

Firms' Locational Choice and Infrastructure Development in Tanzania

Instrumental Variable Spatial Autoregressive Model

Atsushi Iimi

Richard Martin Humphreys

Sevara Melibaeva



WORLD BANK GROUP

Trade and Competitiveness Global Practice Group

June 2015

Abstract

Agglomeration economies are among the most important factors in increasing firm productivity. However, there is little evidence supportive of this in Africa. Using the firm registry database in Tanzania, this paper examines a new application of the logit approach with two empirical issues taken into account: spatial autocorrelation and endogeneity of infrastructure placement. The paper finds significant agglomeration economies. It is also found that

firms are more likely to be located where local connectivity and access to markets are good. The paper finds that dealing with infrastructure endogeneity and spatial autocorrelation in the empirical model is important. According to the exogeneity test, infrastructure variables are likely endogenous. The spatial autoregressive term is significant. As expected, therefore, there are positive externalities of firm location choice around the neighboring areas.

This paper is a product of the Trade and Competitiveness Global Practice Group. 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://econ.worldbank.org>. 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.

**FIRMS' LOCATIONAL CHOICE AND INFRASTRUCTURE DEVELOPMENT IN TANZANIA:
INSTRUMENTAL VARIABLE SPATIAL AUTOREGRESSIVE MODEL**

Atsushi Iimi,[¶] Richard Martin Humphreys, Sevara Melibaeva

Sustainable Development Department, Africa Region
The World Bank

Key words: Agglomeration economies; Conditional logit model; instrumental variable spatial autocorrelation; infrastructure investment

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

[¶] Corresponding Author.

I. INTRODUCTION

The new economic geography literature suggests that agglomeration economies are a key factor in increasing firm productivity (Krugman, 1991; Fujita et al. 1999). Despite the falling costs of distance in recent years, firms still prefer to be located close to one another at a particular locality in order to share the common input markets of labor and intermediate inputs and thus minimize trade and transaction costs. This is referred to as economies of agglomeration, and the literature is generally supportive of this (e.g., Procher, 2011; Lee et al., 2012; Mare and Graham, 2013). In many developed countries and rapidly growing emerging economies, firm agglomerations and industrial clusters have been established (e.g., Yusuf et al. 2008).

However, in Africa there are only a few examples of this occurring, such as the textile sector in East Africa (Otsuka and Sonobe, 2011). It is well recognized that Africa has been lagging behind in the global manufacturing market since the 1960s. A number of constraints exist. Among others, low labor productivity and lack of land access are significant constraints in Africa (Dinh et al. 2012). Infrastructure is also a binding constraint. Electricity access is limited to about 30 percent of total population in the region. Many roads, other transport infrastructure, and services are poor in terms of delivery and/or condition.

Importantly, unlike general households, firms have more flexibility in choosing their locations, depending on available infrastructure services at each locality. Firms are clearly selective in this regard. Particularly, the literature suggests that transport connectivity is crucial to the locational choices of firms. In Hungary, road availability is found to be a significant determinant of investment (Boudier-Bensebaa, 2005). Foreign direct investment tends to be concentrated in countries with good transport infrastructure (Cieřlik and Ryan, 2004; Milner *et al.*, 2006). Access to ports is also important for investment decisions (Belderbos and Carree, 2002; Deichmann *et al.*, 2005).

The current paper aims at examining the firm location behavior in Tanzania, which is a country with a land area of more than 900,000 km². Inequality of infrastructure access is significant, and the economic structure varies markedly across regions. While the coastal areas have good access to one of the major regional seaports at Dar es Salaam, the inland areas are more than 1,000 km away from the coast. Dar es Salaam is the primary city with a population of 4.4 million, while the regions around Lake Victoria are agriculturally very fertile. This is clearly posing a challenge to promote industrial clusters, while ensuring a balanced growth path of the whole economy. The paper analyzes how these characteristics affect firms' location decisions.

From the methodological point of view, the paper also addresses several important empirical issues, such as endogeneity of infrastructure placement. Firms may be located based on infrastructure endowments; infrastructure investments are also dependent on where they are. In addition, spatial externalities may potentially matter, because of the spillover effects of infrastructure as well as agglomeration economies. To address these issues, the paper relies on the instrumental variable spatial autoregressive model (Drukker et al. 2011). The remaining sections are organized as follows: Section II provides an overview of recent economic and infrastructure developments in Tanzania. Section III discusses our empirical methodology and data. Section IV presents main estimation results, and Section V discusses some policy implications. Section VI concludes.

II. RECENT ECONOMIC AND INFRASTRUCTURE DEVELOPMENTS IN TANZANIA

The Tanzanian economy has been growing fast at 6.5 to 8 percent for more than a decade, supported in part by favorable international commodity prices (Figure 1). Manufacturing and services have been particularly strong contributors, partly because of the country's new discovery and production of natural resources. This is contributing to a significant improvement in the electricity supply, which has long been a crucial constraint on the

economy (IMF 2013).¹ By contrast, agriculture growth remains relatively modest. Tanzania is traditionally a large exporter of coffee, tea and tobacco, but despite high international crop prices, the supply response has been slow. The vast majority of agricultural production remains at the subsistence level in Tanzania (Figure 2).

Following the Asian experience, the Government of Tanzania has been implementing strong industrial policies to promote manufacturing investments, particularly in medium-technology industries, natural gas-based and agro-processing. The global experience is generally supportive of the importance of boosting manufacturing, which can bring various positive externalities to the economy. Tanzania is a traditional agriculture-based economy, but the economic structure may be changing, especially in recent years. The share of manufactured exports increased from 10 percent in 2005 to 20 percent in 2010 (Ministry of Industry and Trade, 2012).

