Spatial determinants of poverty in rural Kenya

Paul O. Okwi[a,b], Godfrey Ndeng'e[c], Patti Kristjanson[a], Mike Arunga[a], An Notenbaert[a], Abisalom Omolo[a], Norbert Henninger[d], Todd Benson[e], Patrick Kariuki[a], and John Owuor[a]

[a]International Livestock Research Institute (ILRI), P.O. Box 30709, Nairobi 00100, Kenya; [c]Central Bureau of Statistics (CBS), P.O. Box 30266, Nairobi 00100, Kenya; [d]World Resources Institute (WRI), Washington, DC 20002; and [e]International Food Policy Research Institute (IFPRI), Washington, DC 20006

Edited by Partha Sarathi Dasgupta, University of Cambridge, Cambridge, United Kingdom, and accepted August 6, 2007 (received for review February 6, 2007)

This article investigates the link between poverty incidence and geographical conditions within rural locations in Kenya. Evidence from poverty maps for Kenya and other developing countries suggests that poverty and income distribution are not homogenous. We use spatial regression techniques to explore the effects of geographic factors on poverty. Slope, soil type, distance/travel time to public resources, elevation, type of land use, and demographic variables prove to be significant in explaining spatial patterns of poverty. However, differential influence of these and other factors at the location level shows that provinces in Kenya are highly heterogeneous; hence different spatial factors are important in explaining welfare levels in different areas within provinces, suggesting that targeted propoor policies are needed. Policy simulations are conducted to explore the impact of various interventions on location-level poverty levels. Investments in roads and improvements in soil fertility are shown to potentially reduce poverty rates, with differential impacts in different regions.

Poverty, income inequality, and natural resource degradation are severe problems in Kenya, especially in rural areas. Kenya poverty rates are among the highest in the developing world. National poverty prevalence is estimated at 45% (1), and natural resource degradation is reported to be increasing (2). In the recent past, there have been several studies on poverty and income distribution in Kenya (3). Some of these studies have focused on the poverty profile, but this is limited in its usefulness because it shows how poverty levels are correlated with one characteristic at a time.

This study examines the determinants of poverty prevalence for small, spatially defined populations in rural locations[f] of Kenya. Evidence from poverty maps for East Africa and other developing countries shows that poverty and income distribution are not homogenous and vary widely across space. Some of these differences are caused by differences in geographic and agroclimatic conditions, infrastructural access to markets and public facilities, the presence or absence of natural resources such as forests or water bodies, and political and historical factors. Even though these factors have been identified as major contributors to differences in standards of living of populations in different areas, there has been little empirical work to ascertain the exact relationship between welfare levels and these factors. This type of analysis has been limited largely because of data deficiency and lack of appropriate analytical tools. Recent advances in spatial analytical software now allow such analyses.

Thus in this study, we attempt to explore the link between empirical welfare information and Geographical Information System (GIS)-based environmental data. An important aspect in developing this link is taking into account the fact that the dependent variable is of a different data type and form of spatial aggregation than most of the independent spatial variables. Data of different types and from different sources are used to generate the variables used in the analysis, some of which are spatially autocorrelated or derived from socioeconomic variables that typically exist in a spatially discrete format based on administrative units and differ from environmental data that have a spatially continuous nature. This poses methodological challenges described in *Methods* and in the supporting information (SI) *Appendix*.

The key research questions in this study are: (*i*) What spatial factors account for the spatial variation in Location-level poverty across rural Kenya? (*ii*) Does the relationship between agroclimatic and other spatial variables with poverty differ significantly among provinces? (*iii*) What are the potential poverty impacts of investment/changes in some of the spatially related factors found to influence poverty in different areas of Kenya?

To answer the first question, we use a global spatial regression analysis to examine the determinants[g] of the prevalence of poverty incidence in rural areas of Kenya. For the second question, we use similar analysis for each of the seven rural provinces. And for the third question, we conduct simulations of the impact of possible investments in roads and soil improvement on poverty levels in three provinces.

Table 1 shows the key selected independent variables for the analysis and how they are hypothesized to affect poverty incidence. The variables are divided into two categories. Exogenous variables are those variables that are unlikely to be affected by the level of economic activity or poverty. On the other hand, endogenous variables are those that may both influence poverty and be influenced by poverty.

Developing a better local-level understanding of poverty determinants, together with knowledge about how household-level factors and broader national policies affect household welfare, will assist policy makers and development practitioners in their efforts to enable rural Kenyans improve their livelihoods and welfare.

