Review of Economic Studies (2016) **83**, 1263–1295 doi:10.1093/restud/rdw020

Advance access publication 23 April 2016

Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa

ADAM STOREYGARD
Tufts University

First version received June 2012*; final version accepted January* 2016 (*Eds.*)

This article investigates the role of intercity transport costs in determining the income of sub-Saharan African cities. In particular, focusing on fifteen countries whose largest city is a port, I find that an oil price increase of the magnitude experienced between 2002 and 2008 induces the income of cities near that port to increase by 7% relative to otherwise identical cities 500 km farther away. Combined with external estimates, this implies an elasticity of city economic activity with respect to transport costs of −0.28 at 500 km from the port. 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 centres.

Key words: Urbanization, Transport costs, Infrastructure, Roads, Sub-Saharan Africa

JEL Codes: F15, O18, R11, R12, R4

1. INTRODUCTION

Sub-Saharan Africa has notoriously high transport costs compared with 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 article, 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 have larger increases in income than those farther away or with poorer road connections when oil prices rise. 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 article provides novel measures of both. First, night time lights satellite data (Elvidge *et al*., 1997; Henderson *et al*., 2012) are used to construct a 17-year annual panel

1. This article generally defines sub-Saharan Africa as all countries of the African mainland with no Mediterranean coastline, plus Madagascar.

of city-level measures of economic activity for 289 cities in 15 countries. Secondly, 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 500 km farther away from the primate than initially identical city B, its economy is roughly 7% smaller than city B's at the end of the period. At a differential of 2360 km, the largest in the data, this rises to 29%. Further evidence is consistent with an explanation based on transport costs, but broadly inconsistent with several explanations based on commodity income and the generation and cost of electricity. I then explore the role of connections to secondary cities, heterogeneity with respect to road quality, as proxied by paving status, and effects on city population. The results are consistent with gains from trade predicted by a broad class of models, but the specific mechanisms cannot be distinguished in the present empirical context.

This article 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 seat of elites and often of the government, the largest domestic market, the chief manufacturing centre, and the primary trading connection with the rest of the world. This last role is critical because most African trade is transoceanic. Trade among the contiguous eight members of the West African Economic and Monetary Union (WAEMU) represented $<3\%$ 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 neighbours, 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 15–20% of World Bank lending, with almost three quarters of that amount going to roads (World Bank, 2007a).

Formal manufacturing activity is highly concentrated in the largest cities of sub-Saharan Africa, so most hinterland cities have essentially no export manufacturing that might be protected by higher transport costs. For example, as of 2002, the Dar es Salaam administrative region contained 0.16% of mainland Tanzania's land area, and 8% of its population, but 40% of its manufacturing employment and 53% of manufacturing value added. As of 2008, 55% of manufacturing establishments, and 66% 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).[2] Tanzania has relatively low primacy, so if anything, these fractions would likely be even larger in other countries.[3] Although mainland Tanzania has a coastline of over 1,400 km and three other ports, Dar es Salaam also handled 95% of its port traffic as of 1993 (Hoyle and Charlier, 1995), suggesting that other cities are not important sources of imported manufactured goods either.

This article relates primarily to two bodies of work. The first is on the effect of transport costs on the size and growth of cities and regions. Empirical evidence on this topic has been

2. These manufacturing statistics are based on establishments with more than 10 employees.

3. To the extent that manufacturing is substantial in some hinterland cities, this would work against the results I find below in empirical work.

found primarily using cross-country data (e.g. Limão and Venables, 2001) or the construction of very large national transport networks in the U.S. (Baum-Snow, 2007; Chandra and Thompson, 2000; Atack *et al*., 2010; Duranton and Turner, 2012; Donaldson and Hornbeck, forthcoming), India (Donaldson, forthcoming; Alder, 2015), China (Banerjee *et al*., 2012; Faber, 2014), and Indonesia (Rothenberg, 2013).[4] The existing literature has substantially different findings in different contexts, and applies a variety of models to interpret these findings. For example, Donaldson (forthcoming) finds that Indian colonial districts benefited when railroads were built through them. These results, along with several others noted above, are consistent with an Eaton and Kortum (2002)-type model. In contrast, Faber (2014) finds that peripheral counties in China, especially smaller and more connected ones, were hurt by the construction of a large new highway system in recent decades, consistent with New Economic Geography models following Krugman (1991).

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 (forthcoming), who consider the construction of colonial railroads in Ghana.[5]

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. Teravaninthorn and Raballand 2009). This article suggests that transport costs per se have an important effect on income.

This aritcle takes the road network as given, and relies on the plausibly exogenous annual changes in transport costs induced by world oil price fluctuations, which allows me to determine the short run impact of shocks. Short run does not mean small, however, as interannual price changes averaged over 20% (in absolute value) 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 U.S. 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, 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 *et al*., 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 the present results, these secondary towns form primarily as "consumption cities" where farmers sell their products and buy services and imported goods, as opposed to manufacturing centres 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

4. For a comprehensive overview, see Redding and Turner (2015).

5. Also, Buys *et al*. (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. Atkin and Donaldson (2015) consider the effect of transport costs on *prices* in hinterland cities. Gollin and Rogerson (2010) find that in Uganda, internal transport costs for crops can exceed their farm gate price. 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 2006–09 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.

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 two strands of the literature on geographic determinants of growth and development. Gallup *et al*. (1999) and Collier (2007), among others, emphasize coastal access and the problems of being landlocked, respectively. Nunn and Puga (2012) and Alsan (2015) argue that the relationship between geography and growth and development differs between Africa and the rest of the world, focusing on ruggedness and disease environment, respectively.

The remainder of the article has the following structure. In Section 2, I describe the lights and roads data and the methods used to integrate them, with further details in Appendix A. In Section 3, I describe the econometric specification used, and in Section 4, I report results. Section 5 reports on two extensions: heterogeneity based on road surface, city size and market access, and effects on population. Section 6 concludes.

2. DATA AND SPATIAL METHODS

Empirical work 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.[6] 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 Appendix A.

2.1. *City lights*

To date, very little economic data, especially for income and especially as a panel, have been available for individual African cities. 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 U.S. 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 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.[7] No lights are visible in the overwhelming majority of Africa's land area. In Figure 1, a close view of Dar es Salaam 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 *et al*. (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

6. 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.

7. 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 U.S. National Oceanic and Atmospheric Administration (NOAA), http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html (accessed 22 January 2010).

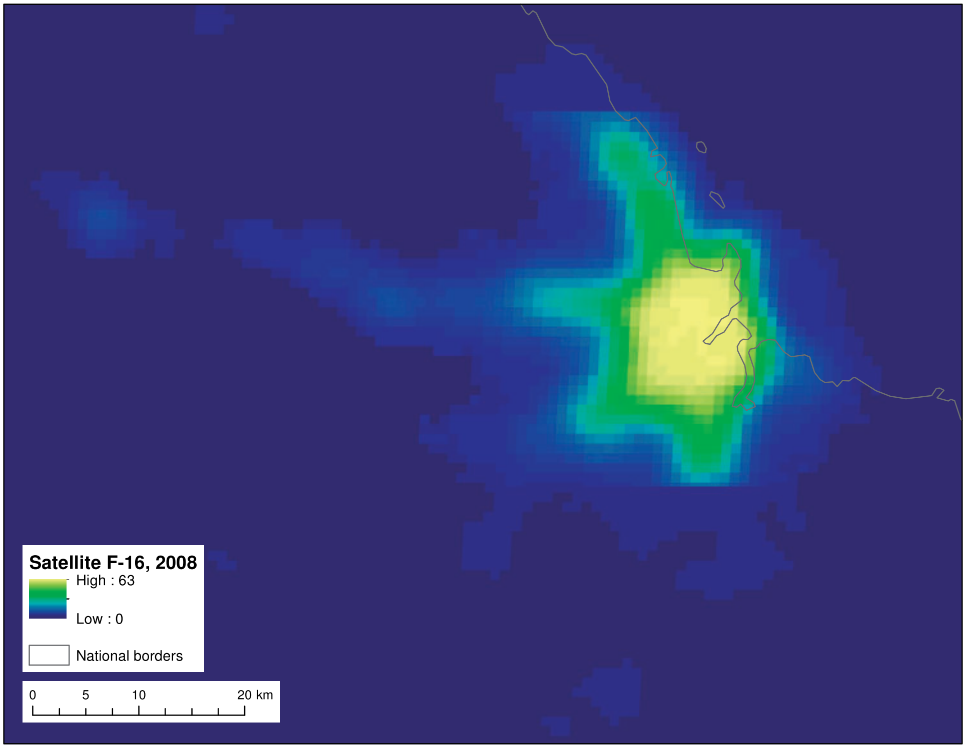

FIGURE 1
Lights DN in and around Dar es Salaam, Tanzania from satellite F-16, 2008

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% of the variation in log GDP net of country and year fixed effects. Table 1, column 1, reports the ordinary least squares (OLS) estimate of the global lights–GDP elasticity from Henderson *et al*. (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 forty-one 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 fifteen 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 quality income data for a short panel of regions.

TABLE 1
Relationship between lights and economic activity

	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) Δ ln(GDP)	(5) Δ ln(GDP)
ln(light)	0.284***	0.264***	0.270***		
	[0.0301]	[0.0342]	[0.0307]		
1(SSA)*ln(light)		0.0702			
		[0.0491]			
1(coastal primate)*ln(light)			0.0907		
			[0.0680]		
Δ ln(light)				0.249**	0.248*
				[0.0995]	[0.141]
Observations	3,015	3,015	3,015	212	120
Countries	188	188	188		
Sample	Global	Global	Global	China	China
Unit	Country	Country	Country	City proper	Prefecture
Type	Panel	Panel	Panel	LD	LD
Country FEs	Yes	Yes	Yes	No	No
Year FEs	Yes	Yes	Yes	No	No
Years	1992–2008	1992–2008	1992–2008	1990/92–2005	1990/92–2005

Notes: 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, they 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 DN, summed across all pixels in the unit shown (having removed gas flares), and averaged across satellite-years within a year when applicable. 1(SSA) is a dummy for sub-Saharan Africa, and coastal primate means the fifteen countries considered in the remainder of the article. Robust standard errors are reported in brackets (clustered by country in columns 1–3, and by province in columns 4–5). *, **, and *** mean significance at the ten, five, and one percent level, respectively.

China has panel GDP data for two relevant types of subnational regions: cities proper and prefectures.[8] 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 1990/92–2005 long difference specification.[9] 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. 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. These ever-lit areas were then converted to polygons: contiguous ever-lit pixels were aggregated, and their DNs 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 13–16% of total DN in the fifteen-country sample.[10] Lights arising from gas flares, as delineated by Elvidge *et al*. (2009), were also removed.

Figure 2 shows the lit polygons and city points for Tanzania. 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.[11] In most countries, census information about cities with populations as small as 10,000 was available, but in some, the

8. I am grateful to Vernon Henderson and Qinghua Zhang for providing the China evidence based on Baum-Snow *et al*. (2015).

