IN SEARCH OF A SPATIAL EQUILIBRIUM IN THE DEVELOPING WORLD*

Douglas Gollin†
University of Oxford

Martina Kirchberger§
Trinity College Dublin

David Lagakos‡
UCSD and NBER

May 2018

Abstract

In most developing countries, there is a large gap in average consumption per capita between urban and rural areas. One appealing interpretation of this gap is that it reflects a spatial equilibrium, in which the higher consumption levels of urban areas are offset by lower non-monetary amenities. This paper draws on new high-resolution evidence to document how non-monetary amenities vary across space within 20 Sub-Saharan African countries. We focus on measures of public goods, crime and pollution. We find that in almost all countries, and for almost all measures, the quality of these amenities is non-decreasing in population density.

*Email: † douglas.gollin@qeh.ox.ac.uk, § martina.kirchberger@tcd.ie, ‡ lagakos@ucsd.edu. For helpful comments we thank Deborah Balk, Joe Cummins, Banu Demir-Pakel, Stefan Dercon, Dave Donaldson, Ben Faber, Marcel Fafchamps, Ed Glaeser, Darby Jack, Kelsey Jack, Matt Kahn, Pat Kinney, Pramila Krishnan, Horacio Larreguy, Ethan Ligon, Matthew Neidell, Natalia Ramondo, Chris Small, Daniel Sturm, Luke Valin, David Weil, and seminar participants at Barcelona, Bristol, Columbia, Copenhagen, Edinburgh, the Ethiopian Development Research Institute, IMF Research Department Development Macroeconomics division, LSE, Notre Dame, Nottingham, Sussex, Trinity College Dublin, UC Riverside, as well as conference participants at the Edinburgh Structural Change Conference, the NEUDC, CEPR and CSAE conferences. For excellent research assistance, we thank Jessie Wang, Eoin Dignam and Abed Mutemi, as well as the Bill and Melinda Gates Foundation for supporting Mutemi's work on this project. We would like to thank Jeff Geddes and Aaron van Donkelaar for sharing the pollution data and their advice. Kirchberger is grateful for many helpful conversations and suggestions while hosted by the Center for International Earth Science Information Network at Columbia University. This paper was partly supported through the Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. All potential errors are our own.

1. Introduction

A body of recent evidence documents that urban-rural consumption gaps are particularly large in developing countries (Ferré, Ferreira, and Lanjouw, 2012; Young, 2014), as are gaps in average labor productivity between non-agricultural and agricultural workers (Gollin, Lagakos, and Waugh, 2014). In an accounting sense, these gaps are important for understanding why developing countries have such low aggregate income, because a large fraction of workers in developing countries live in rural areas and work in agriculture (Caselli, 2005; Restuccia, Yang, and Zhu, 2008; Vollrath, 2009; McMillan and Rodrik, 2011). Furthermore, urban-rural wage gaps may reflect a misallocation of resources (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009), with too many workers in less-productive rural areas. Thus, from a macroeconomic perspective, it is important to understand the determinants of these gaps.

What explains these large rural-urban gaps? An obvious potential explanation is that cities are more expensive places to live, so that nominal differences in income, wages, or productivity might be offset by living costs, leading to an equalization of real consumption levels between rural areas and cities. This issue is addressed to a significant extent in the consumption measures of Young (2014), which cover *real* outcomes, such as housing quality and ownership of durable goods. A second potential explanation of the gaps is that they might arise through the sorting of heterogenous workers. Evidence suggests that urban areas attract the most educated and productive workers (Herrendorf and Schoellman, 2014; Young, 2014; Hamory Hicks, Kleemans, Li, and Miguel, 2017). Still, experiments have shown that inducing workers to migrate to urban areas leads to substantial increases in their households' consumption, at least temporarily (Bryan, Chowdhury, and Mobarak, 2014; Akram, Chowdhury, and Mobarak, 2017). Thus, it seems unlikely that the entire rural-urban gap is due to sorting.

In this paper, we address a different hypothesis that has been at the heart of urban economics for decades, namely that non-monetary amenities offset the higher real consumption levels of urban areas. This idea underlies the concept of *spatial equilibrium* – a theoretical construct used to explain how economic agents locate through space (Rosen, 1979; Roback, 1982). The rationale is simple: if any region offered a better bundle of consumption and amenities than the rest, then utility-maximizing agents would move into the better region until any arbitrage opportunities were gone. Although the concept of a spatial equilibrium has proven useful in a wide variety of applications in advanced economies (see e.g. Glaeser and Gottlieb, 2009; Kline and Moretti, 2014; Redding and Rossi-Hansberg, 2017) few studies have attempted to assess whether amenities differ systematically between rural and urban areas in developing countries. Thus, it is an open question how much, if any, of the rural-urban gaps in developing countries can be accounted for by lower amenities of city life relative to rural life.

We help fill this gap by drawing on new detailed spatial evidence on three categories of amenities that are frequently invoked as prime candidates for rural-urban differences in real consumption in developing countries. These categories are: public goods provision, crime, and air pollution. All three of these could plausibly generate urban disamenities. Vast sections of urban areas in the developing world lack access to public goods like electricity and piped water, and in fact the absence of these amenities is often used to define urban "slums." Urban areas are also often thought to have higher rates of crime (as is the case in the United States). Air pollution could relate to the presence of more automobiles or more concentrated industrial activity. Indeed, in the United States and Western Europe, urban areas have suffered from air pollution; even in pre-industrial times, rural areas were seen as places where people could escape the odors and "miasmas" of urban air (Cutler and Miller, 2005; Costa and Kahn, 2006; Kesztenbaum and Rosenthal, 2016). In this sense, better air quality has long been viewed as an important compensating differential of rural life.

To measure these amenities consistently across countries and at a fine level of geographic detail, we link nationally representative household survey data and satellite-derived measures of pollution with grid cell-level data on population density. An advantage of our approach is that we do not rely on administrative definitions of "urban" and "rural." These binary categories often reflect arbitrary boundaries and cut-offs that differ across space and time. Our data allow us to examine the distribution of amenities and disamenities across locations that differ in population density.

We find that, across the countries that we study, almost all amenities that we consider improve at least weakly with population density. Publicly provided goods are systematically more available in areas with higher population density. Property crime and violent crime are, unfortunately, high throughout density space in most countries. These categories of crime appear to be marginally higher in urban areas, but the differences are statistically insignificant in most countries. The same is true for the percent of adults reporting fear of crime in their homes and feeling unsafe in their neighborhoods. Using plausible values of willingness to pay to avoid crime, the differences between rural and urban areas are dwarfed by the much larger differences in average income.

Our analysis of air quality distinguishes between outdoor and indoor pollution. For outdoor air quality, we consider two pollutants particularly detrimental to human health: fine particulate matter (PM2.5) and nitrogen dioxide (NO2). Both of these are linked to a variety of respiratory diseases and other health problems. Perhaps surprisingly, we find that outdoor air pollution is largely unrelated to population density in sub-Saharan Africa. Fine particulate matter is most prevalent in areas near the Sahara, due to dust and sand in the air. Even excluding dust and sea salt, PM2.5 concentrations are quite similar in urban areas and rural areas, on average, as are NO2 concentrations. Overall magnitudes are also quite low, particularly relative to other developing countries such as India and China. This surprising finding largely reflects the lack of manufacturing activity and the relatively low levels of economic activity in Africa. Arguably more relevant for many Africans is *indoor* air pollution, which primarily comes from cooking in unventilated spaces with solid fuels such as coal or wood. In each of our countries, we find that the fraction of households using solid fuels to cook indoors falls sharply with population density. Thus, people in the rural areas of our countries face worse air quality, on average, than their urban counterparts. Although we have no comparable measures on water pollution, child health indicators – which can be quite sensitive to water quality – are systematically better in densely populated areas.

Thus, at least for these three categories of amenities, it seems unlikely that rural-urban income gaps simply compensate for disamenities of urban life. Still, we obviously cannot address all possible amenities, and there could be some other undesirable non-monetary feature of urban life that compensates for its higher real wages. To address this possibility, at least partially, we consider two additional measures of well-being: subjective welfare measures and rural-urban migration rates. We show, for a smaller set of countries, that reported happiness and life satisfaction are generally higher on average in urban areas than in rural areas. In addition, people are voting with their feet: in almost every country we examine, net migration flows show strong movements of people from rural areas to towns and cities.

Our findings suggest that the urban-rural consumption gaps in the developing world do not fit neatly into a static spatial equilibrium framework in which real consumption differences are offset by amenities and disamenities. Instead, a more plausible explanation may be that spatial wage gaps may reflect high migration costs and lifecycle effects, as in the dynamic models of Eckert and Peters (2017) and Morten and Oliveira (2017), and consistent with the ideas of Topel (1986). Future work should focus on understanding the types of frictions that impede movement of workers to areas with higher average productivity.