Results

All of the models estimated used the location-level poverty rate (the proportion of individuals falling below the national rural poverty line of Kenya shillings (KShs) 1,239 per adult equivalent per month) as the dependent variable. Analysis was undertaken at the national level first (for 2,232 rural locations), followed by models at the provincial level (i.e., for each of Kenya's seven rural-based provinces). Details of the concept of spatial dependence and its associated diagnostic tests are provided in *Methods* and *SI Appendix*.

Author contributions: P.O.O. and P. Kristjanson designed research; P.O.O., G.N., P. Kristjanson, M.A., A.N., and P. Kariuki performed research; T.B. contributed new reagents/analytic tools; P.O.O., G.N., P. Kristjanson, M.A., A.N., A.O., N.H., P. Kariuki, and J.O. analyzed data; and P.O.O., P. Kristjanson, M.A., and A.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: CBS, Central Bureau of Statistics; GIS, Geographical Information System; OLS, ordinary least squares.

[b]To whom correspondence should be addressed. E-mail: pokwi@cgiar.org.

[f]Kenya's administrative units are Province, Division, District, Location, and sub-Location.

[g]We follow the ''risk chain'' theoretical approach taken by Benson that implies the spatial variables used as independent variables are largely exogenous to the outcome (a consumption-based indicator, such as the poverty indicator used here) and therefore can be interpreted as determinants and not merely correlates of poverty. For further discussion of this distinction, see Benson (4).

This article contains supporting information online at www.pnas.org/cgi/content/full/0611107104/DC1.

Table 1. Explanatory variables used in spatial regression analysis

Variables	Expected relationship to poverty
Exogenous variables	
Rainfall	Negative (low rainfall, higher poverty)
Rainfall variation	Positive (high variation, higher poverty)
Elevation	Positive (high elevation, higher poverty)
Slope	Positive (steeper slope, higher poverty)
Type of land cover	Not known
Distance to towns/ municipalities/cities	Positive (greater distance, higher poverty)
Length of growing period (LGP)	Negative (longer LGP, lower poverty)
Soil type	Negative (good soil, lower poverty)
Possible endogenous variables	
Population	Not known
No. and density of markets	Either direction
Transport time to towns, markets, and cities	Either direction
Density of roads	Either direction

Table 2 shows the results of the national model[h] (spatial error model) with 23 independent variables based on the expectations of Table 1 and including dummy variables for the seven provinces at 1%, 5%, and 10% significance levels. The model explains more than half of the variation in rural poverty rates but >15 of the 23 variables are statistically significant at the 5% level. Several variables returned the expected sign although the significant levels varied. Soil quality, elevation, length of growing period, different categories of land use, and locational variables are significant.

To address the question of how sensitive poverty is to quality of soil, a dummy variable for soil quality was included. We expect that locations with good soils are likely to have high agricultural potential and thus have absolute advantage in producing high-value perishable vegetables and other crops. Indeed, we found that locations with good soils are associated with less poverty. The magnitude of effect is not large, ≈1%, i.e., improving soil fertility (from poor to good soil) would reduce poverty by up to one percentage point in rural areas of Kenya's provinces. This strongly points to the policy of improving soil quality through the use of fertilizers and soil conservation techniques.

Given our expectations, and the findings of related studies (5) of a strong relationship between slope of land and poverty, it is not entirely surprising that two of the four estimated slope parameters are significant. Thus, we find that relative to the very flat areas (0–4% slope), Locations that have a high percentage of land made up of steep slopes have higher poverty levels. The coefficient is largest for locations with a >30% slope area, a result that is consistent with theoretical explanations that point toward serious erosion, cultivation, and irrigation-related problems associated with steep land.

Land-use variables emerge as strong determinants of poverty among rural locations in Kenya. The coefficients for the "percentage of the location under particular land uses" show mixed results. As expected, locations that have large areas that are built up (occupied by buildings) tend to have lower rates of poverty. This suggests that built-up areas represent tendencies toward urbanization, and more urbanization is expected to result in lower poverty. In general, the poverty maps (6) show that urban areas are richer than rural areas in Kenya. Our results suggest that locations with large areas under grassland are likely to have lower poverty rates, a somewhat nonintuitive result (6). It may be that this result is reflecting the fact that there are very few people in grasslands areas, or it may indicate that this variable is capturing something else. With respect to the percentage of wooded area, another nonintuitive finding is that locations with more wooded areas are associated with higher poverty rates in rural areas (given that woodlands often provide nuts, fruits, and firewood for poor families).

