

Firm Location, Transport Connectivity, and Agglomeration Economies

Evidence from Liberia

Atsushi Iimi
Kulwinder Rao



WORLD BANK GROUP

Transport and Digital Development Practice

April 2018

Abstract

Transport connectivity is among the most important factors in increasing firm productivity and accelerating economic development. The literature generally supports the idea of agglomeration economies, although there is little evidence of their effectiveness in Africa. There are often empirical challenges, such as spatial externalities and endogeneity of infrastructure development. Using firm registry data in Liberia, this study used the instrumental variable spatial autoregressive model to examine the effects of transport

connectivity on firms' decisions on where to locate. The study found significant spatial autocorrelation and possible endogeneity related to transport infrastructure, and that firms are more likely to be located where market accessibility is better. The data indicated strong agglomeration economies, indicating that the primary city, Monrovia, is likely to continue to grow and attract more people and firms, and that secondary cities can also grow with greater transport connectivity between populated areas, such as district centers.

This paper is a product of the Transport and Digital Development Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at aiimi@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

**Firm Location, Transport Connectivity, and Agglomeration Economies:
Evidence from Liberia**

Atsushi Iimi,[¶] Kulwinder Rao

Transport & ICT Global Practice
World Bank Group

Key words: agglomeration economies; conditional logit model; instrumental variable spatial autocorrelation; infrastructure investment

JEL classification: H54; H41; R32; C21; C26.

[¶] Corresponding Author.

I. INTRODUCTION

Transport infrastructure is among the most important factors in increasing firm productivity and supporting local businesses. In particular, transport connectivity plays an important role in influencing firms' decisions on where to locate. Proximity to markets is essential to firms, and transport infrastructure is its main determinant. In Hungary, road availability is considered to be a significant determinant of firm investment (Boudier-Bensebaa 2005). Access to good port facilities is also essential for private investors (Belderbos and Carree 2002; Deichmann et al. 2005). Foreign direct investment tends to be concentrated in countries with good transport infrastructure (Cieřlik and Ryan 2004; Milner et al. 2006; Procher 2011; Lee et al. 2012; Mare and Graham 2013).

The traditional inventory model (e.g., Arrow et al. 1951) also suggests that firms can reduce inventory costs by lowering transportation costs. In general, it is financially costly to firms to carry inventory, restraining their operational flexibility. In the United States, the level of firm inventory declined by 7 cents for every additional dollar of investments in the highway capital (Shirley and Winston 2004). In China, each dollar of road spending has been estimated to reduce firms' held inventories by 2 percent (Li and Li 2013). Using the Golden Quadrilateral project linking India's four largest cities as a quasi-natural experiment, Datta (2012) found that firms in cities located along the improved highway reduced their inventory by 7 operating days' supply on average.

When firms are located near one another, transport and transaction costs can be minimized. The new economic geography literature indicates that agglomeration economies can make an industrial cluster or a city attractive (e.g., Krugman 1991; Fujita et al. 1999). This creates a dynamic effect that continues to influence further investment decisions of firms. Therefore, although the cost of distance has been declining in recent years because of new technologies such as information and communications technology, firms still prefer to be located near one another to share the common input markets of labor and intermediate inputs and thus minimize trade and transaction costs.

As a result, in many developed countries and rapidly growing emerging economies, firm agglomerations and industrial clusters have been established (e.g., Yusuf et al. 2008). In Africa, there are only a few examples of this occurring, such as the textile sector in East Africa (Otsuka and Sonobe 2011). Africa has lagged behind in the global manufacturing market since the 1960s. Among others, low labor productivity and lack of land access are significant constraints in Africa (Dinh et al. 2012). Poor infrastructure is another constraint (Harrison et al. 2011).

Empirically, there are methodological challenges to exploring the relationship between firm location, productivity, and transport infrastructure. First, infrastructure investment is often endogenous. Firms tend to be located where transport connectivity is better, but causality is less obvious than it would appear. Firms can be more productive because of better mobility of goods and people, but at the same time, governments may invest more in transport infrastructure where firms already exist. Thus, causality remains unclear. Recent publications propose different instrumental techniques to avoid this endogeneity problem (e.g., Chandra and Thompson 2000; Banerjee et al. 2012; Jedwab and Moradi 2012).

Another challenge is spatial autocorrelation. Transport infrastructure, by nature, forms a network. Thus, any improvement in transport connectivity has local and spillover effects. The extent to which firms would be affected is normally considered to depend on proximity to the improvement. Tobler's (1970) first law of geography says that, "everything is related to everything else, but near things are more related than distant things." Spatial autoregressive models address this problem. Wang, Kochelman, and Damien (2012) used the spatial autoregressive probit model to analyze land use in Austin, Texas. Other authors have incorporated possible spatial autocorrelation into the count data regression context (e.g., Bhat, Paleti, and Singh 2014).

The current study examined whether transport connectivity affects firms' decisions on where to locate and how firms' decisions affect those of other firms. Using the firm registry data in

Liberia, the study used the instrumental variable spatial autoregressive model (Drukker et al. 2011), which allows the above-mentioned empirical issues to be addressed simultaneously. Policy implications will be developed from the estimation results: Where should investments be concentrated, in the primary city or other urban areas? What kind of transport connectivity affects firm location? The remaining sections are organized as follows: Section II provides an overview of recent developments of formal businesses and road networks in Liberia. Section III develops an empirical methodology and describes our data. Section IV presents main estimation results and discusses some policy implications. Section V concludes.

II. CURRENT ROAD NETWORK AND FIRM LOCATION IN LIBERIA

Liberia is one of the poorest countries in the world and has a population of approximately 4.4 million. Gross domestic product per capita was US\$457 in 2014.¹ Many people live below the poverty line, depending on subsistence farming without access to markets. According to the national census, the national poverty rate was estimated to be 54.1 percent in 2008 (LISGIS 2009). There is significant urban–rural inequality. Monrovia has a poverty rate of 31.6 percent, but poverty is generally more prevalent in the inland north central region (72 percent), such as Lofa County, which was most severely affected by the recent Ebola outbreak.