9. The lights are from 1992 but the GDP data are from 1990—the closest year with good data.

10. Forest and agricultural fires are removed from the data by NOAA based on their short duration. If a fire lasts longer than the threshold used, it will appear in the data.

11. Light pixels for a given satellite-year actually represent the average light from several slightly larger overlapping pixels 3–5 km on a side from many orbits within the satellite-year. Because of this, the lit area of a given city tends to

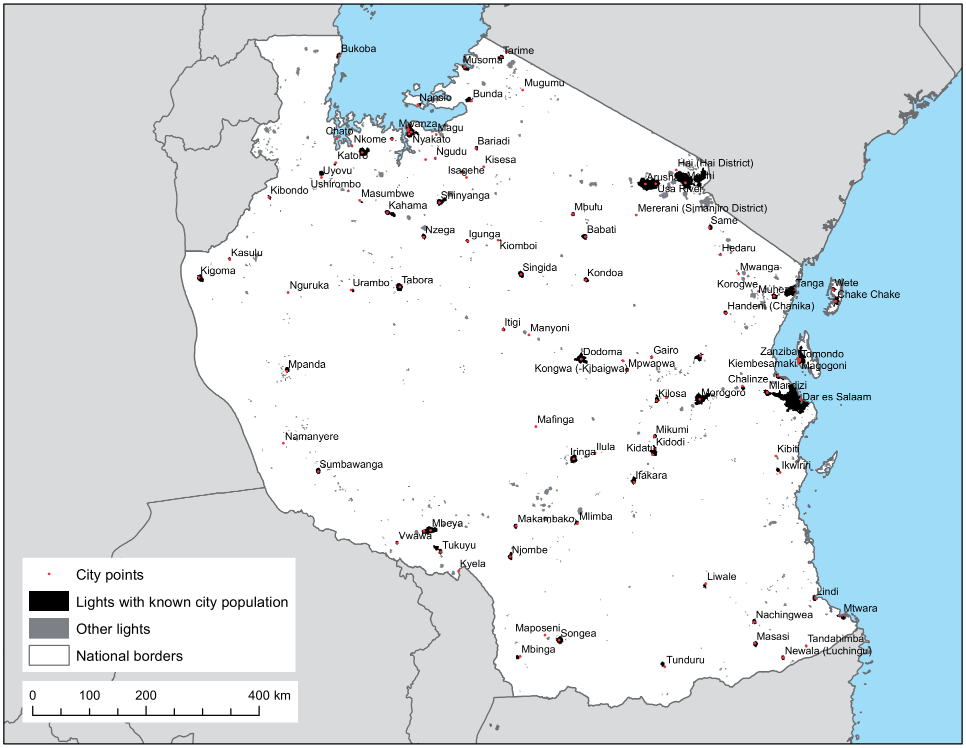

FIGURE 2

Binary lit areas and cities with known populations, Tanzania

cut-off was higher. For all regressions below, I restrict to city polygons with combined population over 20,000 and lit in at least 2 years; in practice 95% of city-years in the sample are lit. The total DN was recorded for each city for each year, averaging across multiple satellite-years where available. The light in each country with the largest associated population in 1992 is designated the primate.[12]

2.2. *Transport costs*

To measure transport costs, I focus on fuel costs and decompose them into two components: (1) fuel prices, which vary across time but not in a generally observable way across cities, and (2) the road distance between a city and its country's primate, which varies across space but not time. Using data from a survey of truckers along several major African intercity corridors, Teravaninthorn and Raballand (2009) estimate that fuel represented roughly 35% of transport costs for trucks in 2005, when oil prices were roughly the mean of the minimum and maximum

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 U.S. 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, but this does not affect the present sample. And even if the area of a given city is overestimated, the light summed for that city is still coming from that city or its outskirts—it may just be partially displaced a pixel or two from where it actually originates.

12. In practice, the primate designation does not change over the course of the sample period in any sample country.

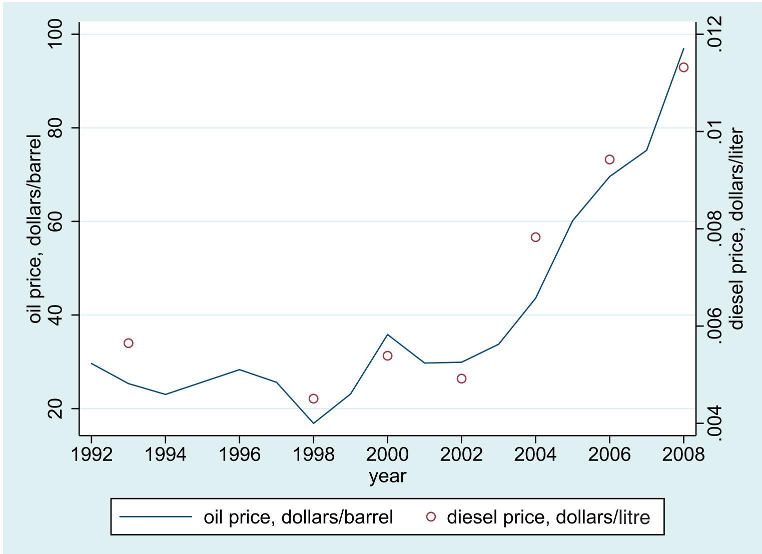

FIGURE 3

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

annual price for the period. For the two routes they study that are primarily within countries in the present sample, fuel represents on average 60–74% of variable operating costs, and 49–66% of total operating costs.The price of oil is a convenient proxy for the price of refined fuels, because no countries in the sample are individually capable of influencing its price substantially. Figure 3 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 1992–4, 1996–8, and 2000–1.

Intercity distances are calculated as the shortest path along a country's road network. GIS roads data are adapted from the World Bank's Africa Infrastructure Country Diagnostic data set (World Bank, 2010), which contains information on over a million kilometres of roads in thirty-nine countries. For over 90% of this length, a measure of the surface type is recorded. The comprehensiveness of the coverage varies by country, but all countries have information on main intercity routes. Roads go through all the lit cities from Figure 2. 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 all other cities in the country, with an emphasis on the distance between each city and its country's primate city. Baseline estimates assume constant speed on all roads, while further work below shows that results are robust to alternative assumptions about the relative travel speed of paved versus unpaved roads. Plausible primate city routes were found for 289 out of 301 cities in the fifteen-country sample. Figure 4 shows all roads and primate routes for Tanzania.

3. EMPIRICAL SPECIFICATION

The baseline specification is:

$$\ln Y_{ict} = \beta p_t x_{ic} + \lambda_{ct} + \gamma_{ic} + \omega_{ic} t + \epsilon_{ict} \tag{1}$$

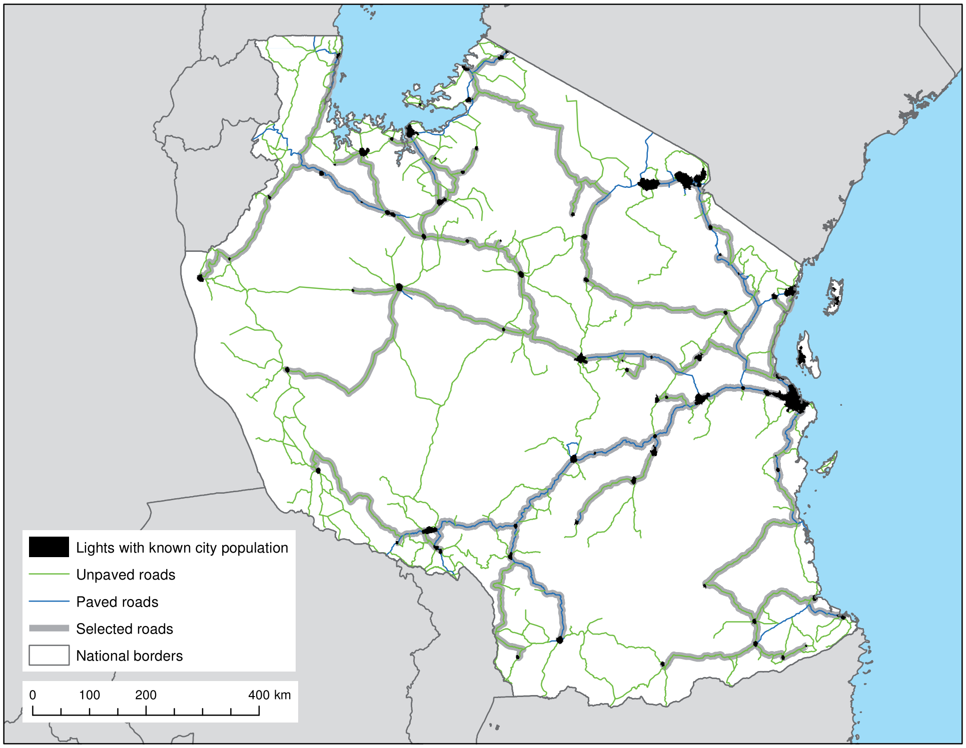

FIGURE 4

Shortest road routes from cities with known populations to Dar es Salaam, Tanzania

where Y_{ict} is light output for city i in country c in year t, p_t is the price of oil, x_{ic} is the distance between city i and its country's primate city along the road network, λ_{ct} is a country-year fixed effect (FE), γ_{ic} is a city fixed effect, and $\omega_{ic}t$ is a linear city-specific time trend. Standard errors are clustered at the city level.[13] 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 <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 balanced panel of 289 cities in 17 years are in Appendix Table A.1. Distances are measured in thousands of kilometres, and prices are in dollars.

When oil prices increase, transport costs increase more for cities farther away from their country's core. Thus, I 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.

Country-year fixed effects control for any national-level time-varying economic conditions constant throughout each country, including the level of industrialization, oil production, terms of trade, and prices in the primate, as well as policies, including gasoline subsidies and preferential trade pacts with developed countries like the American Growth and Opportunity Act (AGOA)

13. If, alternatively, the methods of Conley (1999), as implemented by Hsiang (2010), are used to account for spatial and temporal autocorrelation, the resulting standard errors are smaller.

and trade portions of the European Union's Cotonou Agreement.[14] They also control for global macroeconomic fluctuations, including commodity price shocks, as well as differences across satellites in the lights data. City fixed effects control for initial size and all other fixed city characteristics, including preferences and local agricultural technology, to the extent that these vary slowly. City-specific time trends allow each city to be on its own growth path.

The identifying assumption for β 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,

$$E(\epsilon_{ict}|p_s x_{ic}, \lambda_{cs}, \gamma_{ic}, \omega_{ic} s) = 0, \qquad s,t = 1992, 1993, ..., 2008 \tag{2}$$

The three chief concerns about identification relate to functional form, measurement error in the variable of interest, and omitted variables. I discuss each in turn below in Section 4.2.

