Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa

Adam Storeygard∗

January 2014

Abstract

How does isolation affect the economic activity of cities? Transport costs are widely considered an important barrier to local economic activity but their impact in developing countries is not well-studied. This paper investigates the role of inter-city transport costs in determining the income of sub-Saharan African cities. In particular, focusing on fifteen countries whose largest city is a port, I ask how important access to that city is for the income of hinterland cities. The lack of panel data on both local economic activity and transport costs has prevented rigorous empirical investigation of this question. I fill this gap with two new datasets. Satellite data on lights at night proxy for city economic activity, and new road network data allow me to calculate the shortest route between cities. Cost per unit distance is identified by plausibly exogenous world oil prices. The results show that an oil price increase of the magnitude experienced between 2002 and 2008 induces the income of cities near a major port to increase by 6.6 percent relative to otherwise identical cities one standard deviation farther away. Combined with external estimates, this implies an elasticity of city economic activity with respect to transport costs of -0.25 at that distance. Moreover, the effect differs by the surface of roads between cities. Cities connected to the port by paved roads are chiefly affected by transport costs to the port, while cities connected to the port by unpaved roads are more affected by connections to secondary centers.

JEL classification: F15, O18, R11, R12, R4
Keywords: Urbanization, Transport costs, Infrastructure, Roads, Sub-Saharan Africa

∗Department of Economics, Tufts University. Email: adam.storeygard@tufts.edu. I thank Managing Editor Francesco Caselli and three anonymous referees for comments that substantially improved the paper. I am grateful to David Weil, Nate Baum-Snow, and especially Vernon Henderson for their advice and support. I thank Katherine Casey, Uwe Deichmann, Leo Feler, Andrew Foster, Delia Furtado, Jake Goldston, Doug Gollin, Walker Hanlon, Erik Hurst, Yannis Ioannides, Blaise Melly, Sriniketh Nagavarapu, Mark Roberts, Nathan Schiff, Enrico Spolaore, Michael Sulzer, Ken-suke Teshima, Ishani Tewari, Matt Turner, Nicholas Wilson, Junfu Zhang, and various conference and seminar participants for helpful comments and suggestions, Uwe Deichmann and Siobhan Murray for access to and guidance on the roads data. Financial support from the North American Regional Science Council and the World Bank’s Research Support Budget and Knowledge for Change Program is gratefully acknowledged. All errors are my own.
1 Introduction

Sub-Saharan Africa has notoriously high transport costs compared to other major regions of the world.\(^1\) Population density is relatively low, with a substantial fraction of people residing far from the coast. Ocean-navigable rivers, which provide cheap transport to the interior of most other regions, are virtually non-existent. And road networks are sparse and poorly maintained, on the whole.

In this paper, I argue that these substantial transport costs play an important role in determining the economic size of cities in sub-Saharan Africa. Specifically, I ask whether periphery cities with lower transport costs to their country’s main port grew faster than those farther away or with poorer road connections, in the context of dramatically rising oil prices over the 2000s decade. A typical problem with testing this kind of question in poor countries is that relevant data on city incomes (or population) and transport costs do not exist. This paper provides novel measures of both. First, night time lights satellite data (Elvidge et al., 1997; Henderson, Storeygard and Weil, 2012) are used to construct a 17-year annual panel of city-level measures of economic activity for 287 cities in 15 countries. Second, a new set of roads data provides information about route length and surface material. Transport costs are thus identified by the interaction between world oil prices and distance along these routes. Because I have data on many cities per country over a substantial time period, I can control for annual shocks separately for each country, as well as initial characteristics and even average growth rates of individual cities.

Focusing on countries whose largest, or primate, city is also a port, I find that as the price of oil increases from $25 to $97 (as it did between 2002 and 2008), if city A is 465 kilometers (1 standard deviation) farther away from the primate than initially identical city B, its economy is roughly 6.6 percent smaller than city B’s at the end of the period. At a differential of 2360 kilometers, the largest in the data, this rises to 29 percent. Further evidence shows that this effect is due to transport costs, not commodity income or the generation or cost of electricity. I then determine that this effect falls disproportionately on cities that are connected to the primate by paved roads, most likely because they are initially more engaged in trade. Cities connected to the primate by unpaved roads appear to be more affected by transport costs to secondary cities. This suggests a funneling of trade to small cities through nearby secondary cities, though no direct data on intercity trade is available.

---

\(^1\)This paper generally defines sub-Saharan Africa as all countries of the African mainland with no Mediterranean coastline, plus Madagascar.
The majority of Africa’s population growth is expected to be in cities over the next few decades (National Research Council, 2003). Indeed, of the approximately 2.5 billion net gain in global population expected by 2050, over 30 percent is expected to be in African cities (United Nations, 2008). Much less is understood, however, about which cities will experience the bulk of that growth. This is the first paper that has the data to systematically address an important determinant of the economic size of cities in sub-Saharan Africa. Which cities grow and which do not will have major implications for the future of the region. If conditions favor large coastal cities, agglomeration economies may improve, but urban infrastructure needs, congestion, and the risks associated with sea level rise will all increase. Many countries have pursued decentralization policies at least in part because of these concerns. If growth is more balanced across a large group of cities, intercity infrastructure may be more important. If cities near international borders grow more than others, they may affect international migration and trade. Regional cities that grow enough may become political power bases constituted along social or industrial lines.

This paper focuses on transport to a major port, in countries where that port is the largest city. While these large port cities do not always contain a large fraction of the overall urban population, they typically play a very important role in the economy, as the largest domestic market, the chief manufacturing center, the primary trading connection with the rest of the world, and the seat of elites and often of the government. Ports have a special role because most African trade is transoceanic. Trade among the contiguous eight members of the West African Economic and Monetary Union (WAEMU) represented less than 3 percent of their total trade for each year in the 1990s (Coulibaly and Fontagné, 2006). If anything, one would expect more trade among these countries than other sets of neighbors, because they share a common currency and thus lack one important trade friction. Other cities in the periphery have relationships with their country’s core that are potentially critical to their success. And countries spend to improve those links or simply to reverse decay. Almost $7 billion is invested per year on roads in sub-Saharan Africa, with a substantial portion funded by donors (World Bank, 2010). Worldwide, transport accounts for 15 to 20 percent of World Bank lending, with almost three quarters of that amount going to roads (World Bank, 2007a).

This paper relates primarily to four bodies of work. The first is on the effect of

---

2The center of some ten sub-Saharan African capital cities is within 20 km by land of an international border. Such proximity is extremely rare in other regions, except among countries less than 40 km in diameter. In six of these cases, the built-up area of the city directly abuts the border or border-river. More than half of the countries in the region have capitals within 100 km of a border by air.
transport costs on the size and growth of cities and regions. This work has been done primarily using cross-country data (e.g. Limão and Venables, 2001) or the construction of very large national transport networks in the United States (Baum-Snow, 2007; Chandra and Thompson, 2000; Atack, Bateman, Haines and Margo, 2010; Duranton and Turner, 2011), India (Donaldson, 2010), China (Banerjee, Duflo and Qian, 2009; Faber, 2013), and Indonesia (Rothenberg, 2012). The existing literature has substantially different findings in different contexts, and applies a variety of models to interpret these findings. For example, Donaldson (2010) finds that Indian colonial districts benefited when railroads were built through them. These results are consistent with an Eaton and Kortum (2002)-type model. In contrast, Faber (2013) finds that peripheral counties in China were hurt by the construction of a large new highway system in recent decades, consistent with new economic geography models following Krugman (1991). The model below incorporates effects in both directions.

Little comparable work has been done in sub-Saharan Africa, which has worse roads, lower urbanization, lower income, and much less industry, and consists of many countries, as opposed to one unitary state. The most comparable work in this respect is Jedwab and Moradi (2012), who consider the construction of colonial railroads in Ghana. Similarly ambitious transport infrastructure projects have not been carried out in post-independence sub-Saharan Africa. A policy literature based on engineering models has argued that transport prices in sub-Saharan Africa are high primarily due to the structure of the transport services market, not transport costs per se (e.g. Tavaninithorn and Raballand 2009). This paper suggests that transport costs per se have an important effect on income.

This paper takes the road network as given, and instead relies on the plausibly exogenous annual changes in transport costs induced by world oil price fluctuations, which allow me to determine the short run impact of shocks better than previous work. Short run does not mean small, however, as average annual oil prices varied by a real factor of 5.8 (deflated using the United States Consumer Price Index, CPI-U) during the period of study (1992–2008). These shocks are also of interest because they are more likely to be repeated in the future, in the same places, than is the construction of the United States Interstate Highway System or the Indian railroad network. The fact that short run shocks can have such a large effect is also interesting in its own right. While infrastructure decisions are properly made on the basis of long run effects,

---

3 Also, Buys, Deichmann and Wheeler (2010) consider the possible effects of road upgrading on international trade in sub-Saharan Africa, interpreting the relationship between cross-country trade and road routes between the largest cities in the context of a gravity model.
households and firms must make decisions incorporating transport costs on an annual basis.

The second related literature is on the scope and drivers of urbanization and the evolution of city systems in Africa. This literature is almost exclusively cross-country in nature, so that unobserved country-level factors may be confounding results (Fay and Opal, 2000; Barrios, Bertinelli and Strobl, 2006). An exception is Jedwab (2013), who looks at districts within two countries, Ghana and Côte d’Ivoire, and argues that local production of cash crops, specifically cocoa, spurred urbanization outside of the few largest cities. In his setup, consistent with mine, these secondary towns form primarily as “consumption cities” where farmers sell their products and buy services and imported goods, as opposed to manufacturing centers as is often assumed in models of urbanization and city formation. Unlike all these papers, the present outcome of interest is a proxy for economic activity (lights) that is available for individual cities on an annual basis, as opposed to population, which is typically only available for censuses carried out at most every ten years. This allows me to observe short-run (annual) changes and to control for all potentially confounding country-level variation with country*year fixed effects.

In stressing the role played by the largest city in each country, this work also has implications for the study of urban primacy (Ades and Glaeser, 1995; Henderson, 2002) and decentralization. Finally, in focusing on the importance of coastal cities, this work relates to the literature on geographic determinants of growth, including Gallup, Sachs and Mellinger (1999) and Collier (2007), which emphasize coastal access and the problems of being landlocked, respectively.

The remainder of the paper has the following structure. Section 2 provides a simple conceptual framework to facilitate interpretation. In section 3, I describe the lights and roads data and the methods used to integrate them. In section 4, I describe the econometric specification used, and in section 5, I report results. Section 6 concludes. An appendix provides further details on the data and methods used.

2 Conceptual Framework

Economists sometimes think about the role of intercity transport costs in city growth in the context of two-region New Economic Geography (NEG) models following Krugman (1991), or one of many variants, including one that adds agricultural transport costs (Fujita, Krugman and Venables, 1999), one that adds a foreign sector accessible from a port in one city (Behrens, Gaigné, Ottaviano and Thisse, 2006), and one that is tailored
to African urban primacy (Pholo Bala, 2009). While changes in transport costs drive urban growth in these models, they do so by inducing manufacturing firms to change their location from the (ex-post) periphery city to the (ex-post) core, with its larger home-market effect.