Measured in meters above sea level, elevation has a significant negative effect on Location-level welfare: Communities at higher elevation are likely to be less poor. This is expected because many communities living in the highlands are much better off than their counterparts in many parts of dry lowlands of Kenya (6).

Several variables meant to capture agroclimatic conditions were tested in this model. Rainfall and its coefficient of variation and length of growing period were among those variables. As expected, the variation in poverty among rural communities in Kenya is strongly influenced by agroclimatic factors. The results show that locations with longer growing periods are likely to have lower poverty rates relative to areas with shorter growing periods. The effect here is clear, because most crops such as (maize, beans, millet, sorghum, peas) require >60 days to mature.

For the livestock-related variables, the analysis shows that communities living in rangelands are likely to have higher poverty levels. Our results suggest that there appears a strong positive relationship between poverty and living in the rangelands. Recent studies have shown that the rangelands have some of the highest poverty rates in Kenya (6). This is fairly intuitive, because they are also the areas with poorest access to roads, services (education and health), and general infrastructure in the country. We further explore the determinants of poverty in livestock-keeping areas (rangelands) in a related study.

The demographic variable (population density) has significant negative effects on poverty in rural areas. Areas with high population densities are associated with lower poverty rates. Population density influences labor intensity of agricultural production, including the choice of commodities as well as production technologies and land management practices, by affecting the land–labor ratio. This result implies that people tend to settle in areas where they can enhance their incomes, for example, through farming, and such areas end up having relatively low poverty levels.

Better roads and/or access to markets are expected to favor production of high-value products and nonfarm activities that will contribute to higher incomes or lower poverty. The results of this study show that longer travel times to tarmac and murram roads significantly increase poverty levels. The standard explanation here is that the greater the travel time to a good road, the more difficult it is to access markets, limiting livelihood options. Conversely, communities that have greater access to markets, good infrastructure (health and education), and public administration face lower transaction costs and more livelihood options, leading to lower poverty levels. The above results point toward the need for investment in improved rural roads if poverty is to be reduced in Kenya.

Finally, we investigate the evidence of regional heterogeneity regarding the effects of different spatial determinants on poverty. Thus, for all rural locations, we test for equality of parameter estimates for all provinces except Nairobi and find that the homogeneity hypothesis is strongly rejected. Compared with the Rift Valley Province (which is the reference or base category/province) the results show that with the exception of Central Province, all other provinces are associated with higher poverty levels relative to Rift Valley Province. It is worth noting that the provincial dummies may be capturing a number of factors in the different regions (such as security, administration, infrastructure, culture) which are not captured in the other spatial variables. This heterogeneity strongly justifies the need for province-specific estimations, the results and discussion of which are presented below.

We also explored the effects of spatial factors when we restricted the national-level rural poverty regression to include only variables that are likely to be exogenous to poverty, referred to as selective models. Restricting the model in this way helps us to explore the relative importance of the spatial explanatory variables. Variables representing distance and demographic characteristics were not

[h]Data for North Eastern Province is included in this regression.

Table 2. Results of the spatial-error model

Dependent variable	Poverty incidence coefficient	SE	P
Constant	0.86067	0.03176	0.00000
Demographic			
Population density	−0.00006	0.00001	0.00002
Provincial dummy variables			
reg2 (Central)	−0.14724	0.01140	0.00000
reg3 (Coast)	0.06534	0.01443	0.00001
reg4 (East)	0.10896	0.00883	0.00000
reg5 (North Eastern)	0.23308	0.01480	0.00000
reg6 (Nyanza)	0.13937	0.00917	0.00000
reg8 (Western)	0.09460	0.01181	0.00000
Distance and travel time			
Mean distance to nearest town of 200,000 people	−0.00003	0.00000	0.00000
Average travel time to tarmac or murram road (minutes)	0.00001	0.0000	0.0776
Land use			
Percentage location under grass	−0.00144	0.00026	0.00000
Percentage location under farmland	0.00008	0.00015	0.57360
Percentage location wooded	0.00036	0.00014	0.01020
Percentage location that is built up	−0.01330	0.00282	0.00000
Rangeland (dummy)	0.01262	0.00651	0.05267
Natural factors			
Average elevation (meters above sea level)	−0.00174	0.00077	0.02279
Percentage of location with 4–8% slope	0.00118	0.00020	0.00000
Percentage of location with 8–15% slope	0.00002	0.00024	0.94537
Percent of location with 15–30% slope	−0.00019	0.00031	0.54445
Percent of location with >30% slope	0.00213	0.00035	0.00000
Percentage of location with LGP <60 days	0.00026	0.00013	0.05257
Percentage of location with LGP 180 days	−0.00032	0.00010	0.00169
Good soil (dummy)	−0.01132	0.00516	0.02820
λ	0.19945	0.08684	0.02163
Observations	2,232		
Pseudo R^2	0.5320		
Akaike information criterion:	−3891.33		
Log likelihood	1,968.665		

