An investigation of the spatial determinants of the local prevalence of poverty in rural Malawi

Todd Benson*, Jordan Chamberlin, Ingrid Rhinehart

International Food Policy Research Institute, 2003 K St. NW Washington, DC, 20006 USA. Tel: +1 202-862-5600. Fax: +1 202-467-4439. E-mail addresses: t.benson@cgiar.org; j.chamberlin@cgiar.org; i.rhinehart@cgiar.org

Abstract

Of Malawi's rural population, 66% have a level of consumption below the national poverty line. We examine the spatial determinants of the prevalence of poverty for small spatially defined populations there. A theoretical approach based on the risk-chain conceptualization of household economic vulnerability guided selection of a set of potential risks and coping strategies—our analytical determinants—that could be represented spatially. We used these to develop

---

*Corresponding author.

global and local models of poverty prevalence. In our global spatial error model, only eight of 24 determinants selected for analysis proved significant. In contrast, all determinants considered were significant in at least some of the local models developed using geographically weighted regression. Moreover, these models provided strong evidence of the spatial non-stationarity of the relationship between poverty and its determinants. This result implies that poverty reduction efforts in rural Malawi should be designed for and targeted at district and subdistrict levels.

Keywords: Poverty reduction; Small area estimation; Poverty mapping; Risk chain; Malawi

Introduction

Our research seeks to identify key spatially explicit determinants of differing poverty levels in local areas in rural Malawi. Such an understanding can effectively guide the efforts of government and others to assist rural communities attain higher levels of welfare, particularly for the 66% of the rural population with a level of consumption below the national poverty line (NEC, 2000). This research is undertaken based on the theoretically informed expectation that certain agro-ecological and aggregate socio-economic characteristics of where an individual or household lives can be important determinants of whether those residents will attain an adequate level of welfare to meet their basic needs. Such a local-scale understanding of the significant spatial determinants of local welfare, if coupled with knowledge of how individual and household-specific and broader national and sub-national factors affect household welfare, will contribute to the success of poverty reduction efforts.

We focus on small, local populations of rural Malawi, rather than all of Malawi, to simplify the analysis. Virtually all rural Malawian households employ principal livelihood strategies based on agriculture or the use of other natural resources. Agro-ecological conditions are important elements of these livelihoods and of the risk chains in which they are enmeshed. In contrast, we expect that a much broader range of risk and coping variables would need to be included to adequately capture the determinants of poverty prevalence in urban neighbourhoods.

As is common with much economic research on poverty and welfare, here we define welfare as the level of consumption of an individual or household (Deaton and Zaidi, 2002). The welfare and poverty content of our analysis is based on the computation of a welfare measure for each individual or household in the 1997-98 Malawi Integrated Household Survey (IHS) sample. To determine whether or not an individual or household is poor, we compare the welfare measure to a cost of basic needs poverty line that incorporates the daily basic food and non-food requirements of Malawians. We then evaluate the welfare measure against the poverty line to determine whether one is poor or non-poor.

For the rural sample of the IHS as a whole, 73.5% of the value of their consumption is food (NEC, 2000). From an analytical standpoint at least, in rural Malawi the poor are food insecure and the food insecure will be poor. Consequently, the analysis here is as relevant to issues of household food insecurity in rural Malawi as it is to poverty.

Vulnerability to poverty – the risk chain

We drew the theoretical understanding to guide our analysis from the literature on household economic vulnerability and particularly the concept of the risk chain. Vulnerability to poverty is usually defined in the economics literature as, "having a high probability of being poor in the next period" and is determined by the ability of households and individuals to manage the risks they face (Dercon, 2001). Although vulnerability is a dynamic concept in that it is concerned with the potential future welfare status of individuals and households, it also provides useful insights in accounting for why households and individuals or, as here, aggregations of households are predominantly poor or not poor at a particular point in time.

