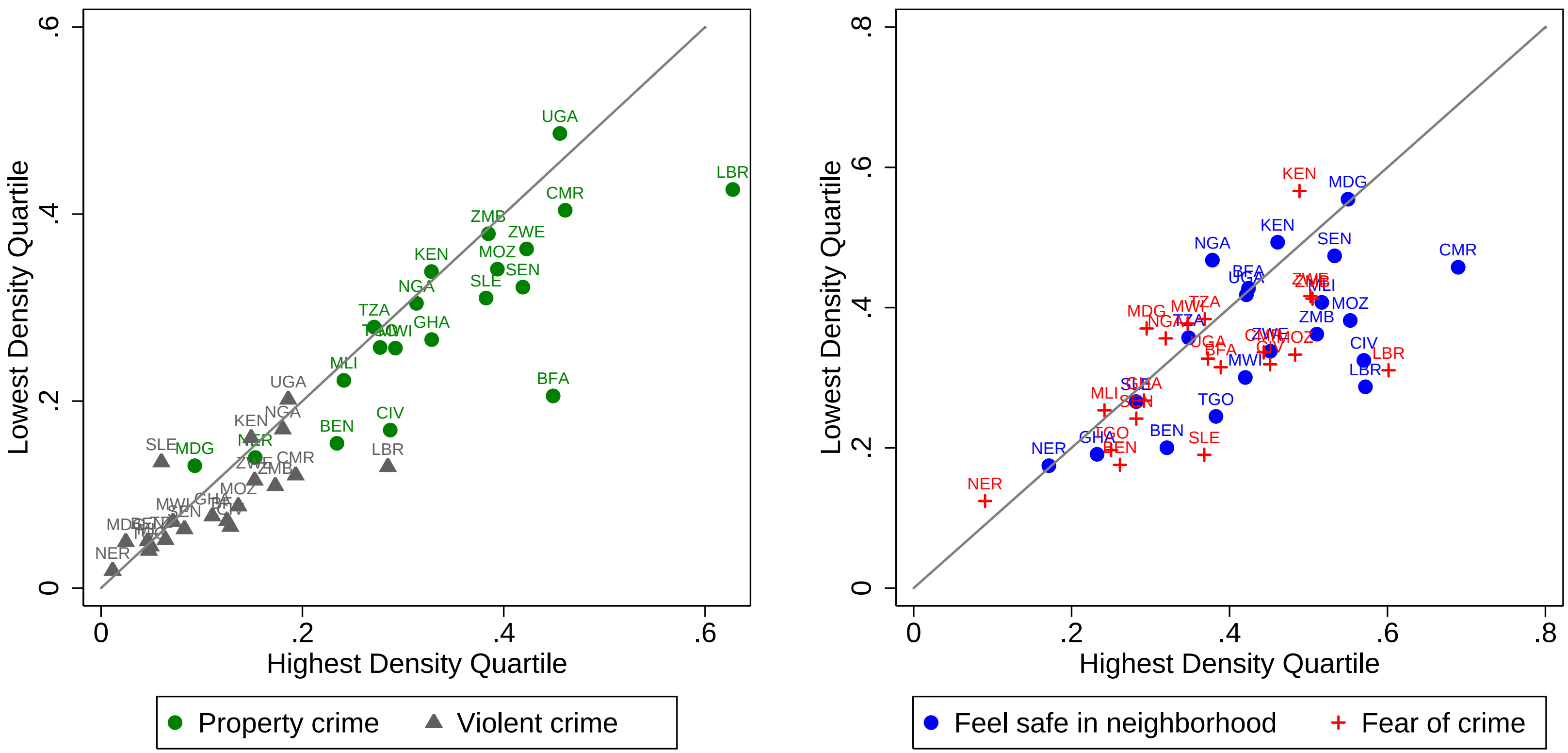
Notes: The figure shows a kernel-weighted local polynomial regression of the a binary variable that is equal to one and zero otherwise if a household uses a solid source of cooking fuel on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