This is unlikely to be the driving force behind the current growth of most cities in Africa, because formal manufacturing activity is already highly concentrated in the largest cities. For example, as of 2002, the Dar es Salaam administrative region contained 0.16 percent of mainland Tanzania’s land area, and 8 percent of its population, but 40 percent of its manufacturing employment and 53 percent of manufacturing value added. As of 2008, 55 percent of manufacturing establishments, and 66 percent excluding food, beverages and tobacco, were in Dar es Salaam (National Bureau of Statistics, 2009; National Bureau of Statistics and Ministry of Industry, Trade and Marketing and Confederation of Tanzanian Industries, 2010).\(^4\) Tanzania has relatively low primacy, so if anything, these fractions would likely be even larger in other countries.\(^5\) Although Tanzania has a coastline of over 1,400 kilometers and three other ports, Dar es Salaam handled 95 percent of its port traffic as of 1993 (Hoyle and Charlier, 1995).

If a periphery city has minimal manufacturing and primarily performs local (untraded) services, the effect of transport costs will depend on how demand for these services relates to demand for manufactured goods. If they are complements such as local retail and imported consumer goods, high transport costs will decrease local urban activity. If they are substitutes like locally produced and imported furniture, high transport costs will increase local urban activity.\(^6\)

The following model, based on Henderson, Storeygard and Deichmann (2014), embeds this intuition in a simple framework in which changes in transport costs drive urban growth. It describes a region consisting of an urban area and its rural (agricultural) hinterland, with populations \(N_u\) and \(N_a\), respectively. Total regional population is fixed by

\[
N = N_a + N_u \quad (1)
\]

with totally differentiated form

\[
dN_a = -dN_u.
\]

All production uses labor only. Agricultural workers receive their average product

\(^4\)These manufacturing statistics are based on establishments with more than 10 employees.
\(^5\)To the extent that manufacturing is substantial in some hinterland cities, this would work against the results I find below in empirical work.
\(^6\)It is also possible that these hinterland cities are essentially in autarky: because they already have very high transport costs, they are largely insulated from transport cost changes. This is inconsistent with the empirical results below.
Urban workers produce a local service with constant returns. Free migration ensures equal per capita incomes across sectors:

\[ y = p_a f(N_a) = b p_u \]  

with log totally differentiated form

\[ d \ln y = d \ln p_a + d N_a f'/f = d \ln p_u \]

where \( p_u \) is the price of the urban service and \( p_a = p_{a0}(1 - x p_o) \) is the local price of the agricultural good. \( p_{a0} \) is the price of the agricultural good in the country’s international trade hub, \( x \) is the distance between the region and the hub, and \( p_o \) is the price of oil. The local service is not traded outside the region. All urban and rural workers in the region consume the agricultural good and the local service.

The model is closed with a market clearing condition for the local service:

\[ b N_u = ND(y, p_u, p_o). \]  

Assuming for now that the cross-price elasticity is zero, the demand elasticities of income (\( \eta_y \)) and own price (\( \eta_{p_u} \)) must be equal in magnitude and opposite in sign, so this has log totally differentiated form

\[ d \ln N_u = \eta_y (d \ln y - d \ln p_u). \]

Substituting the differentiated form of (2) into the differentiated form of (3) implies that \( d \ln N_u = 0 \), and therefore the differentiated form of (2) implies \( d \ln p_a = d \ln y \).

The comparative static of interest is the effect of transport costs on total city income:

\[ \frac{d \ln (y N_u)}{d (x p_o)} = \frac{d \ln y}{d (x p_o)} + \frac{d \ln N_u}{d (x p_o)} = \frac{d \ln p_a}{d (x p_o)} = \frac{-1}{1 - x p_o} < 0. \]  

Higher transport costs decrease farmers’ income and therefore their demand for the urban service, decreasing urban incomes. Since rural and urban incomes decrease the same, no one moves and the urban population doesn’t change.

The empirically identifiable comparative static relies on the interaction between oil price and distance:

\[ \frac{d^2 \ln (y N_u)}{d x d p_o} = \frac{-1}{(1 - x p_o)^2} < 0. \]  

Increased oil prices decrease city income, purely by decreasing per capita income, without changing city population, and their effect is larger far from the port. While \( y \) here
is nominal income, and the price of the local service will fall proportionally with it \((d\ln p_u = d\ln y)\), as long as agents consume any of the agricultural good, higher transport costs will make them worse off. The empirical results below are consistent with this.

With a more flexible demand function, higher transport costs can increase or decrease city population and income. Specifically, adding an imported manufactured good with local price \(p_m = p_m(1 + x p_o)\), and allowing nonzero cross-price elasticities of demand for the local service \(\eta_{p_a}\) and \(\eta_{p_m}\), equation (3) becomes:

\[
d\ln N_u = \eta_y d\ln y + \eta_{p_u} d\ln p_u + \eta_{p_a} d\ln p_a + \eta_{p_m} d\ln p_m. \tag{6}
\]

Using \(d\ln y = d\ln p_u - dN_u f'/f = d\ln p_u\) from Equations (1) and (2), substituting \(d\ln p_a = \frac{-d(x p_o)}{1-x p_o}\) and \(d\ln p_m = \frac{d(x p_o)}{1+x p_o}\), and rearranging,

\[
\frac{d\ln N_u}{d(x p_o)} = \frac{2\eta_{p_m} y}{[y + (\eta_y + \eta_{p_u})p_a N_u f'][1 - (x p_o)^2]}. \tag{7}
\]

Stability of migration implies that the denominator is always positive unless the local service is a Giffen good.\(^7\) Therefore the sign of the effect of transport costs on city population depends on the cross-price elasticity of demand between the manufactured good and the local service. Higher transport costs will shift consumption away from the manufactured good. If the local service is a complement to the manufactured good, demand for the local service will decrease, driving urban service workers to leave the city for the farm. This would be the case, for example, if the manufactured good is processed food and the service is retail sale of that food. If, conversely, the local service is a substitute for the manufactured good, higher transport costs will increase demand for the local service, driving farmers to the city. This would be the case for imported versus locally-made cookware.

The effect on per capita income, derived similarly, is also ambiguous:

\[
\frac{d\ln y}{d(x p_o)} = \frac{-2\eta_{p_m} p_a N_u f'}{[1 - (x p_o)^2][y + (\eta_y + \eta_{p_u})p_a N_u f'][1 - (x p_o)^2]} - \frac{1}{1 - x p_o}. \tag{8}
\]

When agriculture has decreasing returns \((f' < 0)\), the role of \(\eta_{p_m}\) is in the same direction as its role in the urban population effect. The effect on per capita income will however be unambiguously negative if transport costs \(x p_o\) are large enough.

\(^7\)Stability requires that, holding \(p_u\) and \(p_m\) constant, any perturbation resulting in a rural income surplus \((p_a f(N_a) - bp_u > 0)\) cannot be strengthened by migration to the rural area, or equivalently \(\frac{1}{y} \frac{1}{dN_a} [p_a f(N_a) - bp_u] f(N_a) = bp_a (1 + \eta_y/\eta_{p_u}) N_a f' + \frac{\eta_u}{\eta_{p_u}} < 0\). Assuming the local service is not a Giffen good, \(\eta_{p_u} < 0\), so \(y + p_a (\eta_{p_u} + \eta_y) N_a f' > 0\).
3 Data and spatial methods

In order to test this model, attention is restricted to a set of 15 coastal primate countries in which the largest city is also the main port, so transportation to the primate city is important for trade with both the largest domestic market and the rest of the world (Figure 1).\textsuperscript{8} Counterclockwise from the northwest, these countries are Mauritania, Senegal, Guinea, Sierra Leone, Liberia, Côte d’Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Gabon, Angola, Mozambique, and Tanzania. Further details about all data used are in the Appendix.

3.1 City lights

To date, very little economic data, especially for income and especially as a panel, have been available for individual African cities.\textsuperscript{9} In most national household surveys, if any city is individually identifiable, it is only the largest city in a country. The largest program of firm surveys, the World Bank Enterprise Surveys, rarely collects data in more than four cities per African country, and survey documentation suggests that these cities were not chosen randomly.\textsuperscript{10} Censuses often report populations for many cities, but they are almost always carried out at intervals of at least ten years, which limits their usefulness. Specifically, in the present sample of countries, only two, Mozambique and Benin, carried out multiple censuses during the study period.\textsuperscript{11} In order to fill this gap, I propose a novel data source as a proxy for city-level income: satellite data on light emitted into space at night.

Satellites from the United States Air Force Defense Meteorological Satellite Program (DMSP) have been recording data on lights at night using their Operational Linescan System sensor since the mid-1960s, with a global digital archive beginning in 1992. Since two satellites are recording in most years, 30 satellite-years worth of

\textsuperscript{8}Five other countries in sub-Saharan Africa fit this criterion but are not included in analysis. Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia are excluded because they lack (at least) roads data. Using the city definitions below, The Gambia has only one city, and therefore it provides no information in the presence of country*year fixed effects.

\textsuperscript{9}Some administrative data on economic indicators such as employment are collected for subnational regions in some countries, including Tanzania, but assessing their comparability is a challenge, and they are typically only available for large regions and not for multiple years.

\textsuperscript{10}Among African coastal primate countries, the Enterprise Surveys only have multiple (two) years for Angola (3 cities sampled in both years), Cameroon (3 cities sampled in both years, 1 with only 7 firms surveyed in one year), Nigeria (no cities surveyed in both years), and Senegal (2 cities, 1 has only 3 firms in one year). So the full sample would be 6-8 cities in three countries in two years. Only in Senegal are both of these years in the 17-year sample period used in the rest of the paper.

\textsuperscript{11}Population effect estimates for this small subsample of 76 city-years are reported at the end of Section 5.
data are available for the 17-year period 1992–2008. Each 30-arcsecond pixel in each satellite-year contains a digital number (DN), an integer between 0 and 63, inclusive, that represents an average of lights in all nights after sunlight, moonlight, aurorae, forest fires, and clouds have been removed algorithmically, leaving mostly human settlements. Figure 2 shows the lights data for one satellite-year for Tanzania. No lights are visible in the overwhelming majority of the land area. In Figure 3, a closer view of Dar es Salaam, Tanzania’s largest city, shows a contiguously lit area 20–30 km across, extending farther in a few directions along main intercity roads just as the city’s built up area does.

Henderson, Storeygard and Weil (2012) show that light growth is a good proxy for income growth at the national level. Annual changes in gross domestic product (GDP) are correlated with changes in DN, with an elasticity of approximately 0.3 for a global sample as well as a sample of low and middle income countries. In both samples, the lights explain about 20 percent of the variation in log GDP net of country and year fixed effects. Table 1, column 1, reports the estimated global lights-GDP elasticity from Henderson, Storeygard and Weil (2012). One might expect that the elasticity in sub-Saharan Africa is lower, if changes in lights there are starting from a much lower base. Conversely, because the lights are less topcoded in poorer regions, the elasticity could be higher. Column 2 shows that the lights-GDP relationship is not significantly different for 41 sub-Saharan African countries than it is for the rest of the world. If anything, the point estimate on the interaction term suggests that the African lights-GDP elasticity is higher (though it is not significantly different from the rest of the world). Column 3 repeats the same exercise for the 15 coastal primate African countries, with very similar results.