The risk chain decomposes household economic vulnerability into three links—risk or risky events (shock), responses to risk, outcome in terms of welfare. The level of economic vulnerability of households is dependent on the degree to which they are exposed to negative shocks to their welfare and on the degree to which they can cope with such shocks when they occur. Their current welfare status (whether they are poor or not) is the outcome. Although it might be described in different ways, the risk chain is a common conceptual framework in a range of sub-disciplines, including development and welfare economics, the food security literature, hazards and global climate change research, and in health and nutrition (Alwang et al., 2001). Here we provide a brief overview of the sorts of components we consider as making up each link in the risk chain – the determinants of local poverty prevalence in our analysis.[1]

To what extent households or individuals are exposed to shocks to welfare is an important consideration in assessing their likelihood of being vulnerable to falling into poverty. These risks may be events that affect the population broadly (covariate risks) or those that affect individuals or households in a more random fashion (idiosyncratic risks). Covariate risks that affect specific areas or broad and, ideally, spatially defined segments of the population are the easiest to bring into a spatial analysis such as ours. Such shocks (epidemics, drought, flooding) can be mapped. Idiosyncratic risks, in contrast, are less easily managed analytically within a spatial context.

Whether exposure to a risky event results in a decline in welfare depends on the degree to which the household or individual is susceptible to harm from that shock. Their resilience depends on whether they have access to necessary resources or assets to cope effectively with the shock so that no lasting damage is done to their well-being. Households can employ a broad range of risk management strategies in the face of shocks.

The welfare outcome for a household or individual faced with a negative shock to their economic well-being could be measured in several ways—most commonly, a consumption-based welfare indicator. In the analyses here, we use the aggregate poverty headcount for a local area, based on such a welfare indicator, as our dependent variable.[2] Child malnutrition rates, food consumption levels, any manner of human development or welfare indices and so on could also be used.

Methods and data

Poverty mapping

We computed the dependent variable, the poverty headcount for rural aggregated enumeration areas (EA), using the poverty mapping method developed by Elbers, Lanjouw and Lanjouw (Elbers et al. 2000, 2003, 2005).[3] Poverty mapping involves discovering relationships between household and community characteristics and the welfare level of households as revealed by the analysis of a detailed living standards measurement survey. A model of these relationships is then applied to data on the same household and community characteristics contained in a national census in order to determine the welfare level of all households in the census. The resulting estimates of poverty derived from the census can be spatially disaggregated to a much higher degree than is possible using survey information. Moreover, estimates are provided of the error in the calculated poverty measures.

A poverty map for Malawi was completed in early 2002 based upon the 1997-98 Malawi IHS and the September 1998 Malawi Population and Housing Census.[4] Twenty-three separate strata models were developed to construct the poverty map (Benson et al., 2002). For the 23 models, the mean adjusted R^2 is 0.380 and ranges from 0.248 to 0.594.

New analytical geography

The developers of the poverty mapping method have demonstrated that reliable poverty estimates can be generated for quite small populations. While the desired level of statistical precision in the poverty estimates will determine the minimum population size to use, early assessments of the minimum population threshold to which poverty mapping methods could reasonably be applied were as low as 500 households (Elbers et al., 2000).

We sought to exploit this feature of poverty mapping. The EA in Malawi, with an average household population of about 250 households, is too small for reliable use in poverty mapping. Consequently, we developed a new analytical geography by agglomerating EAs into units with populations just above 500 households.[5] The new spatial units respected the boundaries of the poverty mapping strata to enable the application of the models to households resident in the aggregated EAs.[6]

The 3004 aggregated EAs used in the analysis exclude those from the four major urban centres of Malawi, all forest reserves and national parks and some rural areas in Nkhata Bay District for which agricultural data were missing. Urban areas in rural zones are included in the analysis, as it is expected that agriculture is the dominant livelihood strategy for the populations there.

Fig. 1 shows poverty headcount estimates for the rural aggregated EAs and their standard errors. The weighted mean poverty headcount in the rural aggregated EAs is 65.7%. While the 95% confidence interval mean is ±16.0 and the median is ±14.4 percentage points, 10% of all rural aggregated EAs have a confidence interval exceeding ±25.5 percentage points. However, while recognizing the presence of large numbers of outliers, the error terms for most of the estimates are reasonable.

Selection of independent variables

The risk-chain framework guided selection of the independent variables. From all spatial data sets available for Malawi we created a subset of potential independent variables for the analysis. A necessary characteristic of these variables is that they could be aggregated meaningfully and display variation across the country at the aggregated EA scale.

Table 1 describes the 26 independent variables selected for the analysis. They are categorized by their general nature and a priori assessments are provided both as to the position each is assumed to play in the risk chain of economic vulnerability and as to the nature of its relationship to the level of poverty prevalence in a rural aggregated EA. Note that we judged several of the variables to be both a risk factor and a coping factor. For example, good agricultural soils imply lower risk of crop failure and more reliable recovery from a shock to household welfare. For several variables, the assumed relationship between the level of the independent variable and that of the dependent variable is not clear a priori.