Nonetheless, Tanzanian manufacturing is still concentrated on low-tech manufacturing.² Agrobusiness, such as food and beverages, accounts for nearly half of total manufacturing value added, followed by the nonmetal and textile industries. To sustain industrialization and maintain industrial competitiveness in the global market, further diversification is required. Although the government has been making efforts toward creating many industrial parks, such as Special Economic Zones and Export Processing Zones, supporting infrastructure developments are still lagging behind and constraining firm activities (Ministry of Industry and Trade, 2012).

In Tanzania, electricity used to be identified as one of the most crucial constraints on businesses (Figure 3). Only 13.9 percent of total households are estimated to have access to electricity in Tanzania (Table 1). Of course, firms have relatively better access to electricity than residential users do. Still, self-power generation is costly, and the quality of the supply

¹ “United Republic of Tanzania: Sixth Review Under the Policy Support Instrument, Second Review Under the Standby Credit Facility Arrangement, and Request for Modification of Performance Criteria.” 2013. IMF.

² Tanzania is ranked 110th (out of 118) according to the UNIDO’s Competitive Industrial Performance (CIP) index.

remains an issue. Since Tanzania is a large country, infrastructure inequality is significant. Particularly in rural areas, access to electricity is minimal. Only 2,059 villages are electrified out of 10,652 village in Tanzania (Figure 4).

Tanzania's road network is in relatively good condition by regional standards (Figure 5). The primary road network in particular is maintained well. However, transport connectivity in the country is in general still limited. Road density is low (Figure 6), and as a result only 24 percent of total rural population has access to an all-weather road within 2 km (Figure 7). The estimated transport costs vary significantly across locations. For instance, firms based in Dar es Salaam have easy access to the port and thus international markets, but the transport cost from Mwanza (1,000 km away) to the port is \$150 per ton. From Kigoma, it is about \$180 per ton (Figure 8). But these inland areas are much closer to major actual and potential agricultural production areas (Figure 9).

Given these infrastructure and economic conditions, firms are, in theory, supposed to decide where to locate. In Tanzania, there are about 45,000 enterprises that are formally registered according to the Central Register of Establishments data in 2010. Firms are currently concentrated in Dar es Salaam and other major cities. The largest firm agglomeration lies in Dar es Salaam, in which 4.4 million people or about 10 percent of total population live, and accounts for about 25 percent of total firms (Table 2). This is followed by other major regional cities, such as Arusha, Moshi and Mbeya. These secondary agglomerations are located dispersedly in the country (Figure 10). The geographic distribution of firms is relatively less skewed in Tanzania than in Kenya (Figure 11). In Kenya, nearly 60 percent of total firms are based in Nairobi. In the East African region, the firm distribution is much more even in Rwanda, though the size of the country is different (Figure 12).

Notably, however, the primary concentration of firms in Dar es Salaam seems to be accelerating. About 2,300 firms were newly created or formally registered in the three districts in Dar es Salaam in 2010 alone. Particularly, other manufacturing sectors than

agrobusiness are increasingly concentrated in Dar es Salaam (also see Figure 10). The following analysis aims at exploring the possible motivation behind firms' location decisions.

Figure 1. Tanzania: Growth contribution by sector

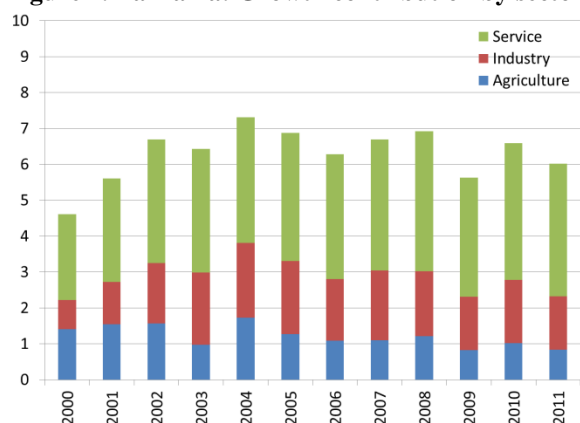


Figure 2. Tanzania: Crop production and prices

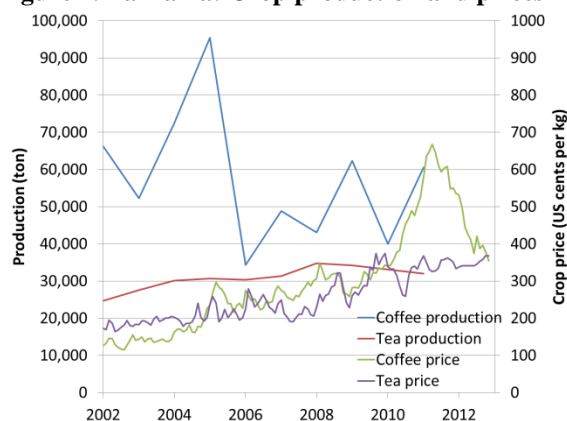
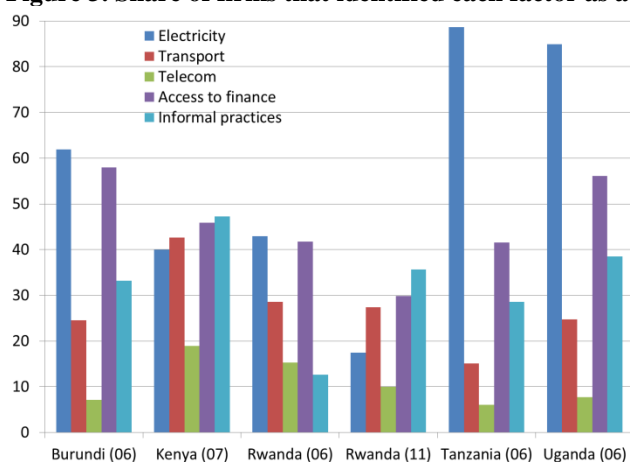


