SPATIAL INEQUALITY: OVERCOMING NEIGHBORHOOD EFFECTS IN AFRICA

NGELEZA, Guyslain K.*
FLORAX, Raymond J.G.M.
MASTERS, William A

Abstract
Spatial inequality in global economic development has left Africa with the least progress in improving living standards among developing regions of the world. Moreover, there are strong neighborhood effects within Africa. This paper revisits the explanation of unequal growth across countries in an African context. We argue that some of the lingering disagreements over the channels through which institutions and geography may explain differences in income per capita across countries could be resolved by accounting for neighborhood effects often overlooked in past analyses. Through simultaneous equations we test how trade, urbanization, and agricultural productivity are affected by a country's policies and factor endowments and the degree to which each aspect of economic development is affected by spillovers from neighboring countries. We use both limited and full information estimators, based partly on a generalized moments estimator for spatial autoregressive coefficients, which allow for spatial error correlation, correlation across equations, and the presence of spatially lagged dependent variables. With this specification, after controlling for spatial proximity, both institutions and geography variables exert, through trade and urbanization, an independent effect on income.

Keywords: Agriculture, economic growth, geography, institutions, simultaneous equations, spatial econometrics.
JEL codes: C31, C33, I18, O13, R12

1. Introduction

Starting with the "Great Divergence," Africa remains the world's poorest continent.[1] In 2000, the African continent was eight times the level for developed countries (Sala-i-Martin 2006). However, the overall statistics obscure the strong *neighborhood effects* within Africa, including spatial heterogeneity and dependency (Anselin et al. 1996). Income data indicate geographic clustering into subregions of developing and least developed countries. Figure 1.1 shows estimated growth profiles for African subregions between 1960 and 2005: in 1960, income per capita for (least developed) western, central,

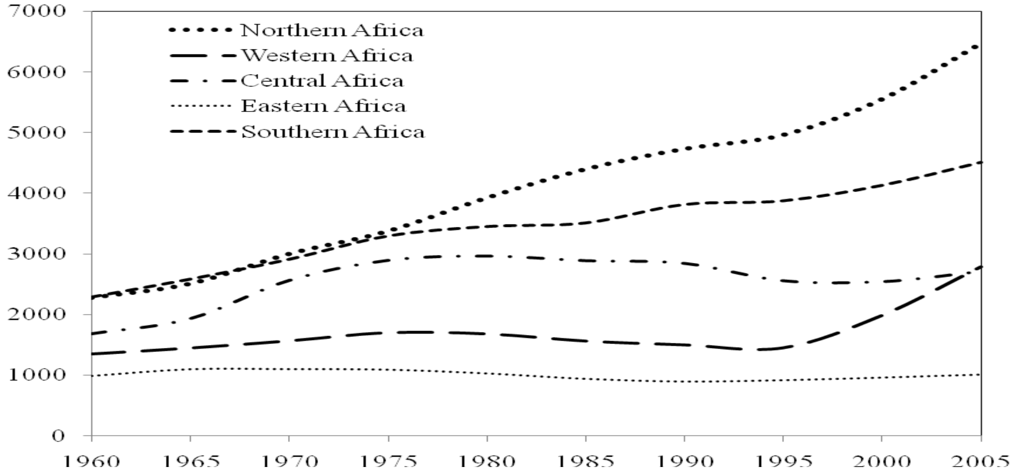
---

* Guyslain K. Ngeleza, International Food Policy Research Institute, 2033 K Street, NW, Washington, DC 20006-1002, USA, Email: g.ngeleza@cgiar.org, Raymond J. G. M. Florax,[2] Department of Agricultural Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907–2056, USA, Email: rflorax@purdue.edu, Department of Spatial Economics, VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: rflorax@purdue.edu, and William A. Masters, Friedman School of Nutrition Science and Policy, Tufts University, 150 Harrison Ave. Boston, MA 02111, USA, Email: William.Masters@tufts.edu

[1] "Great Divergence" refers to the period beginning in the eighteenth century in which western Europe (namely Britain, followed closely by the rest of western Europe) clearly emerged as economically and politically the most powerful region of the world.

and eastern Africa was approximately two-third the level in the (developing) northern and southern African regions. By 2005, the income gap had widened, with the least developed regions at around one-third of the northern African subregion and one-half of the southern African subregion.

Figure 1.1—Per capita income distribution in Africa (in Purchasing Power Parity [PPP] terms, in 2000 U.S. dollars)



Source: Constructed by the authors, using 2010 Summers and Heston dataset

What accounts for this region effect? Following North (1990) and others, Acemoglu et al. (2001) argue that key factors determining income and well-being are *economic and political institutions*, including property rights, market infrastructure, and price incentives as drivers of investment and economic growth. An alternative view, pioneered by Diamond (1997) and Sachs (2001), uses geographic and technological factors to explain income differences—and possibly differences in economic institutions. Bleaney and Dimico (2010) demonstrate that geographic factors influence per capita income trends directly, as well as indirectly through the quality of institutions.

A number of factors have been identified contributing to the African continent's poor economic performance, including external conditions, social conditions, trade specialization, and low urbanization. External conditions include the legacy of slave trading, colonial rule, and political effects of the Cold War; in addition, ethnic divisions, religious diversity, and low levels of social capital all contribute to poor institutions (Bloom and Sachs 1998; Easterly and Levine 1997; Cinyabuguma and Putterman 2011).

Trade and urbanization also affect income levels and are strongly influenced by geographic conditions. Trade affects income growth through various channels: specialization according to comparative advantage; the exploitation of increasing returns from larger markets; exchange of ideas through communication and travel; and the spread of technology through investment and exposure to new goods. In Africa, heavy dependence on a small number of primary exports under volatile terms of trade limits the contribution of trade to economic development. Trade is also determined by geographic factors (Tinbergen 1962; Anderson [1979] 2000).

Cities are generally associated with development (Bradshaw and Noonan 1997) and urbanization is often associated with the level of country income (Fay and Opal 2000). The level of urbanization in Africa is low but growing rapidly, from an average 16 percent in 1960 to 35 percent in 2000.

Despite advances in understanding the channels through which trade and urbanization affect income levels, finding the best estimates and testing for how institutions, trade, and urbanization interact and together affect income per capita remains a challenging task. A principal difficulty is the problem of endogeneity: Any observed correlation between institutions and income could be due to reverse causality, or to other, omitted variables that affect both of them. Some widely cited studies use a system of equations as a way of account for reverse causality: Rodrik, Subramanian, and Trebbi 2004; Easterly and Levine 2003; and Acemoglu, Johnson, and Robinson 2001. These studies find that geographic location is correlated with income *only* through its influence on institutions, with no additional effect. Other researchers, however, using other variables and specifications, have arrived at different results. For example, Gundlach (2004) finds a large and robust influence on income of the degree of malaria transmission, a factor that is independent of a country's institutions but highly correlated with geography. Spatial correlation is a third important issue in testing the effects of institutions, urbanization, and trade on income per capita.

This issue has not been addressed sufficiently in the literature, despite the obvious geographic clusters of rich countries and poor ones. Geographic clustering could be due to spatially correlated attributes, such as climate or access to transport, or to demonstration effects, or to interactions among neighbors, as through trade or migration (see, for example, Magrini 2004; Abreu, de Groot, and Florax 2005b). Although data for some geographic attributes and for some interactions among countries are available, there are inevitably omitted variables that could account for geographic correlations, similar to the synchronized growth fluctuations in Latin America and elsewhere documented by Temple (1999).

This paper builds on the spatial estimator developed by Kelejian and Prucha (2004) to control for very general kinds of neighborhood and spatial spillover effects, while allowing for endogeneity of key regressors. This new estimator has been used to model, for instance, spatial effects in farmland values by Livanis et al. (2006); it serves to raise the bar for each hypothesis by testing against a wider range of alternative processes.

The remainder of this paper is organized as follows. Section 2 situates the study in the literature and introduces a particular simultaneous equation model to accounting for several possible channels by which institutions and technology could influence income. Section 3 presents the data and provides an exploratory empirical assessment of the dynamics over space and time of the key variables in the system. In Section 4 we presents the econometric method. In Section 5, we compares the estimation results of a nonspatial version of the system to the results based on the Kelejian and Prucha (2004) estimator, allowing for spatial error autocorrelation, correlation across equations, and the presence of spatially lagged dependent variables. Section 6 presents conclusions.

2. Determinants of Income across Countries: From Theory to Empirics

The neoclassical representation of economic growth assumes the following production function:

$$Y = \tilde{F}\left(K, \tilde{N}, A\right), \quad (1)$$

where the effective labor is $\tilde{N} = NH$, (with N representing the labor input and H the stock of human capital), K is a vector of physical capital, and A the scalar state of technology. In this production function, only physical capital is accumulated. Population growth and technical change are exogenous.

