WPS8089

Policy Research Working Paper 8089

Spatial Autocorrelation Panel Regression

Agricultural Production and Transport Connectivity

Atsushi Iimi
Liangzhi You
Ulrike Wood-Sichra

WORLD BANK GROUP
Transport and ICT Global Practice Group
June 2017

Abstract

Spatial analysis in economics is becoming increasingly important as more spatial data and innovative data mining technologies are developed. Even in Africa, where data often crucially lack quality analysis, a variety of spatial data have recently been developed, such as highly disaggregated crop production maps. Taking advantage of the historical event that rail operations were ceased in Ethiopia, this paper examines the relationship between agricultural production and transport connectivity, especially port accessibility, which is mainly characterized by rail transport. To deal with endogeneity of infrastructure placement and autocorrelation in spatial data, the spatial autocorrelation panel regression model is applied. It is found that agricultural production decreases with transport costs to the port: the elasticity is estimated at –0.094 to –0.143, depending on model specification. The estimated autocorrelation parameters also support the finding that although farmers in close locations share a certain common production pattern, external shocks, such as drought and flood, have spillover effects over neighboring areas.

This paper is a product of the Transport and ICT 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.

Spatial Autocorrelation Panel Regression: Agricultural Production and Transport Connectivity

Atsushi Iimi,[1 ¶] Liangzhi You,[2] Ulrike Wood-Sichra[2]

[1] Transport & ICT Global Practice, World Bank

[2] International Food Policy Research Institute (IFPRI)

Key words: Spatial autoregressive model; agriculture production; transport infrastructure.

JEL classification: H54; H41; Q12; C21; C34.

[¶] Corresponding author.

I. INTRODUCTION

Spatial data and techniques are becoming increasingly important in applied economics. A wide variety of new spatial data have recently been developed using innovative data mining techniques, such as high resolution satellite imagery and geo-referenced road network data. For instance, there exist several detailed global population distribution data, including the Gridded Population of the World and the WorldPop. The latter has the highest resolution of 100 meters and is continuously updated over time. OpenStreetMap locates many roads over the world. Such spatial data allow to develop a useful tool to identify potential economic opportunities and bottlenecks in developing countries (e.g., World Bank, 2009 and 2010).

In Africa, agriculture remains an important sector, generating \$31 billion or nearly half of the total GDP of the region (World Bank 2013). Africa has great potential according to global spatial data, such as the Global Agro-ecological Zones (GAEZ) system developed by the FAO and the International Institute for Applied Systems Analysis (IIASA). Theoretically, crop suitability based on climate and soil conditions is not necessarily low in Africa. However, the potential has not been fully explored yet. For instance, the ratios of potential to actual agricultural outputs are estimated at 1.5 for cassava, 1.9 for rice, 2.7 for maize and 5 for wheat in West Africa (World Bank 2012).

There are a number of constraints: Africa's agriculture is mostly small-scale subsistence farming. Irrigation and fertilizer are among the most important missing inputs (e.g., Gyimah-Brempong, 1987; Bravo-Ortega and Lederman, 2004; Xu et al., 2009; Dillon, 2011). Access to input and output markets is also important. In India, roads and fertilizer are found to be two important inputs to increase agricultural production (Binswanger, Khandker and Rosenzweig, 1993). In Africa, rural accessibility measured by the proportion of the rural residents within a 2-km walking distance from an all-weather road is less than 30 percent (Gwilliam 2011). Transport access has a crucial role to play to reduce input prices and increase agricultural production (e.g., Khandker, Bakht and Koolwal, 2009; Donaldson, 2010).

From the food security and regional integration points of view, the literature also shows that road quality and distance play a significant role in rationalizing agricultural commodity prices and advancing regional integration (Brenton, Portugal-Perez and Régolo, 2014). In Africa, crop price differentials decrease with the quality of roads, measured by the share of paved roads. Removing trade and transport barriers can improve market integration, and therefore, food security in the region (Hoering, 2013).

The current paper aims at examining the impact of transport access, especially rail connectivity, on agricultural production in Ethiopia. Ethiopia is a landlocked country and more than 95 percent of its total exports and imports depend on the Port of Djibouti, about 800 km away from the capital city, Addis Ababa. The country is traditionally an agricultural exporter, such as coffee, and an importer of a large amount of fertilizer and other inputs. As in other landlocked countries, port connectivity tends to be crucial to the economy. The cost of importing a 20-foot container of goods to Ethiopia is about US$2,960, which is much higher than Djibouti, which has a regional hub port (US$910). For Malawi, which is also landlocked, the importing cost is US$2,895, and it takes about 39 days. Both are unfavorably compared with a regional gateway country, Tanzania (US$1,615 and 26 days, respectively). The paper casts light on these connectivity issues.

The paper addresses methodological challenges, such as autocorrelation in spatial data and endogeneity of infrastructure placement, by applying the spatial autocorrelation regression model to spatial panel data in Ethiopia. With different types of spatial data combined, a Cobb-Douglas production function is estimated. In the literature, there are only a few studies that use spatial data to examine the relationship between transport infrastructure and agriculture development (e.g., Dorosh, Wang, You and Schmidt, 2012).