Figure 3. Share of firms that identified each factor as a major constraint (percent)



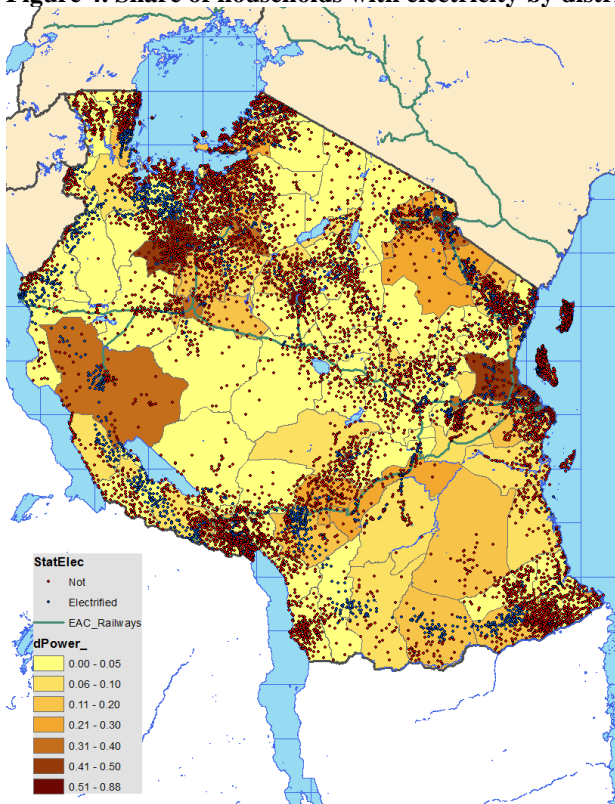
Source: BEEPS.

Table 1. Electrification rates

	Electrification rate (%)	Population without electricity (million)
Burundi	10.8	8.2
Kenya	16.1	33.4
Rwanda	11.0	9.6
Tanzania	13.9	37.7
Uganda	9.0	28.1
Sub-Saharan Africa	30.5	585.2

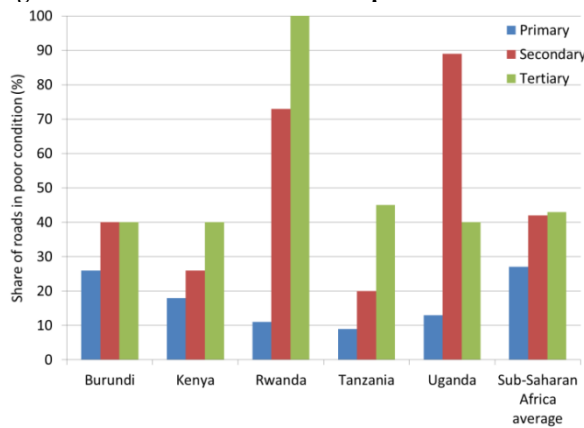
Sources: IEA WEO 2011 for Kenya, Tanzania and Uganda; and estimated based on DHS 2010 for Burundi; and DHS 2010 for Rwanda.

Figure 4. Share of households with electricity by district and electrification status by village



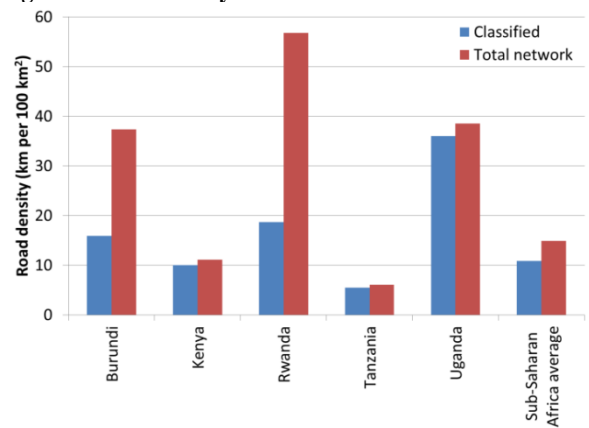
Sources: LSMS 2010 and Rural Energy Agency, Tanzania.

Figure 5. Classified road network in poor condition



Source: Gwilliam (2011).

Figure 6. Road density



Source: Gwilliam (2011).