The chief strength of the lights lies in their geographic specificity—they are highly local measures. To proceed with lights as a measure of city-level GDP, it must first be shown that the strong national relationship holds for subnational regions. This is problematic because of a mismatch in data availability. Rich countries tend to have good local economic data, but the lights data are heavily topcoded in their cities. Lights topcoding is less of a problem in most poorer countries, and especially in sub-Saharan Africa, where almost no pixels are topcoded (15 per 100,000, or 3 per 100,000 outside of South Africa and Nigeria). However, good local economic data are rarely available. China represents a good compromise, with relatively little topcoding but relatively high

---

A 30-arcsecond pixel has an area of approximately 0.86 square km at the equator, decreasing proportionally with the cosine of latitude. The data are processed and distributed by the United States National Oceanic and Atmospheric Administration (NOAA), http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html. Accessed 22 January 2010.
quality income data for a short panel of regions.

China has panel GDP data for two relevant types of subnational regions: cities proper and prefectures.\textsuperscript{13} Columns 4 and 5 in Table 1 show, at the city proper and prefecture level, respectively, that the elasticity of GDP with respect to light is significantly positive in a 1990/1992–2005 long difference specification.\textsuperscript{14} The point estimate is very similar to the one for the global sample.

For the present study, several steps were taken to convert the pixel-level lights data into cities. Figures 4–6 document these steps for Tanzania. The 30 satellite-years of lights data were first combined into one binary grid encoding whether a pixel was lit in at least one satellite-year (Figure 4). These ever-lit areas were then converted to polygons: contiguous ever-lit pixels were aggregated, and their digital numbers were summed within each satellite-year. Polygons not corresponding to a known city, based on census populations with latitude-longitude pairs from Brinkhoff (2010), were dropped. The dropped lights most likely correspond to forest fires or random noise in the sensor output not flagged by NOAA’s algorithm, or smaller towns/large villages, and contain 13 to 16 percent of total DN in the 15-country sample.\textsuperscript{15} Lights arising from gas flares, as delineated by Elvidge et al. (2009) were also removed.

Figure 5 shows all the lit polygons and city points. For those light polygons that did contain one or more census cities, the population of all such cities were summed to obtain a combined population. Most lights correspond to at most one census city.\textsuperscript{16} Figure 6 shows those lights that correspond to known cities. In most countries, census information about cities with populations as small as 10 thousand was available, but in some, the cutoff was higher. For all regressions below, I restrict to cities with combined population over 20 thousand and lit in at least 2 years. The total DN was recorded for each city polygon for each year, averaging across multiple satellite-years where available.

\textsuperscript{13}I am grateful to Vernon Henderson and Qinghua Zhang for providing the China evidence based on their work in progress with Nathaniel Baum-Snow, Loren Brandt, and Matthew Turner.
\textsuperscript{14}The lights are from 1992 but the GDP data are from 1990—the closest year with good data.
\textsuperscript{15}Forest and agricultural fires are removed from the data by NOAA based on their duration. If a fire lasts longer than the threshold used, it will appear in the data.
\textsuperscript{16}Light pixels for a given satellite-year actually represent the average light from several slightly larger overlapping pixels from many orbits within the satellite-year. Because of this, the lit area of a given city tends to be somewhat larger than its actual size. Among densely populated high and middle income countries, this means, for example, that the majority of land in the United States east of the Mississippi River or in continental Western Europe is contiguously lit, so that cities cannot be defined purely based on light contiguity. In Africa, this is much less of a problem because of sparser light overall. Snow also tends to increase the footprint and magnitude of lights. Again, this is less of a problem in Africa than elsewhere. And even if the area of a given city is overestimated, the light summed for that city in still presumably coming from that city or its outskirts—it may just be partially displaced a pixel or two from where it actually originates.
The light in each country with the largest associated population in 1992 is designated the primate.\textsuperscript{17}

### 3.2 Transport costs

Trade and transport cost data are also not widely available for Africa.\textsuperscript{18} In the international trade literature, trade costs are sometimes estimated from a gravity equation based on trade flows (Anderson and van Wincoop, 2004), or price dispersion (Donaldson, 2010; Atkin and Donaldson, 2013) but trade flow data between cities and city-level price data are also not widely available. Furthermore, city growth may endogenously decrease transport costs. Among other reasons including the allocation of paved roads (discussed below), more transport companies are likely to compete on a route to a growing city than on a route to a stagnant one.

I address this problem by decomposing variable transport costs into two components: 1) the world price of oil, which varies across time but not across cities, and 2) the road distance between a city and its country’s primate, which varies across space but not time. Figure 7 shows the evolution of oil prices during the study period. In general, they were relatively steady until a consistent rise beginning in 2002. However, there was some movement in the previous period, including substantial decreases (as a fraction of the initial price) in 1992–1994, 1996–1998, and 2000–2001.

Oil is a convenient proxy for transport cost per distance because no countries in the sample are individually capable of influencing its price substantially. However, motorists consume refined petroleum products, mostly gasoline and diesel, not oil, and some countries, especially oil producers, subsidize their prices. Country-specific diesel prices, surveyed in November in the main city, are available for most countries roughly every two years (Deutsche Gesellschaft für Technische Zusammenarbeit, 2009). As shown in Figure 7, diesel prices averaged over a balanced panel of 12 countries from the main estimation sample generally rise in parallel with oil prices. Nigeria, Gabon, and Angola, the three sample countries for which oil production represents the largest fraction of GDP, show similar, though somewhat noisier time trends (Appendix Figure A.1) despite the fact that they typically had lower prices than average in most years, most likely because of subsidies. Using data from a survey of truckers in several African countries, Teravaninthorn and Raballand (2009) estimate that fuel represented roughly

\textsuperscript{17}In practice, the primate designation does not change over the course of the sample period in any sample country.

\textsuperscript{18}Teravaninthorn and Raballand (2009) provide figures for several important routes from landlocked country capitals to the ports that serve them.
35 percent of transport costs for trucks in 2005, when oil prices were roughly the mean of the minimum and maximum annual price for the period.

Rudimentary national statistics like road density and percentage of roads paved, which are typically used in cross-national studies, fail to capture the role of roads in connecting cities, and are subject to a great deal of error.\textsuperscript{19} A recent World Bank project on infrastructure in Africa has improved the state of georeferenced roads data for the continent so that they can be used to assign infrastructure to specific cities and routes between cities (World Bank, 2010).\textsuperscript{20} The resulting dataset combines information on road location and surface assigned to specific (and recent) years from each country’s Transport Ministry or equivalent, or a consultant specific to the project.\textsuperscript{21} It contains information on over a million kilometers of roads in 39 countries. For over 90 percent of this length, a measure of the surface type is recorded. The comprehensiveness of the coverage varies by country, but only in that some countries contain more minor roads. Intercity roads are available for all countries. Figure 8 shows these roads data for Tanzania. Roads go through all the cities from Figure 6. Most roads are unpaved, and most paved roads are found along a few major corridors.

The shortest path along the road network was calculated between the centroid of each city-light and three destinations: (1) its country’s primate city, defined as the light with the largest associated population in 1992, (2) the nearest city in the same country with a 1992 population of at least one hundred thousand, and (3) the nearest city in the same country with a 1992 population in the top quintile of the population distribution for that country. Plausible primate city routes were found for 287 out of 299 cities in the 15-country sample. Figure 9 shows all roads and primate routes for

\textsuperscript{19}Countries such as Canada, Australia, and Botswana have low road density relative to their economic peers. But their road systems are not particularly inadequate. In each case, a contiguous region containing half or more of total land area has very low population density, so that the marginal benefit of an additional road there is very low. The chief problem with the percentage of roads paved is that the denominator is affected by the coverage of national roads data systems, which can vary substantially. The World Development Indicators (WDI) reports both of these measures. But of the 120 annual changes in road density available for the 255 country-years in the present sample, 66 are zero, while another 8 are, implausibly, over 10 percent in absolute value. When data are missing for a period of one or more years for a given country, the annual growth rates implied by the values before and after the data gap are even more implausible. Similar statements can be made about the percent paved data. These large changes are likely due to reclassification rather than actual road construction.

\textsuperscript{20}The most comprehensive previous spatial database, Vector Map Level 0 (VMAP0, formerly known as Digital Chart of the World, DCW), is a declassified US military product combining data of unknown quality from 4 decades, with little metadata. In some countries, there are clear gaps in coverage. Most strikingly, the most densely populated areas of Bangladesh, surrounding the capital Dhaka, have essentially no roads.

\textsuperscript{21}I am unaware of any systematic time-varying data on road surface in the countries under study. Burgess et al. (2013) use time-varying data for Kenya.
Tanzania.

Limiting attention to road transport costs might be problematic if rail played a major and independent role. However, roads dominate transport in Africa, carrying 80 to 90 percent of passenger and freight traffic (Gwilliam, 2011). In most countries, rail only exists along a few corridors that are also served by roads. Rail is also less likely to matter than roads because of its higher fuel efficiency and greater dependence on parastatals with long term contracts. Of course, regardless of rail’s importance, for the purposes of this paper, it is still a transportation form that uses energy generated from fossil fuels. Empirically, controlling for rail routes does not substantially affect the main result.

4 Empirical specification

My baseline specification testing the effect of transport costs in Equation (5) is:

\[ \ln Y_{it} = \beta p_t x_i + \lambda_{ct} + \gamma_i + \omega_i t + \epsilon_{it} \]  

(9)

where \( Y_{it} \) is light output for city \( i \) in year \( t \), \( p_t \) is the price of oil, \( x_i \) is the distance between city \( i \) and its country’s primate city along the road network, \( \lambda_{ct} \) is a country-year fixed effect (FE), \( \gamma_i \) is a city fixed effect, and \( \omega_i t \) is a linear city-specific time trend. Standard errors are clustered at the city level.\(^{22}\) The regression sample is limited to cities with a 1992 population of at least 20,000, lit in more than one year, because populations and locations of cities of less than 20,000 are not available for several countries, and cities lit in only one year add no intensive margin information because of the city fixed effects. The time period is limited to 1992–2008 because of the lights data availability. Summary statistics for the resulting sample of 287 cities in 17 years are in Table 2. Distances are measured in kilometers, and prices are in dollars.

As oil prices increased over the course of the last decade, I expect that transport costs increased more for cities farther away from their country’s core. Thus, I can use static distance measures interacted with the exogenous oil price increase to identify the differential change in transport costs faced by near and far cities. The model suggests that the less-connected cities will experience a relative loss with higher oil prices, unless there is substantial demand for imported manufactured goods and these goods are primarily substitutes for local services.