Agriculture is an important sector, employing approximately half of the labor force. A significant number of people work in the informal sector, which accounts for nearly 60 percent of the total labor force (table 1). The industry and service sectors, especially in the formal sector, are thin. A recent labor force survey showed that approximately 17,000 people are employed in the manufacturing sector and 343,000 in the service sector.

¹ According to *World Development Indicators*.

Table 1 Labor Statistics

County	Population	Labor force		Informal employment	
	(1,000)	(1,000)	% of total	(1,000)	% of labor force
Bomi	101	31	30.7	20	64.5
Bong	379	179	47.2	108	60.3
Grand Bassa	201	75	37.3	51	68.0
Grand Cape Mount	120	62	51.7	39	62.9
Grand Gedeh	79	27	34.2	21	77.8
Grand Kru	54	26	48.1	16	61.5
Lofa	234	135	57.7	86	63.7
Margibi	219	75	34.2	41	54.7
Maryland	96	28	29.2	17	60.7
Montserrado	1,152	362	31.4	197	54.4
Nimba	427	164	38.4	77	47.0
Rivercess	70	35	50.0	20	57.1
Sinoe	90	18	20.0	9	50.0
River Gee	55	22	40.0	17	77.3
Gbarpolu	61	34	55.7	23	67.6
Total	3,338	1,273	38.1	742	58.3

Source: Liberia Institute of Statistics and Geo-Information Services.

Economic activity is highly concentrated in Monrovia, where more than 1.2 million people, or about 30 percent of the total population and 55 percent of the total urban population, live. The distribution of firm location is even more skewed (figure 1). The government’s business registry database indicates that more than 41,000 firms were officially registered in Liberia as of 2016, of which approximately 80 percent were located in Montserrado County, where the capital city, Monrovia, is (table 2). Thus, formal firms are almost exclusively concentrated in Monrovia, but even while Liberia has been experiencing rapid urbanization, many people live in rural and remote areas.

The skewness of firm location is higher than in other countries (figure 2). This does not mean that there are few businesses in other areas; there are many informal businesses, but they have not been registered. As discussed above, informality is significant in the Liberian economy, particularly in rural areas.

Table 2 Firms Registered as of 2016

County	Number of firms	Share (%)
Bomi	258	0.6
Bong	1,090	2.6
Gbarpolu	149	0.4
Grand Bassa	1,063	2.6
Grand Cape Mount	289	0.7
Grand Gedeh	611	1.5
Grand Kru	74	0.2
Lofa	499	1.2
Margibi	1,849	4.4
Maryland	475	1.1
Montserrado	33,421	80.2
Nimba	1,255	3.0
River Gee	338	0.8
Rivercess	70	0.2
Sinoe	251	0.6
Total	41,692	100.0

Source: Liberia Business Registry Database

Figure 1 Geographic Concentration in Liberia

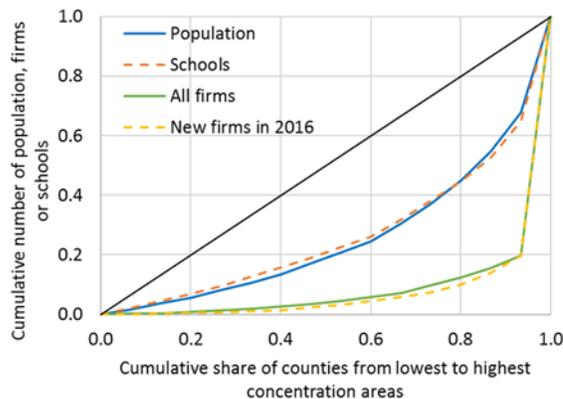
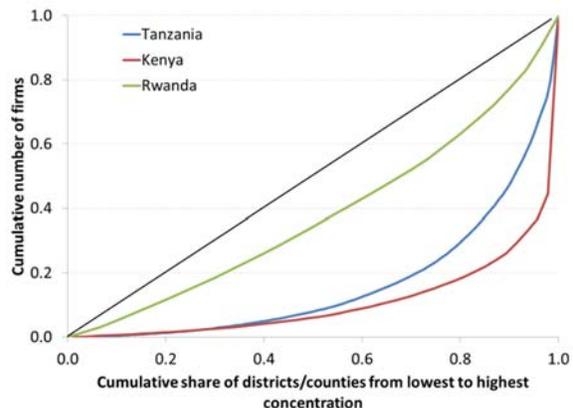


Figure 2 Firm Concentration in Other Countries

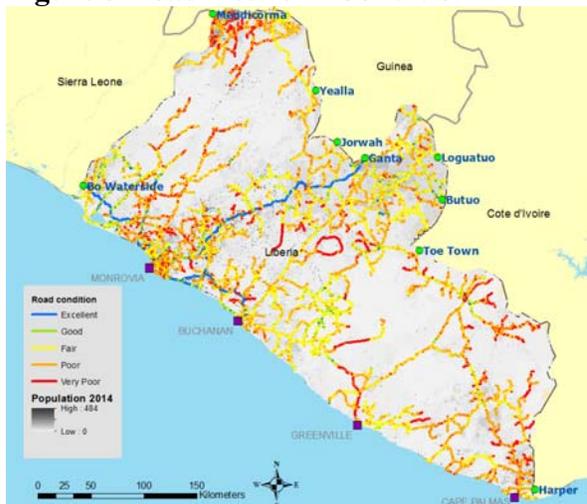


Liberia has approximately 9,916 kilometers of roads, of which only 734 kilometers are paved (Ministry of Public Works 2012). This translates into a road density of 10.3 km per 100 km² of land, which is not particularly high but reasonably comparable with other small countries in the region (12.6 km in Ghana, 11.9 km in Sierra Leone, 8.1 km in Senegal, and 8.0 km in Côte d'Ivoire). Although paved roads are well maintained, the condition of the unpaved road

network is poor.² More than 90 percent of paved roads, especially along the Monrovia–Ganta Corridor, which has recently been rehabilitated, are in good or fair condition, but nearly 60 percent of unpaved roads are in poor or very poor condition (figure 3).