2. Amenities Data

Until recently, measuring amenities and other living-standards metrics across space in the developing world was not feasible. Exploiting progress in surveying and mapping technology, we construct a new dataset that spatially links household micro data, satellite-derived measures of pollution, and gridded estimates of population density. The micro data come from are high-quality nationally representative surveys that cover large numbers of households in developing countries, using consistent methodologies and definitions across countries. The main surveys we employ are the Demographic and Health Surveys (DHS), Living Standards Measurement Surveys (LSMS), World Values Surveys and Afrobarometer surveys. The DHS and LSMS data focus on variables related to population, health, and nutrition, while Afrobarometer focuses on attitudes towards democracy and governance, including the availability of publicly provided goods and services and experiences of crimes. We use the World Values Survey for subjective well-being. For each of these surveys, we have GPS coordinates or location names that allow us to identify the locations of survey households. Our data on outdoor air pollution come from satellite-derived estimates of PM2.5 and NO2 concentrations. To measure population density, we use data from the Gridded Population of the World Version 4 (GPWv4), which provides population density estimates at a resolution of 30 arc-seconds, corresponding to about 1km at the equator (Center for International Earth Science Information Network, 2015). We outline the main choices related to using these data here and in the following section, when we present our results. Further details are found in the Appendix.

Our selection of countries is based on a simple set of criteria. We select low-income African countries, larger in area than 50,000 square kilometers, in which a survey has been done since 2005; we further restrict the sample to those surveys that collected spatial identifiers and for which population density data are available at a sufficiently high spatial resolution (at least 40 regions by country). For the DHS data, we are left with a sample of 276,051 households across 20 African countries, as listed in Appendix Tables 4 and 3, covering countries with a combined population of about 770 million people. We focus on Africa due to its enormous real rural-urban consumption gaps in the world (Young, 2014), and it coverage of so many of the world's developing countries.

We next combine the different sources of data step by step. We link the individual data from the DHS surveys with local measures of population density by taking 5 km buffers around both urban and rural DHS clusters. An important consideration is how representative our samples are across different levels of population density. We discuss the sampling protocol of the surveys in the Appendix. We are able to show that the survey data (both DHS and LSMS) cover a wide range of densities and correspond well to the population density distributions of geo-coded census data. The Afrobarometer surveys did not collect the GPS location for respondents; however, they do record the location name for each respondent. We develop an algorithm that performs a series of exact and fuzzy matches of location names, relying on data from a global gazetteer that contains the latitude and longitude of a location. For each location, we then extract the population density value. Both the pollution data and the population density data are gridded, making it straightforward to link them. We construct a fishnet grid at the same resolution as the pollution data. For each pixel, we compute the average pollution measure and the average population density from the GPWv4.

To present data in a compact fashion for our whole set of countries, we divide locations within each country into quartiles of population density. An advantage of looking at density quartiles is that we avoid the problems of defining "urban" and "rural" locations, which may reflect arbitrary administrative categories and are not standardized across countries.

3. Public Goods

In this section, we present our empirical findings for publicly provided goods and services at different levels of population density. We focus on seven publicly provided goods that are important from a welfare perspective and which together form the bulk of public expenditures: electricity, piped water, sewage, public schools, health clinics, police stations and roads. It is worth noting that the absence of public goods, in particular electricity and piped water, is one of the criteria used to define urban "slums" by the World Bank and United Nations agencies. Thus, to the extent that developing world cities are characterized by pervasive slum conditions, one might expect to find relatively low provision of public goods in urban areas.

Table 1 reports on the allocation of these publicly provided goods across density quartiles in our 20 countries. On average, these publicly provided goods are more prevalent in more densely populated areas, although not all these differences are statistically significant. For electricity, around 39 percent of census enumeration areas in the least dense quartile have access, compared to 42 percent in the second quartile, 48 percent in the third, and 72 percent in the densest quartile. For 10 of the 20 countries, enumeration areas in the least dense quartile have a significantly lower level of electricity coverage than those in the densest quartile, and for none of the 20 countries is the opposite true. Piped water and sewage systems are similarly more prevalent in denser areas, with the second and third quartiles of the population density distribution having moderately higher coverage than the least dense quartile, and the most dense quartile having substantially higher coverage.

Public schools are similarly likely to be present at all ranges of the population density distribution, with an average of about 90 percent coverage in all areas. In none of the 20 countries is there a statistically significant difference between any of the density quartiles. Health clinics are present in 59 percent of the most sparsely populated quartiles and 73 percent in the densest quartiles. Police stations are located in 29 percent of the sparsest quartiles and 47 percent of the densest, on average. In few of the countries are differences in health clinics or police stations statistically different through density space. Paved roads are found in 27 percent of the least dense quartiles and 54 percent of the most dense quartiles, with 6 of 20 countries have

Table 1: Publicly Provided Goods by Density Quartile

	Population Density Quartile			
	Q1	Q2	Q3	Q4
Electricity Grid	0.39	0.42	0.48	0.72
		1-0	0-4	0-10
Piped Water	0.36	0.35	0.42	0.67
		0-0	0-2	0-11
Sewage System	0.14	0.13	0.18	0.37
		0-0	0-0	0-7
Public School	0.91	0.90	0.90	0.90
		0-0	0-0	0-0
Health Clinic	0.59	0.58	0.62	0.73
		1-0	0-0	2-4
Police Station	0.29	0.30	0.33	0.47
		0-0	0-0	1-4
Paved Road	0.27	0.30	0.35	0.54
		1-0	1-2	0-6

Note: This table reports the average fraction of enumeration areas having access to a publicly provided good. Electricity grid, piped water system and sewage system take a value of one if most households in the enumeration area could access them, and zero otherwise. School, health clinic and police station take a value of one if they are present in the enumeration area or within easy walking distance. Paved road takes a value of one if the road taken on the way to the interview was paved, tarred or concrete. Numbers below the averages in each row are the number of countries with a difference from the least dense quartile (Q1) that is statistically significant at the one-percent level. In this table we use within-country quartiles rather than across-country quartiles; this ensures that all countries have enumeration areas across quartiles. This does imply, however, that the quartile densities differ across countries.

statistically significantly more roads in denser areas, and none having statistically significantly fewer roads in the densest areas.[1]

An important limitation of the data is that we can observe only the presence of a public good, but not its attributes. Thus, we observe the availability of clinics and schools, but not how crowded a health facility or school is; we also do not observe whether a sewer is open or closed, or how reliable the electricity off the grid is. We still think that these measures contain valuable information on the distribution of amenities across space.

[1]These findings are consistent with the theory that the high per capita cost of providing public services in remote and sparsely populated locations has led governments to supply lower quantities in these areas.

4. Crime

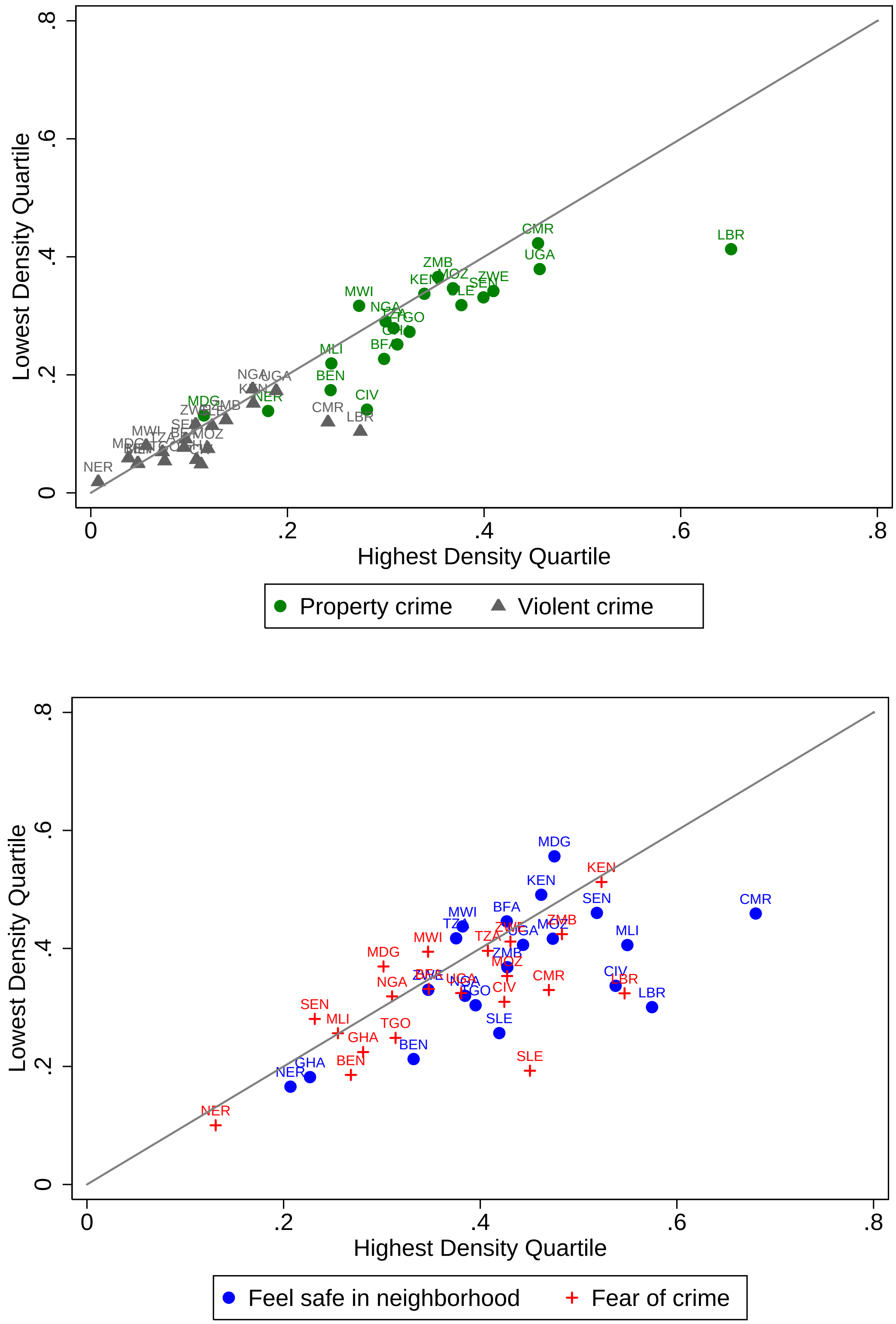
We next turn to measures of crime – arguably one of the most important non-monetary amenities that varies across space. Official crime statistics are not readily available at the local level for most African countries, since administrative records are not always collected centrally. Even if they are collected, they are often unavailable to researchers. Administrative data on crime may also be biased towards areas with greater police presence or better record-keeping capacity. For these reasons, we draw on survey responses about crime. The surveys we use are administered in the same way within and across countries – and also over time. Hence, our data are unlikely to be biased toward any particular geographic area. We consider four main metrics: property crime, violent crime, perceptions of safety in one's neighborhood, and fear of crime in one's home. To measure property and violent crime, we use the survey questions: *"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?"* For each region, we compute the fraction of respondents reporting at least one theft (property crime) or attack (violent crime).