Figure 7. Rural accessibility index

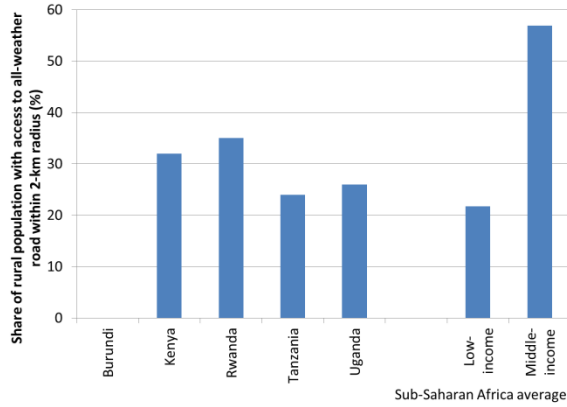
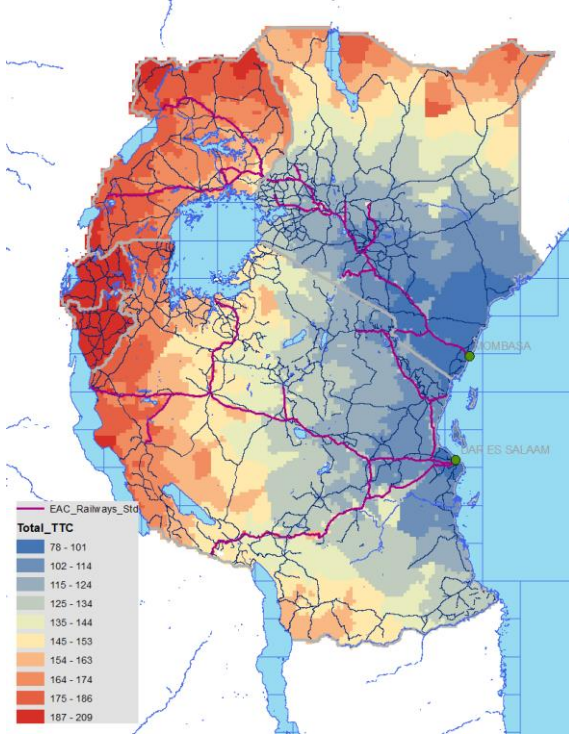
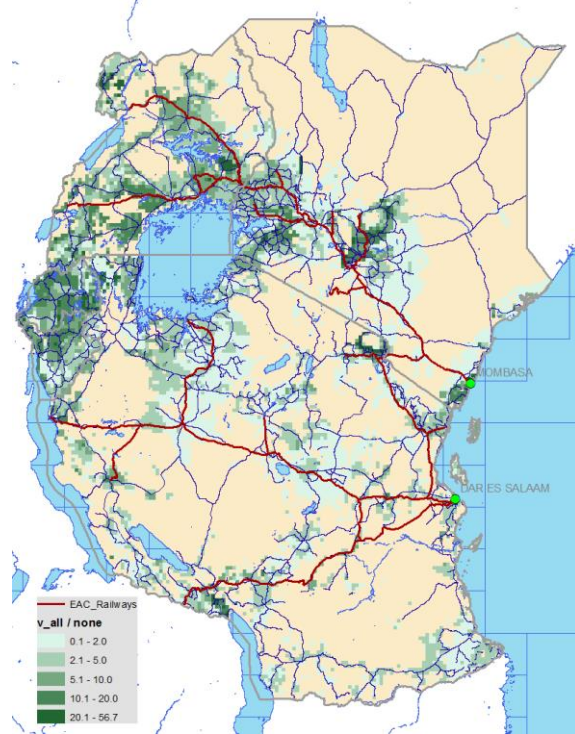


Figure 8. Transport costs to the port (US\$ per ton)



Source: World Bank calculation.

Figure 9. Agriculture production, 2010 (US\$ million)



Source: SPAM update.

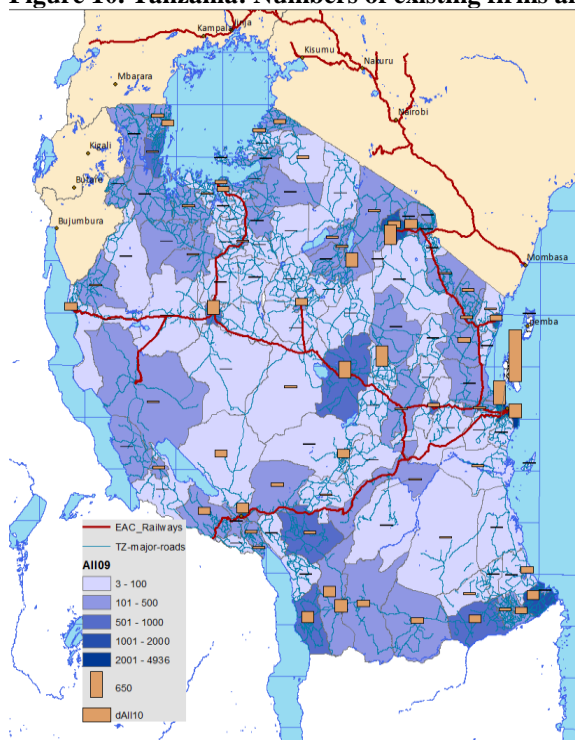
Table 2. Tanzania: Number of existing firms, 2010

	Number of firms registered	Share (%)
Dar es Salaam ¹	11,817	25.8
Arumeru	1,621	3.5
Moshi	1,556	3.4
Mbeya	1,541	3.4
Arusha	1,514	3.3
Songea	1,297	2.8
Bukoba	1,152	2.5
Mbinga	1,124	2.4
Mtwara	1,106	2.4
Tanga	1,032	2.2
Njombe	965	2.1
Dodoma	840	1.8
Other districts	20,316	44.3
Total	45,881	100.0

Source: Tanzania Central Register of Establishments Data.

1/ Dar es Salaam is comprised of three districts: Ilala, Kinondoni, and Temeke.

Figure 10. Tanzania: Numbers of existing firms and new firms



Source: World Bank calculation based on the Central Register of Establishments (CRE) Statistics, 2009, 2010.

Figure 11. Geographic concentration of firms in Kenya and Tanzania

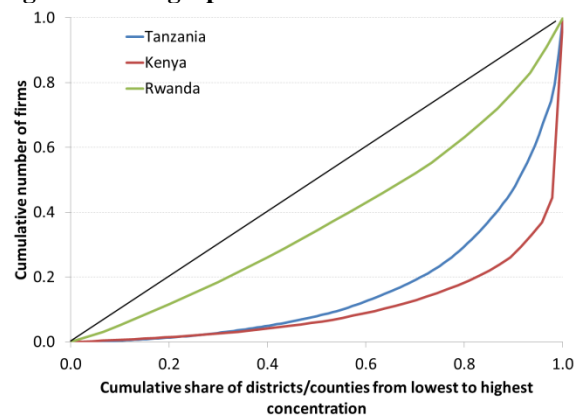
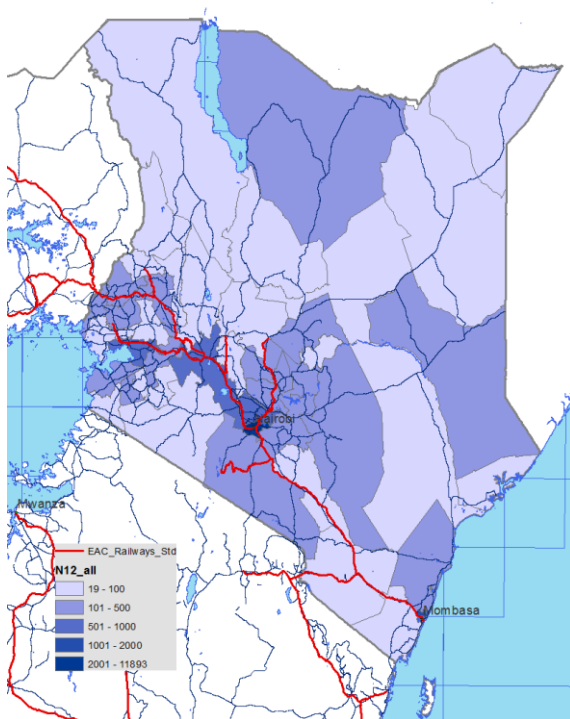