The remaining sections are organized as follows: Section II describes our spatial agriculture model. Section III develops an empirical model, and Section IV describes the data. Section V discusses the main estimation results and some policy implications. Then, Section VI concludes.

II. Spatial Production Allocation Model (SPAM) Update

The paper relies on the spatial production allocation model (SPAM) developed by the International Food Policy Research Institute (IFPRI) for generating geographically highly disaggregated crop production data. The SPAM is a spatial model to allocate crop production derived from large statistics reporting units, such as country, province and district, to a raster grid at a spatial resolution of 5 minutes of arc (approximately 9 km at the equator (normally referred to as 10km x10km pixel for simplicity). To infer likely production locations, the model uses a cross-entropy method (Shannon 1948). Given initial allocations, the cross-entropy method minimizes the cross-entropy distance—entropy is referred to as a measurement of uncertainty of expected information—between different probability distributions of the variables in the analysis, under different spatial constraints. As a result, the SPAM allows the simulation of the most plausible agricultural production locations given all available data at different levels.

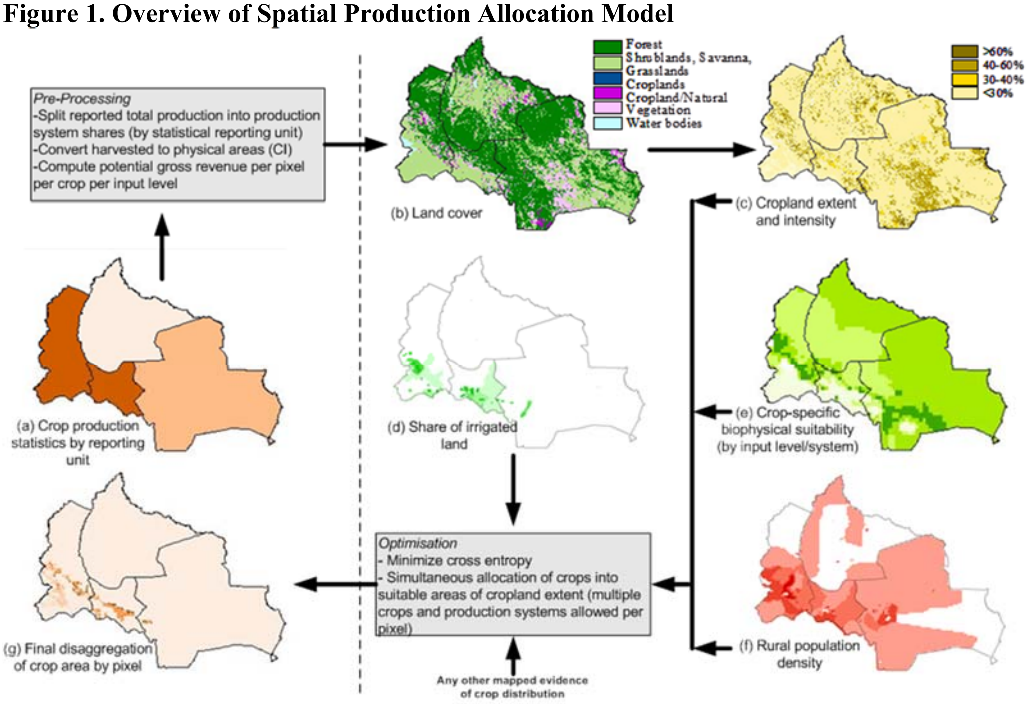
Various input data are taken into account in the cross-entropy procedure, such as sub-national crop production statistics, satellite data on land cover, maps of irrigated areas, biophysical crop suitability assessments, population density, secondary data on irrigation and rain-fed production systems, cropping intensity, and crop prices. Specifically in the SPAM, we start with crop production statistics at the large administrative (geopolitical) units (**Figure 1**).[1] These are typically national or sub-national. Key information to determine where and how much agricultural land exists at the pixel level comes from the existing land cover imagery, which is divided into crop land and non-crop land. With this crop land combined with the crop suitability data based on local climate, terrain and soil conditions, the crop-specific land

[1] See You and Wood (2006), You *et al.* (2009), and You et al. (2014) for further details.

areas can be obtained (e.g., maize in the figure). Together with all these data, the SPAM applies the cross entropy method to obtain the final estimation of each crop distribution.

For the current paper, the SPAM was calibrated for Ethiopia at two points of time: 2005 and 2010. The model produces detailed spatial data of crop production and land area for each of 42 crops. Note that the SPAM model disaggregates crop areas and yields into four different management intensities: (i) irrigated; (ii) high-input rain-fed; (iii) low-input rain-fed; and (iv) subsistence. From the SPAM, therefore, not only production but also inputs, such as harvested land, irrigated land and fertilizer used (under certain assumptions), can be generated at each 10km x 10km area.

Figure 1. Overview of Spatial Production Allocation Model



Source: Authors' illustration.