Several of the independent variables require additional comment. The GINI variable, like the poverty headcount dependent variable, is a product of the poverty mapping exercise. However, we argue that this variable is relatively independent of the poverty headcount measure since it describes the distribution of welfare across the population and is not tied to the poverty line. Its relationship to poverty prevalence is unclear a priori.

The CHEWA_YAO variable serves as a proxy for matrilineality because the Chewa and the Yao are the largest matrilineal ethnic groups in Malawi. Inheritance patterns and associated property rights are among the social institutions that may have developed, among other reasons, to enhance the ability of populations to cope with economic shock. This variable assesses whether, given social trends in recent generations that privilege patrilineal systems, there might be evidence that the matrilineal inheritance system is now dysfunctional in safeguarding welfare.

In the same vein, the OLDPARTY variable points to the role of political organization as a characteristic of economic vulnerability. The Malawi Congress Party (MCP), while not in power at the time of the survey and census, held power in Malawi from 1964 until 1994. Those areas of the country that continued to support the MCP at elections 5 years after the party fell from power may have been motivated by the particular welfare benefits that they had enjoyed through their relatively close association with the former ruling party.

Several issues relating to these independent variables should be highlighted. First, economic vulnerability is a dynamic concept in that it reflects the potential impact on welfare of shocks now and in the future. In contrast, poverty status is a static concept, representing the welfare state of a household or individual at a particular point in time. Our dependent variable is a static poverty measure based on two cross-sectional data sets, the 1997-98 Malawi IHS and the 1998 Census. Moreover, many of the spatial data sets that we employ to account for the determinants of aggregate poverty are themselves cross-sectional and static. Incorporating temporal elements into spatial variables is challenging. We have specifically included spatial variables that either measure the annual variability in a phenomenon or compare the level of a factor at the time of the IHS and the census to its long-term mean. However, we were only able to do so for crop yields and for rainfall. Overall we cannot claim to provide substantive insights on how spatial variables might be altered to reduce the degree of economic vulnerability of households in rural Malawi. The principal contribution that this analysis makes is to identify spatial factors that explain some of the variation in aggregate welfare outcomes. To better understand how these factors contribute to or alleviate household economic vulnerability, they would need to be examined within a dynamic context in which household welfare is traced through time.

Second, the exogeneity of all of the independent variables selected is questionable. Endogeneity arises at two levels. First, some of the independent variables are likely collinear with variables used in some of the poverty mapping models used to estimate the dependent variable. Moreover, poverty status is implicated in the effectiveness with which households can cope with economic shock. The level of several of the independent variables is related to some extent to the relative number of poor individuals resident in an aggregated EA.

Third, in this spatial analysis we are drawing on data that were developed at several different scales. Pooling data from different scales in an analysis poses the risk of the ecological fallacy of drawing inferences about smaller analytical units from the aggregate characteristics of groups of those units. For the analysis here, we are fortunate in having an extensive set of spatial data for Malawi that was collected at more local scales than that of the aggregated EA. However, the agricultural production data are an exception, so any inferences drawn on the basis of these data will necessarily have some error associated with their aggregated character.

Finally, the quality of the data from which we constructed these variables is not uniformly high. While, given the large number of sample points, any outliers likely do not strongly affect the results obtained, they do signal caution. Furthermore, the dependent variable itself is drawn from a survey data set that requires care in analysis.

Analytical methods

To model the prevalence of poverty as a function of spatial variables selected on the basis of the risk chain, we carried out two different analyses: (1) spatial regression to develop a single global model and (2) geographically weighted regression (GWR) to develop local models.

Spatial regression

In this analysis, a preliminary assessment consisted of a simple Ordinary Least Squares (OLS) regression:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{y}$ is a vector of observations on the dependent variable, $\mathbf{X}$ is a matrix of independent variables, $\boldsymbol{\beta}$ is a vector of coefficients and $\boldsymbol{\varepsilon}$ is a vector of random errors. Using OLS, we initially developed a single global model. However, a critical concern here is violation of the OLS assumption that error terms not be spatially correlated with each other, as evidenced by observations from locations near to each other having model residuals of a similar magnitude. The Moran's I statistic is used to assess spatial autocorrelation in the residuals.