Figure 7: Food cooked indoors



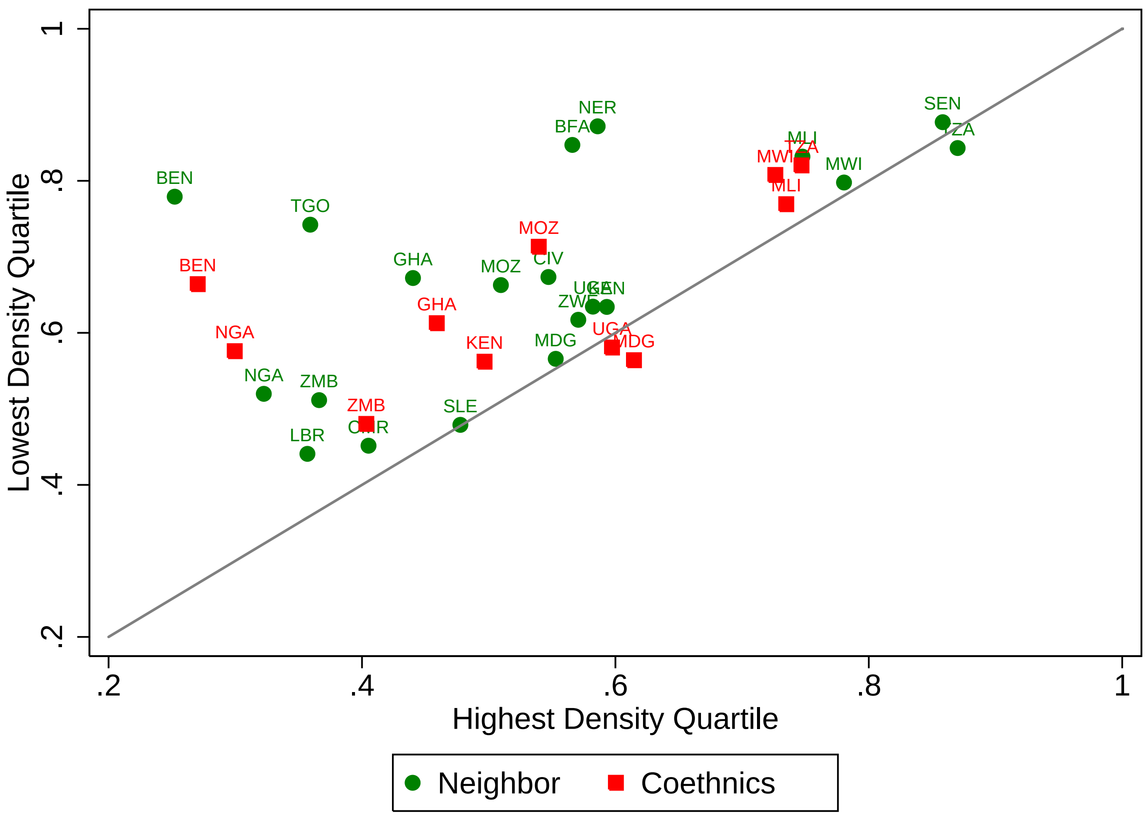
Notes: The figure shows a kernel-weighted local polynomial regression of the a binary variable that is equal to one and zero otherwise if a household household cooks inside on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density distribution.