included in the first selective model. Keeping the provincial dummy variables and excluding these variables (i.e., population density, distance to hospitals and major towns) reduced the explanatory power by five percentage points, to 48%. When the dummy variables representing the seven provinces were also excluded in the second selective model, the explanatory power of the model was reduced to 36%. The exogenous spatial variables, mainly land use and natural factors, on their own are able to explain 36% of the variability in poverty rates that we see across Kenya.

The results of the national-level analysis suggest that there are concerns with the variations and significance of the variables. The provincial-related variables may be picking some omitted variables, and, yet, they explain a high percentage of variation in rural poverty across locations, hence the need for provincial-specific analysis.

Provincial Determinants of Poverty in Kenya. Separate models were run for each of the seven provinces to capture the differences in spatial poverty determinants across these very diverse provinces. Six of the seven provinces showed significant presence of spatial dependence, mainly of the spatial lag type, except Central Province. North Eastern Province showed no presence of spatial autocorrelation, and therefore we discuss their results based on the ordinary least squares (OLS) estimates. The variables that are significant for each of the provinces, as well as at the national level, are summarized in Table 3.

Central Province. Central Province covers an area an area of 13,191 km^2 and is southwest of Mt. Kenya. Rainfall is fairly reliable, falling in two seasons. It has a total population of 3,724,159 (1999 census) inhabitants and is a key producer of coffee. In this Province, 164 locations are considered, and the model fit is 0.50. The results show that limited access to roads is associated with higher poverty levels. The longer the travel time from the location center to the nearest road (track, murram, or tarmac), the poorer it is. Roads provide crucial access to markets, and the result obtained here suggests that areas where it takes people a long time to reach a good road are typically poorer communities. Similarly, our findings show that locations in Central Province that are mainly rangelands or are further away from public forests and on higher elevation are associated with higher poverty levels.

In contrast, a large proportion of the wetland area of a location is associated with lower poverty levels. The wetlands result suggests that people near wetlands may have enhanced livelihoods, and it would be interesting to explore further what ecosystem goods and services they are benefiting from due to presence of these wetlands. Other factors were not significant in this province.

Coast Province. Coast Province covers an area of 83,603 km^2 with a tropical humid climate. It has a population of 2,487,264 inhabitants (1999 census), with tourism as the main source of income. A spatial lag model was estimated for this province, and the results show the percentage of the location under wetlands, the percentage of the location with an 8–15% slope, the probability of flooding, and average length of growing period of 180 days or greater are associated with lower poverty rates. As expected, locations with longer growing periods, and thus much higher cropping potential, are likely to be less poor.

Also among the significant variables, we see that the greater the percentage of the location that is under water (waterlogged), with a slope of 4–8%, and travel time to the nearest road (feeder or murram), the lower the poverty. This reinforces the pattern displayed in the national and Central Provincial poverty results. The greater the distance from a location center to the nearest tarmac or murram road, the higher the poverty. This reflects the importance of access to decent roads to community welfare levels.

Eastern Province. Eastern Province covers an area of 159,891 km^2 and is arid to semiarid in terms of climate, although with areas bordering Mt. Kenya experiencing climate similar to the Central Province. It has a population of 4,631,779 (1999 census), with farming as the main source of income. The Eastern Province model was also a spatial-lag model. Locations that are relatively further from the nearest public forest, have 4–8% and 15–30% slopes, have more area under protected area, and more farmland are poorer. Being far from a public forest has a highly significant influence on living standards. This suggests that many people rely on forest resources such as firewood, fruits, nuts, charcoal, and herbs.

Similar to the findings for Coast Province, variables that were significant and associated with lower poverty rates included elevation, proportion of the location under wetlands, grasslands, and locations with an average growing period of 180 days or greater. These results portray the importance of agricultural potential and land use in poverty reduction.