Figure 12. Kenya: Numbers of existing firms in 2012



III. METHODOLOGY AND DATA

To examine the firm's locational choice in connection to site-specific characteristics, we start with the standard conditional logit framework (McFadden, 1974). Suppose that firm i maximizes the following general profit function at location or district j :

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

where $x'_{ij} \beta + \xi_j$ is the mean profit level at location 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 at district j can be written implicitly by $s_j = \int_{A_j} f(\varepsilon; x, \sigma_\varepsilon) d\varepsilon$. 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:

$$\ln s_j - \ln s_0 = x'_{ij} \beta + \xi_j \quad (2)$$

with the mean profit of the outside option normalized to zero. s_0 is the share of firms that take the outside option. In our context, this is assumed to remain unemployed or work in the informal sector. In Tanzania, there is the significant informal sector. About 2.2 million people or about 47 percent of total nonfarm employment are estimated to work in the informal sector (Adams et al. 2013). The potential number of firms that could be created is defined by the total number of labor force (aged 15 years and above) except for those who are formally employed, divided by the average size of new firms (i.e., five persons). In other words, it is assumed that unemployed people or informal workers can choose to register a

firm and be formalized at any location, if the expected profit is positive. Otherwise, they continue unemployed or informal.

One empirical issue to estimate Equation (2) by the traditional conditional logit model is the assumption of independence of irrelevant alternatives (IIA). The IIA assumption requires that preferences between any pair of two choices are independent of the third option. In our context, this means that the relative attractiveness of one location to another is not changed even if the underlying characteristics of the third location are changed. But this is likely to be violated, because the attractiveness of one location is interdependent on all other locations, particularly surrounding ones. Mathematically, $Cov(\xi_j, \xi_k) \neq 0$ in the equation in n (e.g., Procher, 2011; Mare and Graham, 2013).

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

In the literature, two empirical solutions exist to deal with this problem. The first approach is the nested logit model, which can partly relax the IIA assumption across different clusters. The potential disadvantage is that this is only a partial solution and the nesting structure cannot be known ex ante. The second approach, which the current paper relies on, is to allow correlated errors in ξ_j and eliminate the IIA assumption. This is more flexible with no presumption required and computationally simpler.³

³ Recall that there are 120 choices (or districts) in our case. About 11,000 firms were newly created or registered in 2010 alone. Therefore, it is computationally heavy to apply the standard maximum likelihood procedure for this. The advantage of estimating the share equation (Berry, 1994) is that the possible spatial autocorrelation can be dealt with easily in the linear fashion.

Denoting our dependent variable, $\ln s - \ln s_0$, be s and x be X in the matrix notation, the following general spatial autoregressive model is considered:

$$\begin{aligned} s &= X\beta + \lambda Ws + \xi \\ \xi &= \rho M\xi + v \end{aligned} \tag{3}$$

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

Another empirical issue to estimate Equation (2) is that infrastructure placement is potentially endogenous. Infrastructure investments are normally targeted. There must be good reasons for each infrastructure project. For example, productive firms prefer to be located where transport access is good, road authorities may also invest more in the places where productive firms exist. As the result, infrastructure variables are likely endogenous in the equation.

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

Two types of infrastructure are considered in X : electricity and transport. To measure accessibility to electricity, an economic electricity cost is calculated for each district, using

residential energy spending. As shown in Figure 4, electricity access differs across locations. Therefore, the economic cost of energy should be different across locations. With the option of not having access to power taken into account, the weighted average electricity costs are calculated in a spatial dimension. For Tanzania, four energy options are considered: grid power, solar power, own generator, and kerosene lamps as an option of not having power. While grid power is the cheapest at 6.7 cents/kWh, lighting by kerosene lamps are about 50 times more expensive than grid power when lumens are taken into account.

Transportation is critical for any business activities. The firm location literature particularly focuses on accessibility to fundamental transport infrastructure, such as highway and ports (e.g., Boudier-Bensebaa, 2005; Cieřlik and Ryan, 2004; Milner *et al.*, 2006; Belderbos and Carree, 2002; Deichmann *et al.*, 2005). Three measurements are considered in this paper: (i) local connectivity, (ii) accessibility to the domestic market, and (iii) accessibility to the global market. Local connectivity is measured by road density at the district level. The domestic market accessibility is measured by the transport cost to the nearest city with more than 100,000 populations. Finally, the proximity to the global market is measured by the transport cost to the port (either Mombasa or Dar es Salaam). See Figure 8 above.⁴

In order to instrument these infrastructure variables, two instruments are considered: (i) straight-line distance from each district to the nearest hydropower station, and (ii) straight-line distance from each district to the linear proxy rail line between Dar es Salaam and Kigoma. For the former, it is reasonable to assume that hydro resources are fairly exogenous. And the probability to receive “exogenous” benefits from electrification would likely increase if a district is located closer to hydro resources.