To measure perceptions of safety and fear of crime, the questions are: *"Over the past year, how often, if ever, have you or anyone in your family felt unsafe walking in your neighborhood?"* and *"Over the past year, how often have you or anyone in your family feared crime in your own home?"* The possible answers to these questions on experienced crime and perceived safety are: *"never," "just once or twice," "several times," "many times,"* and *"always."* We define a dummy variable as equal to one if a respondent's reply is anything more than "never."

Overall, we find that crime is quite common in Africa. About one third of respondents report a theft from their house in the previous year. The highest rates of theft are in Liberia (49 percent), Uganda (42 percent) and Senegal (39 percent), and the lowest rates are in Madagascar (13 percent), Niger (18 percent) and Mali (21 percent). 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 reported that they felt unsafe in their neighborhood at least once in the past year, and a similar fraction said that they feared crime in their own home.

Figure 1 shows differences in experienced crime and fear of crime across space. We show both of these categories of variables, as perceptions (or fear) of crime might matter at least as much as experiences of crime for location choices. Both figures illustrate that most countries are located close to the 45-degree line. For fear of crime and perceptions of safety, there are

Figure 1: Crime by Population Density



few differences across population density space. Property crime appears to be slightly higher in denser areas, but the differences for most countries are fairly small. One limitation of the theft variable is that it omits livestock theft, which is common in some rural areas. Including livestock theft would almost certainly reduce the differences in crime rates across density space.

Around 29 percent of households in the least dense quartile reported experiencing property crimes, compared to 33 percent in the densest quartile. Violent crime affects ten percent of households in the least dense quartile compared to 12 percent in the densest quartile. Fear of crime and perceptions of safety are similarly increasing in population density on average, with similarly modest differences by density. Few countries show statistically significant differences related to population density.[2]

This evidence suggests that crime is a plausible contender for an amenity that gets worse with density. Could differing crime rates offset the higher income and consumption levels of more-urban areas? To address this question, we draw on previous studies that have estimated willingness to pay for living in areas with less crime. Using the elasticities calculated in these studies, we ask how much an individual might be prepared to pay to move from a densely populated area, with a 33 percent chance of theft, to a rural area where the annual frequency of theft is just 29 percent. Similarly, we can ask about the willingness to pay to reduce a 12 percent chance of violent crime to a level of just ten percent in the least dense areas.

Relative to the large differences in average income across space, the estimated valuations of crime implied by previous literature are quite modest. For example, using crime and property-value data, Bishop and Murphy (2011) estimate a dynamic model and infer that San Francisco residents are willing to pay $472 per year to avoid a ten-percent increase in violent crime. On an average income per head of $57,276, this amounts to 0.8 percent of average yearly income. Using direct survey questions, Cohen, Rust, Steen, and Tidd (2001) estimate that, in 2000, U.S. residents were willing to pay $120 to reduce the chance of armed robbery by ten percent. This amounts to 0.4 percent of average yearly income ($120 / $34,432). Similarly, Ludwig and Cook (2001) estimate that U.S. households in 1998 were willing to pay $240 per year to reduce the chance of gunshot injury by 30 percent, which amounts to 0.5 percent of average household income ($240 / $51,939). Taken together, these studies suggest that the modest differences in crime rates with density may not be large enough to offset the much higher average incomes in urban areas.

[2]Studies by Fafchamps and Moser (2003) and Demombynes and Ozler (2002) from Madagascar and South Africa point to somewhat higher crime rates in less dense areas. Crime information is also available in some LSMS surveys, though with questions that are harder to compare across countries. Appendix Figure 5 shows the fractions of LSMS households having experienced a crime in the last twelve months, in five countries for which geo-references are available. Visually, a constant rate of crime seems as though it would just about fit within the confidence intervals.

5. Air Pollution

Pollution is a widely studied amenity that varies through space. Sources of outdoor pollution include vehicles, electricity generation, industry, waste and biomass burning, and re-suspended road dust from unpaved roads. Banzhaf and Walsh (2008) find pollution to be an important determinant of locational choice in the United States, and exposure to pollutants significantly affects health, human capital and productivity (Adhvaryu, Kala, and Nyshadham, 2014; Currie and Walker, 2011; Graff Zivin and Neidell, 2013; Kahn and Walsh, 2015).

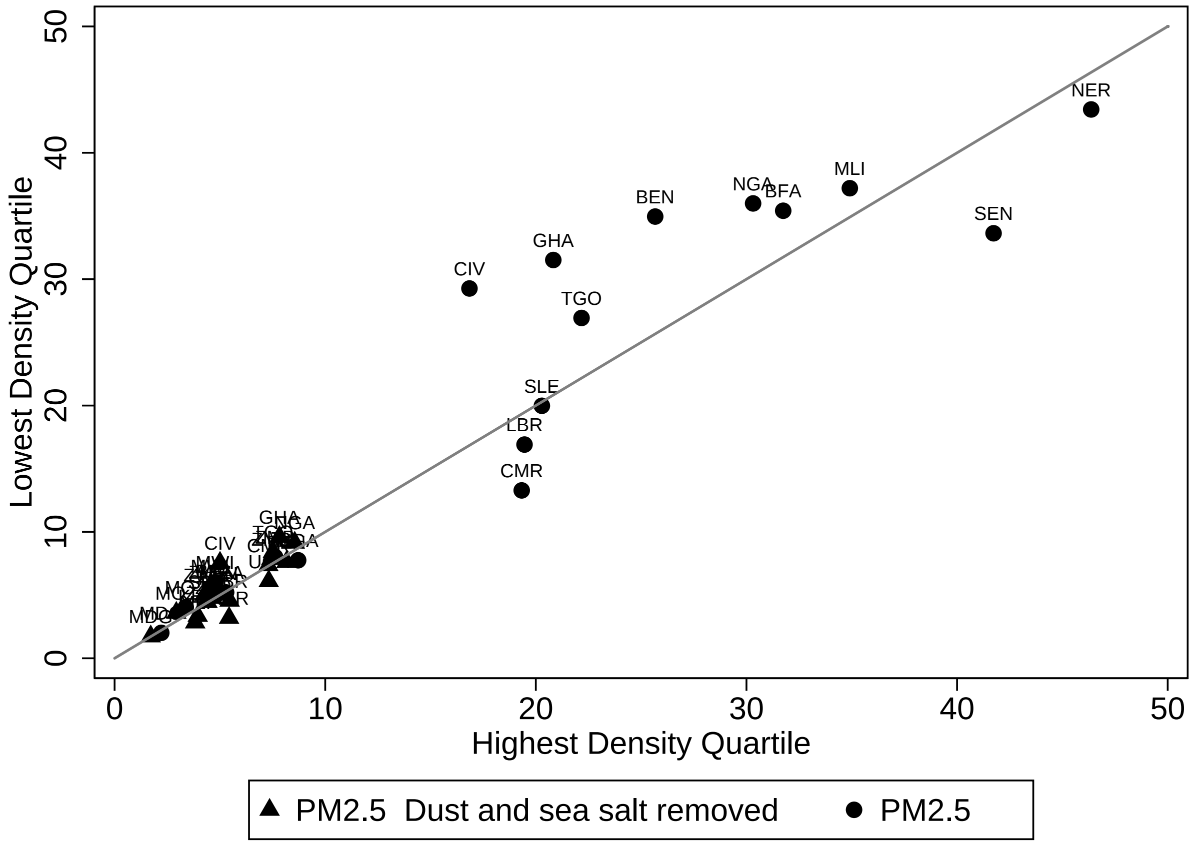
In this section, we use satellite-derived estimates of two measures of outdoor air pollution — fine particulate matter (PM2.5) and nitrogen dioxide (NO2). We show how these vary with population density in our set of countries and several reference countries. We measure both pollutants in micrograms per cubic meter ($\mu g/m^3$). For reference, 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 (Geddes, Martin, Boys, and van Donkelaar, 2016; WHO, 2006). Indeed, Pope and Dockery (2006) conclude from a meta analysis that the relationship between particulate-matter exposure and life expectancy is approximately linear.