Because of the poor condition of the nonprimary road network, transport connectivity is generally limited and varies significantly across the country. For instance, in rural and remote areas, accessibility is extremely limited. The Rural Access Index (RAI), which measures the share of the rural population that lives in within 2 kilometers of a road in good condition, is estimated at 41.9 percent in Liberia (figure 4).

Figure 3 Road Network Condition



Source: World Bank survey.

Figure 4 Rural Access Index



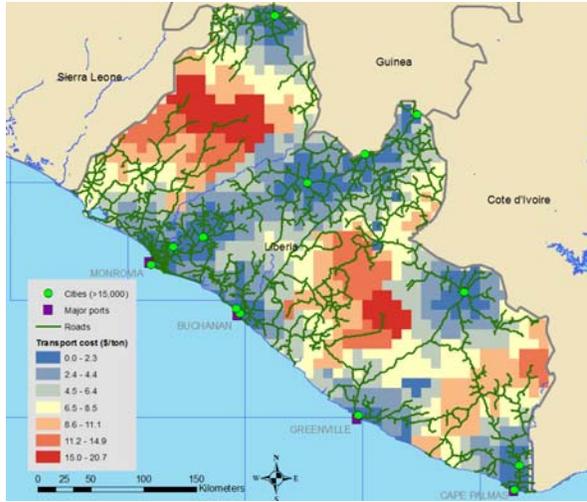
Source: World Bank estimate.

Although access to the road network (RAI in the case of rural areas) is one of the most important economic fundamentals, it is not sufficient from a development point of view. Essentially, people need to be connected to markets to take advantage of economic and social opportunities, which are often located in urban areas. Transport costs to bring one unit of a good to a major market can be estimated using georeferenced road network data (figure 5). Large cities and towns with more than 15,000 people are used as a proxy of market

² To update the road network condition, a road inventory survey using a smartphone application was conducted in the entire country in May 2016; 11,423 kilometers of roads were surveyed.

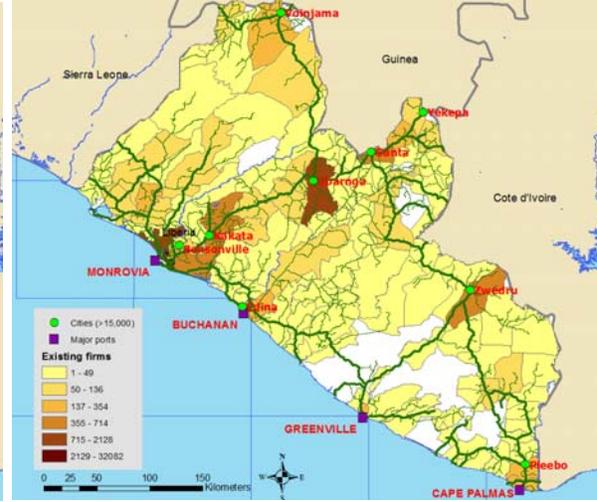
accessibility. Transport costs are relatively low along the Monrovia–Ganta Corridor but are much higher in other parts of the country, particularly in the north, such as Gbapolu County, and the central region, such as River Cess and Nimba Counties. These areas are completely isolated from the domestic market. It is likely that the difference in market accessibility plays an important role in influencing firms’ decisions on where to locate. There is a clear correlation between firm location and transport costs to market (figure 6).

Figure 5 Transport Costs to Market (US\$/ton)



Source: World Bank estimate.

Figure 6 Number of Firms Registered: 2016



Source: Liberia Business Registry Database

III. METHODOLOGY AND DATA

To examine firms’ decisions on where to locate in connection with site-specific characteristics, a standard discrete choice model was considered in the conditional logit framework (McFadden 1974). Suppose that firm i maximizes the following general profit function at location or district j :

$$\pi_{ij} = x'_j \beta + \xi_j + \varepsilon_{ij} \quad (1)$$

where $x'_j \beta + \xi_j$ is the mean profit level at location (district) j , and ε_{ij} is an idiosyncratic error including firm-specific unobservables; x consists of location-specific observables, including a measurement of agglomeration economies and infrastructure variables.

Following Berry (1994), the set of firm-specific unobservables is defined by

$A_j = \{\varepsilon_i \mid \pi_{ij} > \pi_{ik} \text{ for } \forall k \neq j\}$. Then the share of firms choosing to be located in district j can be written implicitly as $s_j = \int_{A_j} f(\varepsilon; x, \sigma_\varepsilon) d\varepsilon$. When ε is assumed to be independently and identically distributed, with the extreme-value distribution $\exp(-\exp(-\varepsilon))$, the log of the probability of a firm choosing location j is:

$$\ln s_j - \ln s_0 = x'_j \beta + \xi_j \quad (2)$$

with the mean profit of the outside option normalized to zero; s_0 is the share of firms that take an outside option, which is assumed to be employment in the informal sector in our context. In other words, people who are working in the informal sector can choose to formalize their businesses in a particular district. They do so if expected profits are positive. Otherwise, they continue to stay in the informal sector or unemployed. Recent labor force statistics were used to calculate the number of informal businesses that choose the outside option, and it is assumed that the average size of informal businesses is five employees.

Transport connectivity can be measured differently. In this analysis, six variables are examined: rural accessibility (*RAI*) in district j , road density (*RDDEN*), share of roads in good condition (*GDRD*), transport costs to a nearby large city (*TCMK*), transport costs to the largest port (Freeport of Monrovia, *TCPT*), and market access index (*MAI*) at location j . The first variable is a global indicator of rural accessibility at the district level, as discussed above. *RDDEN* and *GDRD* are traditional transport indicators. It has long been argued whether they can measure actual connectivity accurately.