Five percent of city-years have a reported DN value of zero. As these cities clearly existed with substantial populations in these years, it is extremely unlikely that they emitted no light on all sampled nights. Instead, these cities are most likely comprised of individual pixels that have DN values <3 and are therefore screened out by NOAA's algorithms as noise. Thus, these are almost certainly not true zeroes, and this is a censoring problem. To address this, I estimate the baseline specification as a tobit regression with a censoring limit of DN $= 5.5$, because 6 is the smallest non-zero value found in the data, and the smallest increment is 0.5.[15] Appendix A provides a more formal model of the data-generating process at the satellite-pixel-year level. While the satellite-pixel-year level relationship could in principle be estimated using maximum likelihood methods, it would be computationally challenging with 5.6 million cases, and the relationship of interest and all of the regressors are at the city-year level. Robustness tests below consider OLS with $\ln \widehat{Y_{ict}}$ as the dependent variable, where

$$\widehat{Y_{ict}} = \begin{cases} 5.5 & \text{if } Y_{ict} = 0; \\ Y_{ict} & \text{otherwise.} \end{cases}$$

4. RESULTS

4.1. *Main results and quantitative interpretation*

Table 2, Column 1, reports a simple "difference-in-difference" version of the main specification. The negative coefficient on $1(x_{ic} > median) * 1(P_{oil} > median)$ shows that on average, net of city and country-year fixed effects and linear city trends, farther cities lose more lights in high oil price years than nearer cities.

Column 2 of Table 2 reports estimates of the baseline specification in equation (1). In this and all subsequent tables, distances are measured in thousands of kilometres, and oil prices in hundreds of dollars. The coefficient of interest, -0.713 on $distance(Primate) * P_{oil}$, 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 500 km farther away from the primate than initially identical city B, it loses 23% more lights than city B by the end of the period. Applying the light-income growth elasticity $\epsilon_{GDP,light} = 0.284$

14. See Table 4 for consideration of the possibility of subnationally varying effects of changing trade policy.

15. Fixed effects tobits are biased for short panels, but this panel is 17 years long and a small percentage of observations are censored. Greene (2004a,b) notes that the bias is small for panels of length 15–20. Using a tobit censoring limit of 1 results in estimates of the coefficient of interest with larger absolute values (not shown).

TABLE 2
Main results and functional form

	(1) $\ln Y_{ict}$	(2) $\ln Y_{ict}$	(3) $\ln Y_{ict}$	(4) $\ln Y_{ict}$	(5) $\ln Y_{ict}$	(6) $\ln Y_{ict}$	(7) $\widehat{\ln Y_{ict}}$	(8) Y_{ict}
$1(x_{ic} > median) * 1(P_{oil} > median)$	−0.131***							
	[0.0459]							
distance(Primate) * P_{oil}		−0.713***			−0.674***	−1.505**	−0.642***	−0.347***
		[0.230]			[0.251]	[0.634]	[0.223]	[0.106]
distance(Primate) * ln(P_{oil})			−0.351***					
			[0.111]					
ln(distance(Primate)+1 km) * P_{oil}				−0.181***				
				[0.0403]				
(distance(Primate)*P_{oil})^2						0.472		
						[0.538]		
($x_{ic} * P_{oil}$ net of city mean)^2						−0.193		
						[0.628]		
Observations	4,913	4,913	4,913	4,913	4,658	4,913	4,913	4,913
Model	Tobit	Tobit	Tobit	Tobit	Tobit	Tobit	OLS	Poisson
Sample	All	All	All	All	No primates	All	All	All
Left censored cases	263	263	263	263	263	263		

Notes: 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, for a balanced annual panel of cities in fifteen coastal primate countries over the period 1992–2008. In columns 1–6, the dependent variable is the log of the lights DN, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. In column 7, the dependent variable is replaced by ln(5.5) when a city-year has no lights. In column 8, the dependent variable is unlogged lights. Distance(Primate) (x_{ic} in columns 1 and 6) is the road network distance to the largest city in the country, measured in thousands of kilometres. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The tobit cut-off is light = 5.5. Robust standard errors, clustered by city, are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

from Henderson *et al*. (2012), this implies a city product differential of 7%. Far cities see their transport costs increase more than near cities, so their income falls more.

The coefficient in column 2 can be interpreted as a semi-elasticity. An elasticity of city product with respect to transport costs is in some respects a more intuitive measure, but since $\ln(p_t x_{ic})$ is equal to $\ln(p_t)+\ln(x_{ic})$, it is collinear with the country-year and city fixed effects and cannot be estimated separately. However, a distance-specific transport cost elasticity ($\epsilon_{GDP,\tau}(x)$) can be calculated. Column 3 reports the coefficient of interest when $p_t x_{ic}$ is replaced with $\ln(p_t)x_{ic}$. It is again negative and significant. This can be translated into a distance-specific elasticity using three additional parameters:

$$\epsilon_{GDP,\tau}(x) = \frac{\epsilon_{GDP,light}\epsilon_{light,P_{oil}}(x)}{\epsilon_{\tau,P_{diesel}}\epsilon_{P_{diesel},P_{oil}}}. \tag{3}$$

A regression of $\ln(P_{diesel})$ on $\ln(P_{oil})$ using the Deutsche Gesellschaft für Technische Zusammenarbeit (2009) data for the sample countries provides an estimate of $\epsilon_{P_{diesel},P_{oil}} = 0.5$. Treating the Teravaninthorn and Raballand (2009) average fuel share as the marginal fuel share implies $\epsilon_{\tau,P_{diesel}} = 0.35$. Combining these estimates implies $\epsilon_{GDP,\tau}(x) = -0.25$ at the median distance from the primate, 439 km, and −0.51 one standard deviation (463 km) farther away. This calculation is meant to be illustrative, as it may suffer from several potential biases, including upward (towards 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.[16]

16. Using the method of Goodman (1960) and assuming independence across samples, the estimate of the product $\epsilon_{GDP,light}\epsilon_{light,P_{oil}} = -0.0997$ has a standard error 0.0334.

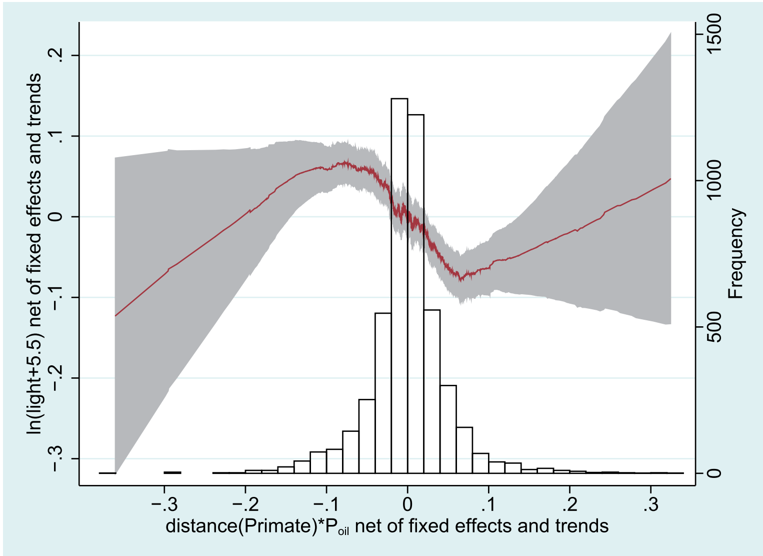

FIGURE 5

Running line smoother of $\ln(\widehat{Y_{ict}})$ on *distance*$(Primate)*P_{oil}$, both net of fixed effects and trends, with 95% confidence interval and histogram of the running variable

4.2. *Robustness*

The baseline estimates may be biased because of an deviations from true functional form, measurement error in the variable of interest, or omitted variables. This section considers each of these in turn.

4.2.1. Functional form.

The baseline functional form of transport costs, $p_t x_{ic}$, is an intuitive combination of its two components because fuel costs per intercity kilometre travelled do not differ substantially with respect to pure distance. However, country size varies dramatically within the estimation sample, so this form may unduly weight large countries. For example, the farthest city in Sierra Leone is only 310 km away from the primate, whereas in Mozambique, the farthest is over 2000 km away. Conversely, if primate cities are different from other cities, their presence in the sample at the shortest distance may be driving results because by definition their $x_{ic} = 0$.

In order to ensure that not all variation is coming from the largest countries, column 4 shows a specification in which $p_t x_{ic}$ is replaced by $p_t \ln(x_{ic})$.[17] This tests for responses linear in proportional increases in distance. The coefficient of interest remains negative and significant. Column 5 drops all primate cities. Results are virtually identical to column 2, suggesting that primate cities themselves are not driving the results.

The term $p_t x_{ic}$ as a whole could also have a non-linear effect on city lights. Figure 5 shows a running line smoothing of $\ln \widehat{Y_{ict}}$ on $p_t x_{ic}$ net of the baseline set of fixed effects and trends, along with a 95% confidence interval. Except in the tails of the distribution, where the confidence interval is extremely wide, the relationship is negative and apparently quite linear.

17. The distance from the primate to itself is arbitrarily redefined as 1 km in the log–log specification.

TABLE 3
Measurement issues

	(1) $\ln Y_{ict}$	(2) $\ln Y_{ict}$	(3) $\ln \widehat{Y_{ict}}$	(4) $\ln Y_{ict}$	(5) $\ln Y_{ict}$
distance(Primate) * P_{oil}	-0.467^{**}			-0.554^{***}	-0.460^{***}
	[0.187]			[0.193]	[0.172]
distance(Primate) * P_{diesel}		-0.362^{**}	-0.557^{***}		
		[0.149]	[0.143]		
Observations	2,240	2,240	2,240	4,913	4,913
Sample	Diesel	Diesel	Diesel	All	All
Model	Tobit	Tobit	IV:oil-unclustered	Tobit	Tobit
Left censored cases	87	87		263	263
Dirt factor				1.5	2

Notes: 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. The dependent variable is the log of the lights DN, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. In the IV specification the dependent variable is replaced by ln(5.5) when a city-year has no lights. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometres. 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 litre. 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. The tobit cut-off is light = 5.5. Robust standard errors, clustered by city except in column 3, are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

Column 6 tests for non-linearity parametrically using a simple quadratic specification. Following McIntosh and Schlenker (2006), it controls separately for within-city and global quadratic terms. Neither is significant. The point estimate of the within-city quadratic term is negative but very noisily measured. Column 7 reports OLS estimates of equation (1) where $\ln Y_{ict}$ is replaced with $\ln \widehat{Y_{ict}}$. Results are very similar to the baseline tobit. Column 8 reports a Poisson specification, which would be appropriate if the city-years with measured lights equal to zero were truly unlit. The coefficient is still negative and precisely estimated, but approximately half as large as the baseline specification.

4.2.2. Measurement error. Oil prices and distance between cities are proxies for fuel cost per distance travelled and distance travelled, respectively. The price of oil is an imperfect measure of fuel costs per distance because in practice, 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 3, 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.

In Table 3 column 1, results are broadly similar when the sample is restricted to country-years with a known diesel price. In column 2, the specification using the diesel price instead of the oil price also has a negative and significant semi-elasticity. 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 3, the effect of the diesel price measure is larger when the oil price is used to instrument for it.[18]

Fuel costs are also imperfectly measured because the baseline intercity distance calculations implicitly assume that costs per kilometre are the same on all roads. This is unlikely to be true 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 *et al*. (1990) find that the fuel efficiency of a 12-ton truck travelling 80 km per hour is 12–13% 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.