Figure 8: Crime - Afrobarometer



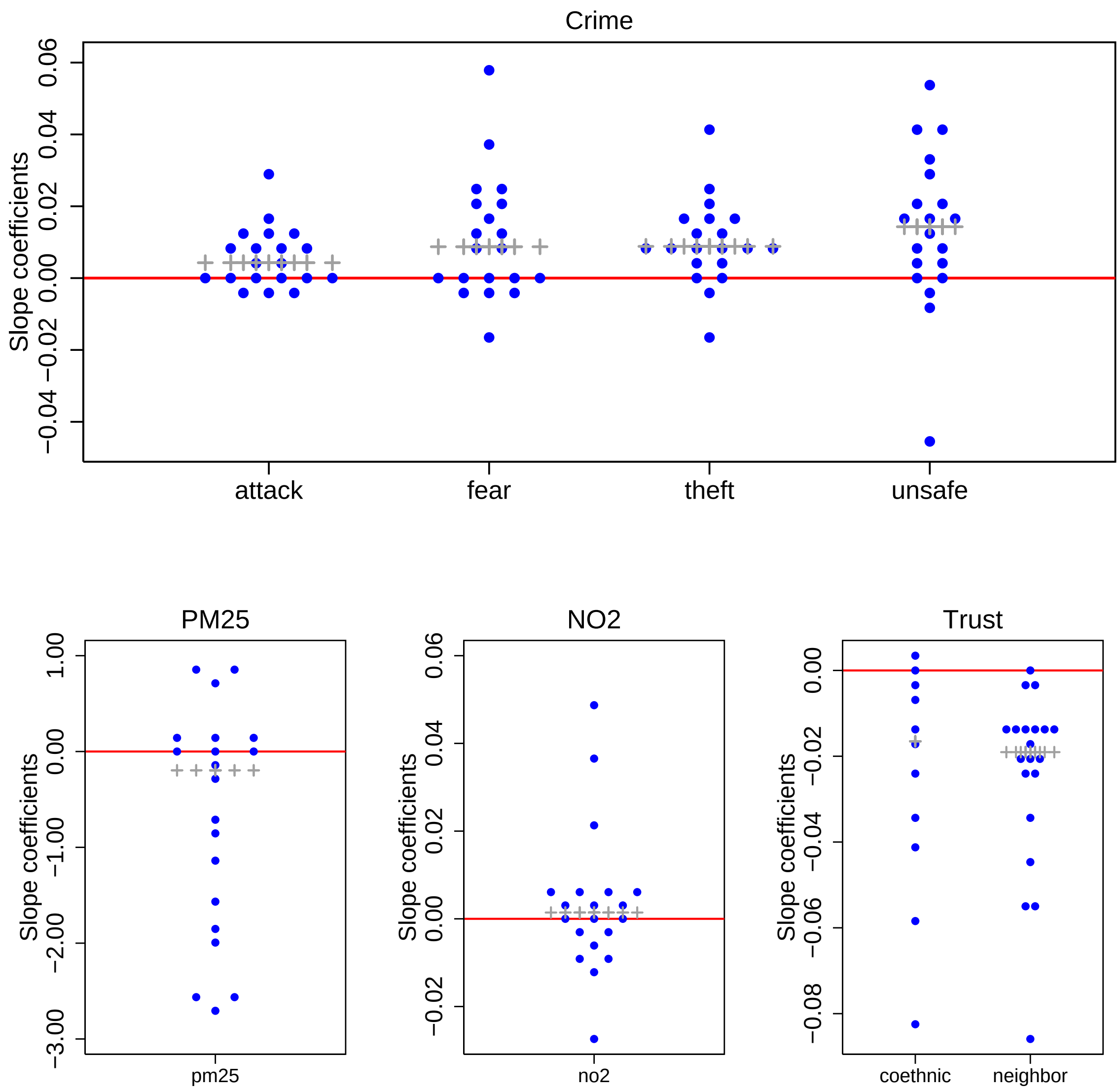
Notes: The figure compares the proportion of individuals who report experiences of crime (shown in the figure on the left) or fear thereof (shown in the figure on the right) in different population density quartiles. The x-axis shows the highest density quartile, and the y-axis shows the lowest density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

Figure 9: Trust - Afrobarometer



Notes: The figure compares the proportion of individuals who report mistrusting their neighbors or coethnics 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 trust levels are.

Figure 10: 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.

Table 2: The relationship between disamenities and density

	Theft	Attack	Fear	Unsafe
	(1)	(2)	(3)	(4)
Log of Pop density	0.008	0.004	0.007	0.017
	(0.003)**	(0.002)	(0.005)	(0.004)***
Obs.	65922	63795	65784	30306
R^2	0.037	0.032	0.029	0.042

	PM25	NO2	Mistrust Neighbor	Mistrust Coethnic
	(1)	(2)	(3)	(4)
Log of Pop density	-.565	0.009	0.027	0.026
	(0.233)**	(0.005)	(0.007)***	(0.009)**
Obs.	65896	65995	46022	15635
R^2	0.752	0.422	0.103	0.082

Notes: Each column shows the estimated population density gradient for the pooled sample, where I include fixed effects for countries and rounds. Robust standard errors are shown in parenthesis; *, **, *** denote significance at 10%, 5% and 1% levels using Benjamini-Hochberg (1995) q-values to correct for multiple hypothesis testing.