The transport cost to market variable (*TCMK*) is considered to be a more-accurate measurement than the above, representing connectivity with the road condition taken into account. Given the georeferenced road condition data, transport costs to bring one unit of a good to a major market are calculated using spatial software. In principle, if road conditions are bad, road user costs tend to be high and travel speed low, which increases time costs. From each location, the optimal route is selected to minimize total transport cost to reach a city with a population of more than 15,000. Similarly, transport costs to Freeport of Monrovia are calculated.

A more-general market accessibility measurement can be calculated. The above transport cost variables are still less general, because they reflect connectivity to a particular destination. In reality, people and firms need to be connected to different locations or potentially all cities and markets. The level of the importance of each market may depend on proximity to it. Thus, with the size of market capacity and distance to the market taken into account, the following simple market access index is considered, based on a conventional gravity framework:³

$$MAI_j = \left(\sum_k Y_k / d_{jk} \right) / \max_j (MAI_j) \quad (3)$$

This is the sum of purchasing power or market capacity, Y , inversely weighted by the degree of impediment between two locations (e.g., distance, d). For Y , city population is used as a proxy for market capacity; d is measured according to transport costs from district j and city k .

To estimate Equation (2) in the logit framework, a critical empirical concern is that it assumes the independence of irrelevant alternatives, which requires that preferences between any pair of two choices be independent of the third option (e.g., Bhat, Paleti, and Singh 2014). In our context, this means that the relative attractiveness of one location (A) over

³ The literature also uses negative exponential weights. See, e.g., Elbadawi, Mengistae, and Zeufack 2006; Lall and Mengistae 2005.

another location (B) is not changed even if the underlying characteristics of the third location (C) are changed, but this is not likely to hold. For instance, suppose that some unanticipated investment is made in location C . It is not in location A or B , but if A is next to C , A may become more attractive than B , because of the spillover effects from C . In general, locational attractiveness is interdependent with all other locations, especially if transport connectivity is good. Thus, $Cov(\xi_j, \xi_k) \neq 0$ in the equation.

Spatial interdependence comes from two sources. First, there is often spatial interdependency between locations. Although infrastructure is a typical network industry (e.g., roads are connected to each other), agglomeration economies themselves imply spatial interdependency with the neighboring areas. Second, from an empirical point of view, some variables are likely to have been omitted from the model. Spatial autocorrelation lies in ξ_j if there are any unobservable location-specific characteristics.

In the literature, there are two empirical solutions to address this problem. The first approach is the nested logit model, which can partly relax the independence of irrelevant alternatives assumption across different clusters. The potential disadvantage is that this is only a partial solution and that the nesting structure cannot be known *ex ante*. The second approach, which the current study relied on, is to allow correlated errors in ξ_j and eliminate the independence of irrelevant alternatives assumption. This is more flexible, with few presumptions required, and computationally simpler. Particularly, taking the share equation (Berry 1994), as in Equation (2), is advantageous, because it is computationally heavy to incorporate spatial autocorrelation into the standard maximum likelihood procedure.

Letting our dependent variable, i.e., $\ln s - \ln s_0$, be s and explanatory variables x be X in the matrix notation, the following general spatial autoregressive model is considered:

$$\begin{aligned} s &= X\beta + \lambda Ws + \xi \\ \xi &= \rho M\xi + v \end{aligned} \tag{4}$$

where λ is a spatial autoregressive dependence in district share s , ρ represents possible autocorrelation in error term ζ , and \mathbf{W} and \mathbf{M} are spatial weighting matrices. For both matrices, inverse distances between two locations (j and k) are used. This follows Tobler's (1970) first law of geography: "everything is related to everything else, but near things are more related than distant things." Two locations are more closely related to each other if they are located near each other. The distance is calculated according to the Euclidean distance between the two locations.

Another empirical challenge to estimating Equation (2) is that infrastructure placement may be endogenous. Firms are productive because of better transport connectivity. At the same time, the government may invest in particular places because firms are located there. For instance, some firms are technically more efficient and prefer to be located where transport access is good. Road authorities may invest more in the places where these productive firms are located. As a result, regardless of transport connectivity, good infrastructure and good firms are likely to coexist.

To address the possible endogeneity of infrastructure variables in \mathbf{X} , the spatial instrumental variable estimator (e.g., Anselin 1988; Drukker et al. 2011) is used, in which (β, λ) is first estimated using the instrumental variable technique in the untransformed model. Using the estimated residuals, ρ is then estimated. Finally, using the results and transforming the equation, the generalized spatial two-stage least-squared estimator can be obtained.

To instrument the transport connectivity variables, three instruments in logarithm are considered: terrain slope at each district center ($SLOP$), the elevation of each district center ($ELEV$), and straight-line distance from each district center to the nearest historical port city ($DIST$). The first two variables are considered to be fairly exogenous because they are geographic conditions. The third variable follows the recent literature focused on the evaluation of infrastructure investment. Chandra and Thompson (2000), examining the effect of U.S. interstate highways on earnings of firms, argue that the nonmetropolitan counties served by highways receive exogenous benefits, because the interstate highways first aimed

at connecting metropolitan areas. Banerjee et al. (2012) apply the same concept to the case of Chinese railways, calculating the distance from counties to straight lines connecting historic cities and ports, which can be arguably treated as exogenous. The validity of our proposed instruments will be examined using conventional test statistics, such as the exogeneity and overidentifying restriction tests, in the following section.

The summary statistics are shown in table 3. There were an average of 52.6 firms registered per district in 2016. There is significant variation in this. To avoid complexity related to another type of endogeneity, the number of existing firms in each district was measured using the 2015 data. On average, there were 320 firms in each district, with a range from none to more than 25,000 firms. As shown above, the RAI varies considerably from district to district. Similarly, other transport variables also differ from district to district. The market access index was normalized to a range from zero to 100.