In order to test the sensitivity of results to this type of measurement error, Table 3 columns 4 and 5 alternatively assume that travel on unpaved roads is 50% or 100% more costly than travel along paved roads, respectively, and replace x_{ic} with the resulting travel cost metric. 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 different, the coefficient of interest still negative and significant and only somewhat smaller. While in principle, great circle distance is another alternative distance measure, in practice road distance and great circle distance are extremely highly correlated in this sample, at 0.982. This implies that other results are not due to missing links in the road network data—in general, most cities have comparably straight routes to the primate.

4.2.3. Omitted variables. In order for an omitted variable to bias the coefficient on $p_t x_{ic}$, it must be correlated with p_t or x_{ic}. The main candidate variables can broadly be divided into two categories: time-varying factors interacted with x_{ic} and geographically-varying factors interacted with p_t. Tables 4–8 consider several potential omitted variables that may be biasing the coefficient of interest, with Tables 4 and 5 focusing on national and global trends potentially correlated with fuel prices, and Tables 6–8 considering spatial variation that may be correlated with distance to the primate city.

Time-varying factors: If factors other than transport costs that differentially affected near and far cities followed a time path correlated with oil prices, these could be driving the results. Figure 3 shows that there were two broad oil price regimes during this period: a relatively flat stretch followed by a steeply rising one. Overall economic growth in sub-Saharan Africa was also generally higher after 2000 after limited growth in the 1990s, though the timing of acceleration varied within the region. If cities near the primate agglomeration disproportionately experienced this period of growth and infrastructure development, that could drive the results previously shown.

Relatedly, the prices of other commodities followed a path broadly similar to the oil price. It could be the case that country governments spent these commodity windfalls disproportionately near the primate/capital city, for example, in the form of increased infrastructure, 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. Private investors may have disproportionately focused their efforts in near regions as well.

While these possibilities cannot be ruled out entirely, Tables 4 and 5 present several pieces of evidence suggesting that these overall trends are not driving the results. Table 4 columns 1, 2, and

18. This specification is estimated using $\ln \widehat{Y_{ict}}$ because maximum likelihood estimation of the analogous tobit estimator did not converge using several techniques and starting values.

TABLE 4
Omitted variables: national trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
distance(Primate) * P_{oil}	−0.591**	−0.921***	−0.601**	−0.234	−0.466*	−0.800***	−0.874***
	[0.236]	[0.282]	[0.246]	[0.157]	[0.258]	[0.259]	[0.323]
distance(Primate)*	0.0705			0.0871			
ln(Gov. spending, PPP)	[0.100]			[0.176]			
distance(Primate)*		−0.00547					
ln(nat. res. income, PPP)		[0.0414]					
distance(Primate)*			−0.613		0.260		
ln(GDP, PPP)			[0.570]		[0.317]		
distance(Primate)*P_{oil}*1						0.107	
(AGOA)						[0.188]	
distance(Primate)*P_{oil}*							0.0159
(AGOA years)							[0.0303]
Observations	4,600	2,819	4,906	1,342	1,438	4,913	4,913
Trends	City	City	City	None	None	City	City
Sample	Gov. spend.	Nat. res.	All	4-year	4-year	All	All
Left censored cases	230	101	263	83	91	263	263

Notes: Each column is a separate tobit regression that includes country-year and city fixed effects. The unit of analysis is the city-year. The dependent variable is the log of the lights DN, summed across all pixels in the city, averaged across satellite-years within a year when applicable. The 4-year samples include data for 1992, 1996, 2000, 2004, and 2008. Each row reports coefficients of variables interacted with the road network distance to the largest city in the country, in thousands of kilometres. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. GDP, Government spending (Gov. spending) and natural resource export income (nat. res. income) are in PPP-adjusted 2005 U.S. dollars. 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. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

3 control for x_{ic} interacted with log government expenditure, log natural resource income, and log GDP, respectively, using national data from the World Development Indicators and other sources (see Appendix). In column 1, the coefficient of interest is slightly smaller than the OLS baseline in Table 2, but still negative and significant, and the government expenditure interaction term is small and positive, suggesting that if anything government expenditure has a greater positive impact on lights farther from the primate, conditional on $p_t x_{ic}$. Column 2 is analogous to column 1, with natural resource export income, which is potentially better-measured but only available for 57% of the sample, replacing government expenditure. Natural resource export income is defined here as mineral and fuel exports. The main effect is larger in this specification, but the distance-natural resource interaction is small and insignificant.[19] In column 3, the natural log of total GDP at purchasing power parity (PPP), which is almost universally available, replaces natural resource income. Its interaction with distance is substantially negative, but noisy, and the main coefficient of interest remains negative and significant.

GDP and its components are notoriously poorly measured in the developing world, especially on short time scales (Deaton and Heston, 2010; Johnson *et al*., 2013). Columns 4 and 5 report the column 1 and 3 specifications on a sample restricted to 1992, 1996, 2000, 2004, and 2008.[20] In this much smaller sample, in column 4, the effect of transport costs is substantially smaller

19. Running the regressions in columns 1 and 2 on the same samples but without the government expenditure and natural resource income regressors results in little change in the coefficients on the main term of interest (not shown).

20. Four years is the longest interval for which five cross sections are available in the data. Results are similar, generally larger, when three or four cross sections at 5-, 6-, or 7-year intervals are used. Because within-city variation is now much more limited, these specifications do not include city-specific time trends. Natural resource income data are generally much more sparse, such that, for example, no country has available data for all five years.

TABLE 5
Omitted variables: timing

	(1) $\ln Y_{ict}$	(2) $\ln Y_{ict}$	(3) $\ln Y_{ict}$	(4) $\ln \widehat{Y_{ict}}$	(5) $\ln Y_{ict}$
distance(Primate) * P_{oil}	0.393	−0.875**	0.492**	−0.549***	−0.459**
	[0.255]	[0.400]	[0.214]	[0.165]	[0.181]
distance(Primate)*lagged P_{oil}					−0.151
					[0.215]
Observations	4,913	2,601	2,312	4,913	4,624
Model	Tobit	Tobit	Tobit	OLS	Tobit
City trends	Quadratic	Linear	Linear	Split linear	Linear
Sample	1992–2008	1992–2000	2001–8	1992–2008	1993–2008
Left censored cases	263	186	77		211

Notes: Each column is a separate regression that includes country-year and city fixed effects. The unit of analysis is the city-year for a balanced annual panel of 289 cities in fifteen coastal primate countries. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometres. 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, 2, 3, and 5 are tobit regressions that include the city-specific time trends shown. The dependent variable is the log of the lights DN, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Column 4 is a two-step least squares regression that additionally includes two city-specific linear time trends per city, with the break point between the two periods estimated separately for each city. In column 4, the dependent variable is replaced by ln(5.5) when a city-year has no lights, and standard errors are boostrapped with 100 replications. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

than in column 1, but still negative. The coefficient on government spending is still positive, small, and insignificant. In column 5, the transport cost coefficient is somewhat smaller and noisier than in the full sample in the presence of the GDP interaction. The GDP interaction is now positive and remains insignificant. The results in this table suggest that oil prices are not *purely* proxying for overall commodity income fluctuations, to the extent that the national statistics used here measure them, when they modulate the temporal variation in the distance-lights gradient as shown in the baseline specification. While the main effect is smaller in smaller samples at wider time intervals, the interactions with government spending and GDP are insignificant and in the opposite direction, so that higher GDP and government expenditure are weakly associated with more lights at distances farther from the primate, conditional on $p_t x_{ic}$.

Trade policy affecting the region, most notably the U.S.' African Growth and Opportunity Act (AGOA), was also changing during the study period.[21] 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 two countries had been declared ineligible. Thus there is some variation in the timing of its implementation across countries, though this timing is clearly endogenous. As noted above, country-year fixed effects control for nationally uniform effects of trade policy, but Coşar and Fajgelbaum (2016) and Atkin and Donaldson (2015) have argued that openness may have a greater impact near trading hubs. Table 4 columns 6 and 7 consider the possibility that AGOA is driving my results. 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.[22] 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. It

21. On AGOA's impact, see Frazer and van 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.

22. If eligibility was effective in October or earlier, the year is counted as eligible.

is possible however that more subtle forces related to international trade had a differential impact within countries.

Differential early and late trends can also be addressed with more flexible functional forms. Table 5 column 1 adds city-specific quadratic time trends to the baseline specification to allow for smooth acceleration of growth. Net of these trends, the transport cost term reverses sign and shows no significant effect. Columns 2 and 3 split the sample into the 1992–2000 period with low growth and fluctuating oil prices, and the 2001–8 period, with faster growth and more uniformly increasing oil prices, respectively, so that each can have its own linear trend. The results suggest that the overall effect is driven by the early period. Column 4 shows an even more flexible specification, in which the single linear trend for each city is replaced with two separate trend lines, with the parameters of each trend line and year of the trend break estimated separately for each city to minimize the variance in city-specific residuals. Thus, the date of the transition between the two regimes is allowed to vary across cities, as well as the level and slope of the underlying trend in each period. The magnitude of the coefficient of interest is somewhat smaller than in the baseline case, but it is still substantially negative.[23] To the extent that these split trend lines can flexibly account for the kinds of unobserved trends differentially affecting near and far cities, this suggests that they are not driving the results.

Finally, the baseline specification includes only contemporaneous oil prices. Since oil prices are autocorrelated over time, this year's price could be acting as a proxy for last year's price. Column 5 adds a lag of the main term of interest. The contemporaneous term decreases in magnitude, not surprisingly given autocorrelation in oil prices, but remains significant. The lagged term is negative as well, but smaller and insignificant. The effects appear to be felt most strongly in the year of an oil price change.

Location-specific factors: Tables 6–8 consider several additional omitted variables potentially correlated with x_{ic} that might modulate the effect of oil prices on city economic activity as measured by lights: the presence of railroads, oil wells, fossil fuel power plants, grid electricity, diesel generators, and access to other cities besides the primate.

Where railroads are present, intercity rail distances are highly correlated with road distances.[24] However, in most countries, rail only exists along a few corridors that are also served by roads. Table 6 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 specification, and the additional rail effect is small and insignificant.

Oil prices could also affect the within-country pattern of economic activity on the oil supply side, because some sample countries are oil producers. 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 column 2, the main effect is interacted with dummies for the twenty-four cities at least partially within 50 km and the twenty-six cities between 50 and 100 km, of an oil or gas field centroid. The main effect is virtually unchanged. The within-50-km interaction term is substantial and negative, while the 50–100 km is positive with a comparable magnitude, and both have large standard errors.[25]

23. Standard errors are bootstrapped by city with 100 replications.

24. Roads generally dominate transport in Africa, carrying 80–90% of passenger and freight traffic (Gwilliam, 2011). Rail transport is also less likely to be affected by oil price fluctuations than roads because of its higher fuel efficiency and greater dependence on parastatals with long term contracts.