In the neoclassical framework, capital accumulation may occur following two paths. One possibility is that savings is exogenous—assumed to be a constant fraction of income $s \in (0,1)$ (Swan 1956; Solow 1957). In this case, the dynamic of capital accumulation is

$$\dot{k}(t) = sf\left(k(t)\right) - (n + g + \delta)k(t), \quad (2)$$

where n is the rate of growth of the population, g is the economic growth rate, δ is the rate of depreciation, $sf\left(k(t)\right)$ is the investment unit of effective labor, $f\left(k(t)\right)$ is the output per unit of effective labor, and $(n + g + \delta)k(t)$ is the break-even investment, that is, the amount of investment necessary to keep the stock capital at its existing level.

The other possibility is that the economy-wide savings is made endogenous and can be determined by the following optimization problem (Cass 1965; Koopmans 1965):

$$\max_{\{c(t), K(t)\}_{t \geq 0}} N(0)\int_0^\infty U\left(c(t)\right)e^{-(\rho - n)t}dt, \qquad \rho > n + \xi \geq 0$$

$$\textit{subject to} \qquad (3)$$

$$\dot{K}(t) = Y(t) - c(t)N(t) - \delta K(t)$$

$$U(c) = \frac{c^{1-\theta} - 1}{1-\theta}, \quad \theta > 0, \qquad Y = \tilde{F}\left(K, \tilde{N}, A\right)$$

where $\dot{K}$ says that capital accumulates from the residual of output, after total consumption and depreciation. The coefficient θ parameterizes the intertemporal elasticity of substitution in consumption, while ρ is the discount rate, restricted to be positive and to exceed the sum of the rates of population growth and technical change. Equation (3) maximizes welfare by choosing the level of consumption and thus of savings and investment.

For the constant savings rate approach (equation 2), a balanced growth path is a positive and time-invariant capital stock, such that

$$\dot{k}(t) = 0 .$$

In contrast, from the first order condition of equation 3, the balanced growth equilibrium is given by

$$\dot{c}(t) = 0 ,$$

Where

$$\dot{c}(t) = f\left(k(t)\right) - (\delta + n + \xi)k(t) \quad \in \left(0, f(k)\right).$$

Under either assumption regarding accumulation of capital, the balanced growth path has similar predictions. For a Cobb-Douglas production function, for instance, for both assumptions it can be shown that the path in the observable per capita income is given by

$$\log y(t) = \log\left(g\left((\delta + n + \xi)^{-1} s\right)\right) + \log A(0) + \xi t \\ + \left[\log y(0) - \left(\log\left(g\left((\delta + n + \xi)^{-1} s\right)\right) + \log A(0)\right)\right] e^{\lambda t} \quad (4)$$

In such a framework, cross-sectional regression analyses come down to different interpretations of equation 4 (Baumol 1986, De Long 1988, Barro and Sala-i-Martin 1992, Mankiw, Romer, and Weil 1992, and Sachs and Warner 1997). In the first interpretation, the term in $e^{\lambda t}$ is considered not yet at its limiting value and the rest of the right-hand side of equation 4 is given, and equation 4 is considered to explain convergence in income. In the second interpretation, the term in $e^{\lambda t}$ is taken to be already at its limiting value, and the first component of expression 4 is taken to explain the cross-sectional distribution of income.

The first interpretation motivated a number of convergence analyses, beginning with the work of Baumol (1986), who established the tendency of the 16 industrialized economies to converge toward a common income level. Regressing output growth from 1870 to 1979 on a constant initial income, Baumol estimated the following equation:

$$\ln\left[\left(\frac{Y}{N}\right)_{i,1979}\right] - \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right] = \alpha + \beta \ln\left[\left(\frac{Y}{N}\right)_{i,1870}\right] + \varepsilon_i, \quad (5)$$

where $\ln(Y/N)$ is the log of income per person, $\beta = -\left(1 - e^{\lambda t}\right)$ is the convergence parameter, ε is an error term, and i represents the index for countries. If there is convergence, β will be significant and negative, showing that countries with lower initial income grow faster. A result $\beta = 0$ would indicate that growth is uncorrelated with initial income (no convergence). Baumol's (1986) results show an almost perfect convergence: the estimate of β is almost equal to -1, and is somewhat precise. Such convergence was indeed found by a large number of studies using the specification pioneered by Barro (1991), showing convergence to be conditional on both geographic and technological variables as well as institutional or policy measures (for example, Sala-i-Martin 1997). However, the estimated coefficients varied widely, and their interpretation remained controversial; see Abreu, de Groot, and Florax 2005a; Dobson, Ramlogan, and Strobl 2006.

Following Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999), attention shifted to the second interpretation of equation 4. The central problem became determining income levels, focusing particularly on the development of new identification strategies such as those introduced by Acemoglu, Johnson, and Robinson (2001) to account for the endogeneity of economic institutions and policy choices. Assuming a Cobb-Douglas production function, Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) specified an estimation equation to measure directly all ingredients of the production function (except technological progress, which is computed as a residual):

$$\ln\left[\frac{Y}{L}\right]=\frac{\alpha}{(1-\alpha)}\ln\left[\frac{K}{L}\right]+\ln\left[\frac{H}{L}\right]+\ln A \qquad (6)$$

Using data from Summers and Heston (1991), they showed that average output per worker in the five richest countries in their sample exceeded the average in the five poorest by a remarkable factor of 31.7. They also found that the ingredients in measuring output level were not independent of each other. Hall and Jones, for instance, found a substantial correlation between physical capital and human capital $(\rho = 0.60)$; a considerable correlation between human capital and the residual A_i $(\rho = 0.52)$; but only a moderate correlation between physical capital and technology $(\rho = 0.25)$—indicating that countries that use one input in large quantity are also likely to use large quantities of all other inputs. This suggests the action of causal forces influencing these proximate determinants of output (Hall and Jones 1999).

Subsequent studies examined these causal factors to explain the differences in determinants of output across countries. The role of *institutions* and *policies* in encouraging investment and production over consumption and diversion of resources (Romer 1996)—an idea going back to Adam Smith—has received attention in a large number of empirical works: Baumol 1986; Murphy, Shleifer, and Vishny 1991; North 1994; Sachs and Warner 1995; Knack and Keefer 1995; Mauro 1995; Olson 1996; and Acemoglu, Johnson, and Robinson 2001.

The argument that institutions and policies are crucial in explaining differences in income across countries is supported by three kinds of evidence. First is the experience of divided countries, such as post-World War II Germany or Korea, and Hong Kong and Taiwan, separated from China (Olson 1996). With largely shared climate, natural resources, initial levels of physical and human capital, and cultural attitudes, the divided countries differed in social and political infrastructure: East Germany, North Korea, and China were Communist, whereas West Germany, South Korea, Hong Kong, and Taiwan had market economies.

A second piece of evidence highlighting the importance of institutions is the observed cross-country differences in capital-output ratio (differences in marginal product of capital, for a Cobb-Douglas production function). The apparent gap between the marginal product of capital and private incentives to invest could be explained by tax rates, corruption level, risk of expropriation, and other institutional obstacles.

A third type of evidence for the influence of institutions on growth comes from statistical assessments of the relationship between social infrastructure and output, focusing mainly on developing new identification strategies to account for endogeneity of economic institution and policy choices (North 1990; Sachs and Warner 1995; Knack and Keefer 1995; Mauro 1995; Murphy, Shleifer, and Vishny 1991; Temple and Johnson 1998; Acemoglu, Johnson, and Robison 2001).

Although persuasive, the institutional and policies argument leaves a number of questions unanswered questions: What are the determinants of the social infrastructure? If good institutions and policies are all that a country needs to increase its output, why don't we see all countries adopt the same types of institutions and policies to increase their levels of output?

An alternative explanation of why some countries use more inputs than others is the impact of geography. Biophysical obstacles to improving agricultural productivity, public health, and transportation infrastructure could explain cross-country differences in average incomes and also influence economic institutions and policies (Diamond 1997; Landes 1998; Sachs 2001; Sachs and Pia 2002; Bloom and Sachs 1998; Miguel and Kremer 2004; Gallup and Sachs 2001; Gallup, Sachs, and Mellinger 1999).

The most compelling evidence for the geographical argument is provided by world income distribution data. Summers and Heston's (2010) data show that per capita GDP in Northern Africa —only 2.1 times larger than that of Eastern Africa in 1960—was 4.4 times larger by 2007. Similarly, per capita GDP in Southern Africa —only 2.2 times larger than that of Eastern Africa in 1960—was more than 3.4 times larger in 2007.

Representing per capita income as of 1995 on a world map, Bloom and Sachs (1998) and others have observed two striking geographical associations with output. First, countries located between 23.45 degrees north and south latitudes are all relatively poor; rich countries are in middle and high latitudes. Second, coastal economies are generally richer than landlocked economies. Specific geographical variables have been used in an increasing number of statistical assessments: latitude, disease ecology, and access to navigable waterways (Masters and McMillan 2001; Masters and Sachs 2001; Sachs 2003; Easterly and Levine 2003, Bleaney and Dimico 2010). Various statistical tests offer support for both the institutional and the technological drivers of economic growth.