Figure 2 plots (as circles) the average PM2.5 concentrations for the highest and lowest population density quartiles across all our countries. Those with the highest overall concentrations of PM2.5 tend to have high concentrations in both sparsely populated areas and densely populated areas. Most of these countries (e.g., Niger, Senegal, Mali and Nigeria) border the Sahel and are thus exposed to dust blown off the Sahara. To account for this, we plot PM2.5 concentrations with dust and sea salt removed (as triangles). There is no evidence suggesting that anthropogenic sources of PM2.5 are more or less harmful for health than are natural sources, but it is possible that individuals perceive anthropogenic sources as more hazardous to their health. When dust and sea salt are taken out of the calculations, the data show far lower levels of PM2.5 across all locations, but it is still the case that there are very modest differences by population density.

Turning to NO2, we find that while there is some variation across countries, there is, again, little apparent difference by density. When we look at average pollution levels by density quartile, we find that NO2 concentrations do not vary in statistically significant ways across density space. NO2 on average is lower in denser areas, though the differences are modest (and statistically insignificant in around half the countries). We conclude that outdoor air pollution concentrations in Africa are at best loosely linked to population density.

We note that this surprising result is highly specific to the countries in our sample. In other parts

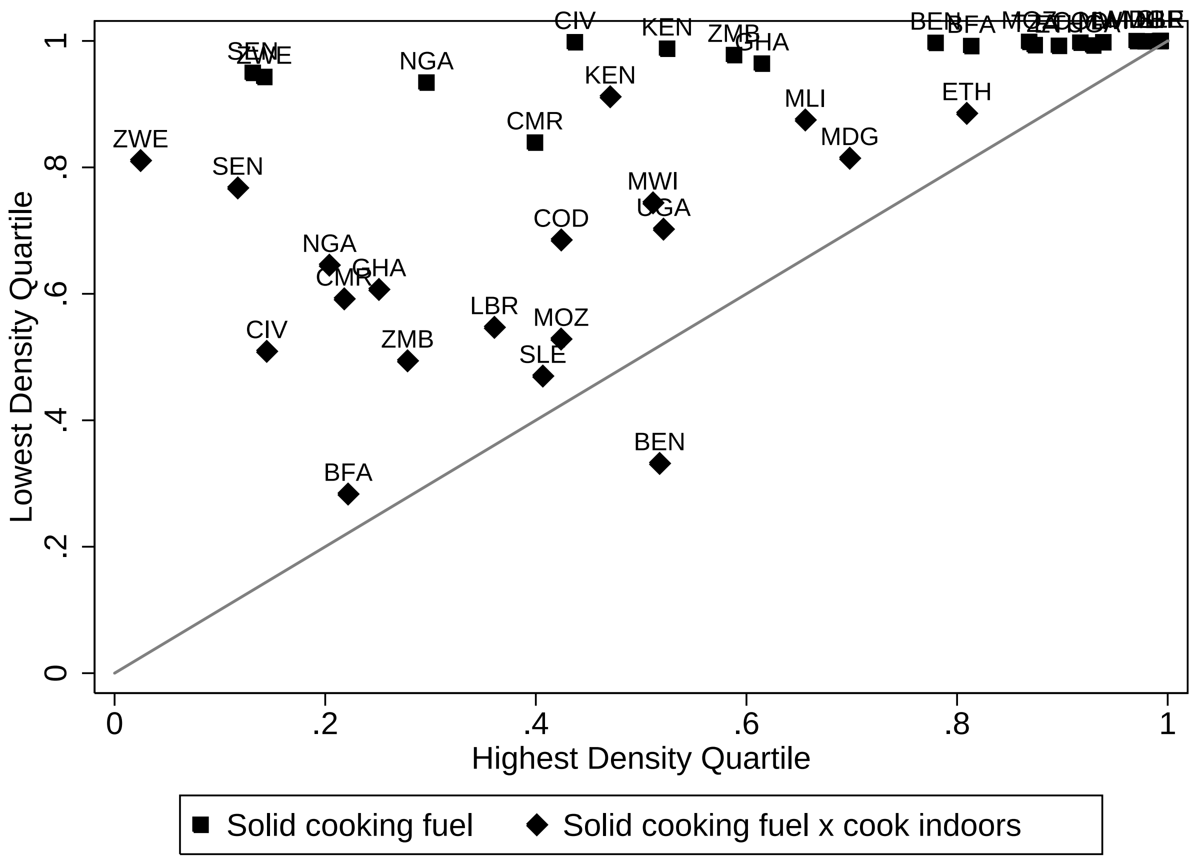
Figure 2: Outdoor Air Pollution



of the world, the same pollution measures display quite strong patterns of (positive) correlation with population density. What is different in Africa is that there is relatively little manufacturing or "dirty" generation of electricity. To see this, we consider PM2.5 concentrations in other countries. We find that in China, India and the United States, pollution gradients with density are strongly positive. 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. These positive gradients are very much what we might expect of a world in which cities have high concentrations of industrial activity and automobile traffic. Although urban areas in Africa are growing rapidly, there is little industrial activity (Gollin, Jedwab, and Vollrath, 2016); consequently, industrial air pollution is relatively low.

To be clear, we are not making a claim that outdoor pollution does not matter in African cities. Our satellite-derived pollution estimates do not capture all dimensions of pollution exposure.

Figure 3: Indoor Air Pollution



At a 10km resolution, our measures are spatially rather coarse. Local effects, such as proximity to roads, may matter significantly.[3] Moreover, satellite-derived measures reflect the column of pollution as observed from space, rather than the concentration experienced on the ground. Nevertheless, as the data from India, China, and the United States illustrate, our metrics seem reasonable and appropriate. What emerges from this analysis is that African cities are neither large enough nor industrialized enough to create large clouds of pollution, and background non-anthropogenic pollution is high. This combination produces different pollution gradients from those observed in more industrialized parts of the world.

Arguably a more serious health risk in developing countries is *indoor* air pollution, largely related to the use of unvented fires for cooking. As a proxy for indoor air quality, we examine the main material used for cooking as reported in the DHS. The World Health Organization esti-

[3]For example, Kinney, Gichuru, Volavka-Close, Ngo, Ndiba, Law, Gachanja, Gaita, Chillrud, and Sclar (2011) find average PM2.5 concentrations at four traffic sites in Nairobi, between 7.30am and 6.30pm, that fall between 58.1 and 98.1 $\mu g/m^3$; by contrast 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).

mates that over four million people suffer from premature deaths due to illnesses attributable to cooking with solid fuels in poorly ventilated indoor spaces (WHO, 2014).

Figure 3 shows the proportion of households using solid fuels for cooking across population density (as squares). Across all of our countries, virtually everyone in the least dense areas uses a solid cooking fuel. In the urban areas there is more variation, with the densest areas ranging from 20 percent solid fuel in Nigeria to nearly half in countries such as Zimbabwe, Cote d'Ivoire and Ghana; for many countries, over 80 percent of households in the most densely populated quartile cook with solid fuels. Overall, however, the pattern is clearly one of much higher use of solid fuels in rural areas than in urban 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. Therefore we also show the interaction effect: among the households using solid fuels, we look at the proportion cooking indoors (as diamonds). The fraction of people cooking primarily indoors with solid fuels is higher in rural areas in every country in our sample other than Benin. On average, the fraction cooking inside with solid fuels is six percentage points lower in the third quartile than in the least dense quartile, and 25 percentage points lower in the densest quartile. All but one country display a statistically significant difference. In summary, indoor air pollution is worse almost everywhere in rural areas than in African cities.[4]

6. Additional Evidence

We now turn attention to two alternative measures of well-being differences between urban and rural areas: subjective well being and net migration rates. While each measure has its limitations, neither suggests that living standards are roughly similar in rural and urban areas in the developing world.