Population density was included in the model to control for differences between districts, which was expected to capture the size of the local market. Firms are normally expected to be located near their customers, holding everything else constant. Labor costs may also affect firms' decisions on where to locate. In theory, firms prefer to locate where wages are low, everything else being equal. Because wage data were not available, poverty and unemployment were included. Presumably, wages tend to be low where poverty and unemployment are high. At the same time, firms may also consider the quality of labor. Firms may want to hire skilled labor regardless of cost. To show this opposite effect, the share of population that attained at least a junior high school education (*EDU*) was included. As discussed, informality of the economy may also be an important factor in firms' decisions about where to locate.

To control for other unobservable subregional characteristics, five region-specific dummy variables are also included (north western (Bomi, Gbarpolu, Grand Cape Mount), north central (Bong, Lofa, Nimba), south central (Grand Bassa, Margibi), south eastern A (Grand Gedeh, Rivercess, Sinoe), and south eastern B (Grand Kru, Maryland, River Gee)), and

Montserrado is used as a baseline. This follows the classification in a recent poverty analysis (LISGIS 2009). Poverty data are not available at a more disaggregated level. Therefore, no poverty measurement can be included in X .

Table 3 Summary Statistics

Variable	Abb.	Obs.	Mean	Std. Dev.	Min	Max
Number of firms newly registered at each district in 2016	<i>dN</i>	136	52.62	407.03	0.00	4264.00
Number of firms registered as of 2015	<i>NI5</i>	136	320.36	2463.31	0.00	25819.00
Rural access index at each district	<i>RAI</i>	136	18.38	20.83	0.00	94.60
Road density (km/100 km ²)	<i>RDDEN</i>	136	15.84	17.23	0.00	143.25
Share of roads in good condition (%)	<i>GDRD</i>	136	6.82	9.65	0.00	58.33
Transport cost to bring one unit of goods to large city with population > 15,000 (\$/ton)	<i>TCMK</i>	134	5.78	3.88	0.00	19.98
Transport cost to bring one unit of goods to Freeport of Monrovia (\$/ton)	<i>TCPT</i>	134	27.11	14.71	0.63	51.39
Market access index from each district center	<i>MAI</i>	134	0.87	8.65	0.01	100.00
Population density (1,000 per km ²)	<i>POPDEN</i>	136	0.10	0.52	0.004	5.92
Share of informal employment (%)	<i>INFO</i>	136	47.36	15.45	20.93	69.70
Unemployment rate (%)	<i>UNEMP</i>	136	40.11	20.38	12.12	91.95
Share of population who attained at least junior high school	<i>EDU</i>	136	17.79	7.11	7.00	48.00
Instruments:						
Slope at each district center (degree)	<i>SLOP</i>	136	0.45	0.53	0.03	3.62
Elevation (m)	<i>ELEV</i>	136	153.86	126.46	5.00	563.00
Distance to nearest historic port (km)	<i>DIST</i>	136	93.60	55.38	0.02	214.39

IV. MAIN ESTIMATION RESULTS

First of all, ordinary least squares (OLS) regression is performed. The results are shown in table 4. Although the results may be biased because endogeneity and spatial autocorrelation are not addressed, they are largely consistent with economic theory. Population density, which is an indication of the size of the local market, has a significant effect. Agglomeration economies are also always strongly positive; firms establish themselves where other firms already exist. For transport connectivity, the coefficients are largely consistent with prior expectations, except for road density, which has a significant negative coefficient. Market access seems particularly important; the coefficients of *TCMK* and *MAI* are statistically significant, although the results are not conclusive, as discussed above.

Table 4 Ordinary Least Squares Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
<i>lnRAI</i>	0.023 (0.114)					
<i>lnRDDEN</i>		-0.183 ** (0.088)				
<i>lnGDRD</i>			0.022 (0.130)			
<i>lnTCMK</i>				-0.341 *** (0.097)		
<i>lnTCPT</i>					-0.544 (0.651)	
<i>lnMAI</i>						0.643 *** (0.140)
<i>lnNI5</i>	0.818 *** (0.073)	0.870 *** (0.076)	0.820 *** (0.074)	0.751 *** (0.074)	0.804 (0.072)	0.726 *** (0.074)
<i>POPDEN</i>	0.2490 * (0.1370)	0.2680 * (0.1520)	0.2490 * (0.1380)	0.3280 *** (0.1130)	0.2740 * (0.1530)	0.2800 *** (0.0929)
<i>INFO</i>	-0.006 (0.027)	-0.010 (0.027)	-0.006 (0.027)	0.003 (0.026)	0.000 (0.027)	0.006 (0.026)
<i>UNEMP</i>	-0.004 (0.015)	-0.005 (0.014)	-0.004 (0.015)	-0.009 (0.014)	-0.001 (0.015)	-0.005 (0.014)
<i>EDU</i>	0.020 (0.098)	0.010 (0.098)	0.019 (0.098)	0.015 (0.095)	0.025 (0.099)	0.012 (0.095)
Constant	-13.370 ** (5.214)	-12.280 ** (5.132)	-13.340 ** (5.280)	-13.090 *** (4.983)	-13.440 ** (5.198)	-12.010 ** (5.032)
Observations	136	136	136	134	134	134
R-squared	0.64	0.643	0.64	0.654	0.643	0.66
F-statistic	103.5	110	103.9	157.04	102.5	186.86
Number of region dummy variables	5	5	5	5	5	5

Instrumental variable regression is performed to address the endogeneity related to transport infrastructure investment (table 5). The results are broadly similar; the data exhibit significant economies of agglomeration and show that market access is important to attracting firms. More firms are located where transport costs to markets are lower and the market access index is higher. The results indicate that transport connectivity seems largely exogenous, although in the model in which *MAI* is regressed, the exogeneity test can be rejected at the 10 percent significance level. In this case, our instruments are found to be

largely valid; the overidentifying restriction test cannot be rejected no matter which transport connectivity variables are used.