25. 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

TABLE 6
Omitted variables: correlates of distance to the primate

	(1) ln(light)	(2) ln(light)	(3) ln(light)	(4) ln(light)	(5) %HH generators
distance(Primate) * P_{oil}	−0.725***	−0.724***	−0.649***	−0.694**	
	[0.236]	[0.239]	[0.238]	[0.320]	
distance(Primate)*P_{oil}*1 (rail to primate)	0.120				
	[0.330]				
distance(Primate)*1 (distance(oilfield) < 50km)		−0.544			
		[0.631]			
distance(Primate)*1(50km < distance(oilfield) < 100km)		0.545			
		[1.231]			
distance(Primate)*P_{oil}*1 (hydro closest)			0.0933		
			[0.276]		
distance(Primate)*1 (electrical transmission line)				−0.444	
				[0.304]	
distance(Primate)					−0.0273
					[0.0261]
Observations	4,913	4,913	4,267	4,454	109
Sample	All	All	Power plant	Tx lines	DHS generators
Model	Tobit	Tobit	Tobit	Tobit	OLS
Left censored cases	263	263	230	245	

Notes: Each column 1–4 is a separate tobit regression that includes country-year and city fixed effects, and city-specific linear time trends. The tobit cut-off is light = 5.5. The unit of analysis is the city-year. The dependent variable is the log of the lights DN, summed across all pixels in the city, averaged across satellite-years within a year when applicable. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometres. 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 connection to the largest city in the country. distance(oilfield) is the distance from a city to the centroid of an oil field. Column 3 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 4 includes only cities in countries with electrical transmission lines data. 1(electrical transmission line) is an indicator for cities with an electrical transmission line passing through them. Column 5 is a cross-sectional OLS regression of the fraction of households with generators on distance to the primate and country fixed effects for 109 cities in the five countries with generator ownership data from DHS surveys. Robust standard errors, clustered by city, are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

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 by driving up the price of grid electricity, because light is produced by electricity, and some electricity is produced by fossil fuels. Two aspects of the relevant electricity markets work against this interpretation. 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 without explicit within-country geographic variation.[26] Secondly, 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 nevertheless increasing the price of electricity differentially within countries, it should do so less where hydro is the most likely source. World Bank (2010) reports the location of power plants, by type. Table 6 column 3 restricts attention to the 251 cities in those countries

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.

26. The one exception to this is a small additional tax on some rates for Abidjan, the primate city of Côte d'Ivoire.

that have both hydro and thermal power plants and adds to the baseline specification a term interacting $distance(Primate) * P_{oil}$ 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 $distance(Primate) * P_{oil}$. Having a hydro plant as the closest generation source 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 fuelled by diesel generators. High oil prices could reduce diesel generator use, reducing lights more in faraway cities than near ones. While I am unaware of systematic data by location on generator use for lighting, some related information can be harnessed to provide suggestive evidence about the role of diesel generators.

First, data on the location of electrical transmission lines are available from World Bank (2010) for thirteen of fifteen sample countries. Transmission lines pass through 185 of 262 cities (71%) in these thirteen countries. In column 4, the sample is restricted to these 262 cities, and the main term is interacted with an indicator for a transmission line passing through. The main effect in the sample of connected cities is similar to the overall effect, and while the interaction term is negative, it is not precisely estimated.

However, even in cities with grid electricity, some households use generators, due to sporadic supply or a limited distribution network, and this may be more common far from the primate. Demographic and Health Surveys (DHS) from five of fifteen sample countries have information on household generator ownership (but not use) for one point in time 2007–11. These surveys are not generally representative at the level of individual small cities, and thus generator ownership is measured with substantial error.[27]

Column 5 reports the results of a descriptive regression of household generator ownership fraction on distance to the primate, with country fixed effects, for the 109 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.

These results provide limited evidence suggesting that diesel generators are not driving my results. Of course, it is possible that even if generator ownership is lower in farther cities as a fraction of the population, the fraction of lights produced by them is higher, or that the DHS ownership data are too noisy to show a relationship.

Access to other cities: Finally, access to other key cities besides the primate may be important in determining outcomes for a given city. If these other key cities are in the same direction as the primate for a substantial subset of cities, this may bias the primate city coefficient upwards. Access to other cities is also potentially of interest in its own right. Several specifications below include access to other cities in combination with distance to the primate. First, I consider nearby foreign primates and ports and two sets of secondary domestic cities: those with an estimated 1992 population over 100,000, and those in the top quintile of the country-specific population distribution of cities. Secondly, I test for the aggregate effect of market access to all other domestic cities. Log market access, as defined, for example, by Donaldson and Hornbeck (forthcoming) in the context of a trade model, is already controlled for by the fixed effects in my baseline

27. The median city in the sample reports generator ownership status for 67 sampled households.

TABLE 7
Distances to other cities

	(1)	(2)	(3)	(4)	(5)
distance(domestic primate)*P_{oil}	−0.759***	−0.645***	−0.594**	−0.523**	−0.453*
	[0.235]	[0.233]	[0.232]	[0.234]	[0.243]
dist.(primate)*P_{oil}*1(foreign primate)	−0.0351				
	[0.467]				
distance(port)*P_{oil}*1(foreign port)	0.333				
	[0.341]				
distance(pop>100k)*P_{oil}		−0.597			
		[0.507]			
dist(pop>100k)*P_{oil}*			−0.958		
1 (nearest pop>100k is not primate)			[0.755]		
distance(poptop20%)*P_{oil}				−1.282***	
				[0.495]	
dist(poptop20%)*P_{oil}*					−1.342**
1 (nearest poptop20% is not primate)					[0.626]

Notes: 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 289 cities in fifteen coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights DN, 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 100,000, 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 primates and ports. Distances are measured in thousands of kilometres. 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 100,000 in column 3, or in the top quintile in column 5) is not the primate. The tobit cut-off is light = 5.5. Robust standard errors, clustered by city, are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

specification.[28] As an alternative, I define city i's log market access to destination set J in year t as $p_t \ln \sum_{j\neq i} Y_j x_{ij}^{-\sigma}$, where $j\in J$, Y_j is the initial (1992) value of city j's lights, x_{ij} is the road distance between cities i and j, and $\sigma = 3.8$ is the "trade elasticity" estimated by Donaldson (forthcoming).

Table 7 reports results controlling for transport costs to other cities besides the domestic primate. Column 1 controls for transport costs to the nearest major foreign port when that is the closest major port, and the nearest foreign primate when that is the closest primate. These distances are calculated as great circle distances because border crossings are not well-defined in the roads data set, and they vary very widely in the average delay truckers face at them (Teravaninthorn and Raballand, 2009). Neither has a significant effect, and the coefficient of interest is virtually unchanged. Using great circle distance for the domestic primate effect as well gives similar results (not shown). Column 2 controls for transport costs to an alternate domestic destination, the nearest city with a 1992 population of at least 100,000. About a third of the cities in the sample (94 out of 289) have a 1992 population of at least 100,000. 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

28. To see this, note that they define log market access of city i in year t as $\ln MA_{it} = \ln \sum_{j\neq i} N_j \tau_{ijt}^{-\sigma}$, where τ_{ijt} is trade costs and N_j is the size of city j. Using the present proxy for trade costs $\tau_{ijt} = x_{ij} p_t$, $\ln MA_{it} = \ln \sum_{j\neq i} N_j (x_{ij} p_t)^{-\sigma} = \ln \left(p_t^{-\sigma} \sum_{j\neq i} N_j x_{ij}^{-\sigma} \right) = \ln p_t^{-\sigma} + \ln \sum_{j\neq i} N_j x_{ij}^{-\sigma}$, which is collinear with the country-year and city fixed effects.

TABLE 8
Market access

	(1)	(2)
distance(Primate)*P_{oil}	−0.765***	
	[0.260]	
ln(MA to non-primate cities) * P_{oil}	−0.0357	−0.0444
	[0.0473]	[0.0483]
ln(MA to primate) * P_{oil}		0.122***
		[0.0325]
Observations	4,641	4,641

Notes: 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 273 non-primate cities in fourteen coastal primate countries over the period 1992–2008. (Liberia drops out of the sample because it only has one non-primate city.) The dependent variable is the log of the lights DN, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Market access to destinations j for city i is $\sum_{j \neq i} Lights_{j1992} * distance_{ij}^{-3.8}$. Distances are measured in thousands of kilometres. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The tobit cut-off is light = 5.5. Robust standard errors, clustered by city, are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

city of at least 100,000 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 cost to the top quintile city is twice as large or more. This result will be explored further when road surface is considered explicitly in Section 5. Still, no two of the primate city coefficients in this table are significantly different from each other.

Table 8 reports results controlling for a log market access measure that aggregates access to all domestic cities other than the primate. The sample excludes primate cities so that log access to the primate city is well defined. Note that because higher distances to large cities decrease market access, in the context of the previous results its expected sign is positive. Column 1 controls for log market access to all cities other than the primate while using the baseline functional form for access to the primate, $distance(Primate) * P_{oil}$. In column 2, access to the primate is defined analogously to access to all other cities, which is equivalent to $-\sigma \ln(distance(Primate)) * P_{oil}$.[29] In each case, access to the primate city is significant and in the expected direction, with its magnitude essentially the same as the equivalent specification in Table 2 that does not control for market access to other cities.[30] While access specifically to secondary cities seems to matter, overall access to the rest of the city system beyond the primate, approximately as parameterized by trade models, does not.

29. To see this, note that when j is just the primate, $p_t \ln\left[\sum_{j \neq i} Y_j x_{ij}^{-\sigma}\right] = p_t \ln(Y_j x_{ij}^{-\sigma}) = p_t \ln(Y_j) - \sigma p_t \ln(x_{ij})$, and $p_t \ln(Y_j)$ is accounted for by the country-year fixed effects.

30. Note that because the column 2 specification drops primates whereas Table 2, column 4 keeps them, the two specifications are not precisely equivalent.

5. EXTENSIONS

5.1. *Road surface*

The results shown so far have made minimal use of 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 data set includes (static) information on road surface type, so each route to the primate 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 short run we might expect a less negative β for the more paved routes, because driving on paved roads is cheaper, in fuel, time, and maintenance costs, than driving on unpaved roads. However, road surface is clearly not randomly assigned, as governments and donors may be more likely to pave a road to a city that is economically important or expected to grow.[31] Even if roads were initially assigned randomly, after assignment better-connected places are 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% 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).