III. Empirical Model

Following the literature (e.g., Gyimah-Brempong, 1987; Bravo-Ortega and Lederman, 2004), a simple production function is considered with transport connectivity included as one of the production inputs:

$$\ln Y_{it} = \beta_0 + \sum_k \beta_k \ln X_{kit} + c_i + u_{it} \tag{1}$$

where y is the total value of crops produced in location i at time t. Four traditional production inputs and two types of transport connectivity are considered: $X_k = \{L, R, I, F, TC1, TC2\}$. The logarithms are taken for all dependent and independent variables.[2] L denotes labor. Land is divided into two types: rain-fed (R) and irrigated (I). F is measured by the total amount of fertilizer. These four variables are common agricultural inputs that are examined in the literature. In Africa, the output elasticity tends to be high for labor, reflecting the fact that agriculture production is labor-intensive. The land elasticity is often modest, reflecting the relative abundance of land in Africa. Fertilizer and irrigation seem to be critical to improve production, but the statistical significance varies across studies (Bravo-Ortega and Lederman, 2004). A growing literature suggests their importance: In Zambia, timely availability of fertilizer could increase maize yields by 11 percent on average (Xu et al., 2009). Improved availability of irrigation could nearly double agricultural productivity in Mali (Dillon, 2011).

The current paper is focused on the possible impacts of transport connectivity, for which two types of variables are examined: (i) Domestic market accessibility measured by transport costs to bring one unit of good to a major city with more than 200,000 population (*TC1*), and (ii) port accessibility, which is measured by unit transport costs to the Port of Djibouti (*TC2*). Transport infrastructure has multiple implications to agricultural production. Better market access can reduce input prices. Khandker, Bakht and Koolwal (2009) find that farm-gate fertilizer prices were lowered by rural road investment in Bangladesh. Better transport infrastructure also provided more opportunities for farmers to engage in cash crop production and market transactions. Agriculture output prices increased by 2 percent and the volume of production was boosted by 22 percent.

---

[2] A small positive number is added if the amount of input used is zero to avoid taking the logarithm of zero. For instance, irrigated land is not used in many observations in our sample.

To estimate Equation (1), there are several empirical issues. First, one of the most important issues is endogeneity of infrastructure placement. While transport investment could generally raise productivity of the economy, transport infrastructure has already been placed where economic productivity is inherently high. In our case, for instance, land productivity may not perfectly be unobserved by the econometrician. Thus, regardless of its real impact, there tends to be positive correlation between agricultural production and transport measurement (*TC1* and *TC2*). Under the additivity assumption, such location-specific unobservables are included as c_i in the equation.

This is a general empirical issue to estimate an unbiased impact of large-scale transport infrastructure. The literature uses instrumental variables based on historical events in the past (e.g., Banerjee et al. 2012; Datta, 2012; Jedwab and Moradi, 2012) and applies panel regression to remove unobserved characteristics (Khandker and Koolwal, 2011). The current paper takes advantage of the latter approach. Taking the difference on both sides of Equation (1), the unobserved time-invariant location-specific effects are eliminated:

$$\Delta \ln Y_i = \beta_0 + \sum_k \beta_k \Delta \ln X_{ki} + \Delta u_i \tag{2}$$

Still, there remains another concern, which is autocorrelation in the error term u. Our primary source of data is the SPAM, which generates spatial data at the approximately10 x 10 km land area level. Thus, even if the difference variables are taken, it is critical to deal with the possible spatial autocorrelation, i.e., $Cov(\Delta u_i, \Delta u_j) \neq 0$. By nature, agricultural land is a continuum of various characteristics, such as soil fertility and water availability. Weather conditions are continuous across near locations. Public infrastructure, such as railway and roads, typically forms a network, which also creates autocorrelation among neighboring areas.

To deal with this problem, the spatial autocorrelation structure is taken into account:

$$\Delta \ln Y_i = \lambda \sum_j w_{ij} \ln Y_j + \sum_k \beta_k \Delta \ln X_{ik} + \Delta u_i \tag{3}$$

$$\Delta u_i = \rho \sum_j w_{ij} \Delta u_j + \varepsilon_i \tag{4}$$

where w is an element of the spatial-weighting matrix. λ and ρ are spatial autoregressive parameters in the dependent variable and error term, respectively. ε is an idiosyncratic error distributed independently and identically. Under the normality assumption, this can be estimated by the maximum likelihood estimation procedure (e.g., Anselin, 1988; Amaral and Anselin, 2011).[3]

For the spatial weighting matrix, inverse distances between two locations i and j are used. The distance is calculated using the Euclidean distance between the two locations. The intuition is that two locations are more closely related to each other, if they are located closely. This follows Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things (Tobler 1970)."

IV. DATA

The summary statistics are shown in **Table 1**. Our primary data source is the SPAM, which is, as discussed above, a model to disaggregate the national and subnational agricultural production and land data into the 10km x 10km pixel level. The model generates data for each of 42 crops, which are aggregated with regional production prices used from FAOSTAT (**Table 2**). Thus, our dependent variable, Y_{it}, is the total value of crops produced at each location i at time t (**Figure 2**).

---

[3] For estimation we applied a STATA command ***spreg*** developed by Drukker, Prucha and Raciborski (2013).