In order to control for spatial autocorrelation, a spatial lag variable can be inserted into the model as a supplementary explanatory variable. This is the weighted mean of a variable for neighbouring spatial units of the observation unit in question. For the dependent variable, the spatial lag variable is generally written as $\mathbf{W}_\mathrm{y}$, where $\mathbf{W}$ is the spatial weights matrix that identifies neighbouring spatial units.

The spatial dependence in the regression model can be modelled in two different ways. First, as a spatial lag model:

$$\mathbf{y} = \rho \mathbf{W}_\mathrm{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

similar to the OLS equation above but with the addition of the $\mathbf{W}_\mathrm{y}$ spatial lag of the dependent variable, which takes the coefficient ρ. Such a model would be used if it were judged that the level of the dependent variable in neighbouring areas affects the level of the dependent variable in the area in question.

Alternatively, the spatial dependence can be attributed to the error term of the model and modelled as a spatial error model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{where } \boldsymbol{\varepsilon} = \lambda \mathbf{W}_{\varepsilon} + \varepsilon \tag{3}$$

Here, the error term is disaggregated into the spatial lag of the error term of neighbouring aggregated EAs, with coefficient λ, and the residual error term for the spatial unit in question. Such a model would be used if it were judged that there was a missing spatial variable for the model that affects an aggregated EA and its neighbours in a similar manner (Anselin, 1992).

Although the two models result from different interpretations of the process accounting for the spatial dependence, in practice, they usually differ very little. In order to choose which to use, a Lagrange Multiplier test is used to assess the statistical significance of the ρ and λ coefficients in each model, respectively. The preferred model is that with the highest test value (Anselin and Rey, 1991).

The choice of spatial weights matrix employed in the analysis is an important analytical decision for which there is little formal guidance (Anselin, 2002). Here we undertook a sensitivity analysis of the results obtained using different weighting schemes and made our choice, a first-order Queen's contiguity-based weighting matrix, based on the resultant explanatory power of the model and the ease of interpretation of the results in light of the spatial weighting scheme.

Geographically weighted regression

In using spatial regression models we assume that the spatial process accounting for poverty headcount levels is the same across rural Malawi. That is, the relationship is spatially stationary. While such an assumption might be reasonable with physical processes governed by universal physical relationships, at least at the generalized level here, few social processes will be found to be so constant over space (Fotheringham et al., 2002). Global models will hide this potential heterogeneity, or spatial non-stationarity, in the determinants of the prevalence of poverty.

GWR provides a method to assess the degree to which the relationship between the potential determinants and the prevalence of poverty varies across space. The method produces local models for each rural aggregated EA in our data. This is done by constructing a spatial weighting matrix and running a weighted regression for each rural aggregated EA.

The global OLS regression model (Eq. 1) can be rewritten as:

$$y = a_0 + \sum_j x_{ij} a_j + \varepsilon \qquad (4)$$

where y is the dependent variable, x is the independent variable, a is the regression coefficient, i is an index for the location, j is an index for the independent variable and ε is the error term. This can be reworked as a local regression model to become:

$$y_i = a_{0i} + \sum_j x_{ij} a_{ij} + \varepsilon \tag{5}$$

in which location dependent coefficients are estimated (Minot et al., 2003). For each location, the neighbouring observations used to estimate the model are chosen and the importance of each for the estimation procedure is weighted based on a distance-based spatial weighting matrix.

The spatial non-stationarity of the relationship of each independent variable to the dependent variable can be assessed to determine whether the GWR method offers any improvement over a global regression model. The variability in the observed GWR estimates for the spatial units is compared to the variability of the GWR results from a large number of random allocations of the analytical data across the units. Where one finds a significant difference between the variability of an observed estimate to those computed using the randomized data, spatial non-stationarity for that independent variable is indicated (Fotheringham et al., 2000).

Spatial autocorrelation and the use of spatial lag variables to control for the autocorrelation do not come into GWR analysis, making the results somewhat easier to interpret in this regard. Spatial autocorrelation is not ignored. However, rather than controlling for spatial dependency, the GWR analysis attempts to explain the nature of this spatial dependence as part of the local analysis (ibid., p. 114-115).

The GWR procedure provides a deluge of information, R^2 values for each spatial unit, coefficients and *t*-statistics for each independent variable, residuals, and so on. Information management in employing the GWR method is most efficiently done using maps.