North Eastern Province. We note that data from this province should be treated with caution. The poverty estimates used for North Eastern are derived estimates from the model for Coast Province, because the Household Budget Survey for 1997, which was used to estimate location-level poverty levels for all of the other provinces, was not implemented in this province because of security-related reasons. The model estimated is an OLS because there was no evidence of spatial autocorrelation. Perhaps not surprisingly in this arid province, the coefficient of variation of rainfall stands out as a major determinant of poverty. In this region, locations with higher

Table 3. Summary: provincial determinants of poverty

Variable	Central	Coast	Eastern	North Eastern	Nyanza	Rift Valley	Western	National
Elevation	*** (+)		*** (−)		*** (+)	NS	NS	**
Distance to forest	*** (+)	NS	*** (+)		** (+)		* (+)	
Perc water		* (−)			** (+)	NS		
Perc built up				NS		*** (−)		***
Perc grass		NS	** (−)		NS	NS	** (−)	***
Perc farmland		NS	** (+)		NS	NS	NS	
Perc wooded			NS	NS				***
Perc wetlands	** (−)	** (−)	** (−)		NS	NS		
Perc protected area			*** (+)				*** (+)	
Perc 0–4% slope	NS							
Perc 4–8% slope		*** (+)	** (+)	*	* (+)	* (+)	NS	***
Perc 8–15% slope		*** (+)	NS		NS	NS	NS	
Perc 15–30% slope			** (+)			NS	** (−)	
Perc ≥30% slope			NS			*** (+)		***
Avg. travel time to tarmac or murram road	*** (+)	* (+)	*** (+)			NS	NS	*
Flood		**(+)						
Mean rain cv	*** (+)			**(+)				
Avg rainfall					NS			
LGP 60 days		NS	NS			NS		***
LGP arid 180		*** (−)	** (−)			*** (−)		***
Distance to district hospital	** (+)	NS	NS		** (+)	NS		
Distance to dispensary			NS	* (+)				
Rangelands	** (+)				NS		NS	*
Mean distance to town of 10,000		NS		** (+)			* (+)	
Mean distance to town of 50,000		NS		NS				
Mean distance to town of 200,000	NS	NS				NS		
Good soil	NS				NS		* (−)	**
ρ/λ	***	***	***		***	***	***	**

***, significant at 1%; **, significant at 5%; *, significant at 10%. Perc, percentage; NS, not significant; CV, coefficient of variation; +, positive effect; −, negative effect.

rainfall variability also tend to be relatively poorer. Two distance-based variables are also key determinants of poverty in this province. Distance to the nearest health center and distance to the nearest town with 10,000 people are both positive and significant, i.e., the further the distance to the nearest health center or town, the higher the poverty incidence. Health-related issues have been found to be a major factor influencing household descents into poverty in Kenya, and this analysis suggests that accessibility to health services is important. Likewise, the distance to the nearest town is very important and likely reflects the fact that basic services, including education, health, trade, and security, are found in these towns.

Nyanza Province. Nyanza province has a population of 4,392,192 (1999 census) within an area of 16,162 km^2. The climate is tropical humid, with fishing as one of the major economic activities and subsistence farming as the other. For this province, a spatial-lag model was estimated. Higher elevation, longer distance to the nearest public forest, a higher percentage of the location under water, and longer distances to the nearest health facility are associated with higher poverty rates in Nyanza Province. Conversely, distance to the nearest city of 200,000 people is significantly and negatively related to poverty, i.e., the further from a city, the lower the poverty, a nonintuitive result that suggests the benefits and services derived from large towns are equally available in small towns in Nyanza; hence, nearness to large towns is not as important as it is in other less-densely populated areas of Kenya.

Rift Valley Province. This province covers an area of 173,854 km^2. It has a population of 6,987,036 inhabitants (1999 census). It has a wide range of climatic zones from arid to humid tropical type and has livestock keeping, tourism, and farming as the major sources of income. A spatial-lag model for Rift Valley's 785 locations was run. The results provide very insightful information, and a number of variables have the expected sign. Variables representing rainfall and associated indicators, namely, length of growing period, are significantly related to poverty. Once again, we found that locations with longer growing periods are associated with lower poverty rates. The higher the percentage of a location that is built up, the lower the poverty incidence. In terms of variables associated with higher poverty levels, the significant indicators include flooding potential and slope. Areas with high potential for flooding are more vulnerable, other things being equal, and, not surprisingly, this has a negative impact on living standards. Larger areas of sloping land may not be as conducive to settlement and farming; thus, areas with larger slopes tend to have higher poverty rates.