⁴ Various assumptions are made to calculate the lowest transport cost from a district to the nearest city or port. It is composed of vehicle operating costs, which are assumed to differ according to road characteristics, such as road class, surface and road conditions, opportunity time costs of drivers, waiting time at transport nodes, such as rail stations and ports, and nodal fees and charges, such as cargo handling and customs fees. An optimal route is selected based on the lowest transport cost, potentially combining multiple transport modes, such as roads and railway.

The second instrument follows the recent literature focused on the evaluation of infrastructure investment. Chandra and Thompson (2000), examining the impact of the U.S. interstate highways on earnings of firms, argue that the non-metropolitan counties served by highways received exogenous benefits, because the interstate highways first aimed at connecting metropolitan areas. Banerjee et al. (2012) also apply the same concept for 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.

Following these studies, a straight-line is drawn from Dar es Salaam, from where the construction of the Central Railway was started in 1905, to Kigoma, a port town situated on Lake Tanganyika, which was reached in 1914 (Figure 13). The railway construction was largely motivated for political and military reasons, especially between two European powers, Britain and Germany (Amin, Willetts and Matheson, 1986). Since other transport infrastructure development followed the railway construction, the areas between Dar es Salaam and Kigoma are considered to be more likely to receive “exogenous” benefits from transport investments.

Other explanatory variables included in X are summarized in Table 3. For obvious reasons, the firm location is likely to depend on the market potential, which is measured by the number of population. Labor market conditions also seem to be important (e.g., Procher, 2011). Both labor costs and quality are included (i.e., average wages and education attainment). The recent literature focuses on the impact of local taxation on firms’ locational choice of investment (e.g., Brulhart, Jametti and Schmidheiny, 2012; Mare and Graham, 2013). Local tax revenue data are available at least at the regional level in Tanzania (denoted by *RENT*). Since agrobusiness is one of the major industries in Tanzania. Presumably, the proximity to agriculture production areas is one possible advantage for agrobusinesses. The total value of agriculture commodities produced in each district is used as a measurement for this.

Figure 13. Instrumental variables for infrastructure variables



Table 3. Summary statistics

	Abb.	Obs	Mean	Std. Dev.	Min	Max
Share of firms deciding to be located at district j	S_j	118	0.000027	0.000046	0.000000	0.000372
Number of firms newly established in district j in 2010		118	93.6	162.2	0.0	1307.0
Number of firms that existed in district j in 2005	NUM	118	137.9	350.4	0.0	2411.0
Road density (km/1,000 km ²) ¹	$RDDN$	118	0.53	1.08	0.02	8.50
Transport cost to the city with more than 100,000 population (US\$/ton) ¹	$TCITY$	118	8.2	15.2	0.0	157.1
Transport cost to the port (US\$/ton) ¹	$TPORT$	118	135.8	21.7	96.4	177.4
District population in 2002	POP	118	283,575	168,541	40,557	1,083,913
Average monthly wage rate (TSh) ²	$WAGE$	118	206,600	270,639	40,333	2,521,750
Share of people who completed at least primary education ²	$EDU1$	118	0.46	0.13	0.14	0.76
Share of people who completed at least secondary education ²	$EDU2$	118	0.05	0.05	0.00	0.24
Land rent (estimated land replacement cost) (TSh million per acre) ²	$RENT$	118	8.71	48.60	0.07	508.00
Agriculture production value (US\$ million) ³	AGV	118	24.0	38.1	0.0	332.6
Local tax (TSh per capita) ⁴	$LTAX$	118	844.9	664.9	212.1	4299.4
Economic electricity cost (US\$ per kWh) ⁵	$POWER$	118	310.3	67.2	96.9	366.5

1/ World Bank calculations based on the Government's road data.

2/ Data are calculated based on the 2010 LSMS.

3/ Based on the National Sample Census of Agriculture in 2002/03, evaluated at the regional commodity prices from FAO statistics.

4/ Based on local tax revenue in 2004. The regional data are available in "Local Government Finance" by President's Office, Regional Administration and Local Government, Ministry of Finance, The United Republic of Tanzania.

5/ World Bank calculations based on the 2010 LSMS data.

IV. MAIN ESTIMATION RESULTS

Instrumental variable spatial autoregressive estimation is performed; the results are shown in Table 4. First of all, agglomeration economies are found to be significant, but the significance seems weak. The elasticity is estimated at 0.14, which is broadly consistent with the existing literature (e.g., 0.18 for French firms in Europe (Procher, 2011); 0.37 for Korean manufactures in the U.S. (Lee et al., 2012); 0.05 for firms in New Zealand (Mare and Graham, 2013)). The policy implication is straightforward: It is important to take advantage

of agglomeration economies to strengthen firm competitiveness. But further efforts are perhaps needed to reinforce the agglomeration effect on the economy.

Second, the estimation results also suggest that transport infrastructure is important to foster firm concentration. Firms are more likely to be located where local connectivity (measured by road density) is high. Similarly, the domestic market connectivity is also found important. The coefficient is significant at -0.37. Firms prefer to be based where the transport costs are lower to buy materials from and send products to the large city.

The global market connectivity is found far more important in terms of magnitude. The coefficient of *TPORT* is significantly negative at -2.44. Therefore, a large number of new firms are expected to be created at a particular locality, if the transport costs from that place to the port are reduced. This indicates the importance of transport corridor investments to connect better local cities to the port as a package.

Regarding other explanatory variables, the local market size is found to be important for firms to decide their location. This is consistent with prior expectation. Education is also important, because this would affect the quality of expected labor supply. Interestingly, the coefficient is significant for primary education, but not for secondary education. The results may be interpreted to mean either that higher education attainment is not demanded by businesses given the level of industrialization or that the quality of secondary education is not sufficient from the business point of view.