6.1. Subjective Well Being

We now compute measures of subjective well being. Measures of subjective well-being are inevitably difficult to interpret, and comparisons across countries may be particularly problematic because of differences in social norms. However, differences between urban and rural areas within a country seem to offer a plausibly valid comparison.

Our analysis draws on data from the World Values Survey, and follows Glaeser (2012) in com-

[4]We note in passing that our pollution measures do not include adequate data on trash and refuse, which might be more abundant and more visible in urban areas than rural areas. We also do not capture measures of water quality directly, although a number of health measures – which are on average better in urban areas – are quite sensitive to certain types of water pollution and sanitation problems.

Table 2: Subjective Well Being

	Average Life Satisfaction			Percent "Happy"		
	Urban	Rural	Difference	Urban	Rural	Difference
Burkina Faso (2007)	5.7	5.5	0.3**	81.0	78.4	2.6
Ethiopia (2007)	5.0	4.4	0.6*	63.6	58.3	5.2
Ghana (2007)	6.5	5.8	0.7***	82.1	75.5	6.5***
Nigeria (2011)	6.3	6.1	0.2**	89.2	74.7	14.4***
Rwanda (2012)	6.5	6.4	0.1	90.2	91.2	-1.0
Uganda (2001)	5.7	5.6	0.1	80.1	78.2	1.8*
Zimbabwe (2011)	6.3	5.6	0.6***	43.4	37.9	2.8*

Note: This table reports average life satisfaction, and the fraction of adults reporting that they are "very happy" or "somewhat happy" (rather than"not very happy" or "not happy at all"), by urban and rural areas. The data come from the World Values Surveys. ***,**,* mean statistically significant at the one-, five- and ten-percent levels.

paring subjective well being in rural and urban areas. We concentrate on two variables: satisfaction, reported on a scale of 1 (least satisfied) to 10 (most satisfied), and happiness, which we measure as the fraction of individuals reporting that they are "quite happy" or "very happy." In both cases we compute the average for people aged 15 and over by urban and rural areas.

Table 2 reports the average life satisfaction rates and percent of people that report being happy in the seven African countries for which we have data both on urban and rural households. In this data set, we lack geo-references for individual households, so we are obliged to use the urban-rural classification of the survey. The results are striking, however. In all seven countries, urban households report higher levels of average life satisfaction than rural households; in five of the countries, the differences are statistically significant. Urban households are happier in six of the seven countries, with the exception being Rwanda, where both rural and urban households report extremely high levels of happiness. Four of the seven countries have statistically significant differences between rural and urban households; in all four of these, the difference is that urban households report being happier. Overall, the evidence for this subset of our countries is consistent with the finding of Glaeser (2012) that residents of urban areas in the developing world are more likely to report being satisfied with their lives, and happy, than those living in rural areas.

6.2. Net Migration Rates

We turn finally to measures of net migration. Net migration is an indirect measure of amenities, since individuals vote with their feet and choose locations with the best mix of amenities and relative income potential. To measure net rural-urban migration rates, we compute, for the subset of countries with appropriate data, the fraction of all surveyed individuals in the DHS that are rural-to-urban migrants and the fraction of all individuals that are urban-to-rural migrants. Ideally, we would know the exact location from which an individual migrated. Unfortunately the DHS data do not contain this information. However, we know if an individual moved from the capital, from a large city or town, or from the countryside. We define an urban-rural migrant as someone who has been residing in the lowest-density quartile for five years or less, and who previously lived in the capital or a large city. Similarly, we define a rural-urban migrant as someone who has been residing in the highest-density quartile for five years or less, and who previously lived in the countryside.

We find that in every country, there are substantially more rural-to-urban migrants than the opposite. On average, 4.4 percent of respondents are rural-urban migrants, and just 0.7 percent are urban-rural migrants. The differences are starkest in Kenya, where 7.6 percent are rural-urban migrants, compared to 0.6 percent urban-to-rural migrants, and Malawi, which has 7.2 percent rural-urban migrants and less than 0.5 percent moving in the opposite direction. All other countries but one have statistically significant net rural-urban migration. This view is consistent with cities being seen as attractive places to live and workers voting with their feet to move there. To most development economists, of course, this may not be a new or controversial claim; the literature has long emphasized the importance of rural-urban migration as one feature of structural transformation. But policy makers continue to worry about excessive urbanization, and many academic economists use models that explicitly or implicitly assume that population movements are associated with some kind of sorting that is consistent with a steady-state distribution of population. For these reasons, we find it useful to emphasize that the net flow of people in these economies is clearly single-directional.[5]

[5] Our finding does not disagree with Young (2014), who reports that urban-to-rural migration as a fraction of the rural population is as large as rural-to-urban migration as a fraction of the urban population. He states that "the difference relative to shares of destination arises because of the smaller average urban population share (0.41 versus 0.59 for rural). Overall, net migration is in favor of urban areas with, on average, 0.126 of the aggregate young adult female population moving to urban areas and only 0.07 to rural areas."

7. Conclusion

One appealing explanation of the higher real consumption levels of urban areas in the developing world is a "spatial equilibrium," where the higher amenities of rural life are enough to offset the greater monetary rewards of living in cities. In this paper, we go searching for a spatial equilibrium in 20 African countries, using new spatially disaggregated evidence on non-monetary amenities. We focus on measures of three of the prime candidate amenities: publicly provided goods, crime, and air pollution. We find that almost all metrics in almost all countries are either unrelated with population density or actually improve with density. The only real exception is crime, although even here the differences are modest in magnitude and dwarfed in monetary value by the size of the rural-urban wage (or consumption) gaps. Net migration is overwhelmingly towards denser areas, and, for a smaller set of countries, subjective well-being measures are generally higher in densely populated areas.

Our findings are hard to reconcile with a simple static spatial equilibrium, in which rural areas offer higher amenities that leave workers indifferent between locations. Instead, they point to a world in which individuals in developing countries migrate on net to cities, which – on average – offer a better mix of consumption and amenities. This kind of mobility may be limited by a range of frictions that prevent utility from being instantaneously equalized. In this way, our conclusions are consistent with the hypothesis of Chauvin, Glaeser, Ma, and Tobio (2016) that a spatial equilibrium emerges only when economies are sufficiently developed. Our findings suggest that to understand better why urban-rural gaps persist in the developing world, researchers should focus on identifying and understanding the frictions that impede migration to cities.

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," Unpublished Working Paper, University of Michigan.

AKRAM, A. A., S. CHOWDHURY, AND A. M. MOBARAK (2017): "General Equilibrium Effects of Emigration on Rural Labor Markets," Unpublished Working Paper, Yale University.

AVERY, S. (2012): "Lake Turkana and the Lower Omo: Hydrological Impacts of Major Dam and Irrigation Developments," Unpublished Working Paper, African Studies Centre, Oxford.

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

BISHOP, K. C., AND A. D. MURPHY (2011): "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach," *American Economic Review Papers and Proceedings*, 101(3), 625–629.

BRYAN, G., S. CHOWDHURY, AND A. M. MOBARAK (2014): "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 82(5), 1671–1748.

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

CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK (2015): "Gridded Population of the World, Version 4 (GPWv4), Preliminary Release 2 (2010)," Accessed 23 September 2015; available: http://www.ciesin.columbia.edu/data/gpw-v4, Columbia University.

CHAUVIN, J. P., E. GLAESER, Y. MA, AND K. TOBIO (2016): "What is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States," *Journal of Urban Economics*.

COHEN, M. A., R. T. RUST, S. STEEN, AND S. T. TIDD (2001): "Willingness-to-Pay for Crime Control Programs," Unpublished Working Paper, Vanderbilt University.

COSTA, D. L., AND M. E. KAHN (2006): "Public Health and Mortality: What Can We Learn from the Past?," in *Poverty, the Distribution of Income, and Public Policy*, ed. by D. C. Alan Auerbach, and J. Quigley. Russell Sage.

CURRIE, J., AND R. WALKER (2011): “Traffic Congestion and Infant Health: Evidence from E-ZPass,” *American Economic Journal. Applied Economics*, 3(1), 65.

CUTLER, D., AND G. MILLER (2005): “The Role of Public Health Improvements in Health Advances: the Twentieth-Century United States,” *Demography*, 42(1), 1–22.

DEMOMBYNES, G., AND B. OZLER (2002): “Crime and Local Inequality in South Africa,” World Bank Policy Research Working Paper 2925.

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.

ECKERT, F., AND M. PETERS (2017): “Spatial Structural Change,” Unpublished Working Paper, Yale University.

FAFCHAMPS, M., AND C. MOSER (2003): “Crime, Isolation and Law Enforcement,” *Journal of African Economies*, 12(4), 625–671.

FERRÉ, C., F. H. G. FERREIRA, AND P. LANJOUW (2012): “Is There a Metropolitan Bias? The Relationship Between Poverty and City Size in a Selection of Developing Countries,” *World Bank Economic Review*, 26(3), 351–382.