In Table 9 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 0.270 and 0.551 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 10 exploit the paving information by including two terms of the main effect in equation (1), one for cities with routes to the primate more paved than the median in their country, and one for cities with less paved routes. Column 1 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

31. For an exception, see Gonzalez-Navarro and Quintana-Domeque (2016).

TABLE 9
Paving and city size

	(1) ln(population)	(2) ln(light)	(3) fraction manufacturing
1(paving > median)	0.270**	0.551**	0.0130***
	[0.105]	[0.273]	[0.00440]
distance(Primate)	−0.0837	0.0577	−0.0158*
	[0.135]	[0.365]	[0.00810]
Observations	274	274	290
Sample	Non-primate cities	Non-primate cities	IPUMS
Model	OLS	Tobit	OLS
Unit	City	City	District
Left censored cases		52	

Notes: 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 274 non-primate cities in fifteen 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 DN, 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 kilometres. The tobit cut-off is light = 5.5. Robust standard errors are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

TABLE 10
Results by paving status

	(1)	(2)
dist(primate)*P_{oil}*1(primate route paving < median)	−0.716***	−0.264
	[0.255]	[0.289]
dist(primate)*P_{oil} *1(primate route paving > median)	−0.707***	−0.706**
	[0.260]	[0.282]
dist(poptop20%)*P_{oil}*1(prim. route pav. < med)*1(near. pop20% not prim.)		−2.039**
		[0.845]
dist(poptop20%)*P_{oil}*1(prim. route pav. > med)*1(near. pop20% not prim.)		0.0699
		[1.073]
Observations	4,913	4,913
Left censored cases	263	263

Notes: 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 289 cities in fifteen coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights DN, 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 kilometres. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. 1(prim. route pav. < med.) is a dummy indicating that a city's route to the primate is less paved than the route of the median city in that country to the primate. 1(near. pop20% not prim.) indicates that the nearest city in the top quintile is not the primate. The tobit cut-off is light = 5.5. Robust standard errors, clustered by city, are in brackets. *, **, and *** mean significance at the ten, five, and one percent level, respectively.

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, natural resource extraction, 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

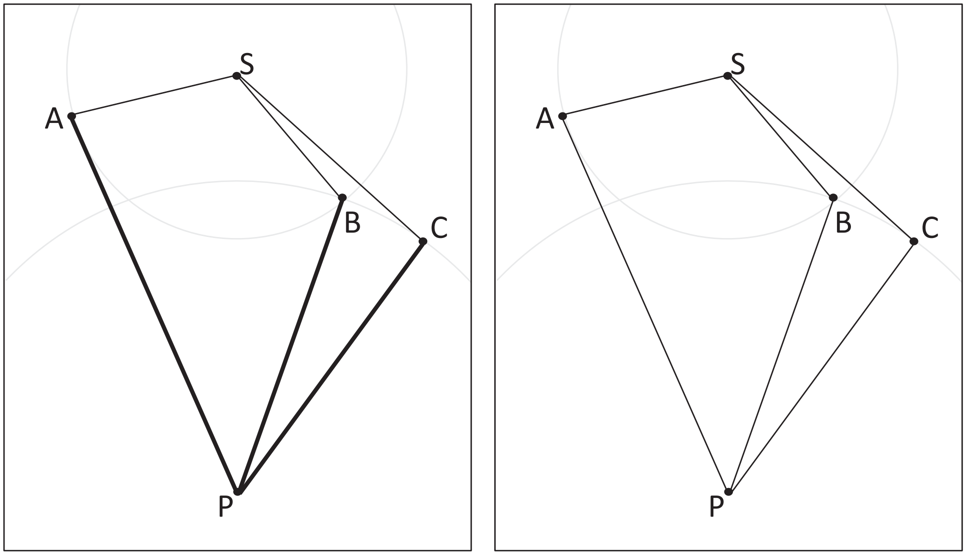

FIGURE 6

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 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

same demand for transport services, cities along them are penalized more per mile by higher fuel prices.

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 funnelled 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% more likely than other non-primate cities to have their connection to the primate more paved than the median city in their country.

The results in column 2 are summarized graphically in Figure 6. 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 the oil price rises, if roads PA, PB, and PC are paved (left figure), A will see its output fall relative to B, which will be affected the same as C. If these three roads are unpaved (right figure), A and B will be equivalently affected, less than C.

5.2. *Other forms of heterogeneity*

As noted in the Introduction, in New Economic Geography models reduced transport costs tend to shift production towards cities that are initially larger and have greater market access. Using

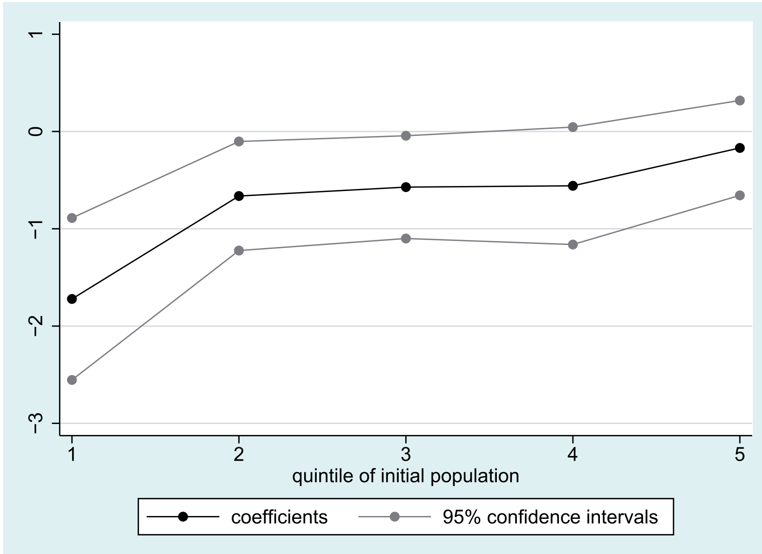

FIGURE 7
Estimates of β for quintiles of the distribution of initial population

the building of a new highway network in China for identification, Faber (2014) finds evidence consistent with this. The present context is limited by the fact that identification of transport cost changes is based on oil price changes that affected the cost of transport in every location directly, as opposed to a highway network built in some locations and not in others. Thus several forms of heterogeneity that are of interest, including by distance to the primate, are not distinguishable from non-linearity considered in Table 2.[32] However, some heterogeneity can be seen. Figure 7 reports estimates of the coefficients of interest and 95% confidence intervals from a variant of the baseline specification in which the transport cost term is interacted with five indicator variables corresponding to quintiles of the distribution of initial city population. Unlike in Faber's case, smaller cities appear to benefit more from reduced transport costs, and an F-test for equality of coefficients can be rejected at 1%.[33] While sectoral information is unavailable for these cities, it is possible that the most affected trading sectors are disproportionately represented in smaller cities. This question is left for future research. Another factor potentially weighing against a large-city advantage is that this analysis is short run, so it may miss some factor reallocation across cities to the extent that such reallocation is slow. The next section considers population over a somewhat longer timescale.

5.3. *Population and the long run*

Changes in economic activity in a city can stem from changes in activity per person, changes in population, or both. In order to explore which is more salient in this context, I consider population

32. Tests for non-linearity in distance, as distinct from oil price, also show no non-linearity. Results are available upon request. Faber (2014) also considers the possibility that this result is due to decentralization within metropolitan areas. This is not possible in the present empirical context, where the cities as identified by the lights are small, and within-city location of lights is measured with greater error than total city lights.

33. In Appendix Figure A.2 heterogeneous effects by quintile of market access are more mixed, and equality cannot be rejected.

TABLE 11
Transport costs and population

	(1) ln(population)	(2) ln(population)
distance(Primate)*P_{oil}	0.0397	−0.127
	[0.122]	[0.0844]
Observations	76	287
Sample	MOZ+BEN	Full
Cities	38	113

Notes: 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 kilometres. P_{oil} is the price of oil (specifically the annual average West Texas Intermediate price) 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 113 cities in the eleven sample countries with city populations from multiple censuses between 1970 and 2008, and an initial population over 20,000.

as an alternative outcome in Table 11. Among sample countries, only Benin and Mozambique had two censuses during the sample period. Even extending the analysis period back to 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 from equation (1). Column 1 shows that in the small Benin-Mozambique sample of 38 cities, the estimated effect of transport costs on population is positive, but very small and imprecisely estimated. In column 2, in the larger eleven-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 in the long run. 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 article provides evidence that transport costs impact urban economic activity in sub-Saharan Africa, by making access to critical primate cities more expensive, with recent increases in oil prices removing several percentage points from the size of far hinterland cities relative to their less remote counterparts in countries where the largest city is on the coast. This is consistent with trade models emphasizing direct gains from trade over the increased competitive pressures faced by hinterland cities. Further evidence suggests that it is not consistent with explanations related to commodity income and the generation of electricity, though this possibility cannot be completely ruled out. 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 impacts on 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 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 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.

The present research design is unable to disentangle the role of the coastal primate city as the largest domestic market versus its role as the gateway to international trade, nor can it quantify net welfare effects or determine the particular responses of firms and individuals to higher intercity transport costs. These questions are left for future research.

APPENDIX

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 nine cases where coordinates or populations 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 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 fifteen countries, nine city points fall below this cut-off for all years 1990–2008. These are included in the sample until explicit population cuts are made.

A.2. *Fuel prices and wells*

The annual average Europe Brent Spot oil price FOB, in dollars per barrel, is from the U.S. Energy Information Administration (http://tonto.eia.doe.gov/; accessed 5 July 2010). Table 11 uses the West Texas Intermediate (WTI) oil price because of its availability in the 1970s and 1980s. Both are deflated by the U.S. Consumer Price Index for Urban Consumers (CPI-U). CPI-U and the WTI price are from the St. Louis Federal Reserve Economic Data (FRED) database (http://research.stlouisfed.org/fred/; accessed 1 March 2013 and 27 January 2013, respectively). Oil and gas field centroid locations were manually georeferenced from Persits *et al*. (2002). Country-specific diesel prices, surveyed in November of selected years in the main city, are from Deutsche Gesellschaft für Technische Zusammenarbeit (2009).

A.3. *Censuses, surveys, and national administrative data*

Census data for Ghana (2000), Guinea (1983), Senegal (1988), and Tanzania (2002) were taken from Minnesota Population Center (2011) and produced by Ghana Statistical Services, National Statistics Directorate, National Agency of Statistics and Demography, and National Bureau of Statistics, respectively. Each is a 10% sample. World Bank Enterprise Surveys for all sample countries were downloaded from http://www.enterprisesurveys.org/ on 22 January 2010. Demographic and Health Surveys (DHS) for Angola (2011), Cameroon (2011), Ghana (2008), Liberia (2007), and Nigeria (2008) were downloaded from http://www.measuredhs.com on 16 August 2013. National administrative data from the World Development Indicators were accessed from http://wdi.worldbank.org on 26 January 2010. Supplemental raw national expenditure data for Angola and Liberia are from the IMF (http://www.imf.org/external/pubs/ft/weo/2013/01/weodata/download.aspx; accessed 8 June 2013). Supplemental raw national expenditure data for Nigeria and Sierra Leone are from their respective central banks (http://www.bsl.gov.sl/Statistics/Fiscal/FISCAL_December_%202012.xlsx; accessed 9 June 2013; http://www.cenbank.org/Out/2012/CCD/2011%20Statistical%20Bulletin-%20Section%20B_FinalWeb.xlsx; accessed 9 June 2013).

A.4. *Power plants, electrical transmission lines and railroads*

Power plant types and locations and electrical transmission lines 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. Railroads are shown on country maps on the same site.