The impact of electricity access turned out unclear. The coefficients are positive, not negative, and statistically insignificant. One possible reason is that our measurement of power infrastructure may be poor, because our economic cost of power is calculated based on the residential power access data. Firms normally have better access than households. Second, the power cost may not have much variation in the spatial dimension, regardless of the differences observed in our variable. Therefore, this may not strongly influence the firm location choice.

Another unexpected result is that the coefficient of local agriculture production value is negative and significant. It means that firms are more likely to be located where agricultural production is lower. This is intuitively acceptable from the urban-rural division point of view. Firms are located in urban areas. However, it also means that firms, partly including agrobusinesses, are not well linked to agriculture production areas. This may raise certain concern and remind policy makers of the importance to strengthen infrastructure investments to connect farmers and firms.

Table 4. Spatial instrumental variable estimation results

	(1)		(2)		(3)
<i>lnRDDN</i>	0.959 *** (0.289)				
<i>lnTCITY</i>			-0.377 ** (0.175)		
<i>lnTPORT</i>				-2.444 * (1.279)	
<i>lnNUM</i>	0.113 (0.086)		0.137 * (0.081)	0.138 * (0.079)	
<i>lnPOP</i>	0.665 *** (0.228)		0.482 ** (0.230)	0.626 ** (0.283)	
<i>lnWAGE</i>	0.300 * (0.165)		0.050 (0.186)	0.203 (0.174)	
<i>lnEDU1</i>	2.338 * (1.297)		2.701 ** (1.156)	2.768 ** (1.187)	
<i>lnEDU2</i>	-3.044 (4.574)		1.273 (5.368)	5.134 (5.560)	
<i>lnRENT</i>	0.128 (0.124)		0.220 * (0.125)	0.197 (0.124)	
<i>lnAGV</i>	-0.062 (0.047)		-0.141 *** (0.047)	-0.178 *** (0.057)	
<i>lnTAX</i>	0.157 (0.264)		0.414 (0.309)	-0.169 (0.382)	
<i>lnPOWER</i>	1.814 (1.882)		1.789 (1.633)	1.513 (1.572)	
constant	-23.998 *** (4.296)		-22.520 *** (4.242)	-9.798 (6.803)	
Obs	118		118	118	
λ	0.440 *** (0.116)		0.341 *** (0.095)	0.339 *** (0.079)	
ρ	0.498 (1.011)		0.057 (1.120)	-1.106 (1.110)	

V. DISCUSSION

Given the above results, one of the important empirical questions is whether spatial autocorrelation really matters to the firm location choice. As shown at the bottom of Table 4, the spatial autocorrelation is found to be significant. The spatial autoregressive term λ is estimated at 0.3 to 0.4. Therefore, as expected, there are positive externalities of firm location choice around the neighboring areas.

On the other hand, the spatial error term ρ is found statistically insignificant. This means that an exogenous shock in a particular district is not likely to affect its neighboring districts. For instance, suppose that some extreme weather events, such as flood, unexpectedly happen to a particular district. The event may affect the firm investment behavior in that district, but not in the neighboring districts.

Another important empirical issue is the endogeneity of infrastructure placement. If the infrastructure variables are exogenous, a more efficient model would be the simple spatial autoregression estimation without any instrument involved. As discussed above, the spatial autocorrelation is significant in our data; therefore, the ordinary least squares estimator is not likely to be unbiased in any case.

The estimated coefficients turned out to be broadly similar to the above (Table 5). However, the conventional Hausman exogeneity tests suggest that the spatial instrumental variable estimator is consistent. The chi-square test statistics are estimated at 9.86, 2.68 and 4.44 for the column (1), (2) and (3) models. For the models with road density and transport costs to the port, the exogeneity hypothesis can be rejected. Therefore, the infrastructure variables are likely to be endogenous in our data.

Table 5. Spatial autoregressive estimation without instruments

	(1)		(2)		(3)	
<i>lnRDDN</i>	0.465 *** (0.161)					
<i>lnTCITY</i>			-0.209 ** (0.094)			
<i>lnTPORT</i>					-2.113 * (1.209)	
<i>lnNUM</i>	0.106 (0.082)		0.120 (0.079)		0.118 (0.077)	
<i>lnPOP</i>	0.490 ** (0.224)		0.417 * (0.219)		0.496 ** (0.239)	
<i>lnWAGE</i>	0.193 (0.157)		0.057 (0.160)		0.138 (0.162)	
<i>lnEDU1</i>	2.404 ** (1.199)		2.615 ** (1.121)		2.656 ** (1.163)	
<i>lnEDU2</i>	-3.549 (3.284)		-1.675 (3.256)		-0.544 (3.479)	
<i>lnRENT</i>	0.166 (0.130)		0.212 * (0.120)		0.198 * (0.119)	
<i>lnAGV</i>	-0.064 ** (0.032)		-0.102 *** (0.028)		-0.120 *** (0.035)	
<i>lnTAX</i>	0.063 (0.277)		0.178 (0.288)		-0.230 (0.377)	
<i>lnPOWER</i>	-0.125 (0.590)		-0.061 (0.759)		-0.516 (0.759)	
constant	-19.656 *** (3.063)		-19.064 *** (3.303)		-7.131 (7.469)	
Obs	118		118		118	
λ	0.347 *** (0.093)		0.289 *** (0.096)		0.318 *** (0.090)	
ρ	0.407 (0.947)		0.615 (1.018)		-0.147 (0.904)	

Finally, from the policy point of view, one may be interested in asking which transport infrastructure is most important to attract more firms. The above analysis includes only one transport variable in each regression, and thus, technically speaking, cannot answer this question. There is an obvious constraint on available instrumental variables.