GEDDES, J. A., R. V. MARTIN, B. L. BOYS, AND A. VAN DONKELAAR (2016): “Long-Term Trends Worldwide in Ambient NO2 Concentrations Inferred from Satellite Observations,” *Environmental Health Perspectives*, 124(3), 281–289.

GLAESER, E. L. (2012): *Triumph of the City*. Penguin Books, New York.

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

GOLLIN, D., R. JEDWAB, AND D. VOLLRATH (2016): “Urbanization with and without Industrialization,” *Journal of Economic Growth*, 21, 35–70.

GOLLIN, D., D. LAGAKOS, AND M. E. WAUGH (2014): “The Agricultural Productivity Gap,” *Quarterly Journal of Economics*, 129(2), 939–993.

GRAFF ZIVIN, J., AND M. NEIDELL (2013): “Environment, Health, and Human Capital,” *Journal of Economic Literature*, 51(3), 689–730.

Hamory Hicks, J., M. Kleemans, N. Y. Li, and E. Miguel (2017): "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata," Unpublished Working Paper, U.C. Berkeley.

Herrendorf, B., and T. Schoellman (2014): "Wages, Human Capital, and Structural Transformation," Unpublished Working Paper, Arizona State University.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, Forthcoming.

Kahn, M. E., and R. Walsh (2015): "Cities and the Environment," in *Handbook of Regional and Urban Economics*, ed. by G. J. Duranton, V. Henderson, and W. C. Strange, vol. 5 of *Handbook of Regional and Urban Economics*, chap. 7, pp. 405 – 465. Elsevier.

Kesztenbaum, L., and J.-L. Rosenthal (2016): "The Democratization of Longevity: How the Poor Became Old in Paris, 1880-1913," in *New Approaches to Death in Cities during the Health Transition*, ed. by D. R. Fariñas, and M. Oris. Springer, Cham.

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 and Policy*, 14(4), 369–378.

Kline, P., and E. Moretti (2014): "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," *Annual Review of Economics*, 6, 629–662.

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), 1–8.

Ludwig, J., and P. J. Cook (2001): "The Benefits of Reducing Gun Violence: Evidence from Contingent-Valuation Survey Data," *Journal of Risk and Uncertainty*, 22(3), 207–226.

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

Morten, M., and J. Oliveira (2017): "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," Unpublished Working Paper, U.C. Berkeley.

Nunn, N., and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101(7), 3221–52.

Perez-Heydrich, C., J. Warren, C. Burgert, and M. Emch (2013): "Guidelines on the use of DHS GPS data," Discussion paper, Demographic and Health Surveys.

Pope, C. A. I., and D. W. Dockery (2006): "Health Effects of Fine Particulate Air Pollution: Lines that Connect," *Journal of Air and Waste Management Association*, 56, 709–742.

Raffo, J. (2015): "MATCHIT: Stata Module to Match Two Datasets Based on Similar Text Patterns," Statistical Software Components, Boston College Department of Economics.

Redding, S., and E. Rossi-Hansberg (2017): "Spatial Development," *Annual Review of Economics*, 9, 21–58.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11(4), 707–720.

Restuccia, D., D. T. Yang, and X. Zhu (2008): "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis," *Journal of Monetary Economics*, 55, 234–250.

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

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

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

van Donkelaar, A., R. Martin, M. Brauer, and B. Boys (2015): "Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter.," *Environmental Health Perspectives*, 123(2), 135–144.

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

Vrijheid, M., D. Martinez, S. Manzanares, P. Dadvand, A. Schembari, J. Rankin, and M. Nieuwenhuijsen (2011): "Ambient Air Pollution and Risk of Congenital Anomalies: A Systematic Review and Meta-Analysis," *Environmental Health Perspectives*, 119(5), 598–606.

WHO (2006): *Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide*. World Health Organization.

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

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

Appendix (For Online Publication)

A. Population Density Data

Our population density data come from the Gridded Population of the World Version 4 (GPWv4), which provides population density estimates at a resolution of 30 arc-seconds (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 than that available in some 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 our analysis, one important consideration is that these other input data might introduce circularity in measurement. For example, if nighttime lights data from satellites are used to assign populations to locations, in an effort to allocate population at a finer geographical resolution, then it would hamper our efforts to estimate the relationship between population density and electrification: by construction, higher densities would be associated with higher rates of electrification. We rule out this circularity by using population density data that are not modeled using further input data. The maximum dispersion assumption of GPWv4 within spatial administrative units therefore biases us towards finding no relationship between population density and outcome variables.

The resolution of the census data underlying the GPWv4 varies across countries due to availability of data. Some countries provide their data at the level of the enumeration area, while others share data only at the second administrative level. We restrict our analysis to countries for which the underlying census data have sufficiently high spatial detail, which corresponds approximately to those for which we have data on more than 40 regions per country.

We use sample weights when computing quartiles and averages across quartiles; when looking at averages across quartiles within countries, we define these within countries; when we aggregate across countries, we define quartiles over the whole sample of countries. All of our results are robust to different uses of the survey weights provided.

B. Demographic and Health Surveys (DHS) Data

2.1. DHS Variables

Indoor cooking is determined using variables hv226 and hv241. Migration status is determined using the years lived in the current location (v104) and the type of the previous residence which is classified into capital, large city; city; town; countryside; and abroad (v105). Following Young (2014), we exclude individuals who lived abroad, and check that all variables are coded consistently across countries; for example, abroad is coded as either 4 or 5.

2.2. Linking DHS data with population density data

To link the individual data from the DHS and Afrobarometer with population density, we would ideally have the GPS location of each household. The DHS readily collects GPS coordinates for survey clusters, but in order to preserve the anonymity of survey respondents, these have been displaced — i.e., reassigned a GPS location that falls within a specified distance of the actual location. Urban DHS clusters are randomly displaced by 0-2km, and rural clusters are randomly displaced by 0-5km, with one percent of clusters randomly selected to be displaced by up to 10km (Perez-Heydrich, Warren, Burgert, and Emch, 2013). We take into account the random offset of DHS cluster locations when linking DHS GPS data with continuous raster data by taking 5 km buffers around both urban and rural DHS clusters, as suggested by Perez-Heydrich, Warren, Burgert, and Emch (2013) and extract the average population density around each cluster. We perform these calculations in WGS1984, since the different areas of the pixel sizes when moving away from the equator has been taken into account when constructing the population density grid, which is defined as the population count divided by the area. Many urban DHS clusters are in proximity closer than 5 km so that buffer polygons around clusters are overlapping. Therefore, we compute our zonal statistics using the Spatial Analyst Supplemental Tools in ArcGIS, a supplemental toolbox that allows computing zonal statistics for overlapping polygons. All computations were performed in ArcGIS 10.4. Table 4 lists the DHS survey countries in our sample and respective populations.

Table 3: Surveys

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

Table 4: Set of Countries Studied

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

C. Afrobarometer

3.1. Afrobarometer variables

We use variables related to feeling unsafe walking in one's neighborhood (q9a), fear of crime in one's own home (q9b); theft (q9b); physical attack (q9c); trust in general (q83); trust towards relatives (q84a); trust toward neighbors (q84b); trust toward one's own ethnic group (q84c); frequency of lack of food (q8a) and medicine (q8c); and anxiety (q96b).

3.2. Geo-locating Afrobarometer respondents

Afrobarometer surveys collect data on attitudes towards democracy and governance, as well as a range of other quality-of-life measures.[6] The Afrobarometer surveys do not record coordinates of respondents, but record the village, district and region names. The 2011 round provides four different administrative names. We use a matching algorithm that matches village names and other provided administrative names to locations as listed in gazetteers; specifically, we follow Nunn and Wantchekon (2011) and use the geonames gazeteer available on www.geonames.com. This website provides a list of locations, each assigned an ID along with several names: the geographical name of the point in utf8 and plain ascii characters; alternative names and the associated latitude and longitude coordinate. There is also auxiliary information such as the modification date of each entry, administrative codes, elevation, and feature classes. If a name is associated with several entries, we keep the most recent entry.

Our matching algorithm uses a mixture of exact matches and fuzzy matches in multiple stages (depending on the survey round, between 13 and 21). Whenever a location name is identified, we assign it the latitude and longitude and remove it from the dataset that is fed into the next stage. In essence, matching is achieved in the following way: first, we perform a series of exact matches based on the village name from Afrobarometer with the asciiname listed in the gazetteer; if there are no exact matches with the village name and the asciiname, we search through the next four alternative names listed in the gazeteer for the specific location. In this first stage we find almost forty percent of locations. We then use the most precise administrative classification. For example, if the data set has information on the village name, district and region, this would be the district. We perform the exact same series of matches on the district name, using again the ascii name and four possible alternative names listed in the gazetteer. In rounds three and four of the survey in which we have only district and region names in addition to the village names, this step finds 49–52 percent of the locations.