This paper allows different types of spatial autocorrelation processes to affect a variety of endogenous variables, in addition to income. For this purpose, we adopt an explicit three-stage least squares (3SLS) approach, using panel data in a system of simultaneous equations. By identifying the entire system, the role of each possibly endogenous determinant of income can be tested through its association with particular exogenous variables. Our identification strategy rests on that exogeneity, together with the exclusion restrictions by which those variables are tied to particular development channels (see, for example, Klein and Vella 2005). These identification assumptions are considered plausible and are not tested here. Rather, our goal is to posit a relatively large and general representative system.

Endogenous Variables

The particular system of equations we use specifies six endogenous variables that jointly influence income, widely cited in the growth literature:

1. *agricultural output*, measured by net production at international prices
2. *infant mortality*, used as a general measure of health
3. *institutional quality*, based on a combination of measures detailed below
4. *urbanization*, the fraction of the population living in towns and cities
5. *trade*, the sum of exports plus imports as a fraction of GDP
6. *income*, real GDP per capita (in constant 2000 USD)

Each of these endogenous variables has been widely used in the growth literature.

Exogenous Variables

The variables that we assume to be exogenous are defined as follows:

a. A set of farmland and climate variables, which help identify the potential influence of *agricultural output*: The specific variables we use are estimates of agricultural land area, average land quality, prevalence of frost in winter, and annual rainfall. (Included as potential determinants of agricultural production in equation 1 but excluded elsewhere.

b. A disease ecology variable, which helps identify the potential influence of disease transmission on health: The specific *variable we use is malaria ecology* (Kiszewski et al. 2004), which captures the ease with which a mosquito-borne disease would spread from person to person, whether or not the disease is actually present. This is included as a potential influence on agricultural (labor) productivity in equation 1 and on infant mortality in equation 2 but excluded from any direct effect elsewhere.

c. A set of *social history* variables, which help identify the role of institutional quality: The specific measures we use are the percentages of the population that are Protestant, Catholic, or Muslim These measures capture the degree to which a country has been influenced by world cultures that spread through migration and military conquest out of northern Europe, southern Europe, or the Middle East, respectively.[2] These variables are included as potential determinants of institutional quality in equation 3 and are excluded elsewhere.

d. A *coastal location* variable, which helps identify the potential impact on growth of either agglomeration in cities or international integration through trade: The specific coastal variable used here is percentage of the population located within 100 kilometers of the ocean or a navigable river. This is included only as a determinant of agglomeration in equation 4 and of trade in equation 5.

e. The *size of the country*, which is used only as a conditioning variable for trade in equation 6: Our specific variable is total population. This variable is included because aA smaller country will have a larger fraction of its transactions classified as international simply because of where its borders are drawn. If that country merged with its neighbors, the same transaction would be classified as domestic. This variable is excluded from other equations because researchers have found very limited-scale effects in most income regressions.

System of Equations

The significance of the endogenous variables is econometrically identified and tested across six equations using the 10 exogenous variables. The resulting system of equations is described below.[3] Our goal is to estimate this representative system against spatially correlated omitted variables and unspecified neighborhood effects. The system is chosen primarily for its size and generality, capturing a wide range of potential growth mechanisms and linkages. We focus here on the cross-sectional properties of the panel and use time dummies in each equation for each five-year period to absorb any global trends.

The first equation (7) uses ecological variables to identify exogenous determinants of **agricultural output**:

$$\begin{aligned} agoutput_{it} = \alpha_1 + \beta_{11} agland_{it} + \beta_{12} landqual_i + \beta_{13} frost_i + \beta_{14} rainfall_i \\ + \beta_{15} malaria_i + \delta_{1t} + \varepsilon_{1it}. \end{aligned} \tag{7}$$

In this equation, agricultural production per capita is a function of land area per capita,

[2] Other social history variables—for example, relating to colonial history—could be explored in further research.

[3] As noted above, although the implied exogeneity and exclusion restrictions are plausible, specification and robustness tests are left to future work.

soil quality, the prevalence of seasonal frost and total rainfall, plus the malaria ecology transmission index. These factors could be associated with exogenously higher agricultural output, which in turn would affect economy-wide income, either positively (Mellor and Johnston 1961) or negatively (Matsuyama 1992).

Equation 8 uses malaria transmission to identify exogenous determinants of **human capital**, health in particular, as measured by infant mortality (*imrate*):

$$imrate_{it} = \alpha_2 + \beta_{21} income_{it} + \beta_{22} malaria_i + \delta_{2t} + \varepsilon_{2it}. \tag{8}$$

Here, infant mortality is subject to feedback effects from income and, potentially, an exogenous effect from malaria ecology. An exogenously driven change in health could affect income in many ways, including acceleration of investment in schooling.

Equation 9 uses **social history** to identify exogenous determinants of the quality of a country's institutions:

$$\begin{aligned} institqual_{it} = \alpha_4 + \beta_{41} income_{it} + \beta_{42} imrate_{it} + \beta_{43} pctcath_i + \beta_{14} pctprot_i \\ + \beta_{45} pctmus_i + \delta_{4t} + \varepsilon_{4it}. \end{aligned} \tag{9}$$

Potential determinants identified here are economy-wide income, the infant mortality rate (a measure of human capital), and social history (percentages of the population that are Catholic, Protestant, and Muslim).

Equation 10 uses **coastal location** as an exogenous driver of opportunities for specialization and exchange:

$$urbanization_{it} = \alpha_5 + \beta_{51} income_{it} + \beta_{52} agoutput_{it} + \beta_{53} coastal_i + \delta_{5t} + \varepsilon_{5it}. \tag{10}$$

Here, urbanization may be driven by feedback from income and agricultural output, as well as by access to coasts or navigable rivers.

An alternative route to specialization is captured in equation 11, which identifies other determinants of **international trade**:

$$trade_{it} = \alpha_6 + \beta_{61} income_{it} + \beta_{62} coastal_i + \beta_{63} population_{it} + \delta_{6t} + \varepsilon_{6it}. \tag{11}$$

Here, trade can be driven by economy-wide income, coastal location, and population size.

The last equation brings the endogenous variables together, without any additional exogenous variables:

$$\begin{aligned} income_{it} = \alpha_7 + \beta_{71} agoutput_{it} + \beta_{72} imrate_{it} + \beta_{74} instqual_{it} + \beta_{75} urbanization_{it} \\ + \beta_{76} trade_{it} + \delta_{7t} + \varepsilon_{7it}. \end{aligned} \tag{12}$$

This system of equations can be estimated using three-stage least squares (3SLS), but the results are likely to be biased due to spatial processes beyond those captured in the regressors. The equations may share spatially autocorrelated errors due to spatially correlated omitted variables, spatially correlated measurement error, or interaction among neighboring countries (Anselin 2003). We therefore allow each endogenous variable to be subject to both spatial dependence and a spatial autoregressive process in the error term (that is, a spatial ARAR model). To do this we utilize a full information estimator based on instrumental variable and generalized moments estimators that allow for correlation across equations (Kelejian and Prucha 2004).[4] Thus, we start with the naïve 3SLS approach and then compare these results to estimates that allow for the potential influence of spatial spillovers and spatially correlated omitted variables.

[4] Kelejian and Prucha (2007) develop an extended estimator that incorporates heteroskedasticity as well, which could be incorporated into any future work in this area.

3. Description and Discussion of the Data

For all time-variant data, we use observations at five-year intervals (1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, and 2000), with an average of five annual observations centered on the year indicated (1963–1967 for 1965, 1968–1972 for 1970, and so forth).[5] For the data on infant mortality, single-year observations are used at the corresponding 10-year intervals.

Regression variables

The agriculture data are: *agricultural output* (real international dollars); and *land used in agriculture* (thousands of hectares) (dataset from Masters and Wiebe 2000, updated using Food and Agriculture Organization (2004) data[6]). The *land quality* index reports the fraction of a country's agricultural land classified in the top three categories of suitability for agriculture in the World Soil Resources classification (Masters and Wiebe 2000).[7] Climatic data were compiled by Masters and McMillan (2001) from data published by the International Panel on Climate Change (IPCC 1999). *Frost prevalence* refers to the proportion land receiving five or more frost days in winter (December through February in the northern hemisphere; June through August in the southern hemisphere). The raw data come from the IPCC's estimated average number of frost days per month over 1961–1990. *Rainfall* is average total annual precipitation for each grid cell averaged over the country's landmass. The country aggregation is based on the Climatic Research Unit 2.0 gridded dataset (Mitchell et al. 2003).

Economic data are drawn from the Penn World Tables 6.2 for *national income* (real GDP per capita, chain indexed) and for *trade share* (exports plus imports as a fraction of GDP). Urbanization data are drawn from the World Development Indicators online (World Bank 2006), as the percentage of the population in urban areas.