With two more instruments constructed: (i) distance to a line between Dar es Salaam and Mwanza, and (ii) distance to the border line between Kenya and Tanzania, the spatial

autoregressive models are estimated (Table 6).⁵ The results are broadly consistent with Table 4. Among transport variables, the global market connectivity, namely, transport costs to the port, turned out most important, followed by local connectivity measured by road density. Therefore, the transport connectivity from cities to the port should be prioritized when public infrastructure investment is considered.

Table 6. Spatial autoregressive estimation with multiple transport variables

	Spatial autoreg.	Spatial IV
<i>lnRDDN</i>	0.469 *** (0.154)	1.044 *** (0.312)
<i>lnTCITY</i>	-0.149 * (0.090)	-0.280 ** (0.135)
<i>lnTPORT</i>	-2.435 ** (1.036)	-2.930 ** (1.181)
<i>lnNUM</i>	0.112 (0.075)	0.104 (0.071)
<i>lnPOP</i>	0.669 *** (0.212)	0.875 *** (0.229)
<i>lnWAGE</i>	0.135 (0.155)	0.151 (0.165)
<i>lnEDU1</i>	2.765 ** (1.183)	2.835 ** (1.351)
<i>lnEDU2</i>	-2.946 (3.581)	-6.665 ** (3.039)
<i>lnRENT</i>	0.135 (0.122)	0.078 (0.115)
<i>lnAGV</i>	-0.109 *** (0.037)	-0.082 ** (0.034)
<i>lnTAX</i>	-0.035 (0.340)	0.085 (0.347)
<i>lnPOWER</i>	0.789 (0.729)	2.001 * (1.056)
constant	-7.646 (6.496)	-7.498 (6.577)
Obs	118	118
λ	0.475 *** (0.084)	0.629 *** (0.112)
ρ	-0.939 (1.239)	-0.880 (1.681)

⁵ The rationale for the first additional instrument is the same as the rail line from Dar es Salaam to Kigoma. The Central Railway aimed at not only Kigoma but also Mwanza. The rationale for the second instrument is based on the historical fact that Britain and Germany were aggressively racing to develop railroads along the de facto border line in the 1890s, to claim colonial rights. Therefore, the areas closer to the border line would be more likely to receive “exogenous” benefits from railways and other transport modes developed subsequently.

VI. CONCLUSION

Agglomeration economies are one of the important factors to increase firm productivity, despite the falling costs of distance in recent years. Firms still prefer to be located close to one another at a particular locality. Unlike other developing countries, Africa has only a few industrial clusters, such as textiles in East Africa. There are a number of potential constraints on businesses in Africa. Among others, labor productivity, land access and infrastructure are among the important common constraints.

The paper recast light on firm location behavior in Africa. Using the recent firm registry database in Tanzania, particular attention was paid to the effects of infrastructure conditions, such as transport connectivity. To deal with the potential endogeneity caused by infrastructure placement as well as spatial autocorrelation among neighboring locations, the instrumental variable spatial autoregressive model was applied. The results indicate significant agglomeration economies. Thus, it is important to foster firm agglomerations and increase industrial competition. It is also found that improved local connectivity measured by road density and transport access to large cities can attract more firms. The global market connectivity is also found to be particularly important. Firms are more likely to select a place where the transport cost to the port is low.

From the methodological point of view, the paper found the importance of dealing with infrastructure endogeneity in the empirical model. According to the exogeneity test, infrastructure variables are likely endogenous. In addition, spatial autocorrelation is also found to matter to the firm location choice. As expected, there are positive externalities of firm location choice around the 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*. The World Bank.
- Amin, Mohamed, Duncan Willetts, and Alastair Matheson. 1986. *Railway across the Equator: The Story of the East African Line*. The Bodley Head Ltd. London.
- 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*, Vol. 16(3), pp. 194-211.
- Berry, Steven. 1994. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, Vol. 25(2), pp. 242-262.
- Boudier-Bensebaa, Fabienne. 2005. Agglomeration economies and location choice: Foreign direct investment in Hungary. *Economics of Transition*, Vol. 13(4), pp. 605-628.
- Brulhart, Marius, Mario Jametti, and Kurt Schmidheiny. 2012. Do agglomeration economies reduce the sensitivity of firm location to tax differentials? *The Economic Journal*, Vol. 122, pp. 1069-1093.
- Chandra, and Thompson. 2000. "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system" *Regional Science and Urban Economics* 30, 457-490.
- 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*, Vol. 8(1), pp. 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. Washington DC: The World Bank.
- Dinh, Hinh, Vincent Palmade, Vandana Chandra, Fances Cossar. 2012. *Light Manufacturing in Africa: Targeted Policies to Enhance Private Investment and Create Jobs*. The 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, The University of Maryland, Department of Economics.
- Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. *The Spatial Economy*. MIT Press.
- Krugman, Paul. 1991. Increasing returns and economic geography. *Journal of Political Economy*, Vol. 99(3), 483-499.
- 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*, Vol. 44, pp. 189-200.
- Mare, David, and Daniel Graham. 2013. Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics*, Vol. 75, pp. 44-56.
- McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105-142. 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*, Vol. 20(2), pp. 193-208.
- Ministry of Industry and Trade. 2012. *Tanzania Industrial Competitiveness Report 2012*. Ministry of Industry and Trade, Tanzania.
- Otsuka, Keijiro, and Tetsushi Sonobe. A cluster-based industrial development policy for low-income countries. Policy Research Working Paper No. 5703, World Bank.
- Procher, Vivien. 2011. Agglomeration effects and the location of FDI: Evidence from French first-time movers. *Annals of Regional Science*, Vol. 46, pp. 295-312.

Tobler Waldo. 1970. A computer movie simulating urban growth in the Detroit region.

Economic Geography, Vol. 46(2), pp. 234-240.

Yusuf, Shahid, Kaoru Nabeshima, and Shoichi Yamashita. 2008. *Growing Industrial*

Clusters in Asia: Serendipity and Science. The World Bank.