Results

Spatial regression model

First, we undertook an OLS regression of poverty headcount on the set of independent variables presented in Table 1. The adjusted R^2 for the OLS model is 0.2856, indicating that much of what determines the level of poverty found in rural aggregated EAs goes unexplained by this model. Moreover, spatial autocorrelation in the model residuals calls into question the validity of the OLS model (Moran's I statistic of 0.5392, $p \leq 0.001$).

We used a spatial regression model to control for this spatial autocorrelation. We chose which spatial dependence model to use (spatial lag or spatial error) using Lagrange Multiplier tests. Although both models exhibited significant spatial dependence, we used the model with the highest test statistic, in this case, the spatial error model.[7]

Table 2 shows results of the spatial error model. The explanatory power of the model increases considerably over the OLS regression, with an unadjusted R^2 of 0.6777. Eight independent variables are significant. Here we review the results by classes of independent variables.

For the agro-climatological and natural hazards variables, only the variable specifying rainfall in the 1997-98 season being higher than normal is just significant and is associated with a lower prevalence of poverty. Higher yields due to increased rainfall during the survey period may be reflected in higher consumption levels at that time.

For the agriculture and livelihood variables, average maize yield is a significant determinant of poverty prevalence. However, contrary to expectations, the coefficient is positive, implying that areas with higher maize yields on average will have higher levels of poverty. This may be a result of in-migration and consequent unprofitably small landholding sizes in these areas of high agricultural potential. The crop diversity and the importance of non-agricultural economic activities variables are also significant.

Surprisingly, all access to services variables are insignificant. Possibly threshold effects operate that govern the effect of access to services on welfare. Alternatively, the welfare effects may only appear in interaction with other variables, such as specific livelihood strategies.

Of the demography variables, only the aggregate dependency ratio is a significant determinant of poverty prevalence. The population density variable is weakly significant ($p \leq 0.10$ level), showing lower poverty prevalence in rural areas with higher population density. This fact complicates our understanding of the counter-intuitive results on maize yield.

For the educational determinants, only average maximum educational attainment is a significant determinant of poverty prevalence. For the other variables, the Gini coefficient of consumption inequality and the CHEWA_YAO proxy for matrilineality are significant. Higher consumption inequality is shown to result in a lower prevalence of poverty. The positive coefficient on the matrilineality proxy suggests higher levels of poverty when a greater proportion of the population follows a matrilineal inheritance system.

The policy implications that we can draw from these results are relatively few and not surprising:

- Irrigate to assure adequate moisture for crops. However, the economics of irrigation in smallholder agriculture poses an important challenge to its profitable use.
- Encourage crop diversification and rural non-farm livelihood strategies.
- Educate the population to the highest level feasible.

It is unclear what actions could be taken in light of the significant but positive association between average maize yields and poverty levels and the significant GINI and matrilineal variables, beyond simply being aware that these factors may interact with whatever actions are taken, forcing modifications if they are to be effective.

Geographically weighted regression

For the GWR we used the same dependent and independent variables as in the previous analysis and a spatial weighting scheme of the 347 nearest neighbours to each aggregated EA, chosen using an optimization procedure.[8] The global adjusted R^2 for the GWR is 0.6993 (0.7452 unadjusted), providing a small improvement over the spatial error model (0.6777 unadjusted). Fig. 2 presents the local R^2 statistic for each rural aggregated EA. Those areas with the lowest R^2s are relatively diverse agro-ecologically and have no obvious socio-economic commonalities. No missing spatial variables for the model are immediately apparent from this pattern.

Turning to the specific estimates of the strength and nature of the local relationship between the determinants and the prevalence of poverty in rural aggregated EAs, as each variable will have 3004 separate coefficients, standard presentations of regression results are difficult to make. Table 3 describes the distribution of the coefficients for all independent variables.

The model results of the GWR can be interpreted in two ways. Those interested in a particular local area in Malawi can use the complete model results for that place to get a multivariate understanding of key local determinants of the level of poverty. We will not do that here. Rather, the second manner in which to examine the results is by considering for each determinant the varying nature across rural Malawi of the relationship (positive, negative or insignificant) between the determinant and local levels of poverty. Doing so will allow us to develop hypotheses on why the global patterns suggested in the spatial error model are not necessarily replicated in the GWR analysis, what might account for counter-intuitive spatial patterns in the parameters and how this analysis might inform efforts to aid households and individuals raise their welfare levels.