Data on *infant mortality* rates are drawn from United Nations Population Statistics (UN Population Division et al. 2007). The estimates of *religious affiliation* (Protestant-Catholic-Muslim) are from the Barro-Lee dataset (Barro 1999). Our *malaria ecology* variable (from Kiszewski et al. 2004) is an index constructed from the physiological characteristics of the dominant mosquito species, combined with temperature data that determine how long a malaria parasite could survive during transmission from person to person. These factors are largely independent of a country's economic activity or its antimalarial efforts. Most important, the index does not include data on the density of mosquitoes or the prevalence of infection, both of which can be reduced in an otherwise malarial region. Our variable for the quality of *national institutions* is a time-varying index, constructed from data reported by Freedom House (2005) and International Country Risk Services (ICRG 2006). The Freedom House data represent an average of their measures forindicate a country's political rights and civil liberties; the ICRG index measures degree of corruption, military role in politics, religion in politics, law and order, and democratic accountability. Data from these two sources are rescaled for comparability and combined to construct a continuous time series from 1960 to 2000.

---

[5] Note that for 1960 and 2000, only three years are available: 1960–1962, and 1998–2000.

[6] Unpublished output series from Jan Poulisse (personal communication).

[7] Agricultural land is defined broadly to include both "cropland" and "cropland plus natural mosaic," according to the International Geosphere-Biosphere Program classification (United States Geological Survey 1999).

Overall, the dataset comprises nine five-year averages from 1960 through 2000 for 30 countries in Africa. Descriptive statistics for each of the variables are provided in Table 3.1. (The list of countries is provided in Appendix Table A.1.)

Table 3.1—Descriptive statistics for regression variables

Variable[a]/Statistic	Mean	Variance	Minimum	Maximum	Skewness	Kurtosis
Agricultural output[b]	0.149	0.039	0.014	1.643	5.332	32.614
Agricultural land (000 ha)	0.314	1.224	0.070	9.531	6.350	46.824
Land quality	8.805	136.076	0.001	56.038	2.720	10.627
Frost[c]	0.122	0.070	0.000	1.000	2.100	6.194
Rainfall (mm)	996.475	321300.400	46.233	2801.867	0.416	3.256
Malaria	9.652	65.469	0.000	30.095	0.616	2.519
Population (× 1,000)	9.562	118000.000	0.042	62.775	0.002	0.008
Infant mortality rate (per 1,000 live births)	122.636	2194.459	13.000	285.000	0.211	2.984
Trade[d]	69.011	3130.034	5.048	541.396	4.572	35.878
Income (× 1,000)[e]	2.000	3482.276	0.384	10.593	0.002	0.008
Institutional quality	0.207	0.007	0.143	0.667	2.546	9.971
Catholic	27.017	560.889	0.100	82.300	0.736	2.650
Protestant	14.213	191.112	0.000	50.000	0.709	2.526
Muslim	28.833	1237.289	0.000	99.400	1.018	2.378
Urbanization	25.470	215.810	2.230	62.790	0.291	2.107
Coastal[f]	21.601	950.293	0.001	100.001	1.609	4.432
Time dummies[g]	0.000	0.223	–1.000	1.000	0.000	4.500

Source: Authors' calculations. Notes: Based on 30 African countries, five-year averages from 1960 through 2000.[a] See text for full discussion of variable definitions. [b] Index of net farm production per capita at international prices in 2000 U.S. dollars. [c] The proportion of land receiving five or more frost days per month in winter.[d] Total trade as a percentage of real GDP (exports plus imports divided by real GDP).[e] Real GDP per capita in PPP terms, expressed in 2000 U.S. dollars. [f] Within 100 kilometers of a seacoast or navigable river. [g] The time dummies allow fixed effects for 1960, 1965, and so forth and are subsequently recomputed as deviations from the omitted category, 1960.

Spatial measure

The geographic distance between countries is captured through a spatial weights matrix, defined a priori and exogenously on the basis of arc distances between the geographical midpoints of the countries. It is a Boolean proximity matrix; elements are coded as inverse distance if the distance between countries is ≤ 2,500 miles. Following Bell and Bockstael (2000), we standardize by enforcing row sums to be equal and setting the diagonal elements to zero. The resulting spatial weight matrix for a single time slice has dimension 30, with 33 percent of the weights being nonzero. The number of links between countries ranges from 1 to 15, with an average of 10. The minimum cutoff distance to ensure that each country would be linked to at least one other country is 1,800 miles. The weight matrix for the pooled dataset is defined as a 270 × 270 block diagonal matrix, with the sequence of nine 30 × 30 matrices on the diagonal. This implies that we assume spatial autocorrelation to be strictly contemporaneous.

Figure 3.1, in the Annex, presents key information for 2000 in maps, specifically for the dependent variables in the system of equations developed in section 2.

The maps indicate that GDP per capita is highest in northern Africa and southern Africa and relatively low in central Africa, especially the Democratic Republic of Congo. Similarly, infant mortality rates are highest in central Africa and comparatively low in the northern and southern African regions. The spatial distribution of institutional quality exhibits a concentration of high-quality institutions in South Africa as well as some western African countries, including Benin, Mali, and Senegal.

For per capita agricultural output, the spatial distribution is much more scattered: Lesotho and the Democratic Republic of Congo show the lowest per capita agricultural output value, while Uganda, South Africa, Tunisia, and Egypt have the highest. The spatial distribution of urbanization level is much more uniform, at 30 percent or higher in most countries. No strong spatial pattern emerges for the trade share of GDP.

Figure 3.2 provides more detail: Moran scatterplots show the spatial distribution of the six dependent variables for the latest available period. The standardized value of each country's variable x_i is plotted against its spatial lag, which equals the spatially weighted average of the x_j values with the set of neighbors defined through the ***i***-th row of the weights matrix (W). W.x represents the neighboring countries' weighted average value of x: W.trade, for example, represents the weighted average of the trade share in the neighboring country. (Weight is given by the inverse distance between countries, defined in the weight matrix.) These scatterplots help identify local clusters of spatial correlation, spatial nonstationarity, and outliers; the gradient of the trend line equals Moran's *I* coefficient (see Anselin 1996).[8]

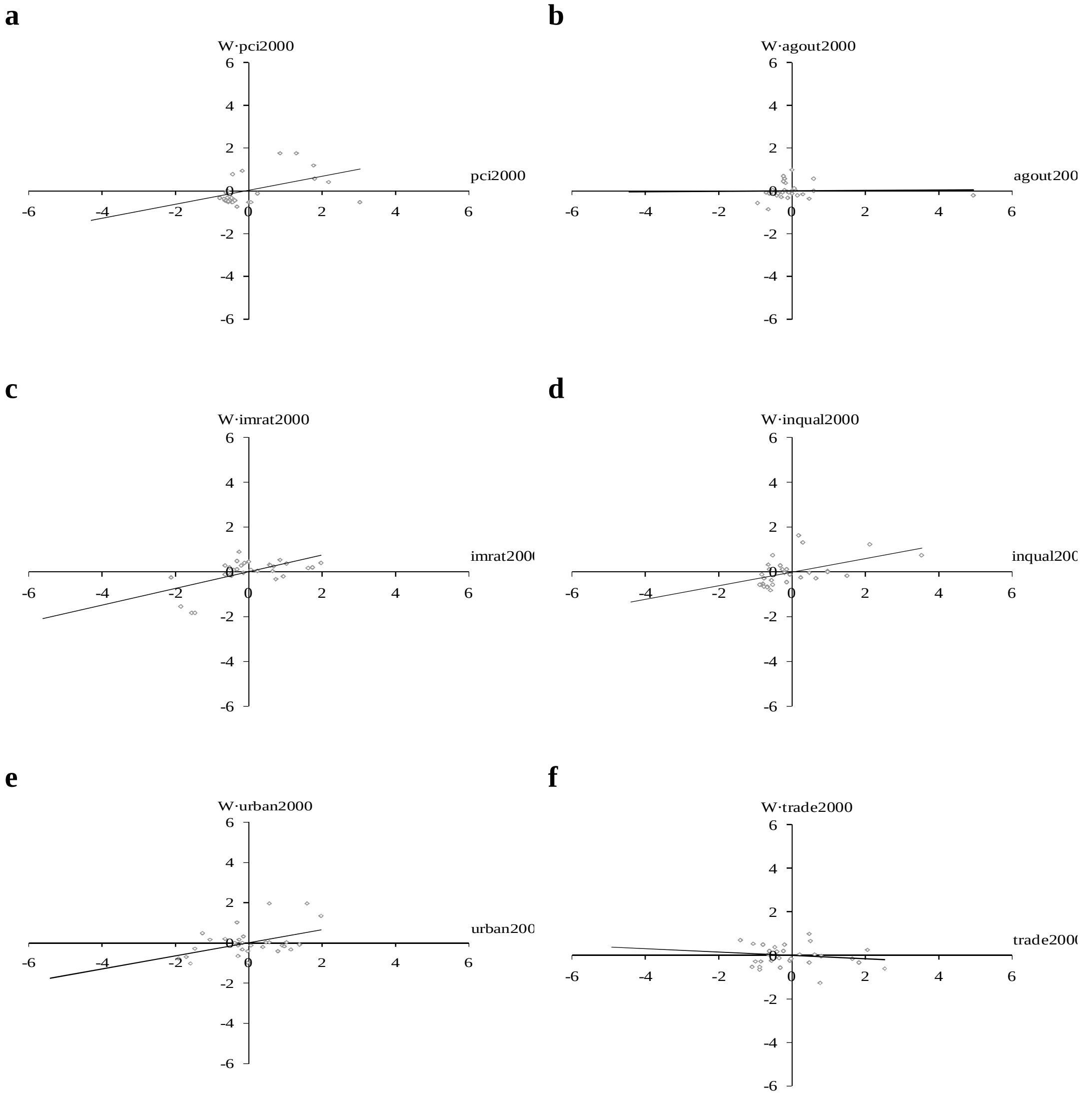
Figure 3.2(a), GDP per capita shows a large clustering of countries in the lower-left quadrant. These are the least-developed countries—that is, lowest-income countries—surrounded by other countries with similarly low per capita incomes. In our sample, only two least-developed countries, Lesotho and Mozambique, are surrounded by neighbors that have incomes above the average: Botswana, and South Africa. In contrast, the relatively high-income countries are surrounded by neighborhoods with above-average incomes (Egypt, Algeria, Tunisia). The outlier (as judged by the 2□□rule) is Seychelles, which is surrounded by neighbors with mostly average per capita income levels. The scatterplot for agricultural output (b) shows a tighter spatial clustering. That is, most countries with low levels of agricultural output are surrounded by other countries with low levels of agricultural output. Uganda, however, is the extreme outlier; it has a high level of agricultural output but is surrounded by neighbors with below-average agricultural outputs. The scatterplot for infant mortality (c) clearly shows two separate clusters of high infant mortality, one comprising most of the countries of central Africa and the other in western Africa. Institutional quality and urbanization (d and e) show a similar degree of spatial clustering, with no obvious outliers. In contrast, the trend line for trade share (f) displays a negative gradient, showing countries with high trade shares as neighbors of countries with low levels of trade share.