Western Province. Western Province covers an area of 8,361 km^2, experiences a mainly tropical climate with variations due to altitude, and has a population of 3,358,776 (1999 census), with agriculture being the main economic activity. With the exception of a few variables, this province has some interesting results that were not significant in the other regions. Locations that are further away from small towns or have long distances to travel to reach public forests and protected areas have higher poverty levels. This is consistent with the earlier notion that distance to facilities and resources is an important determinant of poverty. Locations with majority good soil have lower poverty. Conversely, the findings for Western Province suggest that the larger the percentage of grassland and the more area with slopes of 15–30%, the lower the poverty, both results that are not very intuitive.

These findings suggest that the relationships between poverty and geographic factors do vary significantly and spatially in their effects across rural Kenya. It is clear that some spatial variables are important in influencing poverty in certain provinces and not in others. Such a finding is critical for the formulation and targeting of antipoverty programs. These results can be used to guide local actions aimed at reducing poverty.

Having estimated the poverty determinants, we can now generate simulations to predict reductions or increases in general poverty levels that result from changes in selected spatial characteristics. The purpose of these simulations is to illustrate how changes in levels of the determinants will alter aggregate poverty levels. These changes are such as those that may result from the implementation of specific government policy aimed at reducing poverty. Our simulations involve changing the variables at the provincial level, because the national results may be able to derive accurate inference. We choose to change variables that are significant and amenable to change in three of the seven provinces, namely: Central, Eastern, and Western Provinces.

First, we consider the potential impact of a reduction in the travel time it takes to reach the nearest tarmac (all weather bound) or murram (all weather loose) or track road from the location center in Central Province. We reduce travel time to roads to 1 hour for all locations that have travel times of >1 hour (which is the mean travel time to the nearest road in this province). In this simulation, we are trying to capture improvements in national road infrastructure as a means of improving accessibility of rural communities to markets and general infrastructure. The results show that a reduction in travel time, on average, from >1 h to <1 h to the nearest track, murram, or tarmac roads within all locations in Central Province could potentially lower average location-level poverty rates by 0.8% (or the average province-level poverty rate from 31.3% to 30.5%, which would imply 21,649 poor people escaping poverty. The result for Eastern Province is equally small (0.8%).

Perhaps the disappointing aspect of this simulation is that the expected reduction in poverty is very small. This result holds true in terms of poverty reduction when we look at the sign of the coefficient. However, it should be noted that the small coefficients are a result of a change in only one variable. Roads alone may not be the panacea for the poverty problems in Central Province. There is a need to consider other factors in this simulation. For example, easy access to good roads combined with high agricultural potential (better soils and reliable rainfall) may lead to larger reductions than roads alone.

We also simulate the potential direct impact of a change in soil fertility on poverty incidence in Western Province. Although soil did not show as a significant variable in the spatial models, we simulate its impact based on socioeconomic evidence about Western Province.

The results suggest that the poverty rate for Western Province could be lowered by 9.4 percentage points with investments leading to a change in average soil fertility from poor to good across all locations that have poor soils and poverty levels above the mean for the province (59%). However, this result is indicative of the potentially substantial impact on poverty from improvements to soil fertility levels in western Kenya (a finding supported by numerous other studies). This approach, linked with further research, has the potential for being able to quantify the potential costs, benefits, and impacts on poverty of investments aimed at enhancing soil fertility.

Discussion

In this article, we have sought to improve our general understanding of how (and which) spatial factors are related to poverty and how this varies across Kenya's diverse landscapes, how much of the variation in poverty incidence across Kenya can be explained by environmental/spatial factors, and how this approach can be used to evaluate the potential impact on poverty levels of investments in factors found to have a significant influence on poverty incidence.

The results of the regression models demonstrate the statistical significance of certain spatial variables. At the national level, the set of important variables is diverse and includes regional dummies, land use, elevation, soil conditions/quality, and length of growing period, travel time to roads and towns (market access), and demographic conditions. This suggests the presence of a poverty–environment relationship and hence the impact of environmental factors on the welfare of the poor and on poverty reduction efforts. However, the strength of the provincial dummy variables shows that provinces in Kenya are not homogenous. For this reason, different spatial indicators could be important in different provinces, hence the need for a provincial-level analysis. These region-specific and not national-level variables could be important for designing and evaluating provincial-specific poverty-reduction strategies.

Our simulation results for three provinces suggest that increasing access to roads and improving soil conditions would result in decline in the number of poor people in these provinces. In Western Province, improving soil conditions in locations with poor soil and high poverty rates ($>59\%$) would result in a 9% reduction in poverty levels across locations in Western Province. We find the beneficial impact effect of improved soil quality is robust to whether we consider locations with high or lower rates and proportion of the land that is arable.