Country-year fixed effects control for any national-level time-varying economic con-

\(^{22}\)ff, alternatively, the methods of Conley (1999) are used to account for spatial and temporal autocorrelation, the resulting standard errors are smaller.
ditions. In the context of the model, the relevant factors are prices in the primate city. Empirically, they also include the level of industrialization, oil production, and terms of trade, as well as policies, including gasoline subsidies and preferential trade pacts with developed countries like the American Growth and Opportunity Act (AGOA) and trade portions of the European Union’s Cotonou Agreement.\textsuperscript{23} They also control for global macroeconomic fluctuations, including commodity prices, as well as differences across satellites in the lights data. City fixed effects control for initial size and all other fixed city characteristics. City-specific time trends allow each city to be on its own growth path.

The identifying assumption for $\beta$ is thus that there is no other time-varying within-country variation net of linear growth that is correlated with network distance to the primate times the change in oil price that affects city growth, or more specifically,

$$E(\epsilon_{it}|p_{st}, x_i, \lambda_{cs}, \gamma_i, \omega_is) = 0, \quad s, t = 1992, 1993, ..., 2008$$

In specification checks below, distances to other large cities are tested in combination with distance to the primate to determine whether it is actually the cost to the primate city that matters, as opposed to other correlated transport costs.

In order for these effects to be salient, it must be the case that transport costs substantially affect contemporaneous economic activity, and that oil prices affect contemporaneous transport costs. On the first point, Gollin and Rogerson (2010) find that in Uganda, internal transport costs for crops can easily exceed their farmgate price. It is hard to imagine that this does not affect cropping decisions. Using national trade flows data, Lim\~ao and Venables (2001) find that transport costs affect international trade substantially, with an elasticity of around $-3$. The World Bank Enterprise Surveys of establishments ask respondents whether “transportation of goods, supplies, and inputs...present any obstacle to the current operations of your establishment?” In the most recent (2006–2009) round, in all 15 countries studied, over half of respondents said that transportation was an obstacle, and in 11 countries, at least a quarter said that it was a major or very severe problem.

On the second point, Teravaninthorn and Raballand (2009) report a breakdown of transport costs for truckers and trucking companies along several international corridors from a port to the capital city of a landlocked country. Because these are international journeys, I expect costs to be somewhat higher than for those journeys that remain in the coastal country, but as a fraction of total distance, these routes are overwhelmingly

\textsuperscript{23}See Table 4 for consideration of the possibility of subnationally varying effects of changing trade policy.
in the coastal country. Furthermore, it is not obvious that the international nature of these journeys increases distance-dependent variable costs more than fixed costs, as some components of the latter like insurance and customs logistics may also be higher. On the Accra-Ouagadougou route, over 80 percent of which by distance is in Ghana, variable costs are 9 times the size of fixed costs, and fuel represents 74 percent of variable costs. Tires, which also use petroleum products, represent another 15 percent. For the Mombasa-Kampala route, over 80 percent of which is in Kenya, the analogous numbers are 0.45, 79 percent and 13 percent. For the Douala-N’Djamena route, nearly all of which is in Cameroon, the analogous numbers are 4.8, 60 percent and 17 percent.

Measured light is not strictly the same as light output. Most noticeably, some 5 percent of city-years have a DN value of zero. Roughly half of these are from years before the city was ever lit (“pre-entry”). This is consistent with continuous city growth that eventually passed a threshold above which lights were detectable. The other half are years in which a city has no lights after having been lit in one or more previous years (“late zeroes”).

In the Appendix, I present a model of the satellite-pixel-year level data-generating process that could in principle be estimated using maximum likelihood methods. However, the relationship of interest and all of the regressors are at the city level. Rather than perform this estimation on a dataset with approximately 5.6 million pixel-satellite-years, I instead simply sum lights across pixels within a city, average across satellites within a year, and run tobit regressions with a cutoff value of 5.5, because 6 is the smallest nonzero value found in the data, and the smallest increment is 0.5.

5 Results

5.1 Main results

Table 3, Column 1, reports a simple “difference-in-difference” version of the main specification. The negative coefficient on \(1(distance(Primate) < \text{median}) * 1(P_{oil} < \text{median})\) shows that on average, net of city and country*year fixed effects and and linear city trends, farther cities lose more lights in high oil price years than nearer cities. Figure 10 shows a running line smoothing of \(\ln(lights + 5.5)\) on \(distance(Primate) * P_{oil}\), net of the same fixed effects and trends, along with a 95 percent confidence interval.

---

24 This statistic is calculated after cities that are never lit and cities lit in only one year have already been removed from the sample, as described above.

25 Fixed effects tobits are biased for short panels, but this panel is 17 years long and a small percentage of observations are censored. Using a tobit cutoff of 1 results in estimates of the coefficient of interest with larger absolute values.
Except in the tails of the distribution, where the confidence interval is extremely wide, the relationship negative and quite linear.

Column 2 of Table 3 shows the baseline specification in Equation (9). In this and all subsequent tables, distances are measured in thousands of kilometers, and oil prices in hundreds of dollars. The coefficient of interest, \(-0.715\) on \(distance(Primate) \times P_{oil}\), implies that if the price of oil increased from $25 to $97 per barrel (as it did between 2002 and 2008), if city A is 465 kilometers (1 standard deviation) farther away from the primate than initially identical city B, its lights are 21 percent smaller than city B’s at the end of the period. Applying the light-income growth elasticity \(\epsilon_{GDP, light} = 0.284\) from Henderson, Storeygard and Weil (2012), this implies a city product differential of 6.6 percent. This is consistent with the simple version of the model above. Far cities see their transport costs increase more than near cities, so their income falls more.

The remainder of Table 3 considers several other functional forms. In column 3, 5.5 is added before the DN is logged and an OLS specification is used instead of a tobit. Results are very similar. Figure 7 shows that there were two broad oil price regimes during this period: a relatively flat stretch followed by a steeply rising one. In order to ensure that the results above are not simply driven by the difference between these two regimes, column 4 reports the results of a specification in which two linear splines were fit for each city. The “knot” year, the transition between the two regimes, along with both slopes, are estimated separately for each city to minimize the variance in city-specific residuals. The magnitude of the coefficient of interest is slightly smaller than in the baseline case, but it is still negative, despite the fact that even more temporal variation has been removed.\(^{26}\)

Column 5 adds a lagged version of the main term of interest. The contemporaneous term decreases in magnitude, not surprisingly given autocorrelation in oil prices, but remains significant, while the lagged term is smaller and insignificant. The effects appear to be felt most strongly in the year of an oil price change.

The coefficient in column 2 can be interpreted as a semi-elasticity in the context of the model above. An elasticity of city product with respect to transport costs is in some respects a more intuitive measure, but since \(ln(p_t x_i)\) is equal to \(ln(p_t) + ln(x_i)\), it is collinear with the country-year and city fixed effects and cannot be estimated separately. However, a distance-specific transport cost elasticity can be calculated. Column 6 reports the coefficient of interest when \(p_t x_i\) is replaced with \(ln(p_t) x_i\). It is again negative and significant as expected. This can be translated into a distance-

\(^{26}\)While the standard errors have not been corrected for the first stage estimation of the splines, the relevant comparison is that the point estimate is very similar to that in the baseline specification.
specific elasticity using three additional parameters:

$$\epsilon_{GDP,\tau} = \frac{\epsilon_{GDP,\text{light}}\epsilon_{\text{light},P\text{oil}}}{\epsilon_{\tau,P\text{diesel}}\epsilon_{P\text{diesel},P\text{oil}}}.$$  \hspace{1cm} (11)

A simple regression of $ln(P_{\text{diesel}})$ on $ln(P_{\text{oil}})$ using the (Deutsche Gesellschaft für Technische Zusammenarbeit, 2009) data for the sample countries provides an estimate of $\epsilon_{P_{\text{diesel}},P_{\text{oil}}} = 0.5$. Treating the Teravaninthorn and Raballand (2009) average fuel share as the marginal fuel share implies $\epsilon_{\tau,P_{\text{diesel}}} = 0.35$. Combining these estimates implies $\epsilon_{GDP,\tau} = -0.25$ at the median distance from the primate, 439 km, and -0.52 one standard deviation (465 km) farther away. This calculation is meant to be illustrative, as it may suffer from several potential biases, including upward (toward zero) bias from substitutability of oil in the production of transport and downward (away from zero) bias from substitutability of transport in the production of city activity.\(^{27}\)

Country size varies dramatically within the estimation sample. For example, the farthest city in Sierra Leone is only 310 kilometers away from the primate, whereas in Mozambique, the farthest is over 2000 kilometers away. In order to ensure that not all variation is coming from the largest countries, column 7 shows a specification in which $p_t x_i$ is replaced by $ln(p_t)ln(x_i)$.\(^{28}\) The coefficient of interest remains negative and significant.

Tables 4 and 5 consider several potential omitted variables that may be biasing the coefficient of interest. The prices of commodities other than oil were rising in parallel with the oil price in this time period. It could be the case that country governments spent these commodity windfalls disproportionately near the primate city, either because it is easier for government officials to travel to project sites in cities near the capital, or because governments are more concerned with pleasing the residents of these cities. Table 4 reports results related to this potential effect. Column 1 reports the baseline specification on a sample restricted to the 94 percent of country-years for which national government expenditure data are available. The coefficient of interest is still negative and significant, though a little smaller in magnitude. Column 2 controls for the interaction between distance and the log of the purchasing power parity (PPP) value of government expenditure. The coefficient of interest is essentially unchanged from column 1, and the government expenditure interaction effect is small and insignificant. Columns 3 and 4 are analogous to columns 1 and 2, with natural resource

\(^{27}\)Using the method of Goodman (1960) and assuming independence across samples, the estimate of the product $\epsilon_{GDP,\text{light}}\epsilon_{\text{light},P\text{oil}} = -0.0975$ has a standard error 0.0328.

\(^{28}\)The distance from the primate to itself is arbitrarily redefined as 1 kilometer in the log-log specification.
export income, which is potentially better-measured but only available for 58 percent of the sample, replacing government expenditure. Natural resource export income is defined here as the PPP value of mineral and fuel exports. The main effect is larger for this sample, but the distance-natural resource interaction is small and insignificant. In column 5, the natural log of total PPP GDP, which is almost universally available, replaces natural resource income. Its interaction with distance is substantially negative, but noisy, and the main coefficient of interest, while slightly smaller like in the government expenditure interactions, remains negative and significant. The results in this table suggest that oil prices, not overall commodity income fluctuations, are driving the effect shown in the baseline specification.

As noted above, country*year fixed effects control for nationally constant effects of trade policy. However, Coşar and Fajgelbaum (2013) and others have argued that openness may have a greater impact near trading hubs. In columns 6 and 7, I consider the possibility that the most notable change in trade policy affecting the entire region, AGOA, is driving my results. AGOA came into effect in ten sample countries in 2000, and it had been implemented at one point in all countries as of 2008, though by this time 2 countries had been declared ineligible. Thus there is some variation in the timing of its implementation across countries, though this timing is clearly endogenous. Column 6 includes a differential effect for AGOA-eligible country years, and column 7 allows this differential to be proportional to the number of years since implementation. In neither case is this differential significant, and the point estimates in each case imply that AGOA mitigated the remoteness differential, rather than exacerbating it.