Table 1. Summary statistics

Variable	Abb.	Obs	Mean	Std. Dev.	Min	Max
Total value of crops produced ($ million)	*Y*	8,893	1.51	1.87	0.00	23.65
Population estimated (1,000) [1]	*L*	9,196	11.59	20.22	0.02	1177.36
Rain-fed land area harvested (1,000 ha) [1]	*R*	8,930	2.90	2.57	0.00	16.95
Irrigated land area harvested (1,000 ha) [1]	*I*	1,186	0.31	0.58	0.00	6.49
Fertilizer used (1,000 kg) [1]	*F*	4,731	50.21	80.73	0.00	1017.10
Transport cost to city with pop. >200,000	*TC1*	9,248	105.44	43.20	4.71	199.29
Transport cost to Djibouti	*TC2*	9,248	32.56	20.95	0.08	107.93
Transport cost to city with pop. >100,000	*TC1a*	9,248	21.14	13.66	0.08	95.91

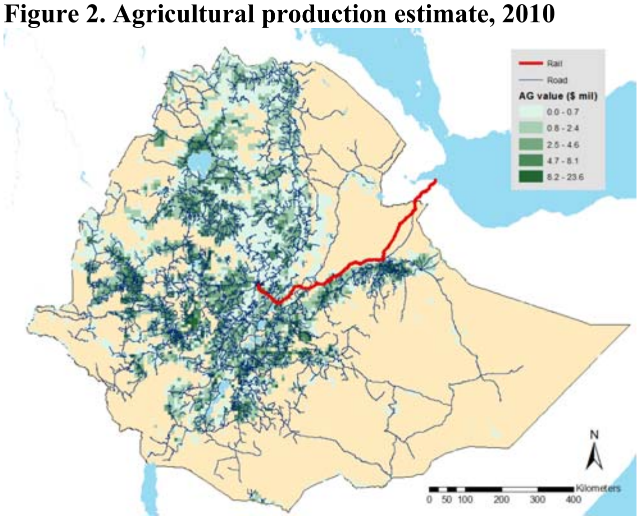
[1] Multiplied by 1000 for presentation purposes.

Table 2. Regional crop prices of selected commodities

Food crop	$/ton	Export crop	$/ton
Maize	260	Coffee	3,192
Rice	856	Tea	1,506
Bananas	536	Cotton	1,429
Plantain	240	Tobacco	1,365
Cassava	294		

Source: Calculated based on FAOSTAT.

Figure 2. Agricultural production estimate, 2010



Source: SPAM

For labor, population engaged at location *i* is measured using the global population distribution data set. The paper takes advantage of the Gridded Population of the World (GPW) data, which estimates the 2005 and 2010 populations for Ethiopia. The data are available at a 2.5 minute resolution, approximately 5km at the equator. Other global population data sets, such as WorldPop, may have higher resolution estimates. The GPW has estimates for both 2005 and 2010, which are needed to match our SPAM panel data.

Land data also come from the SPAM. It provides an estimate of land area used for each crop at each location. Cultivated land is divided into two types: rain-fed (*R*) and irrigated (*I*). Most of the agricultural land is currently rain-fed with little irrigation applied in Ethiopia. For each location, land areas are aggregated across crops.

F represents the quantity of plant nutrients applied to all crops produced at each location. This is calculated based on land area (ha) where fertilizer is used and the national average fertilizer consumption (kg per ha). The size of land areas with fertilizer applied is derived from SPAM, in which land areas under high-input and irrigated production systems are available. The average fertilizer consumption is calculated based on Ethiopia's Agricultural Sample Survey for 2012/13. Our fertilizer variable is the sum of DAP and Urea, although the actual combination may vary across fertilizers (**Table 3**).

Table 3. Average fertilizer use

Crop	kg/ha	Crop	kg/ha
Wheat	111.9	Chickpea	26.5
Maize	100.3	Lentil	59.6
Barley	75.7	Other pulses	48.0
Small millet	74.3	Soybean	52.4
Sorghum	41.7	Groundnut	63.7
Other cereals	89.9	Rapeseed	121.2
Potato	82.7	Sesame seed	44.3
Sweet potato	23.4	Other oil crops	34.8
Other roots	0.3	Vegetables	100.0
Bean	64.3		

Note that only land areas with fertilizer applied are considered.
Source: Authors' estimates based on Ethiopia Agricultural Sample Survey for 2012/13.

The transport connectivity is measured by the unit transport costs from each location to either a nearest large market, which is defined by a city with more than 200,000 population (denoted by *TC1*), or the Port of Djibouti, a major regional gateway port for Ethiopia (denoted by *TC2*). Transport costs are estimated based on underlying road user costs and rail tariff along the optimal route by minimizing the total cost from an origin to a destination. Since road transport is a major mode for relatively short-haul movement of goods, *TC1* is basically the minimum road user costs of bringing one ton of goods to a large city (US$/ton).