Table 5 Instrumental Variable Regression Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
<i>lnRAI</i>	0.028 (0.402)					
<i>lnRDDEN</i>		0.163 (1.277)				
<i>lnGDRD</i>			-0.214 (0.370)			
<i>lnTCMK</i>				-0.240 * (0.141)		
<i>lnTCPT</i>					-0.028 (1.678)	
<i>lnMAI</i>						0.405 ** (0.194)
<i>lnNI5</i>	0.817 *** (0.135)	0.783 ** (0.323)	0.870 *** (0.101)	0.773 *** (0.076)	0.824 *** (0.096)	0.763 *** (0.077)
<i>POPDEN</i>	0.2500 (0.1400)	0.2270 (0.1890)	0.2160 (0.1310)	0.3050 *** (0.1080)	0.2500 (0.1590)	0.2680 *** (0.0891)
<i>INFO</i>	-0.006 (0.027)	-0.003 (0.035)	-0.014 (0.029)	0.001 (0.025)	-0.004 (0.029)	0.003 (0.026)
<i>UNEMP</i>	-0.004 (0.022)	-0.005 (0.014)	-0.015 (0.024)	-0.008 (0.014)	-0.005 (0.020)	-0.005 (0.014)
<i>EDU</i>	0.021 (0.096)	0.025 (0.114)	0.004 (0.094)	0.019 (0.091)	0.027 (0.095)	0.017 (0.091)
Constant	-13.420 ** (5.964)	-13.940 * (8.199)	-11.390 ** (5.605)	-13.250 *** (4.776)	-13.630 *** (5.008)	-12.610 *** (4.803)
Observations	136	136	136	134	134	134
R-squared	0.64	0.631	0.627	0.653	0.641	0.658
Wald chi ²	1247.7	1184.7	1160.7	1669.1	1191.9	1774.8
Number of region dummy variables	5	5	5	5	5	5
Exogeneity test ¹						
Chi ² test statistic	0.000	0.071	0.321	1.152	0.106	2.775 *
P-value	(0.990)	(0.790)	(0.571)	(0.283)	(0.744)	(0.096)
Overidentifying restriction test ¹						
Chi ² test statistic	3.999	3.943	3.695	3.096	4.065	2.909
P-value	(0.135)	(0.139)	(0.158)	(0.213)	(0.131)	(0.234)

Note: Robust standard errors are shown in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1%, respectively.

¹ The instrumental variable estimation is performed using three instruments in logarithm: *SLOP*, *ELEV*, and *DIST*.

The instrumental variable spatial autoregressive model is used to address the possible spatial autocorrelation. The results are shown in table 6, showing that agglomeration economies are significant and market accessibility is important for firms in deciding where to locate. Both coefficients of *TCMK* and *MAI* are statistically significant, but transport costs to Freeport of Monrovia have a negative but insignificant coefficient. Thus, port access per se may not be particularly important to attracting firm investment, possibly because port accessibility is generally poor except for in the Monrovia area. The significance of the coefficient of *MAI* can be interpreted to mean that connectivity to Monrovia is still important both because it is a populated market and because it has the primary port.

The instrumental variable spatial autoregressive results also indicate that spatial autocorrelation affects firms' decisions on where to locate. The spatial autoregressive term λ is estimated to be -0.33 to -0.37 . Contrary to prior expectations, an unanticipated shock of a firm locating in a particular district had negative externalities for neighboring areas. On the other hand, the spatial error term ρ is found to be positive but statistically insignificant. This means that an exogenous shock in a particular district is not likely to affect its neighboring districts. For instance, if an extreme weather event, such as flood, unexpectedly happens in a particular district, the event may affect firm investment behavior in that district but not in neighboring districts.

As discussed above, the spatial autoregressive model is more efficient if the transport connectivity variables are exogenous. The results are unchanged (table 7). Firms are more likely to be located where transport costs to market are lower and general market accessibility measured according to *MAI* is higher. The coefficient of *NI5* is estimated to be approximately 0.75 to 0.8, which is also similar to the above, indicating strong agglomeration economies. Monrovia is likely to continue growing because that is where (formal) enterprises are currently located, although this does not exclude the possibility that other cities will grow. The coefficient of *TCPT*, which essentially measures the effect of proximity to Monrovia, is always insignificant, but *MAI* has a significant coefficient, which measures

more-general connectivity, including not only Monrovia, but also secondary cities.

Therefore, other secondary cities can also grow when their intercity connectivity is improved.

Table 6 Instrumental Variable Spatial Autoregressive Model

	(1)	(2)	(3)	(4)	(5)	(6)
<i>lnRAI</i>	0.115 (0.192)					
<i>lnRDDEN</i>		-0.035 (0.385)				
<i>lnGDRD</i>			-0.123 (0.216)			
<i>lnTCMK</i>				-0.312 ** (0.132)		
<i>lnTCPT</i>					-0.775 (1.087)	
<i>lnMAI</i>						0.503 *** (0.189)
<i>lnNI5</i>	0.788 *** (0.088)	0.828 *** (0.114)	0.845 *** (0.083)	0.752 *** (0.074)	0.788 *** (0.075)	0.741 *** (0.075)
<i>POPDEN</i>	0.1660 (0.1190)	0.1780 (0.1440)	0.1840 (0.1260)	0.2390 (0.1140)	0.1880 (0.1390)	0.1860 (0.0877)
<i>INFO</i>	-0.003 (0.025)	-0.006 (0.026)	-0.009 (0.027)	0.001 (0.024)	0.000 (0.024)	0.003 (0.024)
<i>UNEMP</i>	0.002 (0.014)	-0.003 (0.013)	-0.009 (0.017)	-0.006 (0.013)	0.003 (0.016)	-0.003 (0.013)
<i>EDU</i>	0.021 (0.091)	0.010 (0.091)	0.006 (0.091)	0.003 (0.089)	0.007 (0.092)	0.001 (0.089)
Constant	-15.960 *** (4.991)	-14.630 *** (5.260)	-13.820 *** (5.124)	-14.260 *** (4.662)	-14.540 *** (4.778)	-13.560 *** (4.703)
Observations	134	134	134	134	134	134
Number of region dummy variables	5	5	5	5	5	5
Autoregressive coefficients						
λ	-0.372 ** (0.152)	-0.349 * (0.193)	-0.335 * (0.200)	-0.330 ** (0.160)	-0.376 *** (0.136)	-0.353 ** (0.157)
ρ	0.526 (1.181)	0.759 (1.160)	1.057 (1.103)	0.711 (1.118)	0.390 (1.183)	0.702 (1.120)