The evidence shown so far does not rule out the possibility that oil prices are having an effect for reasons not directly related to intercity road transport costs. However, any confounding effect must also be correlated with distance to the primate city. The oil industry could also have national effects that are correlated with the oil price, but these are removed by the country*year fixed effects. Table 5 provides evidence against four potential alternative mechanisms.

Oil could be affecting transport costs by rail, so that including only road-based cost would confound the two effects. Column 1 considers, along with the main effect, the differential impact of the transport cost measure on cities with a rail connection to their country’s primate city. The main effect is about the same as in the baseline

29On AGOA’s impact, see Frazer and Biesebroeck (2010) and Rotunno et al. (2013). While the European Union’s Cotonou Agreement also affected African trade, it had greater continuity with the Lomé Convention it replaced, and the date of its implementation did not vary by country as much as AGOA’s did.

30If eligibility was effective in October or earlier, the year is counted.
specification, and the additional rail effect is small and insignificant.

Because some sample countries are oil producers, oil prices could also affect the within-country pattern of economic activity on the oil supply side. When oil prices rise, cities near production or exploration areas are more likely to benefit from increased employment and wages. Because oil wells in this set of countries are largely near the coast, oil well proximity is correlated with primate city proximity. However, while transport costs to the primate increase continuously away from the coast, it is unlikely that local oil industry effects persist throughout the country. In columns 2 and 3, I report results for the baseline specification when the 24 cities at least partially within 50 kilometers and 50 cities within 100 kilometers of an oil or gas field are excluded. They are very similar to the baseline specification results. This suggests that local oil industry effects are not driving my results. If other commodities with correlated price series are disproportionately produced near the primate, this could have a related effect, while if they are disproportionately produced far from the primate, this could have an opposite effect. Systematic information on the production location of all other important commodities is not as readily available, and at least in the case of agricultural commodities, the decision to produce in a given location is endogenous to the price.\textsuperscript{31}

It is also possible that the price of oil (and gas and coal, whose prices tend to co-vary with oil’s) is directly reducing the size of distant city lights, because light is produced by electricity, and some electricity is produced by fossil fuels. It is unlikely that this is driving any results, for two reasons. First, nearly all countries in this region have national grids, and many are connected to international ones. Power companies are almost all either state monopolies or former state monopolies wholly or partially privatized as a single entity. Their posted rate structures are characterized by quantity discounts, or more often, premia, and differentiated by sector (residential, commercial, industrial), but with no explicit within-country geographic variation.\textsuperscript{32} Second, to the extent that transmission costs proportional to distance matter in practice, more than a third of electricity in the region is produced from hydropower, with the remainder produced primarily by thermal (oil, gas, or coal) plants (World Bank, 2010). If expensive oil is increasing the price of electricity within countries, it should do so less where hydro is the most likely source. World Bank (2010) reports the location of power plants, by

\textsuperscript{31}Given the multiplicative transport costs assumed in the model, it is also possible that farther cities are only affected due to higher commodity prices, not unit transport prices per se. However, this still means that the change in transport costs is driving my results. I am grateful to an anonymous reviewer for both of these points.

\textsuperscript{32}The one exception to this is a small additional tax on some rates for Abidjan, the primate city of Côte d’Ivoire.
Column 4 of Table 5 restricts attention to the 251 cities in those countries that have both hydro and thermal power plants and adds to the baseline specification a term interacting $\text{distance}(\text{Primate}) \times P_{\text{oil}}$ with an indicator that the closest plant to the city is a hydro plant. The coefficient on the triple interaction is small and insignificant and has little effect on the coefficient on $\text{distance}(\text{Primate}) \times P_{\text{oil}}$. Proximity to a hydro plant has no effect on the relationship between transport costs and lights.

A related concern is that cities far from the primate might not be on the power grid, and therefore might be more likely to rely on electric lights fueled by diesel generators. High oil prices could reduce diesel generator use, lowering lights more in faraway cities than near ones. Data on the location of electrical transmission lines are available from World Bank (2010) for 13 of 15 sample countries. Transmission lines pass through 184 of 260 cities (71%) in these 13 countries. In column 5, when the sample is restricted to these 184 cities least likely to rely on diesel generators, the results are very similar to the baseline. Of course, even in cities with grid electricity, some households have generators, due to sporadic supply. Demographic and Health Surveys from five sample countries have information on household generator ownership. Column 6 reports the results of a regression of household generator ownership on distance to the primate, with country fixed effects, for the 135 sample cities with the relevant information. The slope is not significantly different from zero, and the point estimate suggests that if anything, households are more likely to own a generator when they are closer to the primate. This suggests that diesel generators are not driving my results.\footnote{Kerosene lamps are even less likely to be driving my results. Data on household electricity are available in 24 Demographic and Health Surveys for nine sample countries during the sample period. Weighting by urban population within and then across countries using data and projections from United Nations (2008), 75 percent of urban households have electricity. A simple unweighted average of the 24 surveys and a compound average of the nine countries after averaging across surveys in each country give 62 and 63 percent, respectively. A third of urban residents (as defined for the purposes of this study) in these nine countries live in the primate light, suggesting that at least 63 percent of non-primate urban residents report having access to electricity. According to Mills (2002), locally made kerosene lamps produce 5 to 10 lumens, while store-bought models produce 40 to 50 lumens. Electrical light tends to be cheaper than kerosene, so households with electrical connections are unlikely to use kerosene for lighting. A 60-watt incandescent light bulb produces 800 lumens. So even if all households without electricity had the most advanced kerosene lamp and other households had a single 60-watt bulb, only 2 percent of household light would be from kerosene. This is almost certainly a substantial overstatement, because outdoor light is more likely to come from public or commercial establishments that are less likely to use kerosene.}

Table 6 reports the results of several alternate specifications related to the measurement of transport costs. As noted above motorists use gasoline and diesel, not oil, so as a further check on my results, I can use the price of diesel instead of the price of oil for the subset of countries and years for which it is available. However, countries often
subsidize diesel, and this introduces potential reverse causality because countries may subsidize in part to prevent the isolation of hinterland cities. The oil price is a valid instrument for the diesel price, because it is a very strong predictor and is set on world markets in which no sample country holds sway. In column 1, results for the main specification are broadly similar when the sample is restricted to country-years with a known diesel price. In column 2, the OLS specification using the diesel price instead of the oil price also has a negative and significant semi-elasticity. In column 3, the effect of the diesel price is larger when the oil price is used to instrument for it.

In all results reported so far, the shortest route to the primate was calculated assuming that travel along unpaved roads is equivalent to travel along paved roads. Columns 4 and 5, report results based on the alternate assumptions that travel on unpaved roads is 50 percent and 100 percent more costly, respectively, than travel along paved roads. In each case, the regressor of interest uses effective distance (i.e. route distance with each unpaved segment lengthened by the appropriate factor) instead of simple distance. Although the calculated routes are slightly different, the coefficient of interest is very similar. Road distance and great circle distance are extremely highly correlated in this sample, at 0.988. This confirms that other results are not due to missing links in the road network data - in general, most cities have relatively straight routes to the primate. Unsurprisingly given this, when both oil price-times-road distance and oil price-times-great-circle distance are entered in column 6, standard errors on both become very large.

All results so far have considered transport costs only to their country’s primate city. If large foreign cities or medium-sized cities fill a similar role for some small cities, transport costs to them may also drive economic activity. Table 7 reports results controlling for transport costs to these other cities. Column 1 controls for transport costs to the nearest foreign port when that is the closest port, and the nearest foreign primate when that is the closest primate. Neither has a significant effect, and the coefficient of interest is virtually unchanged. These distances are calculated as great circle distances because border crossings are not well-defined in the roads dataset, and are likely measured with more error also because border crossings vary very widely in the average delay truckers face at them (Teravaninthorn and Raballand, 2009). Using great circle distance for the domestic primate effect as well gives similar results. Column 2 controls for transport costs to an alternate domestic destination, the nearest city with a 1992 population of at least 100 thousand. About a third of the cities in the sample

\[34\] Of course, the oil price also affects the gasoline price, in a very similar way, so the diesel price is best considered as a proxy for diesel and gasoline prices together in this context.
have a 1992 population of at least 100 thousand. The effect of this new cost has a magnitude comparable to the primate cost effect, but with a much larger standard error, and it does not impact the primate cost coefficient substantially. Column 3 refines column 2’s specification slightly by only considering this alternate distance in the case of cities whose nearest city of at least 100 thousand is not the primate, to reduce the correlation between the two measures. The results are similar. Columns 4 and 5 are analogous to columns 2 and 3, with the intermediate destination now the nearest city in the top quintile (by 1992 population) of sample cities in the country. In essence, the absolute size criterion used in columns 2 and 3 is replaced with a relative one. The effect of the primate cost is reduced a little more, but it is still significant, and the effect of the top quintile city is twice as large or more. This result will be explored further when road surface is considered explicitly in Table 9. Still, no two of the primate city coefficients in this table are significantly different from each other.

### 5.2 Road Surface

The results shown so far have not used the available information on road surface. However, road surface helps to explain under what circumstances transport costs to intermediate cities might matter more than transport costs to the primate. The roads dataset includes (static) information on road surface type, so each route can be characterized by the fraction of its length that is paved. For simplicity, this measure is converted to an indicator denoting whether a city’s route is more paved than the route of the median city in its country. If road surface were randomly assigned, in the context of the model above, in the short run we might expect a less negative $\beta$ for the more paved routes, because driving on paved roads is cheaper, in fuel, time, and maintenance costs, than driving on unpaved roads. In a study on South African roads, du Plessis, Visser and Curtayne (1990) find that the fuel efficiency of a 12-ton truck traveling 80 kilometers per hour is 12-13 percent lower on a poor unpaved road (Quarter-car Index, QI=200) than the same truck at the same speed on even a poor paved road (QI=80). This is almost certainly an underestimate, because trucks are unable to maintain high speeds on unpaved roads, and fuel efficiency tends to rise with speed in this range. However, road surface is clearly not randomly assigned, as governments and donors are more likely to pave a road to a city that is economically important or expected to grow.\textsuperscript{35} Even if roads were initially assigned randomly, after assignment better-connected places are

\textsuperscript{35}For an exception, see Gonzalez-Navarro and Quintana-Domeque (2010).
more able to engage in trade.

The road network of a country can change endogenously, in both an extensive and an intensive sense. On the extensive margin, entirely new roads can be built. While this occasionally happens, the overwhelming majority of road improvements take place in the location of existing roads, because this is so much cheaper than purchasing/appropriating, clearing, and grading new land. In rich countries, it is sometimes the case that limited access roads are built away from the existing route between two cities, because the existing road serves a local purpose that would be destroyed by access limitations. But limited access roads are extremely rare in sub-Saharan Africa outside of South Africa.