Appendix

A. Linking survey and population density data

A.1. Geo-locating Afrobarometer respondents

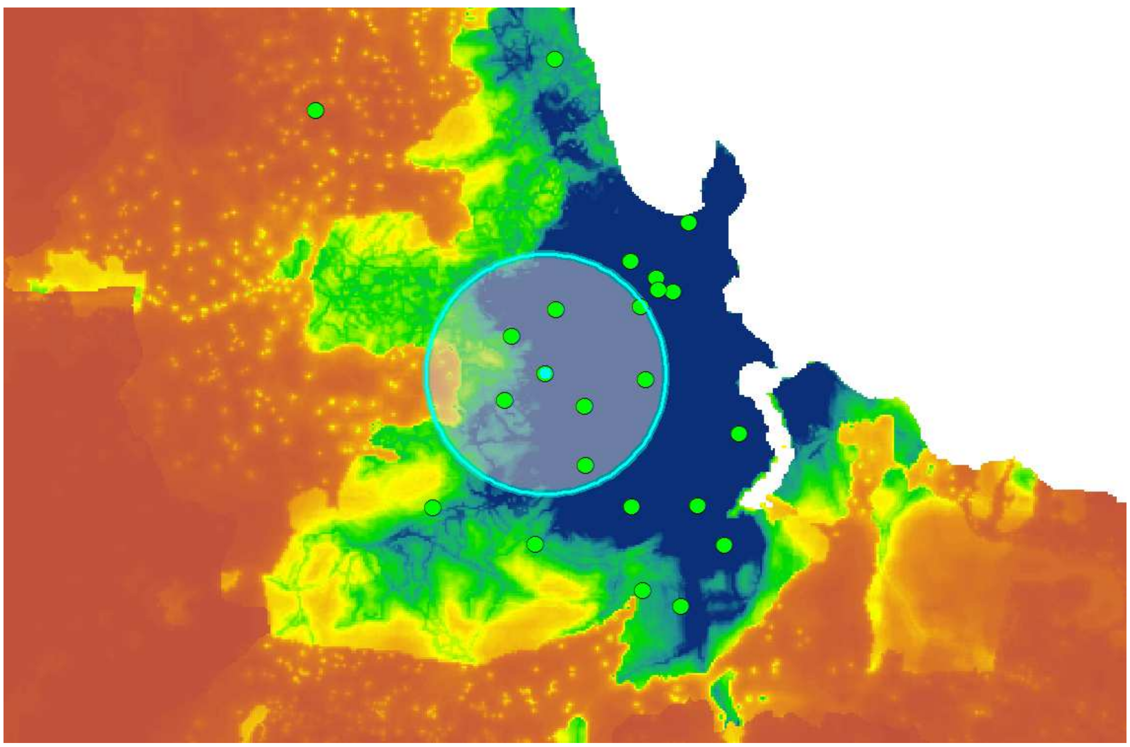
The Afrobarometer surveys do not coordinates of respondents, but record the village, district and region names. The 2011 round provides four different administrative names. I use a matching algorithm that matches village names and other provided administrative names to locations as listed in gazetteers; specifically, I follow Nunn and Wantchekon (2011) and use the fallingrain gazeteer available on www.fallingrain.com. This website provides a list of locations where each location is assigned an id along with several 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 I keep the most recent entry.

The 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, I 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, I perform a series of exact matches based on the village name from Afrobarometer with the asciiname listed in the gazetteer; in this stage I find already between thirty-six and forty percent of locations. 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. This allows me to match another 2–2.5 percent based on the village location. I 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. I 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 I have only district and region names in addition to the village names, this step finds 49–52 percent of the locations. Third, I match on the region name which finds another 7–9 percent of the sample. Finally, to catch any remaining misspellings I perform a fuzzy match based on similar text patterns between the village name and the asciiname using a command developed by Raffo (2015). I 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, I am able to match between 85 and 95 percent of locations in each round.