Note: Robust standard errors are shown in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1%, respectively.

¹ The instrumental variable estimation is performed with three instruments in logarithm: *SLOP*, *ELEV*, and *DIST*.

Table 7 Spatial Autoregressive Model without Instruments

	(1)	(2)	(3)	(4)	(5)	(6)
<i>lnRAI</i>	0.046 (0.107)					
<i>lnRDDEN</i>		-0.132 (0.092)				
<i>lnGDRD</i>			0.055 (0.122)			
<i>lnTCMK</i>				-0.324 *** (0.091)		
<i>lnTCPT</i>					-0.745 (0.634)	
<i>lnMAI</i>						0.653 *** (0.124)
<i>lnNI5</i>	0.806 *** (0.069)	0.853 *** (0.072)	0.808 *** (0.070)	0.749 *** (0.068)	0.789 *** (0.066)	0.718 *** (0.067)
<i>POPDEN</i>	0.1680 (0.1200)	0.2000 (0.1360)	0.1660 (0.1210)	0.2420 (0.1120)	0.1880 (0.1300)	0.1910 (0.0969)
<i>INFO</i>	-0.004 (0.024)	-0.008 (0.025)	-0.004 (0.025)	0.001 (0.024)	0.000 (0.024)	0.005 (0.024)
<i>UNEMP</i>	-0.001 (0.013)	-0.003 (0.013)	0.000 (0.014)	-0.007 (0.013)	0.003 (0.014)	-0.003 (0.013)
<i>EDU</i>	0.016 (0.091)	0.007 (0.090)	0.015 (0.091)	0.002 (0.089)	0.007 (0.091)	-0.003 (0.089)
Constant	-15.260 *** (4.831)	-14.120 *** (4.744)	-15.240 *** (4.908)	-14.200 *** (4.644)	-14.540 *** (4.750)	-13.140 *** (4.667)
Observations	134	134	134	134	134	134
Number of region dummy variables	5	5	5	5	5	5
Autoregressive coefficients						
λ	-0.357 ** (0.160)	-0.330 * (0.181)	-0.358 ** (0.160)	-0.324 ** (0.159)	-0.370 ** (0.147)	-0.348 ** (0.157)
ρ	0.622 (1.111)	0.899 (1.060)	0.556 (1.151)	0.700 (1.118)	0.408 (1.126)	0.674 (1.148)

Note: Robust standard errors are shown in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1%, respectively.

One may wonder to examine whether the agglomeration variable *NI5* is also endogenous. More-lagged values are used to test this. The numbers of firms registered by 2013 and 2014 are denoted as *NI3* and *NI4*, respectively. It is less likely that past decisions directly affect a firm's decision on where to locate. The results confirm robustness of the main estimation results shown above (table 8). Market accessibility is an important determinant of firm

location, and agglomeration economies were still found to be significant. Autocorrelation was also significant.

Table 8 Instrumental Variable Spatial Autoregressive Model with Different Agglomeration Variables

	(1)	(2)	(3)	(4)
<i>lnTCMK</i>	-0.322 ** (0.137)		-0.307 ** (0.124)	
<i>lnMAI</i>		0.512 *** (0.195)		0.468 *** (0.181)
<i>lnNI4</i>	0.694 *** (0.092)	0.683 *** (0.093)		
<i>lnNI3</i>			0.757 *** (0.089)	0.746 *** (0.091)
<i>POPDEN</i>	0.2910 ** (0.1370)	0.2390 ** (0.1110)	0.1700 (0.1320)	0.1280 (0.1070)
<i>INFO</i>	0.002 (0.026)	0.004 (0.026)	0.005 (0.024)	0.007 (0.024)
<i>UNEMP</i>	-0.011 (0.014)	-0.007 (0.014)	-0.009 (0.014)	-0.006 (0.013)
<i>EDU</i>	0.018 (0.095)	0.016 (0.095)	0.044 (0.089)	0.042 (0.089)
Constant	-14.100 *** (4.913)	-13.400 *** (4.967)	-15.210 *** (4.628)	-14.550 *** (4.672)
Observations	134	134	134	134
Number of region dummy variables	5	5	5	5
Autoregressive coefficients				
λ	-0.303 * (0.156)	-0.327 ** (0.152)	-0.294 *** (0.104)	-0.315 *** (0.105)
ρ	0.432 (1.330)	0.372 (1.318)	-1.091 (1.249)	-1.020 (1.250)

Note: Robust standard errors are shown in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1%, respectively.

¹ The instrumental variable estimation is performed with three instruments in logarithm: *SLOP*, *ELEV*, and *DIST*.

V. CONCLUSION

Agglomeration economies are an important factor in increasing firm productivity and promoting economic growth. The evidence shows that firms prefer to be located close to one

another, but unlike in other developing areas, Africa has only a few industrial clusters, such as textiles in East Africa. There is little evidence, although many African countries are currently experiencing rapid urbanization.