The intensive margin is a somewhat thornier problem. Road surfaces can be improved or widened, and they can also deteriorate. However, I expect that the oil price changes in this time period, which include a nominal increase by 760 percent between 1998 and 2008, are large enough that they overwhelm more modest changes in road infrastructure. While the $7 billion annual regional roads investment may be a substantial portion of regional annual GDP, it does not necessarily buy a large length of new or maintained roads. By comparison, China, which has less than half the land area, spent about $45 billion per year between 2000 and 2005 on highways alone (World Bank, 2007b), presumably with higher efficiency.36

In Table 8 columns 1 and 2, we see that empirically, hinterland cities with routes to the coastal primate that are more paved than the median route in that country are 0.280 and 0.584 log points larger on average than places with routes less paved, in terms of population and lights, respectively, after controlling for distance to the primate. In column 3, after controlling for distance to the primate, more paving is correlated with a larger fraction of adults working in the manufacturing sector, in a sample of districts in 4 countries (Ghana, Guinea, Senegal, and Tanzania) for which census data are available from Minnesota Population Center (2011). These results are all consistent with the idea that cities connected by more paved roads could be more hurt by higher oil prices because they are more economically connected to the primate, whereas cities that are connected by mostly unpaved roads are smaller and closer to autarky.

The regressions in Table 9 exploit the paving information by including two terms of the main effect $\beta_p x_i$ in Equation (9), one for cities with routes to the primate more paved than the median and one for cities with less paved routes. Column 1

36I am unaware of any dataset assigning road maintenance expenditures to road segments in any sub-Saharan African country.
demonstrates that the effect of transport cost to the primate is similar in the two categories of cities. While paving status was likely determined before the study period in most cases, it is potentially endogenous to local economic activity, limiting the scope for causal interpretation. It is likely that transport costs affect the two sets of cities in slightly different ways. Routes to some cities were paved for any number of reasons (early manufacturing promise, political or military importance, corruption), and then this paving helped these cities to grow more, at least in part because of transport-sensitive firms that were then penalized by increases in oil prices. On the other hand, unpaved roads require slower and more fuel intensive travel, so given the same demand for transport services, cities along them are penalized more per mile.

However, without a mostly paved road to the primate, firms in a city may seek alternate trading connections, relying on intermediate cities instead. Column 2 adds the transport cost to the nearest city in the top population quintile if that city is not the primate, separately based on the paving status (high or low) of the route to the primate. As in Table 7, higher transport cost to a top quintile city decreases output. However, this effect is limited to cities with relatively unpaved routes to the primate. This suggests that these cities, relatively unconnected to the primate, are essentially consumer cities as in the formulation of Jedwab (2013). In essence, their trade is funneled through a regional hub, not the primate. Conversely, among cities that are relatively well-connected to the primate, it is the primate distance that matters, not the top quintile city distance. Not surprisingly, the intermediate (top quintile non-primate) cities are themselves 20 percent more likely than other non-primate cities to have their connection to the primate mostly paved.

The results in column 2 are summarized graphically in Figure 11. Three cities, A, B, and C, identical except for their locations, are connected to a primate city P and a secondary (i.e. top quintile) city S, with the distance relationships \( d_{SA} = d_{SB} < d_{SC} < d_{PC} = d_{PB} < d_{PA} \). When oil price rise, if roads PA, PB and PC are paved (left figure), A will grow slower than B, which will grow about as fast as C. If these three roads are unpaved (right figure), A and B will grow at the same rate, faster than C.

5.3 Population and the long run

The framework in Section 2 suggests that changes in economic activity in a city can stem from changes in activity per person, a higher population, or both. In order to explore which is more salient in this context, I consider population as an alternative outcome in Table 10. As noted earlier, only Benin and Mozambique had two censuses during the
sample period. Even extending the analysis period back to the 1970, city populations are available for at least two censuses for only eleven countries. Six countries have city populations from three censuses. None have populations from more than three. Because of this limited temporal variation, the population analysis omits the city-specific linear time trends used earlier. Column 1 shows that in the small Benin-Mozambique sample of 38 cities, the estimated effect of transport costs on population is positive, but small and imprecisely estimated. In column 2, in the larger 11-country sample the transport cost coefficient is negative, but not significantly different from zero. This is consistent with some of the overall effect on city activity being in the form of population changes. However, the estimate is imprecise, and since it uses more limited variation over a much longer and different historical period, a more precise conclusion cannot be drawn.

6 Conclusion

This paper provides evidence that transport costs impact urban economic activity in sub-Saharan Africa, by making access to critical core cities more expensive, with recent increases in oil prices removing several percentage points from the size of far hinterland cities in countries where the largest city is on the coast. This is consistent with a simple model in which cities provide local services to farmers, whose output must be shipped to the primate city. It is not consistent with explanations related to commodity income or the generation of electricity. Despite being larger and likely facing smaller absolute changes in costs, cities with more paved routes are no more or less sensitive to changing transport costs, most likely because they are more integrated with national and global markets. However, cities with less paved routes seem to be less affected by transport costs to the primate city than they are by transport costs to a nearer secondary city. While in principle, the overall effect could be decomposed into population and economic activity per capita, limited population data do not provide strong evidence.

While previous work has shown that improvements in transport infrastructure can increase local activity and growth, most of it is based on very large construction projects, and none has been in an African context where industry is highly concentrated in the largest cities. The nature of the variation in the current work, provided simply by changes in oil prices interacted with distance, means that the results are unlikely to be driven by changes in long term investment in non-transport sectors. Instead, they provide clearer evidence of the direct short run effect of transport costs on urban economic activity. Annual city-level measures of economic activity provide evidence net of the country-year level variation used in previous comprehensive work on
urbanization, urban growth, and coastal access in sub-Saharan Africa. More generally, this city-level variation opens up exciting new possibilities for future research.

References


Collier, Paul (2007) *The bottom billion: why the poorest countries are failing and what can be done about it*, Oxford: Oxford University Press.


Gonzalez-Navarro, Marco and Climent Quintana-Domeque (2010) “Street Pavement: Results from an Infrastructure Experiment in Mexico,” Working Papers 1247, Princeton University, Department of Economics, Industrial Relations Section.


Henderson, Vernon, Adam Storeygard, and Uwe Deichmann (2014) “Has climate change promoted urbanization in Sub-Saharan Africa?” mimeo, Tufts University.


Tables and Figures
Table 1: Relationship between lights and economic activity

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(GDP)</td>
<td>ln(GDP)</td>
<td>ln(GDP)</td>
<td>∆ ln(GDP)</td>
<td>∆ ln(GDP)</td>
</tr>
<tr>
<td>ln(light)</td>
<td>0.284***</td>
<td>0.264***</td>
<td>0.270***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0301]</td>
<td>[0.0342]</td>
<td>[0.0307]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(SSA)*ln(light)</td>
<td>0.0702</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0491]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(coastal primate)*ln(light)</td>
<td>0.0907</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0680]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆ ln(light)</td>
<td></td>
<td></td>
<td></td>
<td>0.265**</td>
<td>0.268**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.101]</td>
<td>[0.119]</td>
</tr>
<tr>
<td>Observations</td>
<td>3,015</td>
<td>3,015</td>
<td>3,015</td>
<td>204</td>
<td>146</td>
</tr>
<tr>
<td>sample type</td>
<td>Global</td>
<td>Global</td>
<td>Global</td>
<td>China</td>
<td>China</td>
</tr>
<tr>
<td>unit type</td>
<td>countries</td>
<td>countries</td>
<td>countries</td>
<td>city proper</td>
<td>prefecture</td>
</tr>
<tr>
<td>country FEs</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>year FEs</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Each column is a separate OLS regression. In columns 1-3, these are balanced 17-year panels with country and year fixed effects included. In columns 4–5, these are long difference regressions between the years shown; the lights are from 1992 but the administrative GDP data are from 1990, the closest year with good data. The independent variable is the log of the lights digital number, summed across all pixels in the unit shown (having removed gas flares), and averaged across satellite-years within a year when applicable (and long-differenced in columns 4–5. 1(SSA) is a dummy for sub-Saharan Africa, and coastal primate means the 15 countries considered in the remainder of the paper. Robust standard errors are reported in brackets (clustered by country in columns 1–3. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 2: Descriptives

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>StdDev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lights digital number (DN)</td>
<td>3153</td>
<td>10992</td>
<td>0</td>
<td>152239</td>
</tr>
<tr>
<td>ln(light+5.5)</td>
<td>6.279</td>
<td>1.908</td>
<td>1.705</td>
<td>11.93</td>
</tr>
<tr>
<td>distance(Primate)</td>
<td>560.7</td>
<td>464.9</td>
<td>0</td>
<td>2360</td>
</tr>
<tr>
<td>distance(pop≥100k)</td>
<td>136.7</td>
<td>168</td>
<td>0</td>
<td>898.7</td>
</tr>
<tr>
<td>distance(poptop20%)</td>
<td>162.3</td>
<td>169.7</td>
<td>0</td>
<td>1166</td>
</tr>
<tr>
<td>fraction of path to Primate paved</td>
<td>0.682</td>
<td>0.26</td>
<td>0.115</td>
<td>1</td>
</tr>
</tbody>
</table>

The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. All distances are measured in kilometers. Distance(Primate), distance(pop≥100k), and distance(poptop20%) are the road network distance to the largest city in the country, the nearest city (in the same country) with a population of at least 100 thousand, and the nearest city in the top quintile of the country’s 1992 city population distribution, respectively.
Table 3: Main results and functional form

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>ln(light+5.5)</td>
<td>ln(light+5.5)</td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>ln(light)</td>
</tr>
<tr>
<td>1(dist.(Prim.) &gt; median)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*1(P_{oil} &gt; median)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.132***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0463]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*P_{oil}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.715***</td>
<td>-0.644***</td>
<td>-0.565***</td>
<td>-0.459**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.230]</td>
<td>[0.222]</td>
<td>[0.173]</td>
<td>[0.181]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lagged distance(Primate)*P_{oil}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.154</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.215]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*ln(P_{oil})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.352***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.111]</td>
</tr>
<tr>
<td>ln(distance(Primate))*ln(P_{oil})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0888***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0198]</td>
</tr>
</tbody>
</table>

Observations: 4,879 4,879 4,879 4,592 4,879 4,879 4,879 4,879 4,879 4,879 4,879 4,879 4,879 4,879 263 263 211 263 263

Each column is a separate regression that includes country*year fixed effects. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Distance(Primate) is the road network distance to the largest city in the country, measured in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively. Splines in column 4 have one knot per city. Standard errors have not been corrected for the first stage estimation.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance(Primate)*P_oil</td>
<td>-0.565**</td>
<td>-0.544**</td>
<td>-0.939***</td>
<td>-0.942***</td>
<td>-0.536**</td>
<td>-0.775***</td>
<td>-0.764**</td>
</tr>
<tr>
<td></td>
<td>[0.230]</td>
<td>[0.230]</td>
<td>[0.309]</td>
<td>[0.289]</td>
<td>[0.235]</td>
<td>[0.261]</td>
<td>[0.347]</td>
</tr>
<tr>
<td>distance(Primate)*ln(Gov. spending, PPP)</td>
<td>0.0950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.103]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*ln(nat. res. income, PPP)</td>
<td></td>
<td>-0.00288</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0455]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*ln(GDP, PPP)</td>
<td></td>
<td></td>
<td>-0.584</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.603]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*P_oil*1(AGOA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.161</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.213]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*P_oil*(AGOA years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0118</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0319]</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4,586</td>
<td>4,586</td>
<td>2,819</td>
<td>2,819</td>
<td>4,874</td>
<td>4,879</td>
<td>4,879</td>
</tr>
<tr>
<td>cities</td>
<td>287</td>
<td>287</td>
<td>261</td>
<td>261</td>
<td>287</td>
<td>287</td>
<td>287</td>
</tr>
</tbody>
</table>