Fig. 3 presents only selected results for the GWR analysis. Four variables are chosen because they were shown to be important determinants in the global spatial error model. The fifth, the hospital access variable, is chosen because although insignificant in the global model, improving access to services is a common approach of poverty reduction efforts. The top map in each pair is of the value of the independent variable, while the bottom map portrays the statistical significance and sign of the t-statistic of the coefficient for the variable across rural aggregated EAs, but not the value of the coefficient itself. In the lower map, a three-category legend is used, with legend category breaks at a t-value of ± 1.96 ($p \leq 0.05$) level.

The GWR model intercept term shows how the local prevalence of poverty will differ from the overall mean when all independent variables are held constant. Just as the local R^2 map might point to missing variables, so too with the map of the intercept. Somewhat lower levels of poverty than can be explained by the determinants are found in a band running along the upland plateau area where tobacco is grown. However, whether a tobacco-production factor might be a missing variable for these local models will require additional investigation.

The five selected determinants show that the results of the global model mask considerable heterogeneity in the nature of the relationship between the determinant and the estimated poverty prevalence in small rural populations.

Higher average maize yields tend, non-intuitively, to result in higher poverty levels. This pattern was seen in the global model. Exceptions to this pattern are seen near Lilongwe, Zomba and Blantyre urban centres, where likely urban food market demand enhances the welfare benefit farmers derive from higher productivity.

The variable on non-agricultural economic activities, PCT_NOT_FA, shows a consistent pattern nationally—in virtually no areas does greater participation by the local population in non-agricultural economic pursuits result in a higher prevalence of poverty. Nevertheless, the relationship is not spatially stationary as for a large proportion of the rural population this variable is an insignificant determinant of poverty levels.

The access to hospital and other district services variable, HOSP_HR, highlights the poverty effects of poor access in northern Malawi, in particular. In comparison to the other access variables analysed, this variable is significant over most of rural Malawi, suggesting that access to district-level services is the most critical form of access to services necessary to enhance aggregate welfare. However, this pattern of inaccessibility to district-level services being associated with higher poverty is not uniform.

Education is frequently advocated as a cure for poverty. Consequently, it was expected that the MAXED variable would be significant and negative in the global model. However, in the local analysis, considerable variation in this relationship is seen. The north of the country, in particular, sees a strong positive association between education and poverty. This implies that the relatively well-educated population there is unable to derive any significant welfare benefit from the knowledge they have gained—education is not sufficient in itself to reduce poverty. However, elsewhere, higher general schooling levels are shown to be important in reducing the local incidence of poverty.

Finally, concerning consumption inequality, the broad global pattern of a negative association with poverty levels over most of the country is observed. However, there are unexplained exceptions to this pattern, most notably in the mid-altitude, tobacco areas of Kasungu, Ntchisi and Dowa Districts.

The final column of Table 3 gives spatial non-stationarity assessment results for the independent variables. Of the 26 variables, 18 have a statistically significant probability of their relationship with poverty prevalence being spatially non-stationary. It is primarily the demographic variables that are spatially stationary. This is an interesting result, given our earlier assertion that social processes can be expected to be spatially non-stationary. However, it should be noted that the strength of the relationship of most of these spatially stationary variables in the global model is weak. Generally, this assessment of spatial non-stationarity provides strong support for the use of local models of the determinants of poverty prevalence in designing poverty reduction policies and programmes in rural Malawi.

The guidelines that can be drawn from the GWR analysis for action on the determinants assessed are particularly dependent upon whether or not the relationship of the determinant to the local prevalence of poverty is shown to be spatially stationary. If stationary, as found for most of the demographic variables, and most notably for the road density variable, then a single national approach to modifying local conditions for these variables can be adopted. However, for the others, geographically designed and targeted approaches to change local conditions so that they are more conducive to reducing the local level of poverty will be needed. Which approach is used in a particular locale will depend upon the locally varying relationship between the determinant(s) addressed by a particular action and poverty prevalence. For example, as shown in the *t*-statistic map for MAXED in Fig. 3, efforts to improve general levels of educational attainment will be of greater value in reducing poverty in the southern lakeshore area and in the northern districts of the Central region, than in those areas where the model shows the puzzling positive association between educational attainment and the prevalence of poverty. Similar guidance could be drawn from the maps of many of the other independent variables.