This study explored the relationship between firm location and transport connectivity. Using the recent firm registry database in Liberia, firms' decisions on where to locate were examined with respect to various indicators that measure transport connectivity. The instrumental variable spatial autoregressive model was used to address the potential endogeneity that infrastructure placement causes spatial autocorrelation between neighboring locations. The results indicate significant agglomeration economies. It is likely that Monrovia continues to grow because that is where (formal) enterprises are located. Other cities will also grow when their intercity transport connectivity is improved.

It was also found that market accessibility is a main driver of firm agglomeration; it is important to connect firms to large markets. Other connectivity measures do not have significant coefficients. Traditional indicators such as road density and the share of good roads have no power to explain firms' decisions on where to locate. These measurements may not be important to consider firm location and promote firm investment.

Finally, from a methodological point of view, the study found the importance of addressing spatial autocorrelation in the firm location literature. Spatial autocorrelation is always significant. There are externalities of firm location choice around neighboring areas.

REFERENCES

- Adams, Arvil, Sara Johansson de Silva, and Setareh Razmara. 2013. *Improving Skills Development in the Informal Sector Strategies for Sub-Saharan Africa*. Washington, DC: World Bank.
- Amin, Mohamed, Duncan Willetts, and Alastair Matheson. 1986. *Railway across the Equator: The Story of the East African Line*. London: The Bodley Head Ltd.
- Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.
- Banerjee, A., E. Duflo, and N. Qian. 2012. "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Paper 17897, National Bureau of Economic Research, Washington, DC.
- Belderbos, René, and Martin Carree. 2002. "The Location of Japanese Investments in China: Agglomeration Effects, Keiretsu, and Firm Heterogeneity." *Journal of the Japanese and International Economies* 16 (3): 194–211.
- Berry, Steven. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25(2): 242–62.
- Bhat, Chandra R., Rajesh Paleti, and, Palvinder Singh. 2014. "A Spatial Multivariate Count Model for Firm Location Decisions." *Journal of Regional Science* 54 (3): 462–502.
- Boudier-Bensebaa, Fabienne. 2005. "Agglomeration Economies and Location Choice: Foreign Direct Investment in Hungary." *Economics of Transition* 13 (4): 605–28.
- Brulhart, Marius, Mario Jametti, and Kurt Schmidheiny. 2012. "Do Agglomeration Economies Reduce the Sensitivity of Firm Location to Tax Differentials?" *The Economic Journal* 122: 1069–93.
- Chandra, and Thompson. 2000. "Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System" *Regional Science and Urban Economics* 30: 457–90.
- Cieřlik, Andrzej, and Michael Ryan. 2004. Explaining Japanese Direct Investment Flows into an Enlarged Europe: A Comparison of Gravity and Economic Potential Approaches. *Journal of the Japanese and International Economies* 8 (1): 12–37.

- Deichmann, Uwe, Kai Kaiser, Somik Lall, and Zmarak Shalizi. 2005. "Agglomeration, Transport, and Regional Development in Indonesia." Policy Research Working Paper No. 3477. World Bank, Washington, DC.
- Dinh, Hinh, Vincent Palmade, Vandana Chandra, Fances Cossar. 2012. *Light Manufacturing in Africa: Targeted Policies to Enhance Private Investment and Create Jobs*. Washington, DC: World Bank.
- Drukker, D. M., I. R. Prucha, and R. Raciborski. 2011. A Command for Estimating Spatial-Autoregressive Models with Spatial Autoregressive Disturbances and Additional Endogenous Variables. Working Paper. Department of Economics, University of Maryland, , Maryland.
- Elbadawi, Mengistae, and Zeufack. 2006. "Market Access, Supplier Access, and Africa's Manufactured Exports: An Analysis of the Role of Geography and Institutions." World Bank Policy Research Working Paper No. 3942. World Bank, Washington, DC.
- Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. *The Spatial Economy*. Cambridge, MA: MIT Press.
- Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99 (3): 483–99.
- Lall and Mengistae. 2005. "The Impact of Business Environment and Economic Geography on Plant-Level Productivity: An Analysis of Indian Industry." World Bank Policy Research Working Paper No. 3664. World Bank, Washington, DC.
- Lee, Ki-Dong, Seok-Joon Hwang, and Min-hwan Lee. 2012. "Agglomeration Economies and Location Choice of Korean Manufactures within the United States." *Applied Economics* 44: 189–200.
- LISGIS (Liberia Institute of Statistics and Geo-Information Services). 2009. "Liberia: 2008 National Population and Housing Census Final Results." Liberia Institute of Statistics and Geo-Information Services, Monrovia, Liberia
- Mare, David, and Daniel Graham. 2013. "Agglomeration Elasticities and Firm Heterogeneity." *Journal of Urban Economics* 75: 44–56.

- McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka: 105–42. Academic Press.
- Milner, Chris, Geoff Reed, and Pawin Talerngsri. 2006. "Vertical Linkages and Agglomeration Effects in Japanese FDI in Thailand." *Journal of the Japanese and International Economies* 20 (2): 193–208.
- Ministry of Industry and Trade. 2012. *Tanzania Industrial Competitiveness Report 2012*. Ministry of Industry and Trade, Tanzania.
- Ministry of Public Works. 2012. "National Transport Masterplan." Ministry of Public Works, Monrovia, Liberia
- Otsuka, Keijiro, and Tetsushi Sonobe. "A Cluster-Based Industrial Development Policy for Low-Income Countries." Policy Research Working Paper No. 5703. World Bank, Washington, DC.
- Procher, Vivien. 2011. "Agglomeration Effects and the Location of FDI: Evidence from French First-Time Movers." *Annals of Regional Science* 46: 295–312.
- Tobler Waldo. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46 (2): 234–40.
- Yusuf, Shahid, Kaoru Nabeshima, and Shoichi Yamashita. 2008. *Growing Industrial Clusters in Asia: Serendipity and Science*. Washington, DC: World Bank.