In addition to random checks of the identified locations I use the 2005 data to check the consistency between my algorithm and Nunn and Wantchekon (2011)'s location data. For the subset of locations for which they provide geo-locations, I find that the median distance between their location and my location is 10km. Further, considering the population density data vary largely at the district and region level, I 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 my data is 0.6 with a p-value of 0.000.

A.2. Spatial linking of DHS

Figure A.1: 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.

A.3. Samples

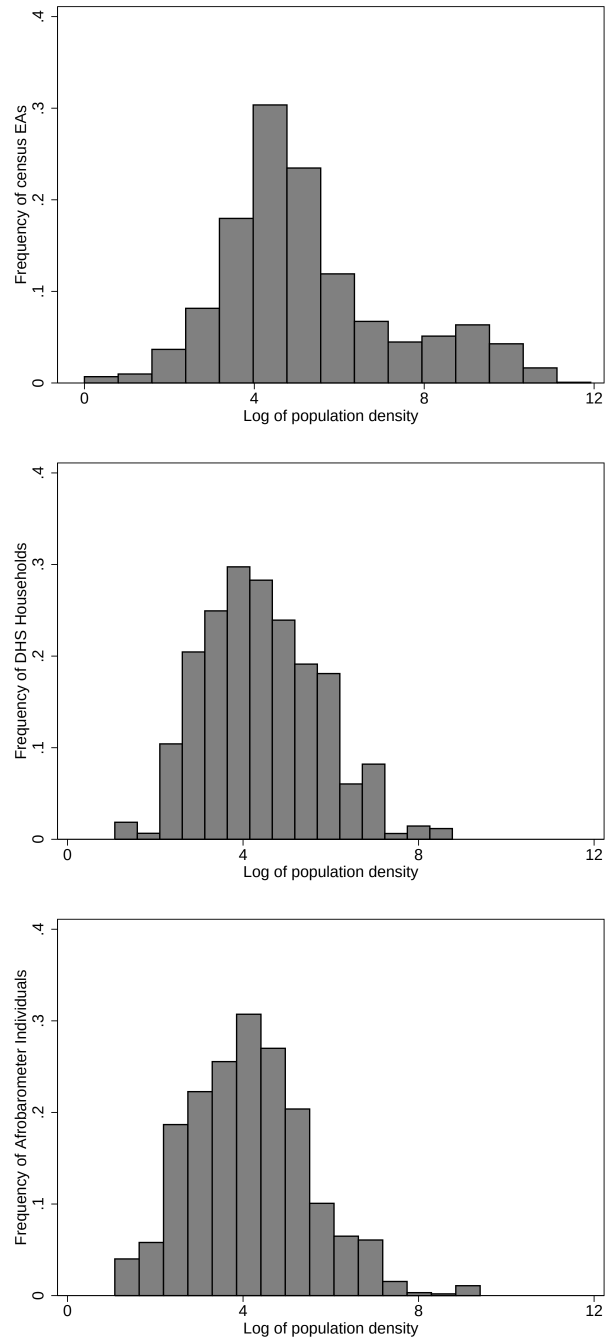
The DHS sample is in general chosen to be representative at the second administrative level, as well as rural/urban within the second administrative level. While the DHS aim to make survey instruments and samples comparable across countries, the exact sampling differs according to the particular survey.[27] The target population of most DHS surveys are women aged 15-

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

49 and children under the age of five living in residential households with the most common sampling following a two-stage cluster sampling procedure (ICF International, 2012). 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. The Afrobarometer sample is in general chosen to be representative at the national level of the voting population of a particular country.