Conclusions

The two models provide somewhat different results. The spatial error model produced global results that one might use with confidence. The set of determinants shown to be significant is relatively restricted. For several of these, the nature of their relationship to the prevalence of poverty was in line with expectations. However, determinants for which we did not have any strong theoretically based expectations also were shown to be significant. Understanding the reasons for the processes that account for these determinants featuring in the model remains a challenge. Finally, the variable on average maize yields was significant but the nature of its relationship to the dependent variable was counter to expectations.

The GWR analysis produced strong evidence that the determinants of poverty prevalence vary spatially in their effects across rural Malawi. The results might most easily be employed to guide quite local action to reduce poverty by examining the local model of the prevalence of poverty for a specific locale.

From the standpoint of guiding broad action to reduce poverty, overall the analyses had quite low explanatory power. In the global spatial error model, most of the more than 24 determinants that we selected for analysis proved non significant. In contrast, most of these determinants were significant in at least some rural areas in the GWR analysis. The implication is that poverty reduction efforts in rural Malawi will need to be targeted at the district and subdistrict levels. A national, relatively inflexible approach to poverty reduction is unlikely to enjoy broad success.

Perhaps more so than with the other determinants considered in our analysis, the agro-ecological variables provided an unclear picture. The strongest relationship observed is that those populations in which non-agricultural livelihood strategies can be widely pursued have fewer poor. The other consistent relationship is that areas in which higher maize yields are attained are also areas where poverty is more prevalent. The six or more other agro-ecological variables examined generally proved to have a weak relationship to poverty prevalence.

There is little evidence in the analysis to permit one to argue that the poor in Malawi are trapped in areas of low agricultural productivity, subject to frequent drought and farming on poor soil. The poor are throughout Malawi, on the best land and the worst land, in areas of relatively high productivity and of low productivity. Extending this idea, we noted that poverty and food insecurity in rural Malawi are closely linked. The fact that agriculture is shown to be positively associated with poverty also implies that agriculture, if not a source of food insecurity, is not serving as an effective means of reducing food insecurity. Subsistence farming dominates the rural economy of Malawi but the evidence here is that such farming is not providing a reliable and sufficient livelihood for most. Moreover, this dismal relationship is not found in isolated pockets but is the dominant pattern observed.

On the role of access to services and infrastructure as a spatial determinant of the prevalence of poverty, the results were less clear than we expected. The most important determinant is travel time to the nearest hospital, a variable that we interpreted as a proxy of access to district-level services. Access to more local services such as at subdistrict markets or to regional services at the larger markets and urban centres were less important as determinants of poverty levels.

Enhancing access to district-level services is a policy prescription emerging from this analysis.

Human capital development, particularly through education, finds support in this analysis. However, the local model shows that the relationship between education and reduced poverty is more complex than we might think. Broad areas of northern Malawi show that higher education is associated with higher poverty. The welfare returns to increased education are not linear in all circumstances. Our findings point to the need to determine just what the necessary circumstances are for increased educational attainment in an area to always result in higher generalized welfare.

Finally, in making use of the results, we must caution about the ecological fallacy of drawing inferences about smaller analytical units from the aggregate characteristics of groups of those units. Our analysis here is of the aggregate characteristics of populations resident in rural aggregated EAs. Consequently, in using this analysis to plan poverty reduction activities, it is important not to assume that the nature of the relationships observed here will be replicated at the level of the household or individual. The aggregate likely masks heterogeneity in characteristics of individuals and households that would render any action at those levels undertaken on the basis of the analysis here to be irrelevant or even harmful for individuals and households targeted. Our analysis is most useful in guiding broad community and other subdistrict level action.

References

Alwang, J., Siegel, P. B., Jørgensen, S. L., 2001. Vulnerability: a view from different disciplines. Social Protection Discussion Paper no. 0115, Human Development Network, World Bank, Washington, DC.

Anselin, L. 1992. Spatial data analysis with GIS: an introduction to application in the social sciences. Technical report 92-10, National Centre for Geographic Information and Analysis, University of California, Santa Barbara.

Anselin, L., 2002. Under the hood: issues in the specification and interpretation of spatial regression models. Agricultural Economics, 27:247-267.