By contrast, port connectivity is influenced more by railways than road transport in Ethiopia. In general, rail transport has comparative advantage for long-haul freight transportation. The freight rates used to be 0.25 to 0.31 Ethiopian birrs (ETB) or 2.9 to 3.5 U.S. cents per ton-km, much lower than typical vehicle operating costs (VOCs) on roads in Ethiopia. Based on the traditional highway engineering model (HDM4), the VOCs vary from 8 to 10 U.S. cents per ton-km, depending on road conditions.[4]

Ethiopia depends on the Port of Djibouti, about 800 km away from the capital city, Addis Ababa, for more than 95 percent of the country's total exports and imports. The country is a traditional agricultural exporter, such as coffee, and largely relies on imports for fertilizer and other inputs. Port connectivity is considered to be critical to determine agricultural productivity, especially for landlocked countries, such as Ethiopia. In Ethiopia, transport costs account for 64-80 percent of fertilizer farmgate prices (Rashid et al. 2013).

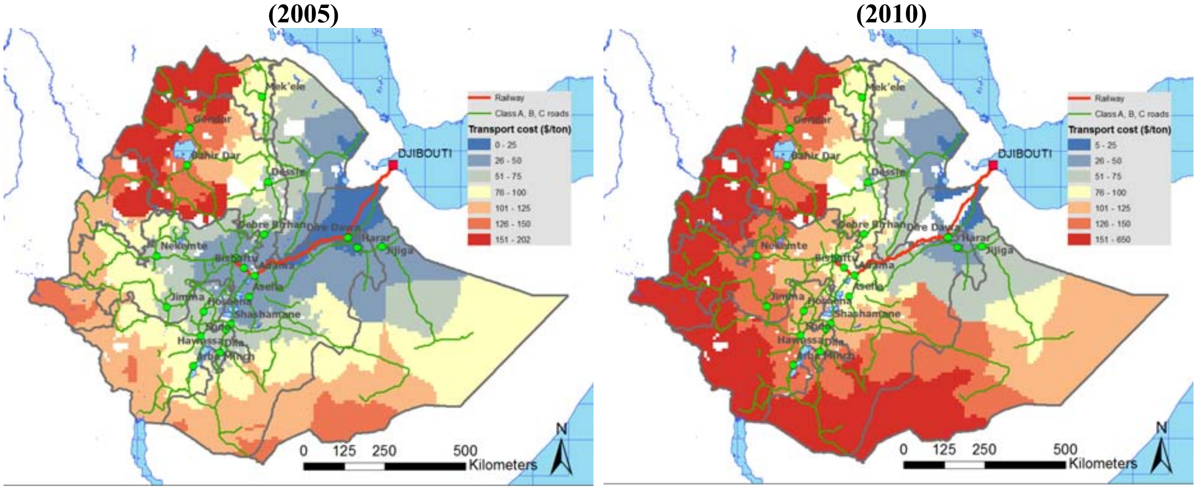
The current paper takes advantage of a historical change in port connectivity between 2005 and 2010. Since its completion in the early 20th century, the Ethio-Djibouti Railways has been a major transport means for Ethiopia to access the global market. Until the 1990s, it carried more than 100 million ton-km of freight. By 2007, however, the rail operations were ceased between Addis Ababa and Dire Dawa, mainly because of insufficiency of financial resources and resultant lack of maintenance. The level of rail services deteriorated further. By 2009, the whole rail line ceased operating. As a result, transport costs to the port (*TC2*)

[4] In theory, with free market entry, the market prices should converge on transport costs, i.e., VOCs. In practice, however, these may not be the same because of the poor quality of the road network and the lack of competition in the trucking industry. Our companion paper, Iimi et al. (2017), uses an alternative transport cost variable, which is based on adjusted VOCs with a 60 percent markup taken into account. The results indicate that VOCs are a good proxy of market transport prices.

are considered to have changed significantly (**Figure 3**). This allows us to identify the impact of rail transport connectivity on agricultural production.

Figure 3. Estimated transport costs to the Port of Djibouti

(2005) **(2010)**



Source: Authors' estimation.

V. Estimation Results and Policy Implications

First of all, the ordinary least squares (OLS) regression is performed separately at each of the two points of time: 2005 and 2010 (**Table 4**). As discussed above, the results are likely to be biased because of possible autocorrelation in spatial data and unobserved location-specific fixed-effects, which may cause endogeneity associated with the transport variables. It is not surprising that some of the estimated coefficients are inconsistent with economic theory. The result indicates that irrigation would be unproductive with the data for 2005. In addition, the coefficient of port accessibility (*TC2*) is negative for 2005 but positive for 2010. All the results indicate that the OLS estimation is not appropriate to estimate our empirical model with our data.

Table 4. OLS regression results

	t=2005			t=2010		
	Coef.	Std.Err.		Coef.	Std.Err.	
ln*L*	0.244	(0.027)	***	0.364	(0.010)	***
ln*R*	0.471	(0.005)	***	0.453	(0.008)	***
ln*I*	-0.011	(0.012)		0.042	(0.004)	***
ln*F*	0.030	(0.011)	***	0.066	(0.004)	***
ln*TC1*	0.130	(0.026)	***	0.131	(0.031)	***
ln*TC2*	-0.259	(0.038)	***	0.090	(0.083)	
constant	-5.233	(0.347)	***	-7.733	(0.399)	***
Obs	4624			4624		
F statistic	3752.5			1634.8		
R-squared	0.829			0.800		

Note: The dependent variable is ln*Y*. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

To mitigate the endogeneity problem, the OLS model with the difference variables is estimated, which is empirically the same as the fixed-effects model with panel data (**Table 5**). Thus, the unobserved location-specific effects are controlled for. Still, autocorrelation remains unsolved. But all the coefficients turned out consistent with our expectation, but one, the effect of domestic market access (*TC1*). All the crop production inputs are productive. The port access (*TC2*) has a negative coefficient as expected: Agricultural production would be greater where transport costs to the port are lower or port accessibility is high.