A.5. *Lights*

The lights data are described in Henderson *et al*. (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 four populated lights in the fifteen-country sample. The resulting contiguous ever-lit areas were converted into polygons, and split by national borders. Only lights within 3 km 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; Tuttle *et al*., 2013). 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 486 lights in fifteen countries. Populations for each point were summed across all points assigned to each light. In fifty lights, more than one city was present; in twenty-six of these, exactly two were present. In thirteen 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 except for Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia. The data sets generally have comparable metadata on road surface, quality, and hierarchy (primary/secondary/tertiary), as well as estimates of traffic. 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 data set, and a topology was built with the rule "no dangles". This means that every dead end was identified. In most cases, dead ends are likely legitimate features of the road network. In other cases, however, they are arteifacts of a data generation process in which some segments that are connected in the real world are not connected in the data set. 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 m if that would cause them to no longer be dangles. In theory, the topology "Trim" tool could be used to remove dangles $<$100 m long. However, a bug in ArcGIS made this infeasible. But extra dangles only affect final results to the extent that they cause additional urban connections to be created (see below).

The Extend operation does not close all gaps of $>$100 m. 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. Using the Spatial Join tool, all dangles were paired with the closest other dangle if it was within 100 m, 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 intercity roads. Unfortunately, this means that in many cases, information on roads within cities was not collected, greatly reducing the connectivity of the data set in many countries. Urban connections 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 urban connections were added to the roads, and the network was "planarized".

Before planarizing, the topology of the network was such that an urban connection could cross a true 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 <10 m each were created in six other places to fix dangles affecting routes to sixteen cities. Recall also that urban connections were created between a centroid and any dangles within a city. Once such connections are created, they are the network location of the centroid, so if they connect only to a dead end, other nearby roads cannot be reached even if they are very close. In these cases, deleting one or more urban connections fixed the problem. Five of these, affecting four 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 data set was built using the roads data set. 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" for the main analysis (all light centroids for the market access analysis), 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 km of its centroid.

Unfortunately, ArcGIS calculates only total length as a true geodesic distance; 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 301 lights, lit in at least two years, with a population of at least 20,000 in 1992, 289 (96%) received plausible routes to the primate. Of the remaining twelve, four were in exclaves or islands, two had centroids more than 5 km from the nearest road, and one received no routes because it was on a road segment disconnected from the primate by a large gap. The remaining five received implausible routes (because of suspicious gaps of longer than 100 m 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 such as censuses and surveys as well.

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

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

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

where j indexes pixels, which nest in cities i, s indexes satellite-years within year y, Y_{jist} is measured pixel-level light, and Y^*_{jist} is true (latent) pixel-level light. Two non-linearities appear here, in addition to rounding to the nearest integer. Processing by NOAA converts to zero nearly all (1) individual pixel values of one or two and (2) clusters of $<$four non-zero 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 non-linearly at the pixel level. Rather than using equation (A.1) to estimate all relationships of interest via maximum likelihood with approximately 5.6 million pixel satellite-years, I instead simply sum measured lights across pixels and satellites within a city:

$$Y_{it}=(1/S_t)\sum_{s=1}^{S_t}\sum_{j\in i}Y_{jist} \tag{A.2}$$

where S_t is the number of satellites active in year t (always one or two), and run a tobit regression with a censoring limit of 5.5. The theoretical minimum non-zero city-year has a DN value of six: in one satellite-year it is unlit, while in the other satellite-year, it consists of four 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.

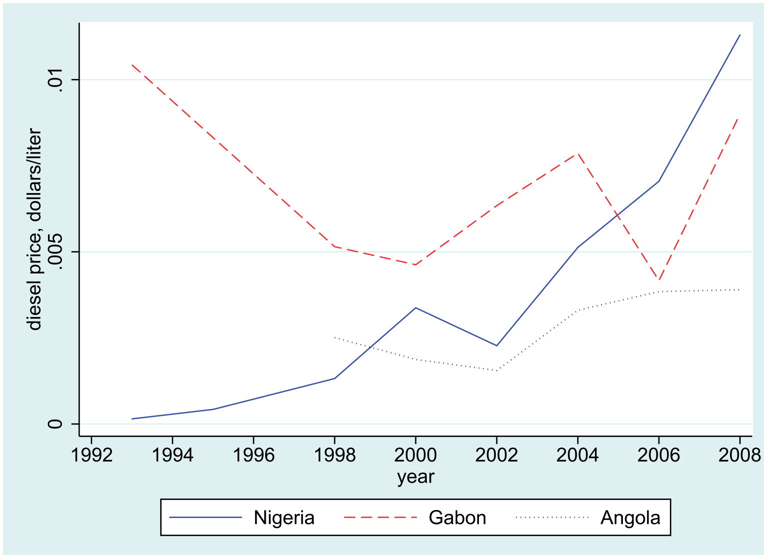

FIGURE A.1

Diesel prices for the main oil producing countries, 1993–2008

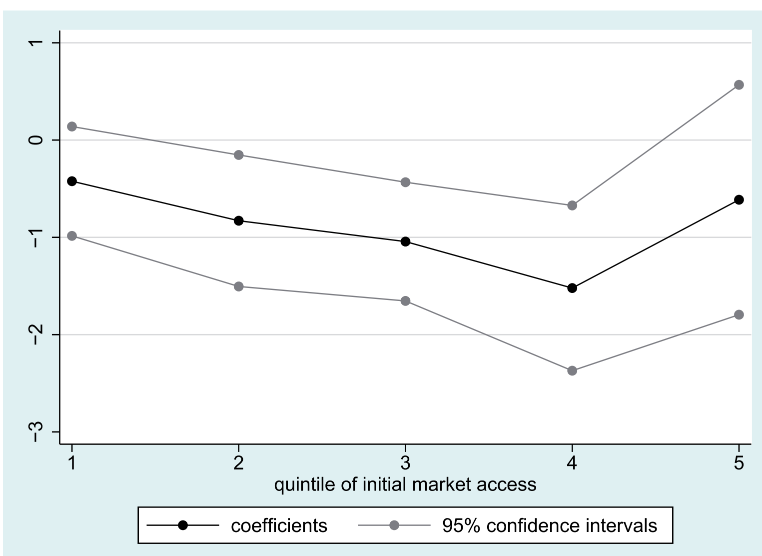

FIGURE A.2

Estimates of β for quintiles of the distribution of initial market access

TABLE A.1
Descriptives

Variable	Mean	Std. Dev.	Min.	Max.	N
Lights DN	3,138.658	10,955.395	0	152,239	4,913
ln(lights), with zeroes set to ln(5.5)	6.256	1.927	1.705	11.933	4,913
P_{oil}, hundreds of U.S. Dollars/barrel	0.395	0.218	0.168	0.969	4,913
distance(Primate)	0.560	0.463	0	2.36	4,913
distance(Primate) * P_{oil}	0.221	0.242	0	2.288	4,913
distance(Primate) * ln(P_{oil})	−0.59	0.596	−4.203	0	4,913
ln(distance(Primate)+1 km) * P_{oil}	−0.449	0.746	−6.696	0.833	4,913
P_{diesel}, U.S. Dollars/litre	0.659	0.353	0.015	1.37	2,240
distance(Primate) * P_{diesel}	0.363	0.406	0	3.234	2,240
dist.(Primate; unpaved 50% more costly)	0.632	0.547	0	2.814	4,913
dist.(Primate; unpaved 100% more costly)	0.673	0.599	0	2.973	4,913
dist.(Primate; unpaved 50%) * P_{oil}	0.249	0.282	0	2.728	4,913
dist.(Primate; unpaved 100%) * P_{oil}	0.265	0.307	0	2.882	4,913
ln(government spending, PPP)	22.087	1.522	18.302	24.72	4,600
ln(natural resource income, PPP)	21.412	2.591	14.33	25.349	2,819
ln(GDP, PPP)	24.206	1.298	19.497	26.397	4,906
1(country is eligible for AGOA in year)	0.461	0.499	0	1	4,913
No. of years since country in AGOA	2.291	2.939	0	9	4,913
distance(Primate)*ln(Gov. spending, PPP)	12.5	10.196	0	49.133	4,600
distance(Primate)*ln(nat. res. income, PPP)	12.31	10.308	0	52.371	2,819
distance(Primate)*ln(GDP, PPP)	13.614	11.164	0	55.629	4,906
distance(Primate)*P_{oil}*1(AGOA)	0.143	0.26	0	2.288	4,913
distance(Primate)*P_{oil}*(AGOA years)	0.857	1.942	0	20.592	4,913
1(rail to primate)	0.235	0.424	0	1	4,913
1(closest power plant is hydroelectric)	0.422	0.494	0	1	4,267
distance(Primate)*P_{oil}*1(rail to primate)	0.041	0.121	0	1.41	4,913
distance(Primate)*P_{oil}*1(hydro closest)	0.095	0.193	0	2.288	4,267
1(nearest primate is foreign)	0.429	0.495	0	1	4,913
1(nearest port is foreign)	0.287	0.453	0	1	4,913
distance to pop>100k	0.139	0.17	0	0.899	4,913
distance to pop in top 20% of country	0.163	0.17	0	1.166	4,913
distance(pop>100k)*P_{oil}	0.055	0.082	0	0.871	4,913
distance(poptop20%)*P_{oil}	0.064	0.084	0	1.13	4,913
dist.(primate)*P_{oil}*1(foreign primate)	0.053	0.082	0	0.751	4,913
distance(port)*P_{oil}*1(foreign port)	0.093	0.129	0	1.077	4,913
ln(market access (MA) to primate)	13.182	3.16	6.033	20.954	4,641
ln(MA to other cities)	15.925	3.102	6.408	22.508	4,641
ln(MA to primate) * P_{oil}	5.201	3.205	1.016	20.313	4,641
ln(MA to other cities) * P_{oil}	6.283	3.74	1.08	21.819	4,641
Primate path fraction paved	0.681	0.262	0.025	1	4,913
dist.(primate)*P_{oil}*1(paving> median)	0.089	0.18	0	2.035	4,913
dist.(poptop20%)*P_{oil}*1(paving> median)	0.015	0.041	0	0.479	4,913

Notes: The unit of analysis is the city-year. Distances are in thousands of kilometres.