[8] With a standardized weights matrix, Moran's *I* is defined as

$$I = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x}) / \sum_{i=1}^{n}(x_i - \bar{x})^2 ,$$

where the variable x is measured in deviations from its mean and w_{ij} are the elements of the weights matrix.

Figure 3.2—Moran scatterplots for (a) GDP per capita, (b) agricultural output, (c) infant mortality rate, (d) institutional quality, (e) urbanization, and (f) trade share, in 2000



Source:Authors .

4. Discussion of Econometric Methods

In two recent papers, Abreu, de Groot, and Florax (2005a, 2005b) stress two important implications of the earlier economic growth literature. First, in a quantitative analysis of over 600 estimates drawn from nearly 50 convergence studies, they find that correcting for endogeneity in the explanatory variables results in significantly higher estimates of the rate of convergence. This is in line with earlier findings in Cho (1996) and Caselli, Esquivel, and Lefort (1996). In addition, they document (2005a) that the use of panel data and concurrent corrections for unobserved heterogeneity in technology levels and/or steady states leads to substantially higher rates of convergence, as confirmed by Dobson, Ramlogan, and Strobl (2006). Abreu, de Groot, and Florax (2005b) review the spatial econometric literature dealing with (regional) economic growth. They find that it has not established a strong link to prevalent economic growth theories and that it tends to restrict the modeling of spatial spillover processes to either a spatial lag or a spatial error model, eventually in combination with spatial regimes, to account for nonstationarity in the mean and variance. Only recently have spatial methods been more rigorously applied, as in Ertur and Koch (2007) and Fingleton and López-Bazo (2006).

Following the approach outlined in Kelejian and Prucha (2004), this paper uses a spatial econometric specification that is less restrictive than previous work in terms of spatial correlation. At the same time, this approach accommodates endogeneity as a system feedback effect, rather than restricting it to spatial spillover effects.

In terms of spatial autocorrelation, the specification allows for spatial spillover effects through the dependent variable as well as for a spatial autoregressive error structure. This specification, known as the spatial ARAR model, reads (for a single equation)

$$\begin{aligned} y &= \rho W y + X\beta + \varepsilon, \\ \varepsilon &= \lambda W \varepsilon + \mu, \end{aligned} \tag{13}$$

where y is an ($n \times 1$) vector of observations on the dependent variable, X is an ($n \times k$) matrix of nonstochastic regressors, W is an ($n \times n$) spatial weights matrix that represents the topology of the spatial system, □ is an ($n \times 1$) vector of iid errors, □ is a ($k \times 1$) vector of regression coefficients, and □ and □ are spatial autoregressive parameters. Substitution and rearrangement of terms in equation 13 leads to

$$y = (I - \rho W)^{-1}\left(X\beta + (I - \lambda W)^{-1}\mu\right). \tag{14}$$

Equation 13 thus implies a rather complex form of spatial autocorrelation evoked by nested spatial multiplier processes pertaining to both the observable and the nonobservable parts of the model (see also Anselin 2003). Testing for spatial autocorrelation is nevertheless rather straightforward and can be based on a Lagrange Multiplier test—generally known as the SARMA test—for which the asymptotic distribution has been derived in a maximum likelihood framework. However, since Lagrange Multiplier tests cannot distinguish between locally equivalent autoregressive and moving average processes (Godfrey 1988), the SARMA test can also be used to detect an ARAR process.[9] A distinct advantage of the Kelejian and Prucha (2004) systems approach is that it explicitly allows for endogeneity to be taken into account. The endogeneity is not necessarily restricted to spatial spillover effects, but it can also include the usual system feedback effects. Kelejian and Prucha derive a full information generalized spatial systems estimator (GS3SLS) in a sequential estimation procedure using the limited information instrumental variable and general moments estimation to provide initial estimates of the spatial autoregressive parameters.

---

[9] Anselin and Kelejian (1997) discuss testing for spatial autocorrelation in a model with endogenous regressors wherein the endogeneity is caused by systems feedbacks or by spatial interaction of an endogenous variable. In the empirical application, we initially use ordinary least squares–based tests, although this ignores the endogeneity of some of the regressors. Testing for spatial autocorrelation can also be based on the general results for Moran's I in Kelejian and Prucha (2001).

Instead of using a purely cross-sectional dataset, we use a panel dataset comprising nine time slices centered on 1960, 1965, and so forth through 2000. Although we treat the data as independent replications of the cross-sectional data, we do include fixed effects for the different time periods, thus accommodating a possible time trend.[10]

5. Estimation Results

The results are generated using the same spatial weights matrix throughout the entire model, accounting only for spatially lagged dependent variables rather than the more general spatially lagged endogenous variables (across equations). Spatially lagged exogenous variables are not incorporated. Following Kelejian and Robinson (1993), we define the instruments as the linearly independent exogenous variables and their first-order spatial lags, although alternatives are available (Lee 2003). Table 5.1 presents the 3SLS results, which account for endogeneity but not spatial effects. Table 5.2 presents the full information results using the feasible GS3SLS estimator discussed above.

Table 5.1—3SLS system estimation results, not allowing for spatial spillovers

Variable	Agricultural Output	Infant Mortality Rate	Institutional Quality	Urbanization	Trade	Income
Agricultural output				–0.139*** (0.029)		–0.072** (0.031)
Agricultural land	0.153* (0.090)					
Land quality	–0.049** (0.021)					
Frost	0.074** (0.040)					
Rainfall	0.649***					
	(0.099)					
Malaria	–0.263***	0.048***				
	(0.031)	(0.007)				
Infant mortality rate			–0.392***			–0.995***
			(0.101)			(0.084)
Income		–0.311***	–0.056	0.705***	0.689***	
		(0.039)	(0.056)	(0.067)	(0.081)	
Institutional quality						0.224
						(0.147)
Trade						0.331***
						(0.098)
Catholic			–0.004***			
			(0.001)			
Protestant			0.002			

[10] Given that some data offer yearly observations, richer models incorporating spatiotemporal dynamics are feasible, but we leave those for future research (see Anselin, Le Gallo, and Jayet 2006).

			(0.002)			
Muslim			–0.002**			
			(0.001)			
Urbanization						0.402***
						(0.073)
Coastal				0.038***	–0.011	
				(0.007)	(0.009)	
Population					–0.115***	
					(0.030)	
Constant	–6.045***	6.938***	0.779	–2.492***	–0.053	9.641***
	(0.542)	(0.290)	(0.853)	(0.527)	(0.668)	(0.528)
R^2	0.40	0.56	0.16	0.50	–0.22	0.30
Goodness of fit	194.89***	390.94***	83.73***	370.12***	103.61***	461.16***

Source: Authors' calculations.Notes: Standard errors are in parentheses. All variables enter in logarithmic form, except for the time dummies and the constant. Note that the R^2 value is not restricted to the usual (–1, +1) interval. The goodness-of-fit test is a Wald test with an asymptotic χ^2 distribution.*p < .10. **p < .05. ***p < .01.

When account for endogeneity but not spatial spillovers, as in the naïve ordinary least squares regression (Table 5.3), per capita income in a system context is not correlated with institutional quality. However, Table 5.1 shows income having a strong and significant link to infant mortality. Agricultural output however appears with a counterintuitive sign. Per capita agricultural output is mainly determined by the availability of agricultural land; the prevalence of winter frost has the same significant effect as in Masters and McMillan (2001), while disease ecology (malaria) also has the expected effect. Controlling for income and the time dummies, disease ecology also shows strong correlation with infant mortality. Institutional quality is associated negatively with infant mortality, but does not have the expected correlation with income. The links to institutions from cultural variables appear to be of minor importance. As expected, the level of urbanization is positively linked to income, agricultural output, and coastal location. Finally, per capita income is strongly negatively affected by infant mortality and positively affected by urbanization.