With not only the location-specific fixed effects but also autocorrelation taken into account, the spatial autocorrelation regression model is estimated. The results are broadly consistent with *a priori* expectations. The estimated autocorrelation parameters, i.e., λ and ρ, indicate that both cross-sectional OLS and fixed-effect models are likely biased. The hypothesis that the spatial autocorrelation parameters are zero can easily be rejected.

Moreover, the estimated spatial autoregressive term λ, which is small but significantly different from zero, shows that spatial concentration is positive. It means that if agricultural production takes place in one location, its neighboring places are also likely to grow a similar amount of agricultural produce. This is a natural and consistent result with **Figure 2**. The autocorrelation coefficient in the errors, ρ, is also significantly positive. This can be interpreted to mean that an exogenous shock—for example, drought and flood—in a given location has a substantial spillover effect on its neighboring places. The magnitude of the coefficient is much larger than the autoregressive term λ, suggesting the practical significance of external shocks affecting agricultural production in a wide range of areas.

Regarding the impacts of transport accessibility, port accessibility has a significant impact on crop production: The elasticity is estimated at -0.143, meaning that a 10 percent reduction in transport costs to the port would increase agricultural production by 1.43 percent, which looks like a relatively modest effect. This looks broadly consistent with our companion paper (Iimi et al. 2017), which estimates the same impact of port accessibility with micro household data in Ethiopia: The elasticity is estimated at 0.276.

On the other hand, the impact of domestic market access remains inconclusive. Local market connectivity ($\Delta TC1$) still has a positive coefficient, meaning that agriculture production is higher where transport costs are high or connectivity is low. But the coefficient is not statistically significant in the autocorrelation model. This may not be surprising because local transport connectivity is less relevant to ceased rail operations, which are used as our main identification strategy in the estimation: The changes in transport costs to a near large city were too small to affect agricultural production.

Regarding other production inputs, our estimation results suggest that labor and land are two important inputs in Ethiopian agricultural production. In the spatial regression, the elasticities are estimated at 0.35 and 0.40 for labor and land, respectively. These relatively large coefficients may reflect the fact that agricultural production in Ethiopia remains labor- and land-intensive. Irrigation and fertilizer are surely productive, as the literature suggests: The coefficients are both positive and significant. But the magnitude of the elasticities remains small because the use of fertilizer and irrigation is limited in our data.

Table 5. Fixed effect and spatial autocorrelation regression results

	OLS (difference)			Spatial regression		
	Coef.	Std.Err.		Coef.	Std.Err.	
$\Delta \ln L$	0.369	(0.015)	***	0.347	(0.007)	***
$\Delta \ln R$	0.405	(0.005)	***	0.403	(0.003)	***
$\Delta \ln I$	0.093	(0.016)	***	0.071	(0.007)	***
$\Delta \ln F$	0.013	(0.002)	***	0.010	(0.001)	***
$\Delta \ln TC1$	0.657	(0.167)	***	0.285	(0.238)	
$\Delta \ln TC2$	-0.318	(0.051)	***	-0.143	(0.039)	***
constant	0.030	(0.061)		-0.303	(0.056)	***
Obs	4624			4624		
F statistic	1220.5					
Wald2				18480.4		
R-squared	0.820					
Spatial parameters:						
λ				0.078	(0.010)	***
ρ				0.305	(0.004)	***

Note: The dependent variable is $\Delta \ln Y$. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

One unexpected result is that the elasticity of labor is high. Ethiopian agricultural production is labor-intensive, but there is a common view that the labor force in the agriculture is too abundant to be productive. The marginal productivity is generally considered low. In Ethiopia, 73 percent of the total labor force is absorbed by agricultural activities.[5] Particularly in rural areas, agricultural employment is dominant at 83 percent, compared to 13.5 percent in urban areas. The estimated high elasticity may have captured some impacts of local consumption of agricultural produce, because our labor measurement comes from the global population distribution data, WorldPop, which includes non-agricultural labor force as well. For obvious reasons, holding everything else constant, agricultural production is larger where more people live.

To examine this effect, the labor variable is adjusted using the region-level agricultural employment rates in 2005 and 2013. The 2005 survey data are applied to our 2005 data, and

[5] According to the recent government statistics: Central Statistical Agency, Ethiopia. (2014). Statistical Report on the 2013 National Labour Force Survey.

(continued)

the 2013 data are used for our 2010 data. In addition, urban and rural areas are differentiated by global spatial urban-rural data, the Global Rural-Urban Mapping Project (GRUMP) data set.[6] Thus, the systematic difference in agricultural employment share across regions and between urban and rural areas is taken into account. Using these data, population data are divided into two: Agricultural labor force, L_{AG}, and other population, L_{NON}.