Unless the sampling frame is specifically selected to match the population along the lines of population density, it is possible that the distribution of the survey sample according to population density might not match that of the entire population. In practice, the cases that I have examined show very little effort to oversample or undersample with respect to population density. For Tanzania I can compare the population density distribution of the Afrobarometer and DHS clusters with those of the overall population from the census data where I weight the population density of enumeration areas by the population. As is evident from Figure A.2, both the Afrobarometer survey as well as the DHS appear to capture a sample that covers a wide range of population densities.

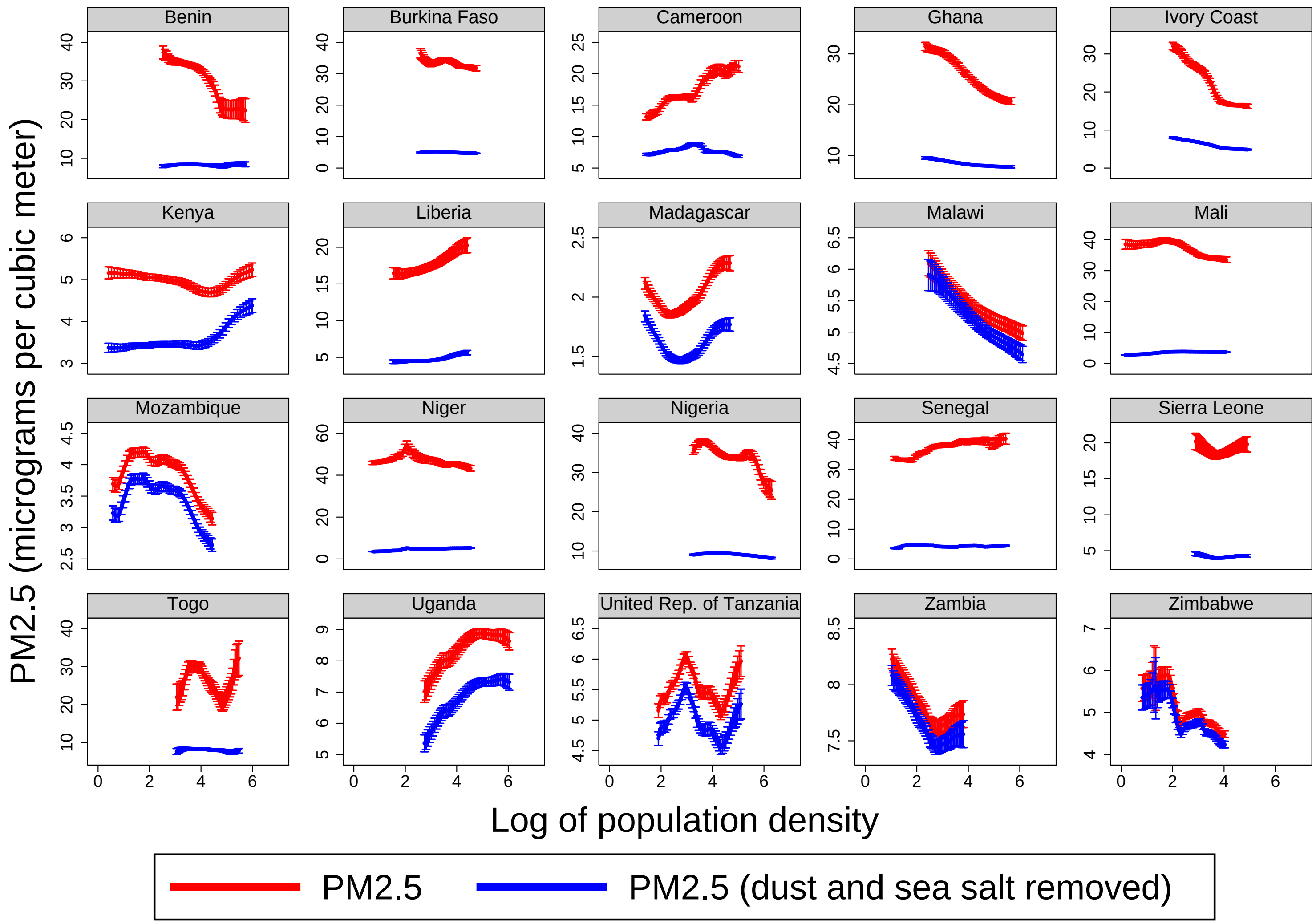
Figure A.2: Distribution of population, DHS and Afrobarometer respondents in Tanzania



Notes: The top figure shows the distribution of the population using the 2002 enumeration area census data and the total population in each enumeration area as sample weights. The middle graph shows the distribution of population densities from the DHS data. The bottom graph shows distribution of clusters from Afrobarometer data. For expositional simplicity the top graph excludes 112 enumeration areas that have a log of population density above 12.