Each column is a separate OLS regression that includes country*year and city fixed effects, and city-specific linear time trends. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable, plus 5.5. The unit of analysis is the city-year. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometers. $P_{oil}$ is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. Gov spending, Nat. res. income and GDP are government expenditures, natural resource income and gross domestic product, respectively, in PPP terms. 1(AGOA) is a dummy for the American Growth and Opportunity Act (AGOA) being operative by October of the relevant country-year, and AGOA years is the number of years it has been operative. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 5: Exclusion of other channels by which oil prices could be affecting lights

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>%HH generators</td>
</tr>
<tr>
<td>distance(Primate)*$P_{oil}$</td>
<td>-0.727***</td>
<td>-0.658***</td>
<td>-0.632**</td>
<td>-0.649***</td>
<td>-0.720***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.236]</td>
<td>[0.251]</td>
<td>[0.264]</td>
<td>[0.238]</td>
<td>[0.269]</td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*$P_{oil}$*1(rail to primate)</td>
<td>0.121</td>
<td></td>
<td></td>
<td></td>
<td>0.0933</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>[0.330]</td>
<td></td>
<td></td>
<td></td>
<td>[0.276]</td>
<td>[0.024]</td>
</tr>
<tr>
<td>distance(Primate)*$P_{oil}$*1(hydro closest)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.024]</td>
</tr>
</tbody>
</table>

| Observations                        | 4,879     | 4,471     | 4,029     | 4,267     | 3,128     | 135       |
| sample                              | full      | oil>50km  | oil>100km | power plant | electrified | DHS generators |
| model                               | tobit     | tobit     | tobit     | tobit     | tobit     | OLS       |
| left censored cases                 | 263       | 247       | 216       | 230       | 127       |           |

Each column 1–5 is a separate regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometers. $P_{oil}$ is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. 1(rail to primate) is a dummy indicating that the city has a rail connect to the largest city in the country. In columns 2 and 3, cities at least partially within 50 and 100 km, respectively, of an oil well are excluded. Column 4 includes only cities in countries with both hydro and other power plants. 1(hydro closest) is a dummy indicating that the nearest power plant to the city is a hydro plant. Column 5 includes only cities with electric transmission lines passing through them. Column 6 is a cross-sectional regression of the fraction of households with generators on distance to the primate and country fixed effects for 135 cities in the five countries with generator ownership data from DHS surveys. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 6: Measurement issues

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(light)</td>
<td>ln(light+5.5)</td>
<td>ln(light+5.5)</td>
<td>ln(light)</td>
<td>ln(light)</td>
<td>ln(light)</td>
</tr>
<tr>
<td>distance(Primate)*$P_{oil}$</td>
<td>-0.459**</td>
<td>-0.683***</td>
<td>-0.670***</td>
<td>-0.373</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.193]</td>
<td>[0.220]</td>
<td>[0.214]</td>
<td>[0.694]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*$P_{diesel}$</td>
<td>-0.357**</td>
<td>-0.575***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.157]</td>
<td>[0.142]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCdistance(Primate)*$P_{oil}$</td>
<td></td>
<td></td>
<td></td>
<td>-0.466</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.908]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations          | 2,226        | 2,226        | 2,226        | 4,879        | 4,879        | 4,879        |
Sample                | diesel       | diesel       | diesel       | all          | all          | all          |
Model                 | OLS          | OLS          | IV:oil-unclustered | tobit       | tobit       | tobit       |
Left censored cases   | 263          | 263          | 263          |              |              |              |
Dirt factor           | 1.5          | 2            |              |              |              |              |

Each column is a separate regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometers. $P_{oil}$ is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. $P_{diesel}$ is the price of diesel in the county’s capital city in November of the given year, in dollars per liter. Columns 1-3 are limited to country-years for which $P_{diesel}$ is available. In column 3, $P_{oil}$ is the instrument for $P_{diesel}$. Columns 4 and 5 are analogous to the baseline specification with route distances calculated differently. Dirt factor is the ratio of the time required to traverse a given length of dirt road and the time required to traverse the same length of paved road, used in calculating shortest routes. GCdistance(Primate) is the great-circle distance to the primate, calculated using the Haversine formula. The tobit cutoff is light=5.5. Robust standard errors, clustered by city except in column 3, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 7: Distances to other cities

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(light) * P_{oil}</td>
<td>-0.741***</td>
<td>-0.646***</td>
<td>-0.597**</td>
<td>-0.526**</td>
<td>-0.456*</td>
</tr>
<tr>
<td></td>
<td>[0.235]</td>
<td>[0.233]</td>
<td>[0.232]</td>
<td>[0.235]</td>
<td>[0.243]</td>
</tr>
<tr>
<td>distance(Primate) * P_{oil} *1</td>
<td>0.165</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.478]</td>
</tr>
<tr>
<td>distance(Port) * P_{oil} *1</td>
<td>0.294</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.341]</td>
</tr>
<tr>
<td>distance(pop ≥ 100k) * P_{oil} *1</td>
<td></td>
<td>-0.618</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.314]</td>
</tr>
<tr>
<td>dist(pop ≥ 100k) * P_{oil} *1</td>
<td></td>
<td>-0.954</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.755]</td>
</tr>
<tr>
<td>distance(poptop20%) * P_{oil} *1</td>
<td></td>
<td>-1.266**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.499]</td>
</tr>
<tr>
<td>dist(poptop20%) * P_{oil} *1</td>
<td></td>
<td>-1.337**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.627]</td>
</tr>
</tbody>
</table>

Observations 4,879 4,879 4,879 4,879 4,879
left censored cases 263 263 263 263 263

Each column is a separate tobit regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Distance(domestic primate), distance(pop ≥ 100k), and distance(poptop20%) are the road network distance to the largest city in the country, the nearest city (in the same country) with a population of at least 100 thousand, and the nearest city in the top quintile of the country’s 1992 city population, respectively. Distance(Primate) and distance(Port) are great-circle distances to foreign ports and primates. Distances are measured in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The dummy variables interacted with these distances in columns (3) and (5) are one if the nearest large city (of 100k in column 3, or in the top quintile in column 5) is not the primate. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 8: Paving and city size

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(population)</td>
<td>ln(light+5.5)</td>
<td>fraction manufacturing</td>
</tr>
<tr>
<td>1(paving &gt; median)</td>
<td>0.280***</td>
<td>0.584**</td>
<td>0.0151***</td>
</tr>
<tr>
<td></td>
<td>[0.106]</td>
<td>[0.275]</td>
<td>[0.00521]</td>
</tr>
<tr>
<td>distance(Primate)</td>
<td>-0.0791</td>
<td>0.0633</td>
<td>-0.0196**</td>
</tr>
<tr>
<td></td>
<td>[0.135]</td>
<td>[0.366]</td>
<td>[0.00943]</td>
</tr>
<tr>
<td>Observations</td>
<td>272</td>
<td>272</td>
<td>293</td>
</tr>
<tr>
<td>sample non-primate cities</td>
<td>OLS</td>
<td>non-primate cities</td>
<td>IPUMS</td>
</tr>
<tr>
<td>model</td>
<td>city</td>
<td>city</td>
<td>city</td>
</tr>
<tr>
<td>unit</td>
<td>city</td>
<td>city</td>
<td>district</td>
</tr>
<tr>
<td>left censored cases</td>
<td>52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each column is a separate regression, with country fixed effects. The independent variable of interest is a dummy indicating that the unit’s path to its country’s primate city is more paved than the average within that country. In columns 1 and 2, the sample is 272 non-primate cities in 15 coastal primate countries in 1992, the initial year. In column 1, the dependent variable is the log of population, while in column 2 it is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. In column 3, the sample is census administrative units in Ghana (2000), Guinea (1983), Senegal (1988), and Tanzania (2002), and the dependent variable is fraction of the employed population over age 10 working in manufacturing. Distance(Primate) is the road network distance to the largest city in the country, measured in thousands of kilometers. The tobit cutoff is light=5.5. Robust standard errors, clustered by city in columns 1 and 2, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 9: Results by paving status

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln(light)</td>
<td>ln(light)</td>
</tr>
<tr>
<td>dist(primate)(\times)(P_{oil}) \times 1(primate route paving &lt; median)</td>
<td>-0.693*** [0.247]</td>
<td>-0.272 [0.283]</td>
</tr>
<tr>
<td>dist(primate)(\times)(P_{oil}) \times 1(primate route paving &gt; median)</td>
<td>-0.663*** [0.255]</td>
<td>-0.671** [0.276]</td>
</tr>
<tr>
<td>dist(poptop20%)(\times)(P_{oil}) \times 1(primate route paving &lt; median)*1(nearest poptop20% not primate)</td>
<td>-1.889** [0.820]</td>
<td></td>
</tr>
<tr>
<td>dist(poptop20%)(\times)(P_{oil}) \times 1(primate route paving &gt; median)*1(nearest poptop20% not primate)</td>
<td>0.116 [1.031]</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4,879</td>
<td>4,879</td>
</tr>
<tr>
<td>left censored cases</td>
<td>263</td>
<td>263</td>
</tr>
</tbody>
</table>

Each column is a separate tobit regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Dist(Primate) and dist(poptop20%) are the road network distance to the largest city in the country and the nearest city (in the same country) in the top population quintile, respectively. Distances are measured in thousands of kilometers. \(P_{oil}\) is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. 1(primate route paving < median) is a dummy indicating that a city’s route to the primate is more paved than the route of the median city in that country to the primate. 1(nearest poptop20% not primate) indicates that the nearest city in the top quintile is not the primate. The tobit cutoff is DN=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.
Table 10: Transport costs and population

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(population)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(population)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance(Primate)*P_{oil}</td>
<td>0.0855</td>
<td>-0.283</td>
</tr>
<tr>
<td></td>
<td>[0.263]</td>
<td>[0.182]</td>
</tr>
<tr>
<td>Observations</td>
<td>76</td>
<td>278</td>
</tr>
<tr>
<td>cities</td>
<td>38</td>
<td>110</td>
</tr>
<tr>
<td>sample</td>
<td>MOZ+BEN</td>
<td>full</td>
</tr>
</tbody>
</table>