The results are shown in **Table 6**: While the main results remain unchanged, the coefficient of agriculture-specific labor L_{AG} is much smaller. As expected, L_{NON}, which is expected to be a proxy of local non-farmer consumption, also has a significantly positive coefficient. This supports the view that the original global population data may not be an accurate measurement of the agricultural labor force. Still, there is a risk of measurement errors and bias. Unfortunately, there are no detailed spatial data to distinguish agricultural and non-agricultural labor force at the more granular level.

Table 6. Fixed effect and spatial autocorrelation regression with agriculture labor force used

	OLS (difference)			Spatial regression		
	Coef.	Std.Err.		Coef.	Std.Err.	
$\Delta \ln L_{AG}$	0.363	(0.043)	***	0.227	(0.043)	***
$\Delta \ln L_{NON}$	0.012	(0.043)		0.132	(0.044)	***
$\Delta \ln R$	0.406	(0.006)	***	0.403	(0.003)	***
$\Delta \ln I$	0.094	(0.016)	***	0.070	(0.007)	***
$\Delta \ln F$	0.013	(0.002)	***	0.010	(0.001)	***
$\Delta \ln TC1$	0.642	(0.166)	***	0.243	(0.238)	
$\Delta \ln TC2$	-0.261	(0.051)	***	-0.096	(0.041)	**
constant	0.017	(0.060)		-0.347	(0.057)	***
Obs	4624			4624		
F statistic	1055.7					
Wald2				18409.1		
R-squared	0.820					
Spatial parameters:						
λ				0.079	(0.010)	***
ρ				0.303	(0.004)	***

Note: The dependent variable is $\Delta \ln Y$. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

[6] SPAM cells are considered as urban areas, when their centroids are within 20 km of the urban areas defined by GRUMP.

In addition, one may be concerned about the robustness of the above results, especially, the positive coefficient of domestic market accessibility (*TC1*). Two different definitions of "market" are examined. First, the market is defined by cities with more than 100,000 population (*TC1a*). Nineteen locations are identified in our data. Second, the market is defined by the regional or zonal capitals. This comprises 59 cities, some of which are relatively small in terms of population (*TC1b*).

The results remain broadly unchanged (**Table 7**). Under the spatial autocorrelation framework, the coefficients of local market, either *TC1a* or *TC1b*, are still positive but statistically not significant. As discussed above, one possible reason is that our identification strategy may be weak to measure the possible impact of local market connectivity on agricultural production. Another possibility is that the result may reflect the fact that agriculture production takes place in rural areas: Agriculture production is higher where it is far from major cities.

Table 7. Robustness check with different definitions of "domestic market"

	Spatial regression			Spatial regression		
	Coef.	Std.Err.		Coef.	Std.Err.	
$\Delta\ln L_{AG}$	0.224	(0.043)	***	0.226	(0.043)	***
$\Delta\ln L_{NON}$	0.134	(0.044)	***	0.132	(0.044)	***
$\Delta\ln R$	0.403	(0.003)	***	0.403	(0.003)	***
$\Delta\ln I$	0.070	(0.007)	***	0.070	(0.007)	***
$\Delta\ln F$	0.010	(0.001)	***	0.010	(0.001)	***
$\Delta\ln TC1a$	0.345	(0.231)				
$\Delta\ln TC1b$				0.227	(0.181)	
$\Delta\ln TC2$	-0.097	(0.041)	**	-0.097	(0.041)	**
constant	-0.350	(0.057)	***	-0.346	(0.057)	***
Obs	4624			4624		
F statistic						
Wald2	18413.3			18478.4		
R-squared						
Spatial parameters:						
λ	0.080	(0.010)	***	0.079	(0.010)	***
ρ	0.303	(0.004)	***	0.303	(0.004)	***

Note: The dependent variable is $\Delta\ln Y$. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

VI. CONCLUSION

Africa has great potential for agriculture from the agro-ecological point of view. However, the potential has not been fully explored yet. One of the important constraints is transport connectivity. In the case of Ethiopia, access to ports as well as markets is crucial. Many farmers do not have access to a market: The latest Rural Access Index is estimated at 21.6 percent, leaving about 63 million rural people unconnected to the road network (World Bank, 2016). Port connectivity is also crucial because Ethiopia is a landlocked country. Significant transport and trade costs are incurred to the economy.

Spatial data and analyses are becoming increasingly important to identify economic opportunities and reveal possible constraints in a geographic manner. There are only a few studies that analyze the relationship between agricultural production and transport infrastructure with spatial data. Taking advantage of a historical event that rail operations were ceased in Ethiopia in the late 2000s, the paper applied the spatial autocorrelation panel regression to estimate an unbiased impact of port connectivity, which is mainly characterized by rail transportation in the case of Ethiopia, on agricultural production. Methodologically, the paper's approach has the advantage of addressing the endogeneity of infrastructure placement as well as autocorrelation in spatial data.

It is found that agricultural production increases with port connectivity or decreases with transport costs to the port: The elasticity is estimated at -0.094 to -0.143, depending on model specification. The results are largely consistent with other available evidence. Among conventional agricultural production inputs, labor and land are found to have relatively high elasticities, which represent the fact that the current crop production is based on extensification of inputs, rather than intensification. Fertilizer and irrigation are productive, but their impacts remain limited in Ethiopia.