Third, we match on the region name which finds another four to six percent of the sample. Finally, to catch any remaining misspellings, we perform a fuzzy match based on similar text patterns between the village name and the ascii name using a command developed by Raffo (2015). We use a similarity score of above 0.70 and a vectorial decomposition algorithm (3-gram). This finds another one to three percent of locations. In total, we are able to match between 92 and 95 percent of individuals in each round. Table 5 shows the number of observations for each country and survey rounds we employ.

In addition to random checks of the identified locations, we use the 2005 data to check the

[6] For further information, see http://www.afrobarometer.org.

consistency between our algorithm and the location data of Nunn and Wantchekon (2011). For the subset of locations for which they provide geo-locations, we find that the median distance between their location and our locations is 12.5 km. Further, considering that the population density data vary largely at the district and region levels, we expect the difference to be even smaller when looking at the resulting population densities. Indeed, the correlation coefficient between the population density from their and our data is 0.65 with a p-value of 0.000.

Table 5: Afrobarometer Sample

	Individuals	Round 3	Round 4	Round 5
Benin	3,543	x	x	x
Burkina Faso	2,255		x	x
Cameroon	1,072			x
Cote D'Ivoire	1,192			x
Ghana	4,089	x	x	x
Kenya	4,659	x	x	x
Liberia	2,282		x	x
Madagascar	3,881	x	x	x
Malawi	4,784	x	x	x
Mali	3,663	x	x	x
Mozambique	4,744	x	x	x
Niger	1,199			x
Nigeria	6,961	x	x	x
Senegal	3,596	x	x	x
Sierra Leone	1,190			x
Tanzania	4,791	x	x	x
Togo	1,056			x
Uganda	7,191	x	x	x
Zambia	3,590	x	x	x
Zimbabwe	4,344	x	x	x

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

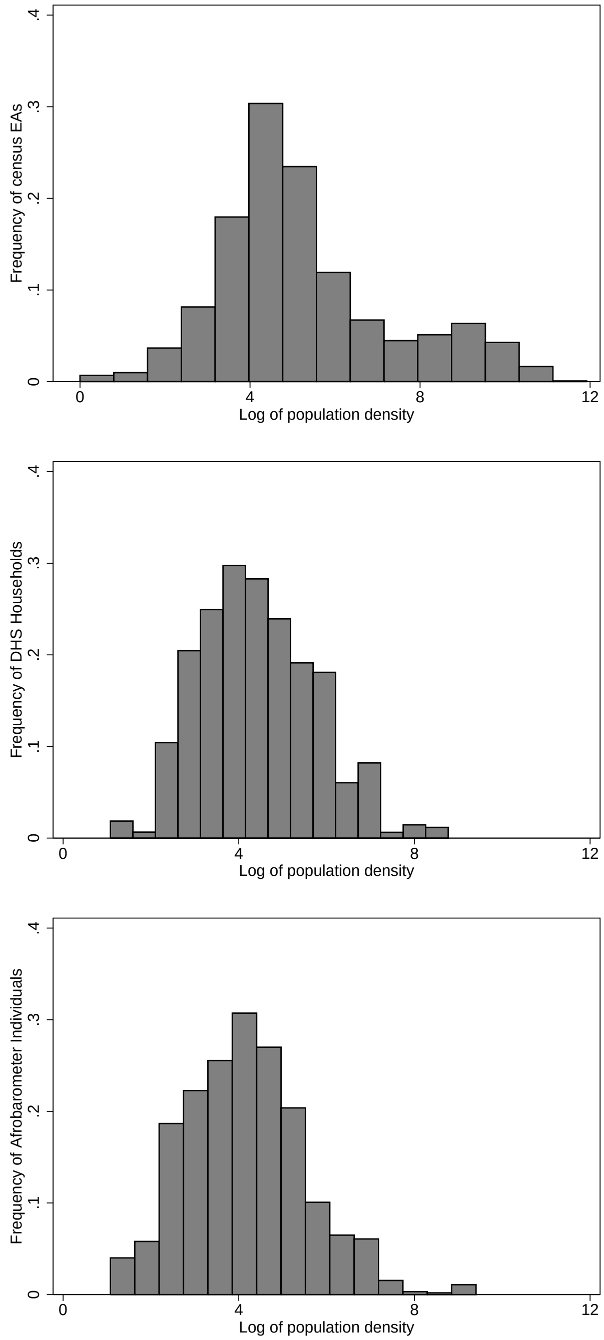
D. Representativeness of samples

While the DHS aim to make survey instruments and samples comparable across countries, the exact sampling differs according to the particular survey.[7] The target population of most DHS surveys are women aged 15-49 and children under the age of five living in residential households with the most common sampling following a two-stage cluster sampling procedure. 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.

If the sampling frame is not specifically selected to match the population along the lines of population density, it is likely that the distribution of the survey sample according to population density might not match that of the entire population. In practice, the cases we have examined show very little effort to oversample or undersample with respect to population density. For Tanzania, we can compare the population density distribution of the Afrobarometer and DHS clusters with those of the overall population from the census data, where we weight the population density of enumeration areas by the population. As is evident from Figure 4, both the Afrobarometer survey and the DHS appear to capture a sample that covers a wide range of population densities.

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

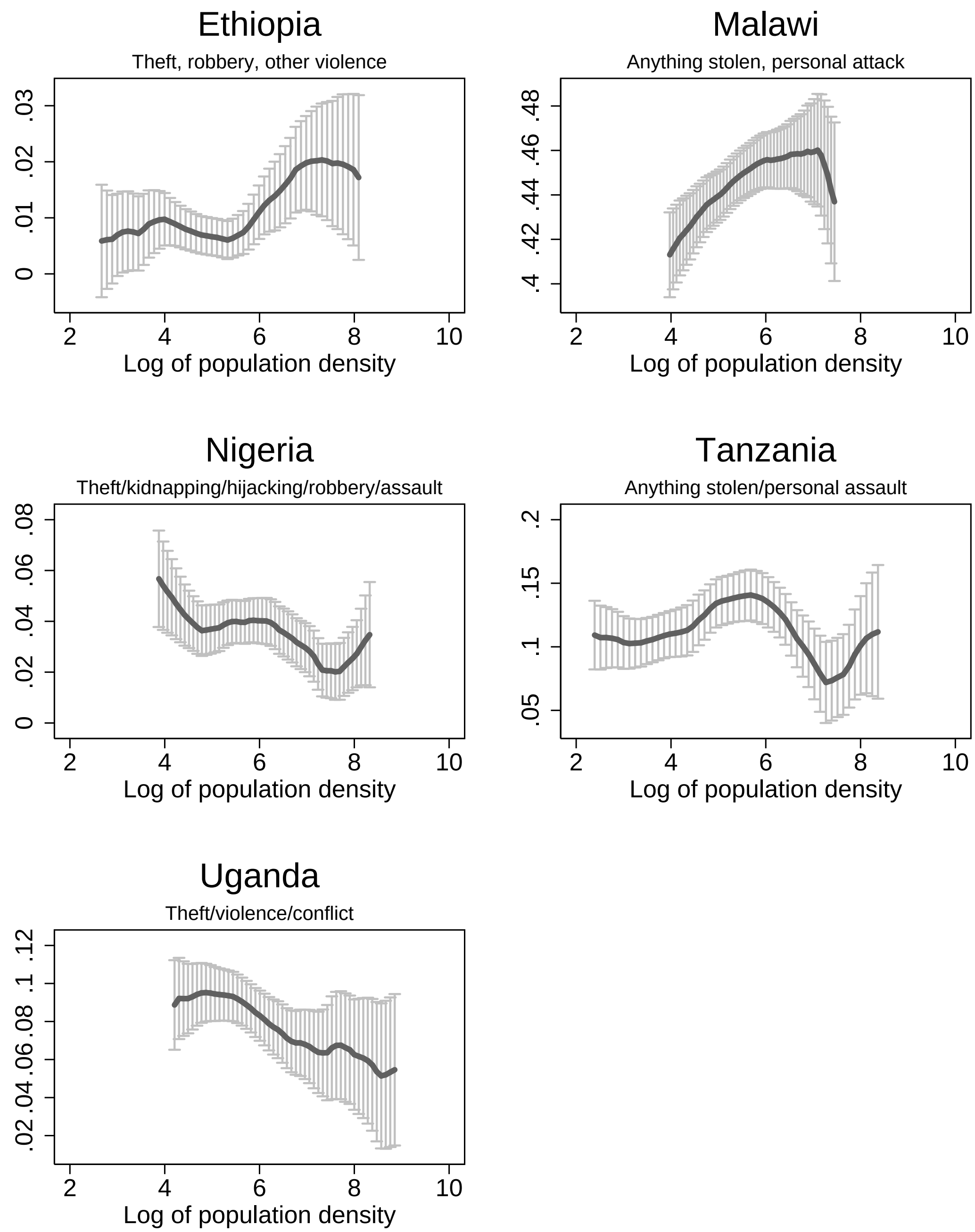
Figure 4: Distribution of population, DHS and Afrobarometer respondents in Tanzania



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.