Tables 5.1 and 5.2 contain the estimation results for the systems estimators. Table 5.1 accounts for endogeneity using 3SLS, whereas Table 5.2 accounts for both endogeneity and neighborhood effects using the spatial ARAR model.

Table 5.2 —Full information system estimator results for the ARAR specification

Variable	Agri cultural Output	Infant Mortality Rate	Institu tional Quality	Urbaniza tion	Trade	Income
W· agricultural output	–0.761*** (0.099)					
W· infant mortality		0.069 (0.110)				
W· institutional quality			0.568*** (0.118)			

W· urbanization				0.544*** (0.057)		
W· trade					0.256** (0.094)	
W· income						0.203*** (0.056)
Agricultural output				–0.048*** (0.022)		0.011 (0.026)
Agricultural land	0.107 (0.087)					
Land quality	–0.063* (0.021)					
Frost	0.155** (0.042)					
Rainfall	0.675*** (0.097)					
Malaria	–0.229*** (0.031)	0.060*** (0.008)				
Infant mortality rate			–0.209** (0.053)			–0.802*** (0.071)
Income		–0.272*** (0.032)	0.003 (0.030)	0.483*** (0.045)	0.213** (0.067)	
Institutional quality						0.537** (0.147)
Trade						–0.092* (0.056)
Catholic			–0.003** (0.001)			
Protestant			0.001 (0.002)			
Muslim			–0.001 (0.001)			
Urbanization						0.483*** (0.044)
Coastal				0.038*** (0.006)	0.016* (0.009)	
Population					–0.145*** (0.031)	

Source: Authors' calculations.Note: Standard errors are in parentheses. All variables enter in logarithmic form, except for the time dummies and the constant. Estimated values and standard errors for □ are based on the general moments estimator in the second step of the stimation procedure used in the Cochrane-Orcutt transformation to obtain full information estimates.*p < .10. **p < .05. ***p < .01.

The results for the spatial system of equations presented in Table 5.2 are broadly similar to those in Table 5.1. However, allowing for spatial dependence changes the results in important ways. First, after controlling for the observed variables, we find significant spatial lags among all of the endogenous variables except infant mortality, which is explained by our data on country characteristics without recourse to unobserved neighborhood effects. For income, institutional quality, urbanization, and international trade there are positive spatial lags, whereas agricultural output shows a negative spatial lag. Controlling for unobserved spillovers and regional characteristics, the measured variables in Table 5.2 show several very interesting correlations. First, for agricultural output, our variables on prevalence of winter frosts, rain, and malaria ecology remain significant and of the expected sign; land quantity, however, becomes insignificant. In the second column, for infant mortality, both malaria ecology and income are significant as expected, and the residual effect of time (not shown here) is quite large and significant, suggesting that important technological improvements may lower infant mortality at a given level of income and malaria ecology. The third variable, institutional quality, has a positive correlation with infant mortality and a small correlation with social history (as measured by the percentage Catholic variable). Urbanization is strongly correlated with local agricultural output, income, and coastal location and has a small positive time trend when controlling for these factors. Trade is negatively correlated with population size (perhaps due to the increased role of nontraded services) and positively linked to income and coastal location, with a small positive time trend.

In our final equation in Table 5.2, all of these endogenous variables have an independent correlation with income, except for agricultural output. In other words, exogenously higher agricultural output drives increased income only by facilitating urbanization. There is also a large residual effect of time on real income, with unmeasured factors driving increases in measured income from 1960 until 1975, followed by decreases through 2000.

Table 5.3 provides the results for an equation-by-equation estimation using ordinary least squares and includes several (spatial) diagnostic test results. The estimated model concerns a rather naïve specification, without any control for endogeneity or spatial lags. It shows that income is closely correlated with a number of endogenous regressors, notably infant mortality, institutional quality, and urbanization. Each of these variables is in turn also correlated with income, when controlling for various other significant determinants. [11]

The results for the Jarque-Bera test indicate that the null hypothesis of normally distributed errors is rejected for nearly all equations. While this provides another reason for interpreting the Lagrange Multiplier diagnostics cautiously, it does not have major implications for the systems estimator, because the estimator does not require the disturbances to be normal. The Breusch-Pagan test results, with random coefficient variation as the alternative hypothesis, show that homoskedasticity is rejected in all equations. This implies that it would be very useful to address this issue in future work. The spatial diagnostics are fairly mixed. For all equations there is evidence that a higher-order model is appropriate for the equations pertaining to agricultural output, infant mortality, urbanization, trade, and income. There is, however, no clear indication of spatial autocorrelation for the institutional quality variable.

[11] The misspecification test results shown here are also only heuristic since they are derived without accounting for the endogeneity of some of the regressors. The condition number shows that multicollinearity does not impair the results.

Table 5.3— Equation-by-equation ordinary least squares estimation results, with diagnostics for spatial effects[a,b]

Variable	Agricultural Output	Infant Mortality Rate	Institutional Quality	Urbanization	Trade	Income
Agricultural output				–0.041*		0.047*
				(0.024)		(0.026)
Agricultural land	0.144					
	(0.116)					
Land quality	–0.064**					
	(0.032)					
Frost	0.200***					
	(0.075)					
Rainfall	0.851***					
	(0.114)					
Malaria	–0.269***	0.070***				
	(0.045)	(0.010)				
Infant mortality rate			–0.175***			–0.727***
			(0.051)			(0.080)
Income		–0.228***	0.02	0.447***	0.083	0.118
		(0.032)	(0.029)	(0.050)	(0.062)	(0.123)
Institutional quality						
Trade						–0.131***
						(0.049)
Catholic			–0.005***			
			(0.001)			
Protestant			0.00003			
			(0.002)			
Muslim			–0.002*			
			(0.001)			
Urbanization						0.508***
						(0.049)
Coastal				0.054***	0.024**	
				(0.007)	(0.010)	
Population					–0.122***	
					(0.031)	
Constant	–7.008***	6.295***	–0.711*	–0.347	4.449***	10.047***
	(0.609)	(0.241)	(0.417)	(0.390)	(0.562)	(0.503)
Condition	30	37	67	36	37	49

number						
Jarque-Bera	23.364***	2.419	243.323***	0.945	48.316***	4.485*
Breusch-Pagan	31.284***	72.052 ***	20.208 ***	57.319 ***	22.419 **	28.573 ***
Moran's *I*	6.179***	–0.784	2.318**	0.313	–1.315	0.172
LM-error	22.187***	2.671*	1.521	0.42	4.645**	0.59
Robust LM-error	0.02	0.156	2.354	7.330***	0.029	17.759***
LM-lag	28.560***	4.619**	3.286*	1.565	4.753**	21.969***
Robust LM-lag	6.393**	2.103	4.119**	8.475***	0.136	39.138***
SARMA	28.580***	4.774*	5.640**	8.895***	4.781*	39.728***
R^2-adjusted	0.34	0.56	0.2	0.54	0.07	0.52
AIC	847.997	159.927	60.494	418.061	559.268	392.481
Log likelihood	–409.998	–68.9633	–16.247	–197.03	–267.634	–182.24

Source: Authors' calculations.Notes: Standard errors are in parentheses.[a] All variables enter in logarithmic form, except for the time dummies and the constant.[b] The Jarque-Bera and the Breusch-Pagan tests are asymptotically χ^2 distributed and test for normality of the errors and homoskedasticity with random coefficient variation as the alternative hypothesis, respectively.
**p* < .10. ***p* < .05. ****p* < .01.

In sum, when controlling for spatial processes in this model, we find support for both the institutionalist and the geographic approaches. Geographic factors such as malaria ecology, coastal location, and seasonal frost do have significant independent effects on the system, influencing institutional quality but not completely determining it. At the same time, a country's institutional quality does have a strong independent role in determining income.

6. Conclusions

Using panel data in a system of simultaneous equations, and controlling for spatial spillovers and unobserved spatial heterogeneity, this paper has explored how measured country characteristics such as institutions, trade, and urbanization might be linked to real income per capita in Africa. This approach uses a new kind of test for how particular types of technologies and institutions might affect income, and then tests the robustness of each variable against various kinds of neighborhood effects.[12]

The endogenous variables associated with income are agricultural output per capita, health status, institutional quality, trade, and urbanization. The exogenous

[12] Throughout this paper we have indicated potential extensions and variations to be explored and addressed in future work: testing for exogeneity and exclusion restrictions; incorporating heteroskedasticity following the procedures developed in Kelejian and Prucha (2007); assessing parameter heterogeneity and other robustness checks; and considering the temporal dynamics of the system.

variables represent climate, malaria ecology, social history, coastal location, and population size. With this specification, after controlling for spatial proximity, *all* of the variables have some independent effect on income. This result provides strong empirical support for both geographic and institutionalist hypotheses. Geographic variables such as land quality, coastal location, and malaria prevalence have strong independent effects on income, primarily by facilitating urbanization and declines in infant mortality. Institutional quality also has a strong independent effect on income, even when controlling for reverse causality and neighborhood effects. Most notably, accounting for these country characteristics still leaves large residual spatial lags, suggesting that our specification has only begun to capture the relevant spillovers and spatial heterogeneity among countries. Understanding these spatial correlations will require not only more precise measurement of the unobserved factors driving local agricultural productivity, public health, and urbanization, but also more complete accounting for the cross-border flows associated with migration, investment, or technology diffusion.