Because the results suggest that different spatial factors are important in different provinces, the design and implementation of any poverty-reduction strategies can be province-specific. However, in interpreting the importance of the results for poverty reduction, one should not assume that these effects are instantaneous, even though we estimated them from static models. Road investments, for instance, have inherently long gestation, whereas soil improvements can have immediate effects during the next planting season. Our results indicate that these variables can have powerful effects in terms of long-term reduction in poverty.

Finally, it should be reiterated that, although this analysis has helped explain the geographic determinants of poverty, there is need to refine and extend this analysis, including more disaggregate analysis following development domains in Kenya as well as incorporating supplementary information from other data sources such as the livestock and agricultural census.

Methods

Data. The poverty estimation makes use of data obtained from the 1997 Welfare Monitoring Survey (WMSIII) and the 1999 Population and Housing Census. The survey questionnaire collected information on household and demographic characteristics, education, assets, employment, income, and expenditures (8). The 1997 Population and Housing Census was conducted by the same institution [the Central Bureau of Statistics (CBS)]. The census questionnaire included information on household members and was administered to all households in the country, with the exception of North Eastern Province. Although the census did not collect information on income and expenditures, it provides information on a number of characteristics that have been shown to be strong correlates of poverty.

The spatial-analysis portion uses a range of spatially referenced variables describing topography, land cover and land use, climate, demography, and market/town access. The data on roads and other topographic data such as land cover, soils, and climate were obtained from Africover,[i] including information on the distribution of road, market and town infrastructure. We used a subset of these variables as our independent variables and the candidate independent variables are aggregated to the location level.

In selecting among potential determinants of welfare, one key consideration was to get variables that are arguably exogenous to welfare or current consumption. Thus, for instance, we exclude several nonspatial characteristics of households such as type of dwelling or value of assets, because some of these items are already

[i]Africover is a Food and Agriculture Organization (FAO) environmental database for environmental resources. More information is available at www.africover.org/system/area.php?place=1.

used in the derivation of welfare levels. Some of these excluded household characteristics may also be, in part, determined by living standards in the area, and would cause endogeneity concerns in the choice of modeling approach.

In the spatial-regression analysis, we use data developed at several different scales. As pointed out in Benson *et al.* (4), pooling from different scales in such an analysis leads to the risk of drawing inferences about smaller analytical units from the aggregate characteristics of a group made up of several of those units. This is not a problem for us because the spatial factors are all collected at more local scales than at location level.

Estimation Strategy. To model the prevalence of poverty as a function of selected spatial variables, we carried out two different analyses: (*i*) a simple ordinary least-squares regression and (*ii*) a global spatial regression. We also analyzed poverty at two different levels, national and provincial.

Generalized OLS regression model. Applied to this context, we estimate the OLS regression model as: $y_i = \beta X_i + \varepsilon_i$, where Y is a vector of observations on the dependent variable; X is a matrix of independent variables, β is a vector of coefficients, and e is a vector of random errors.

Despite the popularity of this approach, problems of spatial autocorrelation limit its application in analyzing spatial relationships. Because poverty in one location may be influenced by poverty in a neighboring location, it is important to consider the nature of the spatial dependence inherent in the data. Another way the problem of spatial autocorrelation manifests itself is through the correlation of error terms. Therefore, unless we correct for spatial autocorrelation, the assumptions of OLS regression are violated, and the estimates derived from this method are likely to be biased. To assess spatial autocorrelation, the clustering of the residuals from the OLS model was examined by using the Moran's I statistic.

Global spatial regression model. To control for spatial autocorrelation in the model, we modify the model by including a supplementary explanatory variable. This variable is meant to represent the spatial dependency of the dependent variable. This is commonly done by using the spatial lag of the dependent variable. In this case, the spatial lag of the dependent variable is defined as the weighted mean of a variable for neighboring spatial units of the observation unit in question (9).

There are two major ways in which spatial autocorrelation can manifest itself, referred to as spatial-lag dependence and spatial-error dependence. Spatial-lag dependence refers to a situation in which the dependent variable in one area is affected by the dependent variable in nearby areas. Such a relationship is modeled as a spatial-lag model and can be written as follows:

$$y_i = \delta \sum_{j \neq 1} w_{ij} y_j + \beta X_j + \varepsilon_j, \quad \textbf{[1]}$$

where y_i is the dependent variable for area i, δ is the spatial autoregressive coefficient, w_{ij} is the spatial weight reflecting the proximity of i and j, y_j is the dependent variable for area j, β is a vector of coefficients, X_j is a matrix of explanatory variables, and ε_j is the error term.