E. Crime from LSMS data

Figure 5: Crime - LSMS

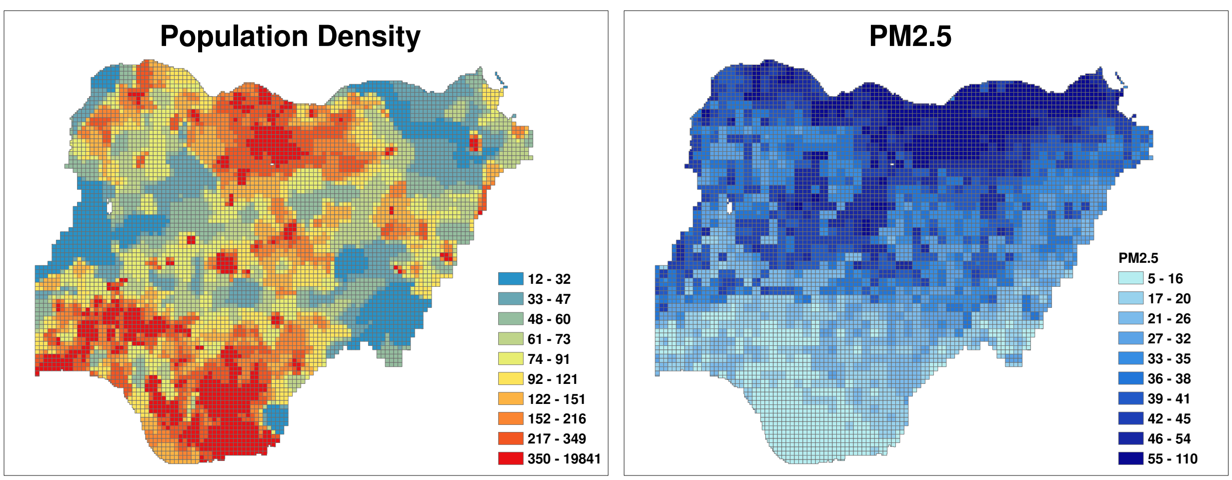


F. Pollution data

Our pollution data for PM2.5 and NO2 concentrations come from van Donkelaar, Martin, Brauer, and Boys (2015) and Geddes, Martin, Boys, and van Donkelaar (2016), respectively. As the date for the Gridded Population of the World v4 (GPWv4) data is approximately 2010, we take the pollution measures that are closest in time: the tri-annual mean (2009-2011) for both PM2.5 series; for NO2, we have the exact year 2010. The estimated PM2.5 and NO2 concentrations are available at a resolution of 0.1 decimal degrees (about 10km at the equator). We construct a fishnet of the same resolution and, for each pixel, compute the average pollution measure and the average population density from the GPWv4. PM2.5 is measured in $\mu g/m^3$, while NO2 is measured in ppb (parts per billion). Following Vrijheid, Martinez, Manzanares, Dadvand, Schembari, Rankin, and Nieuwenhuijsen (2011), we use a conversion of 1ppb= 1.88 f $\mu g/m^3$, which assumes ambient pressure of 1 atmosphere and a temperature of 25 degrees celsius.

Figure 6 illustrates this procedure and shows the distributions of PM2.5 and population density across space in Nigeria. The left graph shows the distribution of population density; the right

Figure 6: Outdoor Air Pollution (PM2.5) 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.

graph shows the PM2.5 distribution; and the two bottom graphs show PM2.5. Warmer and 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.

Moving to the pollution measure, several observations are worth highlighting: first, at least visually, population density does not appear to be strongly correlated with PM2.5 concentrations. PM2.5 levels appear to be driven mainly by dust from the Sahara when inspecting the graph. Removing sea salt and dust produces quite a different distribution, with higher levels in the center and over some cities, but still with little obvious correlation with population density.

It is instructive to look separately at these two indicators for pollution as shown in Figure 7.[8] The pairwise correlation between PM2.5 and NO2 is -0.0085 with a p-value of 0.4633. Across our whole set of African countries, the correlation of these two measures ranges from 0.65 (Cameroon) to -0.47 (Senegal).

[8]This is in line with what Geddes, Martin, Boys, and van Donkelaar (2016) find when they inspect population weighted average PM2.5 and NO2 levels and trends.

G. Net Migration Rates

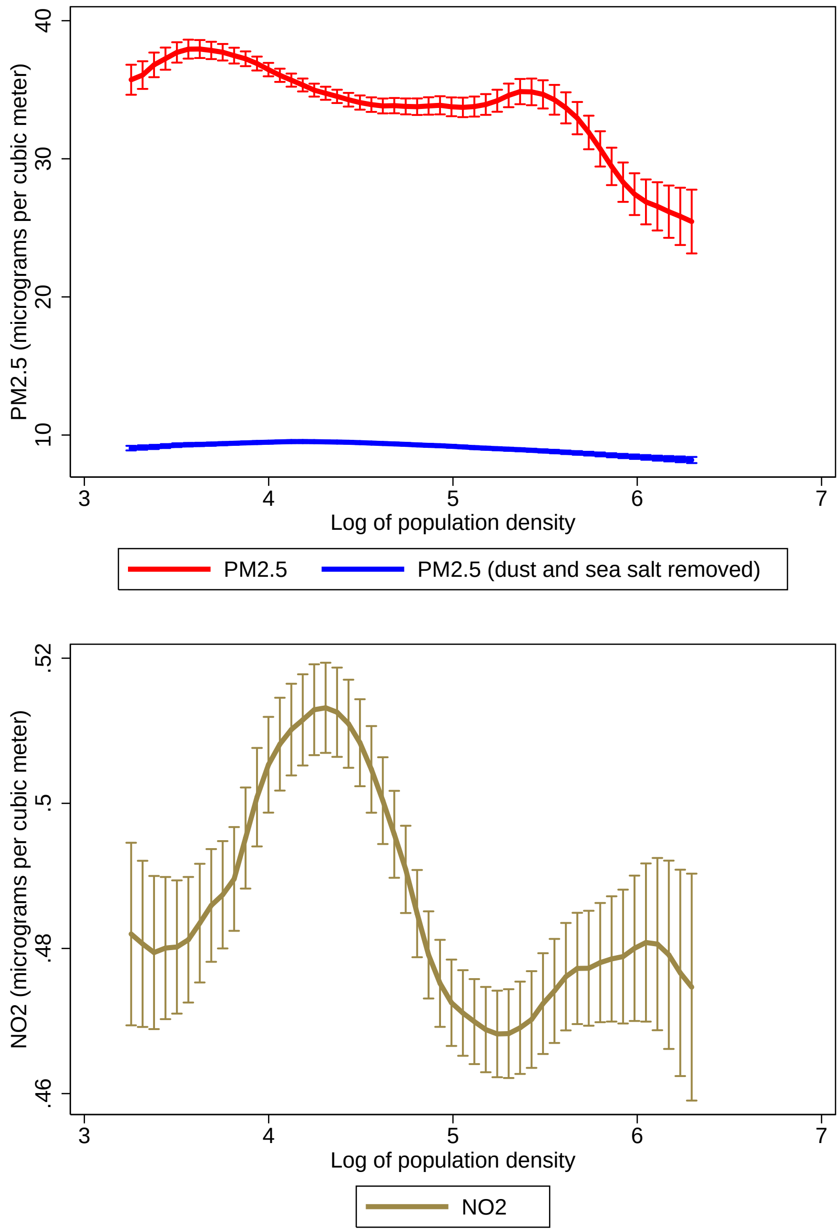
Table 6 displays the fractions of all individuals that are rural-to-urban migrants, urban-to-rural migrants, and their difference.

Table 6: Net Migration

	Rural-to-Urban	Urban-to-Rural	Difference
	Percent of Adults		
Dem. Republic of the Congo (2007)	2.39	0.47	1.92***
Ethiopia (2005)	3.08	0.15	2.93***
Ghana (2008)	4.82	1.18	3.64***
Kenya (2008-2009)	7.60	0.58	7.02***
Liberia (2007)	2.46	2.24	0.23
Madagascar (2008-2009)	4.16	0.19	3.97***
Malawi (2010)	7.23	0.45	6.77***
Mali (2006)	4.46	0.66	3.80***
Nigeria (2008)	4.83	0.37	4.46***
Senegal (2005)	2.75	0.92	1.83***
SierraLeone (2008)	4.44	0.36	4.08***
Zambia (2007)	4.00	0.56	3.44***

Note: The first column lists the country and year of survey. The first two data columns report the percent of adults that are rural-to-urban migrants and urban-to-rural migrants, respectively, in the last five years. The third data column reports the simple difference. ***,**,* mean statistically significant at the one-, five- and ten-percent levels. Test statistics are computed taking into account the stratified sampling design.

Figure 7: 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. We remove the top and bottom five percentiles of the population density distribution.