References

Abreu, M., H. L. F. de Groot, and R. J. G. M. Florax. 2005a. “A Meta-analysis of β-convergence: The Legendary 2%.” *Journal of Economic Surveys* 19 (3): 389–420.

Abreu, M., H. L. F. de Groot, and R. J. G. M. Florax. 2005b. “Space and Growth: A Survey of Empirical Evidence and Methods.” *Région et Développement* 21:13–44.

Acemoglu, D., S. Johnson, and J. Robinson. 2001. “The Colonial Origins of Comparative Development.” *American Economic Review* 91 (5): 1369–1401.

Anderson, J. E. [1979] 2000. “A Theoretical Foundation for the Gravity Equation.” *American Economic Review* 69 (1): 106–116.

Anselin, L. 1996. “The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association.” In *Spatial Analytical Perspectives on GIS*, edited by M. Fischer, H. Scholten, and D. Unwin (111-125). London: Taylor and Francis.

Anselin, L. 2003. “Spatial Externalities, Spatial Multipliers, and Spatial Econometrics.” *International Regional Science Review* 26 (2): 153–66.

Anselin, L., A. Bera, R. Florax, and M. Yoon. 1996. “Simple Diagnostic Tests for Spatial Dependence.” *Regional Science and Urban Economics* 26 (1): 77–104.

Anselin, L., and H. Kelejian. 1997. “Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors.” *International Regional Science Review* 20 (1/2): 153–182.

Anselin, L., J. Le Gallo, and H. Jayet. 2006. “Spatial Panel Econometrics.” In *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice,* edited by L. Matyas and P. Sevestre: 625-660. Dordrecht, The Netherlands: Kluwer.

Barro, R. J. 1991. “Economic Growth in a Cross Section of Countries.” *Quarterly Journal of Economics* 106 (2): 407–443.

Barro, R. J. 1999. “Determinants of Democracy.” *Journal of Political Economy* 107 (2): 158–163.

Barro, R. J. and X. Sala-i-Martin (1992) Convergence, *Journal of Political Economy* (100): 223–51.

Baumol, W. J. 1986. “Productivity Growth, Convergence, and Welfare: What the Long-run Data Show.” *American Economic Review* 76 (5): 1072–1085.

Bell, K., and N. Bockstael. 2000. "Applying the Generalized Moments Estimation Approach to Spatial Problems Involving Microlevel Data." *Review of Economics and Statistics* 82:72–82.

Bleaney, M. and A. Dimico 2010. Journal of African Economies, 19 (5): 635-656

Bloom, D. E., and J. D. Sachs. 1998. "Geography, Demography, and Economic Growth in Africa." *Brookings Papers on Economic Activity* 1998 (2): 207–295.

Bradshaw, Y. W., and R. Noonan. 1997. "Urbanization, Economic Growth, and Women's Labour Force Participation: A Theoretical and Empirical Reassessment." In *Cities in the Developing World: Issues, Theory and Policy,* edited by J. Gugler: 6–22. New York: Oxford University Press.

Caselli, F., G. Esquivel, and F. Lefort. 1996. "Reopening the Convergence Debate: A New Look at Cross-country Growth Empirics." *Journal of Economic Growth* 1 (3): 363–389.

Cass, D. (1965) Optimum Growth in an Aggregative Model of Capital Accumulation, *The Review of Economic Studies* 32(3): 233-240.

Cinyabuguma, M. and L. Putterman 2011. Sub-Saharan Growth Surprises: Being Heterogeneous, Inland and Close to the Equator Does not Slow Growth Within Africa, *Journal of African Economies*. 20(2): 217–262.

Cho, D. 1996. "An Alternative Interpretation of Conditional Convergence Results." *Journal of Money, Credit, and Banking* 28 (4): 669–681.

DeLong, J. B. (1988) Productivity Growth, Convergence, and Welfare: Comment, *American Economic Review* 78(5):1138-1154.

Diamond, J. M. 1997. *Guns, Germs and Steel: The Fate of Human Societies.* New York: Norton.

Dobson, S., C. Ramlogan, and E. Strobl. 2006. "Why Do Rates of □-convergence Differ? A Meta-Regression Analysis." *Scottish Journal of Political Economy* 53 (2): 153–173.

Easterly, W., and R. Levine. 2003. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *Journal of Monetary Economics* 50 (1): 3–39.

Ertur, C., and W. Koch. 2007. "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence." *Journal of Applied Econometrics* 2: 1033–1062

Fay, M., and C. Opal. 2000. *Urbanization without Growth: A Not So Uncommon Phenomenon.* World Bank Policy Research Working Paper No. 2412. Washington, DC: World Bank.

Fingleton, B., and E. López-Bazo. 2006. "Empirical Growth Models with Spatial Effects." *Papers in Regional Science* 85 (2): 177–198.

Food and Agriculture Organization. 2004. *FAOStat.* Rome, Italy: Food and Agriculture Organization of the United Nations. http://faostat.fao.org.

Freedom House. 2005. "Freedom in the World Country Ratings." http://www.freedomhouse.org.

Gallup, J.L. and J.D. Sachs (2001) The economic burden of malaria, *The American Journal of Tropical Medicine and Hygiene* 64 (1−2): 85–96.

Gallup, J.L., J.D. Sachs and A.D. Mellinger (1999) Geography and Economic Development, *International Regional Science Review* 22(2): 179–232.

Godfrey, L. G. 1988. *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches.* Cambridge, UK: Cambridge University Press.

Gundlach, E. 2004. "The Primacy of Institutions Reconsidered: The Effects of Malaria Prevalence in the Empirics of Development." Kiel Institute for World Economics Working Paper No. 1210. Kiel, Germany: Kiel Institute for World Economics.

Hall, R., and C. I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (1): 83–116.

ICRG. 2006. "Country data, International Country Risk Guide." http://www.countrydata.com.

International Panel on Climate Change. 1999. "IPCC Data Distribution Data Centre CDROM, April 1999." http://ipcc-ddc.cru.uea.ac.uk.

Kelejian, H., and D. Robinson. 1993. "A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model." *Papers in Regional Science* 72:297–312.

Kelejian, H. H., and I. R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40 (2): 509–533.

Kelejian, H. H., and I. R. Prucha. 2001. "On the Asymptotic Distribution of the Moran *I* test Statistic with Applications." *Journal of Econometrics* 104 (2001): 219–257.

Kelejian, H. H., and I. R. Prucha. 2004. "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations." *Journal of Econometrics* 118 (1/2): 27–50.

Kelejian, H. H., and I. R. Prucha. 2007. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." Unpublished manuscript, University of Maryland, US.

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs. 2004. "A Global Index Representing the Stability of Malaria Transmission." *American Journal of Tropical Medecine and Hygiene* 70:486–498.

Klenow, P. J. and A. Rodriguez-Claire (1997) The Neoclassical Revival in Growth Economics: Has It Gone too Far? in *1997 NBER Macroeconomics Annual*. Cambridge, MIT Press, 83–103.

Klein, R., and F. Vella. 2005. "Estimating a Class of Triangular Simultaneous Equations Models without Exclusion Restrictions." Centre for Microdata Methods and Practice Working Paper No. CWP08/05. London: Centre for Microdata Methods and Practice.

Knack, S. and P. Keefer (1995) Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures, *Economics and Politics* 7(3):207–27.

Koopmans, T. C. (1965) On the Concept of Optimal Economic Growth, in *The Econometric Approach to Development Planning,* Pontif. Acad. Sc. Scripta Varia 28,pp. 225-300; reissued North-Holland Publ.

Landes, David (1998), *The Wealth and Poverty of Nations*, New York, NY: W. W. Norton.

Lee, L. F. 2003. "Best Spatial Two-stage Least Squares Estimation for a Spatial Autoregressive Model with Autoregressive Disturbances." *Econometric Reviews* 22:307–335.

Livanis, G., C. B. Moss, V. E. Breneman, and R. F. Nehring. 2006. "Urban Sprawl and Farmland Prices." *American Journal of Agricultural Economics* 88 (4): 915–929.

Maddison, A. 1995. *Monitoring the World Economy: 1820–1992*. Paris: Organization for Economic Co-operation and Development.

Maddison, A. 2004. "World Population, GDP and Per Capita GDP, 1-2001 AD." http://www.eco.rug.nl/~Maddison/.