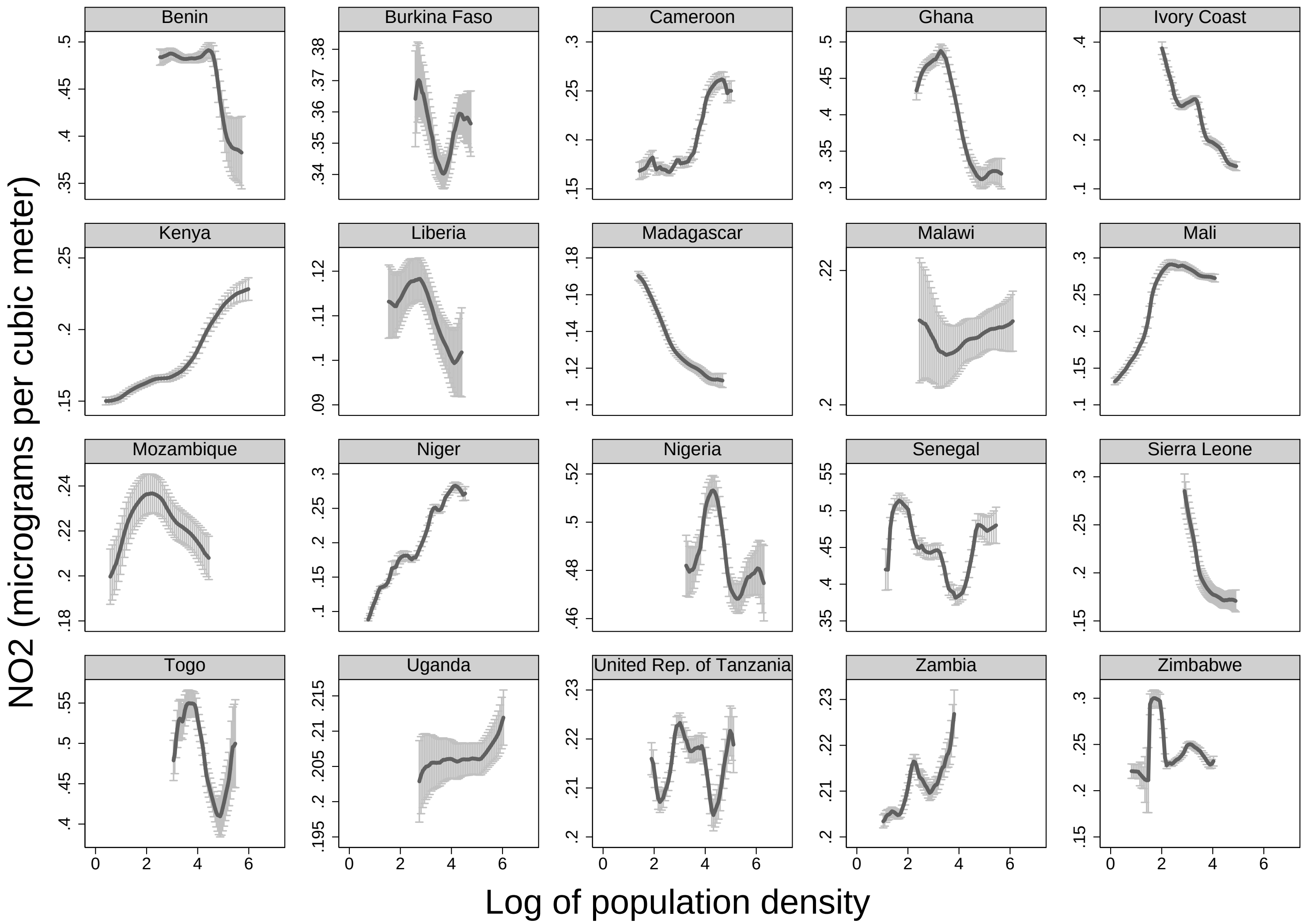
B. Additional Tables and Figures

Figure B.1: PM2.5 concentration



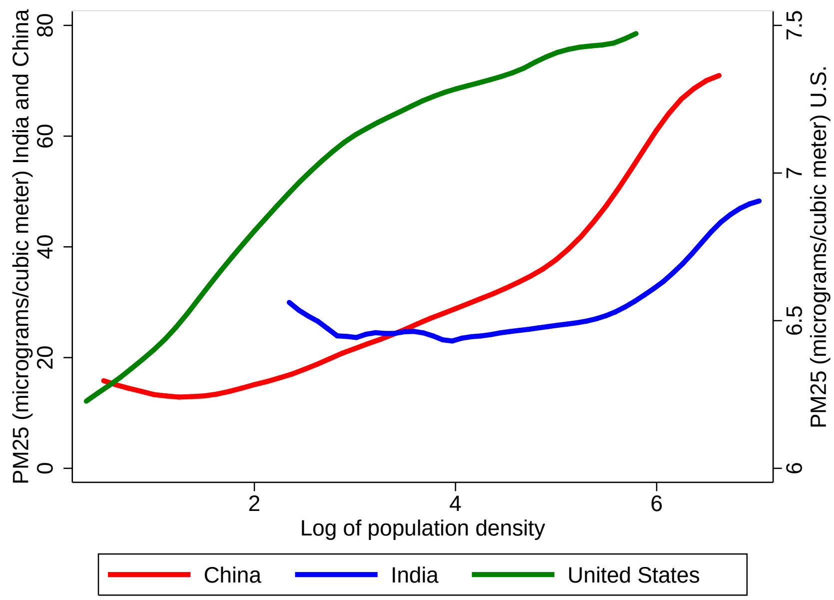
Notes: The figure shows a kernel-weighted local polynomial regression of the level of PM2.5 on the log of population density, including 95 percent confidence interval. Taking the log of population density removes