Acknowledgments. I thank Managing Editor Francesco Caselli and three anonymous referees for comments that substantially improved the article. I am grateful to David Weil, Nate Baum-Snow, and especially Vernon Henderson for their advice and support. I thank Katherine Casey, Uwe Deichmann, Kyle Emerick, Leo Feler, Ivan Fernandez-Val, Andrew Foster, Delia Furtado, Jake Goldston, Doug Gollin, Walker Hanlon, Erik Hurst, Yannis Ioannides, Margaret McMillan, Blaise Melly, Sriniketh Nagavarapu, Mark Roberts, Nathan Schiff, Enrico Spolaore, Michael Suher, Kensuke Teshima, Ishani Tewari, Matt Turner, Nicholas Wilson, Jeff Zabel, Junfu Zhang, and conference and seminar participants at Brown, CIESIN, Clark, Columbia, Georgia State, Maryland, MIT, Notre Dame, NYU, Syracuse, Toronto, Tufts, UC Irvine, Wharton, EconCon, MWIEDC, NEUDC, Toulouse Workshop on Procurement and Infrastructure, the UC Berkeley-IDB Impact Evaluation Collaborative, the World Bank conference on "Urbanization in Africa: Trends, Promises and Challenges" and the annual meetings of the AEA, CSAE, and NARSC for helpful comments and suggestions, Uwe Deichmann and Siobhan Murray for access to and guidance on the roads data, and Qinghua Zhang for help with the China data and analysis. Financial support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development award number T32HD007338, the North American Regional Science Council, the World Bank's Research Support Budget and Knowledge for Change Program, and a Global Research Program on Spatial Development of Cities funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the U.K. Department for International Development, is gratefully acknowledged. All errors are my own.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

ADES, A. F. and GLAESER, E. L. (1995), "Trade and Circuses: Explaining Urban Giants", *The Quarterly Journal of Economics*, **110**, 195–227.

ALDER, S. (2015), "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development" (Mimeo, University of North Carolina).

ALSAN, M. (2015), "The Effect of the TseTse Fly on African Development", *American Economic Review*, **105**, 382–410.

ATACK, J., BATEMAN, F., HAINES, M., *et al*. (2010), "Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850-1860", *Social Science History*, **34**, 171–0197.

ATKIN, D. and DONALDSON, D. (2015), "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs" (Working papers 21439, National Bureau of Economic Research, Inc).

BALK, D., POZZI, F., YETMAN, G., *et al*. (2004), "The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents" (Mimeo, CIESIN, Columbia University).

BANERJEE, A., DUFLO, E. and QIAN, N. (2012), "On the Road: Access to Transportation Infrastructure and Economic Growth in China" (Working Papers 17897, National Bureau of Economic Research, Inc).

BARRIOS, S., BERTINELLI, L. and STROBL, E. (2006), "Climatic change and rural-urban migration: the case of sub-Saharan Africa", *Journal of Urban Economics*, **60**, 357–371.

BAUM-SNOW, N. (2007), "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics*, **122**, 775–805.

BAUM-SNOW, N., BRANDT, L. HENDERSON, J. V. *et al*. (2015), "Roads, Railroads and Decentralization of Chinese Cities" (Mimeo, Brown University).

BRINKHOFF, T. (2010), "City Population" http://www.citypopulation.de (accessed: 19 March 2010).

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

CHANDRA, A. and THOMPSON, E. (2000), "Does Public Infrastructure Affect Economic Activity?: Evidence from the Rural Interstate Highway System", *Regional Science and Urban Economics*, **30**, 457–490.

COŞAR, A. K. and FAJGELBAUM, P. D. (2016), "Internal Geography, International Trade, and Regional Specialization", *American Economic Journal: Microeconomics*, **8**, 24–56.

COLLIER, P. (2007), *The Bottom Billion: Why the Poorest Countries Are Failing and What Can be Done About It* (Oxford: Oxford University Press).

CONLEY, T. G. (1999), "GMM Estimation with Cross Sectional Dependence", *Journal of Econometrics*, **92**, 1–45.

COULIBALY, S. and FONTAGNÉ, L. (2006), "South–South Trade: Geography Matters", *Journal of African Economies*, **15**, 313–341.

DEATON, A. and HESTON, A. (2010), "Understanding PPPs and PPP-Based National Accounts", *American Economic Journal: Macroeconomics*, **2**, 1–35.

DEUTSCHE GESELLSCHAFT FÜR TECHNISCHE ZUSAMMENARBEIT (2009), *International Fuel Prices 2009* (6th edn. Eschborn, Germany: Deutsche Gesellschaft für Technische Zusammenarbeit).

DONALDSON, D. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure", *American Economic Review*, (forthcoming).

DONALDSON, D. and HORNBECK, R. "Railroads and American Economic Growth: A 'Market Access' Approach", *The Quarterly Journal of Economics*, (forthcoming).

DURANTON, G. and TURNER, M. A. (2012), "Urban Growth and Transportation", *The Review of Economic Studies*, **79**, 1407–1440.

EATON, J. and KORTUM, S. (2002), "Technology, Geography, and Trade", *Econometrica*, **70**, 1741–1779.

ELVIDGE, C. D., BAUGH, K. E., KIHN, E. A. *et al*. (1997), "Mapping of City Lights using DMSP Operational Linescan System Data", *Photogrammetric Engineering and Remote Sensing*, **63**, 727–734.

ELVIDGE, C. D., SAFRAN, J. M., NELSON, I. L. *et al*. (2004), "Area and Position Accuracy of DMSP Nighttime Lights Data", in Lunetta, R. S. and Lyon, J. G. (eds.) *Remote Sensing and GIS Accuracy Assessment* (Boca Raton, FL: CRC Press), 281–292.

ELVIDGE, C. D., ZISKIN, D., BAUGH, K. E. *et al*. (2009), "A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data", *Energies*, **2**, 595–622.

FABER, B. (2014), "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System", *The Review of Economic Studies*, **81**, 1046–1070.

FAY, M. and OPAL, C. (2000), "Urbanization Without Growth: A Not-So-Uncommon Phenomenon" (Policy Research Working Paper Series 2412, The World Bank).

FRAZER, G. and VAN BIESEBROECK, J. (2010), "Trade Growth under the African Growth and Opportunity Act", *The Review of Economics and Statistics*, **92**, 128–144.

GALLUP, J. L., SACHS, J. and MELLINGER, A. (1999), "Geography and Economic Development", *International Regional Science Review*, **22**, 179–232.

GOLLIN, D. and ROGERSON, R. (2010), "Agriculture, Roads, and Economic Development in Uganda" (Working Papers 15863, National Bureau of Economic Research, Inc).

GONZALEZ-NAVARRO, M. and QUINTANA-DOMEQUE, C. (2016), "Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects", *The Review of Economics and Statistics*, **98**, 254–267.

GOODMAN, L. A. (1960), "On the Exact Variance of Products", *Journal of the American Statistical Association*, **55**, 708–713.

GREENE, W. (2004a), "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model", *Econometric Reviews*, **23**, 125–147.

—— (2004b), "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects", *Econometrics Journal*, **7**, 98–119.

GWILLIAM, K. M. (2011), *Africa's Transport Infrastructure* (Washington, DC: World Bank).

HENDERSON, VERNON (2002), "Urban Primacy, External Costs, and Quality of Life", *Resource and Energy Economics*, **24**, 95–106.

HENDERSON, V., STOREYGARD, A. and WEIL, D. N. (2012), "Measuring Economic Growth from Outer Space", *American Economic Review*, **102**, 994–1028.

HOYLE, B. and CHARLIER, J. (1995), "Inter-port Competition in Developing Countries: An East African Case Study", *Journal of Transport Geography*, **3**, 87–103.

HSIANG, S. M. (2010), "Temperatures and Cyclones Strongly Associated with Economic Production in the Caribbean and Central America", *Proceedings of the National Academy of Sciences*, **107**, 15367–15372.

JEDWAB, R. (2013), "Urbanization without Industrialization: Evidence from Consumption Cities in Africa" (Mimeo, The George Washington University).

JEDWAB, R. and MORADI, A. (2016), "The Permanent Economic Effects of Transportation Revolutions in Poor Countries: Evidence from Africa", *The Review of Economics and Statistics*, **98**, 268–284.

JOHNSON, S., LARSON, W., PAPAGEORGIOU, C. *et al*. (2013), "Is newer better? Penn World Table Revisions and Their Impact on Growth Estimates", *Journal of Monetary Economics*, **60**, 255–274.

KRUGMAN, P. R. (1991), "Increasing Returns and Economic Geography", *Journal of Political Economy*, **99**, 483–499.

LIMÃO, N. and VENABLES, A. (2001), "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade", *The World Bank Economic Review*, **15**, 451–479.

McINTOSH, C. T. and SCHLENKER, W. (2006), "Identifying Non-Linearities in Fixed Effects Models" (Mimeo, UC-San Diego).

MINNESOTA POPULATION CENTER. (2011), "Integrated Public Use Microdata Series, International: Version 6.1 [Machine-Readable Database]" http://international.ipums.org (accessed: 28 April 2011).

NATIONAL BUREAU OF STATISTICS (2009), *Tanzania Statistical Abstract 2006* (Dar es Salaam: National Bureau of Statistics).

NATIONAL BUREAU OF STATISTICS AND MINISTRY OF INDUSTRY, TRADE AND MARKETING AND CONFEDERATION OF TANZANIAN INDUSTRIES. (2010), *Annual Survey of Industrial Production and Performance, 2008: Statistical Report* (Dar es Salaam: National Bureau of Statistics).

NUNN, N. and PUGA, D. (2012), "Ruggedness: The Blessing of Bad Geography in Africa", *The Review of Economics and Statistics*, **94**, 20–36.

PERSITS, F., AHLBRANDT, T., TUTTLE, M. *et al*. (2002), "Map Showing Geology, Oil and Gas Fields, and Geologic Provinces of Africa, version 2" (Open File Report 97-470A, version 2.0, United States Geological Survey).

DU PLESSIS, H. W., VISSER, A. T. and CURTAYNE, P. C. (1990), "Fuel Consumption of vehicles as Affected by Road-Surface Characteristics", in Meyer, W. E. and Reichert, J., (eds), *Surface Characteristics of Roadways: International Research and Technologies* (Philadelphia: American Society for Testing and Materials), 480–496.

REDDING, S. J. and TURNER, M. A. (2015), "Transportation Costs and the Spatial Organization of Economic Activity", in Duranton, G., Henderson, V. and Strange, W., (eds), *Handbook of Urban and Regional Economics*, vol. 5. (Amsterdam: North-Holland), 1339–1398.

ROTHENBERG, A. D. (2013), "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways" (Mimeo, RAND Corporation).

ROTUNNO, L., VÉZINA, P.-L. and WANG, Z. (2013), "The Rise and Fall of (Chinese) African Apparel Exports", *Journal of Development Economics*, **105**, 152–163.

TERAVANINTHORN, S. and RABALLAND, G. (2009), *Transport Prices and Costs in Africa: a Review of the Main International Corridors* (Washington, DC: World Bank).

TUTTLE, B. T., ANDERSON, S. J., SUTTON, P. *et al*. (2013), "It used to be dark here: geolocation calibration of the defense meteorological satellite program operational linescan system", *Photogrammetric Engineering and Remote Sensing*, **79**, 287–297.

UNITED NATIONS (2008), *World Urbanization Prospects: The 2007 Revision* (New York: United Nations).

WORLD BANK (2007a), *A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005* (Washington, DC: World Bank).

—— (2007b), "An Overview of China's Transport Sector-2007" (EASTE Working Paper 15, Transport, Energy and Mining Sector Unit, Sustainable Development Department, East Asia and Pacific Region, The World Bank).

—— (2010), *Africa's Infrastructure: a Time for Transformation* (Washington, DC: World Bank).