Magrini, S. 2004. “Regional (Di)Convergence.” In *Handbook of Urban and Regional Economics*, edited by V. Henderson and J.-F. Thisse 4: 2741-2796. Amsterdam: Elsevier.

Mankiw, N. G. D. Romer, and D. N. Weil (1992) A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics* 107(2) 407-37.

Masters, W. A., and M. S. McMillan. 2001. “Climate and Scale in Economic Growth.” *Journal of Economic Growth* 6 (3): 167–186.

Masters, W. A., and K. Wiebe. 2000. “Climate and Agricultural Productivity.” Paper presented at the Conference on Biophysical Constraints in Tropical Agriculture, Harvard University, Cambridge, Manuscript processed October 15, 200. MA, US.

Masters, William A., Jeffrey D. Sachs (2001). Climate and Development. Paper presented at the annual meetings of the AEA, New Orleans, January (mimeo).

Matsuyama, K. 1992. “Agricultural Productivity, Comparative Advantage, and Economic. Growth.” *Journal of Economic Theory* 58 (2): 317–334.

Mauro, P. (1995) Corruption and growth, *Quarterly Journal of Economics* 110(3):681–712.

Mellor, J. W., and B. F. Johnston. 1961. “The Role of Agriculture in Economic Development.” *American Economic Review* 51 (4): 566–593.

Miguel, E. and M. Kremer (2004) Worms: Identifying Impacts on Health and Education in the Presence of Treatment Externalities, *Econometrica* 72(1): 159–217.

Mitchell, T. D., T. R. Carter, P. D. Jones, M. Hulme, and M. New. 2003. “A Comprehensive Set of High-resolution Grids of Monthly Climate for Europe and the Globe: The Observed Record (1901–2000) and 16 Scenarios (2001–2100).” *Journal of Climate*. http:// earthscape.org.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1991) The Allocation of Talent: Implications for Growth, *The Quarterly Journal of Economics*, 106(2): 503–30.

Natural Resources Conservation Service. 1999. “World Soil Resources Database.” Washington, DC: Natural Resources Conservation Service, USDA. http://www.nhq.nrcs.usda.gov/WSR/.

North, D. C. 1990. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.

North, D. (1994) Economic Performance through Time, *American Economic Review* 84(3) 359–368.

Olson, M. (1996), Big Bills Left on the Sidewalk: Why Some Nations are Rich and Others are Poor, *Journal of Economic Perspectives* 10(2):3–24.

Rodrik, D., A. Subramanian, and F. Trebbi. 2004. “Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development.” *Journal of Economic Growth* 9 (2): 131–165.

Romer, D. (1996) *Advanced Macroeconomics*, McGraw-Hill: New York.

Sachs, J. D. 2001. “Tropical Underdevelopment.” NBER Working Paper No. 8119. http://www.nber.org/papers/W8119.

Sachs, J. D. (2003) Institutions Don't Rule: Direct Effects of Geography on Per Capita Income, *NBER Working Paper*, 9490, February.

Sachs, Jeffrey D. and M. Pia (2002) The Economic and Social Burden of Malaria, *Nature* 415 (6872): 680–685.

Sachs, J. D. and A. M. Warner (1997) Fundamental Sources of Long-Run Growth, *The*

American Economic Review, American Economic Association 87(2):184–88.

Sala-i-Martin, X. 1997. “I Just Ran Two Million Regressions.” *American Economic Review* 87 (2): 178–183.

Sala-i-Martin, X. 2006. “The World Distribution of Income.” *Quarterly Journal of Economics* 121 (2): 351–397.

Solow, R. 1957. Technical Change and the Aggregate Production Function, *Review of Economics and Statistics* 39(3):312-320.

Summers, R. and A. Heston (1991) The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988, *The Quarterly Journal of Economics* 106(2):327–368.

Swan, T. W. (1956) Economic growth and capital accumulation, *Economic Record* 32(2): 334–362.

Temple, J. 1999. “The New Growth Evidence.” *Journal of Economic Literature* 37:112–156.

Temple, J. and P.A. Johnson (1998) Social Capability and Economic Growth, *The Quarterly Journal of Economics* 113 (3):965–990.

Timbergen, J. 1962. *Shaping the World Economy*. *Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund.

UN Population Division, UNICEF, WHO, and World Bank. 2007. *Levels and Trends of Child Mortality in 2006: Estimates developed by the Inter-agency Group for Child Mortality Estimation.* New York: United Nations Children’s Fund.

United States Geological Survey. 1999. “Global Land Cover Characterization.” http://edcwww.cr.usgs.gov/landdaac/glcc/ glcc.html.

World Bank. 2006. “World Development Indicators.” http://publications.worldbank.org/WDI

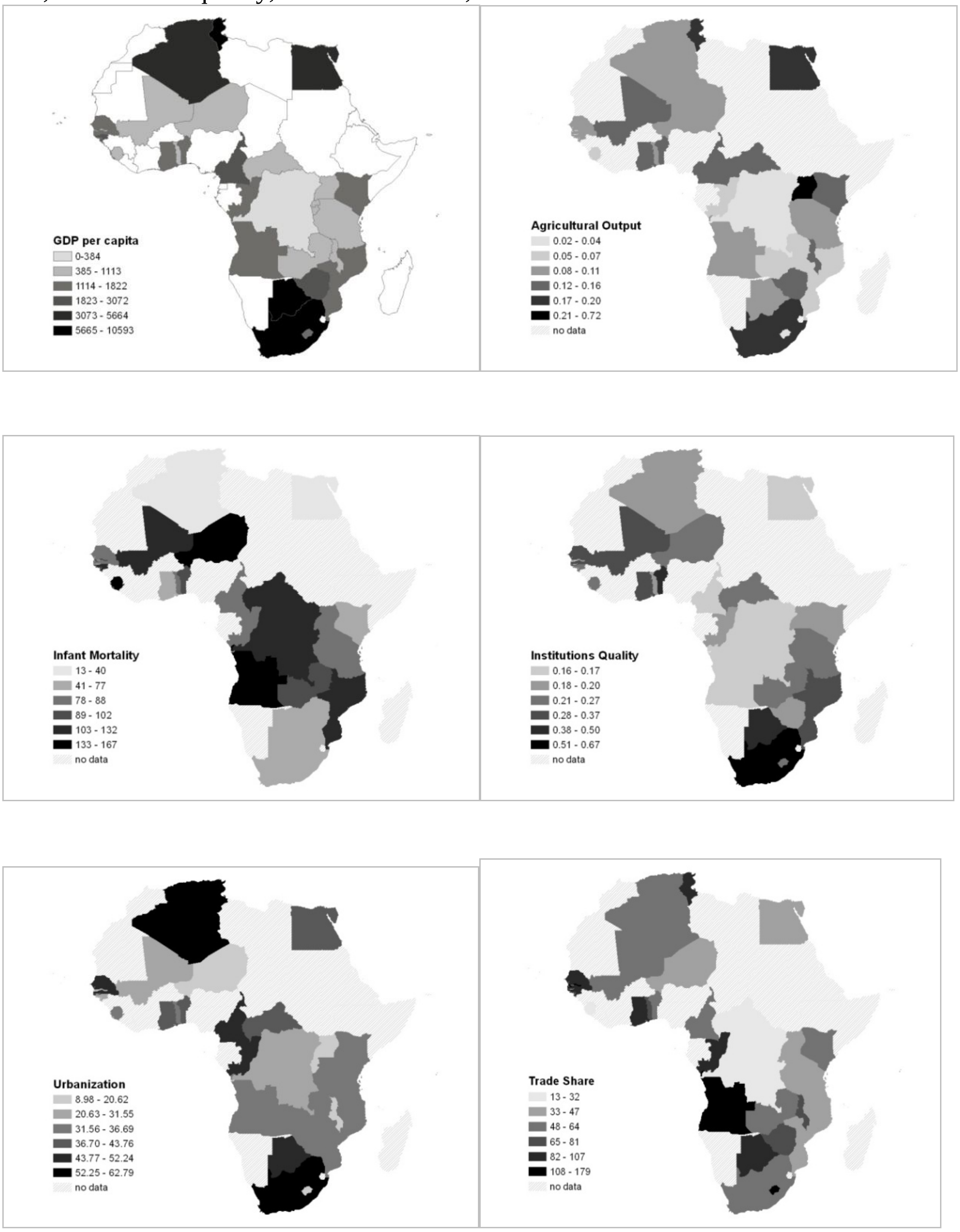
Appendix: Supplementary Table and Maps

Table A.1—Countries included in the sample

Algeria	Malawi
Angola	Mali
Benin	Mozambique
Botswana	Niger
Burundi	Rwanda
Cameroon	Senegal
Central African Rep	Seychelles
Congo	Sierra Leone
Congo, Dem Rep	South Africa
Egypt	Tanzania
Gambia	Togo
Ghana	Tunisia
Guinea-Bissau	Uganda
Kenya	Zambia
Lesotho	Zimbabwe

Source: Authors.

Figure 3.1—Maps of GDP per capita, agricultural output, trade share, infant mortality rate, institutional quality, and urbanization, in 2000



Source: Authors.