Finally, the paper confirms that the traditional OLS or even fixed-effects models tend to be biased without spatial autocorrelation and infrastructure endogeneity taken into account. Spatial autocorrelation matters. The estimated autocorrelation parameters indicate that the autoregressive coefficient is about 0.078, and the autocorrelation coefficient in the errors is about 0.305. This can be interpreted to mean that farmers in close locations share a certain common production pattern, while external shocks, such as drought and flood, have spillover effects over neighboring areas.

REFERENCES

Amaral, Pedro, and Luc Anselin. 2011. Finite sample properties of Moran's *I* test for spatial autocorrelation in Tobit models. Working Paper No. 07, GeoDa Center for Geospatial Analysis and Computation, Arizona State University.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* 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.

Binswanger H, Khandker S and Rosenzweig. 1993. How infrastructure and financial institutions affect agricultural output and investment in India. *Journal of Development Economics*, Vol. 41, pp. 337-336.

Bravo-Ortega, and Lederman. 2004. Agricultural productivity and its determinants: Revisiting international experiences. *Estudios de Economia*, Vol. 31(2), pp. 133-163.

Brenton, Paul, Alberto Portugal-Perez, and Julie Regolo. 2014. Food prices, road infrastructure, and market integration in Central and Eastern Africa. Policy Research Working Paper No. 7003, World Bank.

Datta, Saugato. 2012. The impact of improved highways on Indian firms. *Journal of Development Economics*, Vol. 99(1), pp. 46-57.

Dillon, Andrew. 2011. Do differences in the scale of irrigation projects generate different impacts on poverty and production? *Journal of Agricultural Economics*, Vol. 62(2), pp. 474-492.

Donaldson, Dave. 2010. Railroads and the Raj: The economic impact of transportation infrastructure. NBER Working Paper No. 16487.

Dorosh, Wang, You, and Schmidt. 2012. Road connectivity, population, and crop production in Sub-Saharan Africa. *Agricultural Economics*, Vol. 43, pp. 89-103

Drukker, David, Ingmar Prucha, and Rafal Raciborski. 2013. Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *The Stata Journal*, Vol. 13(2), pp. 221-241.

Gwilliam, Ken. 2011. *Africa's Transport Infrastructure: Mainstreaming Maintenance and Management.* The World Bank.

Gyimah-Brempong. 1987. Scale elasticities in Ghanaian cocoa production. *Applied Economics*, Vol. 19, pp. 1383-1390.

Hoering, Uwe. 2013. Alternatives to food import dependency. FDCL Policy Paper, Centre for Research and Documentation Chile-Latin America.

Iimi, Atsushi, Haileyesus Adamtei, James Markland, and Eyasu Tsehaye. 2017. Port rail connectivity and agricultural production: Evidence from a large-sample of farmers in Ethiopia. Unpublished.

Jedwab, R. and A. Moradi. 2012. Colonial investments and long-term development in Africa: Evidence from Ghanaian railroads, unpublished paper, George Washington University; STICERD, London School of Economics; and University of Sussex.

Khandker, Shahidur, and Gayatri Koolwal. 2011. Estimating the long-term impacts of rural roads: A dynamic panel approach, Policy Research Working Paper No. 5867, The World Bank.

Khandker, Shahidur, Zaid Bakht, and Gayatri Koolwal. 2009. The poverty impact of rural roads: Evidence from Bangladesh. *Economic Development and Cultural Change*, Vol. 57(4), pp. 685-722.

Rashid, Shahidur, Nigussie Tefera, Nicholas Minot, Gezahengn Ayele. 2013. Fertilizer in Ethiopia. IFPRI Discussion Paper No. 01304, International Food Policy Research Institute.

Shannon, Claude. 1948. A mathematical theory of communication, Bell System Technology Journal, Vol. 27, pp. 379-423.

Tobler Waldo. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, Vol. 46(2), pp. 234-240.

Xu, Zhiying, Zhengfei Guan, T.S. Jayne, Roy Black. 2009. Factors influencing the profitability of fertilizer use on maize in Zambia. *Agricultural Economics*, Vol. 40, pp. 437-446.

You, L. and S. Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems, Vol. 90(1-3), pp. 329-347.

You, L., S. Wood, U. Wood-Sichra. 2009. Generating plausible crop distribution and performance maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. Agricultural System, Vol. 99(2-3), pp. 126-140.

You, L., S. Wood, U. Wood-Sichra, W. Wu. 2014. Generating global crop distribution maps: From census to grid. Agricultural System, forthcoming.

World Bank. 2009. *World Development Report 2009: Reshaping Economic Geography.* World Bank.

World Bank. 2010. *Africa's Infrastructure: A Time for Transformation.* World Bank.

World Bank. 2012. *Africa Can Help Feed Africa: Removing Barriers to Regional Trade in Food Staples.*

World Bank. 2013. *Growing Africa: Unlocking the Potential of Agribusiness*.

World Bank. 2016. *Measuring Rural Access: Using New Technologies.*