The spatial weights matrix, w, represents the degree of proximity between each pair of spatial observations.[j] It is a binary variable if the two areas are contiguous or else a continuous variable based on a function of the distance between the two areas or locations. Omitting this adjustment will result in the coefficients being biased and inconsistent.

A second type of spatial dependence can be attributed to the error term of the model. In this case, the error for the model in one area or location is correlated with the error terms in its neighboring locations (9). This kind of spatial dependence occurs if there are variables that are omitted from the regression model but do have an effect on the dependent variable and they are spatially correlated. Such a relationship can be modeled as a spatial-error model:

$$y_i = \beta X_j + \lambda \sum_{j \neq 1} w_{ij} y_j \varepsilon_j + \varepsilon_i, \quad \textbf{[2]}$$

where y_i is the dependent variable for area i, λ is the spatial autoregressive coefficient, w_{ij} is the spatial weight reflecting the proximity of i and j, y_j is the dependent variable for area j, β is a vector of coefficients, X_j is a matrix of explanatory variables, and ε_j is the error term.

Here, the error term is disaggregated into the spatial lag of the error term of neighboring locations and the residual error term for the spatial unit in question. When there is spatial error dependence, OLS coefficients will be unbiased but not efficient (the standard errors will be larger than if there were no omitted variables), making interpretation of the significance results difficult (10).

Spatial autocorrelation can be detected by using standard global and local statistics that have been developed, including Moran's Index, Geary's C, G statistics, and LISA (10). Whenever there is either spatial error or spatial dependence, an appropriate model can be used to correct for the problem. For spatial dependence, the spatial-lag model is used. In the case of spatial error, we use the spatial-error model. In practice, there is usually very little difference between the two spatial models. However, to select which model to use, a Lagrange multiplier test is used to assess the statistical significance of the coefficients in each model. Where spatial autocorrelation is likely, usually the result of the test on each will be significant. The preferred model in such a case is the one with the highest Lagrange multiplier test value (10). The spatial regression models therefore correct for spatial autocorrelation, and their estimates are unbiased, efficient, and consistent. Details of the estimation procedure are provided in *SI Appendix*.

We thank John Lynam for his initiative and contribution. The authors also express their appreciation to CBS, the Food and Agriculture Organization (FAO), and other organizations that made this data available. The contributions and comments from the staff of CBS and the Government of Kenya are acknowledged and appreciated. Special thanks are also due to Russ Kruska of the GIS Unit of the International Livestock Research Institute (ILRI) for help in GIS mapping and analysis, Luis Carlos Rodriguez and Maren Radeny for their comments, Douglas Ikong'o for logistical support, and many others whom we have not mentioned here. This report is a publication of a project jointly implemented by the CBS and ILRI and funded by The Rockefeller Foundation.

[j]Inverse distance weights were used to perform these tests. We carried out the analysis with another commonly used form of distance weighting, squared inverse distance, and obtained similar results.

1. Kenya National Bureau of Statistics (2007) *Basic Report on Well-Being in Kenya* (Regal Press Kenya, Nairobi, Kenya).
2. National Environment Management Authority (2003) *Kenya* (Government of Kenya, Nairobi, Kenya).
3. Mwabu G, Kimenyi MS, Kimalu PK, Nafula N, Manda DK (2003) *African Dev Rev* 15(1):77–85.
4. Benson T, Chamberlin J, Rhinehart I (2005) *IFPRI Discussion Papers* (International Food Policy Research Institute, Washington, DC) IFPRI Paper 198.
5. Minot N (2000) *World Dev* 28:19–331.
6. Central Bureau of Statistics (2003) *Geographic Dimensions of Well-Being in Kenya* (Regal Press Kenya, Nairobi, Kenya), Vol 1.
7. Kristjanson P, Radeny M, Baltenweck I, Ogutu J, Notenbaert A (2005) *Food Policy* 30:568–583.
8. Central Bureau of Statistics (1998) *Welfare Monitoring Survey* (Regal Press Kenya, Nairobi, Kenya).
9. Anselin L (2002) *Agr Econ* 27:247–267.
10. Anselin L (1988) *Spatial Econometrics: Methods and Models* (Kluwer, Dordrecht, The Netherlands).