uninhabited pixels. I remove the top and bottom five percentile of the population density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

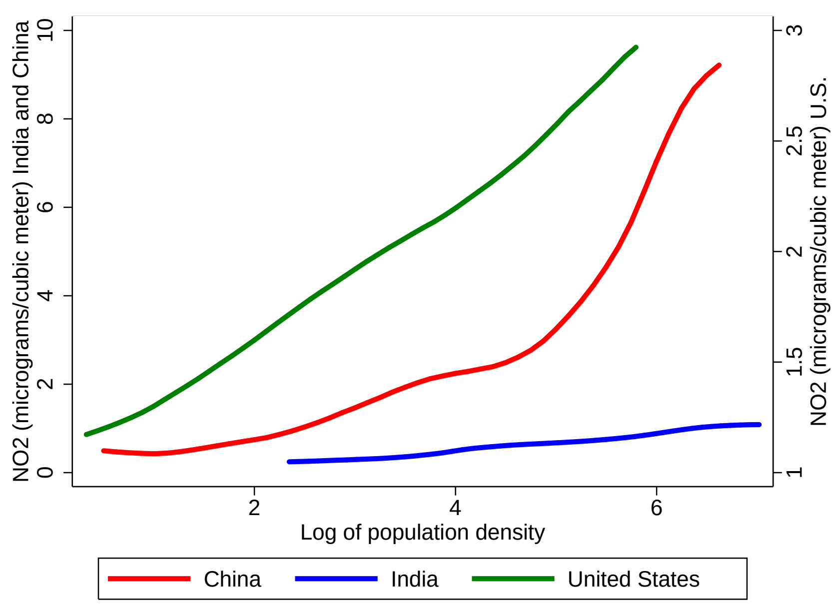
Figure B.2: Nitrogen dioxide concentration across Africa



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 B.3: 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.