Each column is a separate regression that includes country*year and city fixed effects. The unit of analysis is the city-year, for an unbalanced panel of census years. The dependent variable is the log of city population. Distance(Primate) is the road network distance to the largest city in the country, measured in thousands of kilometers. $P_{oil}$ is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. Robust standard errors, clustered by city, are in brackets. Column 1 is restricted to the two countries with multiple censuses between 1992 and 2008, Benin and Mozambique. Column 2 includes 110 cities in the 11 sample countries with city populations from multiple censuses between 1970 and 2008.
Figure 1: African countries classified by their coastal access, the coastal access of their primate city, and availability of comparable roads and lights data.
Figure 2: Lights digital number in Tanzania from satellite F-16, 2008
Figure 3: Lights digital number in and around Dar es Salaam, Tanzania from satellite F-16, 2008
Figure 4: Binary lights (pixels lit in at least one satellite-year, 1992-2008), Tanzania.
Figure 5: Binary lights and cities with known populations, Tanzania
Figure 6: Binary lights, restricted to cities with known populations, Tanzania.
Figure 7: Oil and diesel prices (averaged across the 12 countries in the sample with data for all 7 years shown), 1992–2008. Diesel prices were surveyed in November, while oil prices are averaged across the whole year.
Figure 8: Roads in Tanzania
Figure 9: Shortest road routes from cities with known populations to Dar es Salaam, Tanzania.
Figure 10: Running line smoother of $\ln(\text{light}+5.5)$ on $\text{distance}(\text{Primate}) \cdot P_{oil}$, both net of fixed effects, with 95 percent confidence interval and histogram of the running variable.
Diagram of the paving results. Three cities, A, B, and C, identical except for their locations, are connected to a primate city P and a secondary (i.e., top quintile) city S, with the distance relationships $d_{SA} = d_{SB} < d_{SC} < d_{PC} = d_{PB} < d_{PA}$. When oil price rises, if roads PA, PB, and PC are paved (left figure), A will grow slower than B, which will grow about as fast as C. If these three roads are unpaved (right figure), A and B will grow at the same rate, faster than C.
A  Data and spatial methods

A.1  City points

City locations (latitude and longitude) and census populations were collected from Brinkhoff (2010) and spot-checked with official sources where available. In 9 cases where coordinates were unavailable from Brinkhoff (2010), coordinates from Google Earth or World Gazetteer (http://www.world-gazetteer.com) were used. Using city-specific growth rates based on multiple censuses where available, or national urban growth rates from United Nations (2008) otherwise, I estimated populations for all years for each city.

For all 15 countries in the sample except Angola, City Population claims to list all cities above a given (country-specific) population, typically 5,000, 10,000, or 20,000. However, it does not explicitly cite the year for which this claim is made. Of the 738 cities with location and population information in these 15 countries, 9 city points fall below this cutoff for all years 1990 to 2008. These are included in the sample until explicit population cuts are made.

A.2  Oil prices and wells

The annual average Europe Brent Spot Price FOB, in dollars per barrel, are from the United States Energy Information Administration (http://tonto.eia.doe.gov/; accessed 5 Jul 2010). They are deflated by the United States Consumer Price Index for Urban Consumers (CPI-U). Oil and gas field centroid locations were manually georeferenced from Persits et al. (2002).

A.3  Censuses and surveys


A.4  Power plants

Power plant types and locations are from the African Infrastructure Country Diagnostic database (AICD; http://www.infrastructureafrica.org/). This analysis excludes three
plants, one in Nigeria and two in Tanzania, characterized as neither thermal nor hydro. All three are part of sugar or paper mills.

A.5 Lights

The lights data are described in Henderson, Storeygard and Weil (2012). The sensors are designed to collect low light imaging data for the purpose of detecting moonlit clouds, not lights from human settlements. For the present study the 30 satellite-years of lights data were first combined into one binary grid encoding whether a pixel was lit in at least one satellite-year. Lights arising from gas flares, as delineated by Elvidge et al. (2009) were also removed. These affected only 4 populated lights in the 15-country sample. The resulting contiguous ever-lit areas were converted into polygons, and split by national borders. Only lights within 3 kilometers of one of the city points described above with a known census population were kept in the sample. The 3-km buffer is used because of georeferencing error in both the points and the lights (Balk et al., 2004; Elvidge et al., 2004). While some of the other lights are likely to be small settlements, some are noise from the sensor or from fires lasting for too long to be excluded by the data cleaning algorithm, mines or other facilities. In general, they are smaller and weaker lights as well.

The resulting sample is 485 lights in 15 countries. Populations for each point were summed across all points assigned to each light. In 49 lights, more than one city was present; in 26 of these, exactly two were present. In 13 cases, a point fell within 3 km of multiple lights. In such cases, the point’s population was only retained by the light to which it was closest. The light with the largest 1992 population within each country was designated the primate. In most countries, this corresponded to the historical political capital. The only exception is Douala, Cameroon, which is larger than the capital Yaoundé. The historical political capitals that are not current formal political capitals are Dar es Salaam, Tanzania, which was replaced by Dodoma, Abidjan, Côte d’Ivoire, replaced by Yamoussoukro, and Lagos, Nigeria, replaced by Abuja.

A.6 Roads

The AICD database contains comparable roads data for all countries of the African mainland with no Mediterranean coastline, plus Madagascar, except for Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia. Two datasets were produced for Sudan: one for the north and one for the south. The datasets generally have comparable metadata on road surface, quality, and hierarchy (primary/secondary/tertiary), as well as
estimates of traffic. Still, before they could be used for the current analysis, several changes had to be made. All steps below were carried out using ArcGIS 9.3 software, except for some tabular cleaning done in Excel and Stata. Nearly all steps in ArcGIS, and all steps in Stata, were automated in Python, Arc Macro Language (AML), or Stata scripts.

The roads data were cleaned tabularly, to ensure that the relevant fields were coded consistently, and projected to a sinusoidal projection, with a central meridian of 15 degrees east longitude. This reduced distance distortions with respect to their native plate carrée (latitude and longitude). Next, roads from all countries were combined into one large dataset, and a topology was built with the rule “no dangles”. This means that every dead end was flagged. In most cases, dead ends are legitimate features of the road network. In other cases, however, they are artifacts of a data generation process in which some segments that are connected in the real world are not connected in the dataset. This is critical in the network analysis to follow.

Problematic dangles were fixed in several ways. First, using the topology “Extend” tool, dangles were extended up to 100 meters if that would cause them to no longer be dangles. In theory, the topology “Trim” tool could be used for the opposite task to dangles less than 100 meters long. However, a bug in ArcGIS made this infeasible. But extra dangles only affect final results to the extent that they cause additional “spiders” to be created (see below).

The Extend operation does not close all gaps of less than 100 meters. To see this, imagine the forward slash and backslash characters typed with a space between them: \\
. Extending either character individually, even by doubling its length, would not make the two touch, because they are pointed in the wrong direction. To deal with cases like these, “bridges” were created as follows. All dangles were paired with the closest other dangle if it was within 100 meters using the Spatial Join tool, and connecting lines were created between these pairs of dangles. These bridges were added to the rest of the roads.

The AICD roads database was gathered with explicit reference to inter-city roads. Unfortunately, this means that in many cases, information on roads within cities was not collected, greatly reducing the connectivity of the dataset in many countries. “Spiders” were created to model missing city roads. For every dangle falling within a city, a (paved) road was created between the city centroid and the dangle. The implicit assumption is that radial road travel within cities is comparatively easy. The resulting spiders were added to the roads, and all spiderlegs and any roads that intersected them were “planarized”. Before planarizing, the topology of the network was such that a
spiderleg could cross a road without being connected to it. Planarizing ended this.

Manual edits were necessary for several reasons. Recall that Extend did not close holes if a dangle was not pointed at another road, and that bridges were only created between two nearby dangles, not a dangle and a non-dangle. So extra segments of less than 10 meters each were created in 6 other places to fix dangles affecting routes to 16 cities. Recall also that spiders were created between a centroid and any dangles within a city. So once a spider is created, the spiderleg is the network location of the centroid, so if it connects to a dead end, other nearby roads cannot be reached even if they are very close. In these cases, deleting one or more spiderlegs fixed the problem. Five spiderlegs affecting 4 cities were deleted.

A.7 Route calculation

To prepare for building the network, the roads were intersected with all land borders, so that the resulting border posts could be used as barriers—non-traversable points on the network. Coastlines were not treated as borders in this operation, because the only reason a road would cross a coastline is because of misalignment—the resulting route is most likely legitimate.

A network dataset was built using the roads dataset. The “Closest Facility” solver was used with the following settings. All light centroids were used as the “Incidents”, centroids of primate cities were used as “Facilities”, and the intersections of the roads and the land borders were used as “Barriers”. Each city was assigned a network location on the closest road within 5 kilometers of its centroid.

Unfortunately, because of a quirk in the program, the total calculated length is a true geodesic distance, while distances by paving status are projected distances. However, this never causes a discrepancy of more than a few percent, and because the same projection is used for the whole continent, these errors are highly correlated within countries.

Of 485 populated lights, 464 (96 percent) received plausible routes. Of the remaining 21, 6 were in exclaves or islands, and 2 had centroids more than 5 km from the nearest domestic road. Three received no routes because they were on road segments disconnected from the primate by a gap of at least 100 meters. The remaining ten received implausible routes (because of suspicious gaps of longer than 100 meters in the road network) and were removed. To the extent that these cities are in fact less connected than others, or that government officials have not mapped their roads correctly or at all, they are more likely to be excluded from traditional data sources like
censuses and surveys as well.

**A.8 Pixel-level data-generating process**

The pixel-level data-generating process can be modeled as follows:

\[
Y_{jist} = \begin{cases} 
0 & \text{if } Y^*_{jist} < 2.5 \text{ or } \sum_{k \in i} 1\{Y^*_{kist} \geq 2.5\} < 4 \\
63 & \text{if } Y^*_{jist} > 62.5 \\
\text{int}(Y^*_{jist} + 0.5) & \text{otherwise}
\end{cases}
\]  \tag{12}

where \(j\) indexes pixels, which nest in cities, \(s\) indexes satellite-years within a year, \(Y_{jist}\) is measured pixel-level light, and \(Y^*_{jist}\) is true (latent) pixel-level light. Two nonlinearities appear here, in addition to rounding to the nearest integer. Processing by NOAA converts to zero nearly all (1) individual pixel values of 1 or 2 and (2) clusters of less than 4 nonzero pixels. In both cases, NOAA’s algorithm interprets these patterns as random noise.

The relationship of interest is at the city level, as are all of the regressors, but the lights data are generated nonlinearly at the pixel level. Rather than estimate Equation (12) via maximum likelihood with approximately 5.6 million pixel-satellite-years, I instead simply sum lights across pixels and satellites within a city:

\[
Y_{it} = \frac{1}{S_t} \sum_{s=1}^{S_t} \sum_{j \in i} Y_{jist}
\]  \tag{13}

where \(S_t\) is the number of active satellites (always 1 or 2), and run a tobit regression with a cutoff value of 5.5. The theoretical minimum non-zero city-year has a DN value of 6: in one satellite-year it is unlit, while in the other satellite-year, it consists of 4 pixels, each with a DN of 3. In practice, this is also the minimum non-zero city-year DN value in the estimation sample. The smallest increment in city DN is 0.5 because satellite-year pixel values are integers but there are up to two satellite per year, so averaging across two satellites sometimes produces half-integer values.
Figure A.1: Diesel prices for the oil producing countries, 1993–2008