Anselin, L., 2003. GeoDa 0.9 User's Guide. June 15, 2003 revision. Spatial Analysis Laboratory, University of Illinois and Centre for Spatially Integrated Social Science, Urbana-Champaign, Illinois.

Anselin, L., Rey, S., 1991. The performance of tests for spatial dependence in linear regression. Technical report 91-13, National Centre for Geographic Information and Analysis, University of California, Santa Barbara.

Benson, T., Kaphuka, J., Kanyanda, S., Chinula, R., 2002. Malawi – an atlas of social statistics. International Food Policy Research Institute, Washington, DC and National Statistical Office, Zomba, Malawi.

Deaton, A., Zaidi, S., 2002. Guidelines for constructing consumption aggregates for welfare analysis. Living Standards Measurement Study Working Paper no. 135. World Bank, Washington, DC.

Demombynes, G., Elbers, C., Lanjouw, J., Lanjouw, P., Mistiaen, J., Özler, B., 2002. Producing an improved geographic profile of poverty: methodology and evidence from three developing countries. Discussion Paper no. 2002/39, United Nations University/World Institute for Development Economics Research, Helsinki, Finland.

Dercon, S., 2001. Assessing vulnerability to poverty. Paper prepared for the Department for International Development, Economics Department, Oxford University, Oxford.

Elbers, C., Lanjouw, J., Lanjouw, P., 2000. Welfare in villages and towns: micro-measurement of poverty and inequality. Tinbergen Institute Discussion Papers 2000-029/2, Amsterdam.

Elbers, C., Lanjouw, J., Lanjouw, P., 2003. Micro-level estimation of poverty and inequality. Econometrica 71(1), 355-364.

Elbers, C., Lanjouw, J., Lanjouw, P., 2005. Imputed welfare estimates in regression analysis. Journal of Economic Geography 5(1), 101-118.

Foster, J., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 52(3):761-765.

Fotheringham, A. S., Brunsdon, C., Charlton, M., 2000. Quantitative geography: perspectives on spatial data analysis. SAGE Publications, London.

Fotheringham, A. S., Brunsdon, C., Charlton, M., 2002. Geographically weighted regression – the analysis of spatially varying relationships. John Wiley, Chichester.

Minot, N., Baulch, B., Epprecht, M., 2003. Poverty and inequality in Vietnam:

spatial patterns and geographical determinants. International Food Policy Research Institute, Washington, DC and Institute of Development Studies, Brighton.

NEC (National Economic Council), 2000. Profile of poverty in Malawi, 1998 – poverty analysis of the Malawi Integrated Household Survey, 1997-98. Government of Malawi, Lilongwe.

---

[1] We use the term "determinants" for our independent variables because they were selected on the basis of a theory of the determinants of household welfare – the risk-chain concept. The analysis here was not done to simply identify correlates of poverty prevalence. Rather, we are interested in examining the strength and nature of the relationship between what theory suggests might be potentially important determinants of household welfare and local poverty prevalence in rural Malawi.

[2] In this report, we use poverty headcount, the prevalence of poverty, $p0$, and FGT_0 interchangeably. All mean the proportion of the population whose level of welfare is below the poverty line. Formally, the measure is one of three Foster-Greer-Thorbecke poverty measures – the other two being the depth and the severity of poverty measures (Foster et al., 1984).

[3] Elbers et al. (2005) assess the use of imputed welfare estimates, particularly those from poverty mapping analyses, in regression analyses. Some caution in the interpretation of significance levels of coefficients is necessary when such estimates are used as the dependent variable for a model, as here – one should be somewhat conservative in the interpretation of significance. However, Elbers et al. find that such estimates can be used as independent variables in regressions in a relatively straightforward manner, as in the use of the GINI variable in this analysis.

[4] We are grateful to the Commissioner for Statistics of the Malawi National Statistical Office for allowing us to use these data here. We also thank his staff for their efforts in developing the census

data set.

[5] A later assessment in several countries in which poverty maps have been developed led to an upwards revision in this threshold but still shows reasonably precise poverty headcount estimates for populations down to about 1000 households (Demombynes et al., 2002).

[6] Note that neither the EA nor the aggregated EA geographies are administrative units. Although their boundaries respect administrative boundaries, the units are established by the National Statistical Office purely for data collection purposes.

[7] Lagrange Multiplier test results are not presented here. We developed and assessed the spatial regression models using GeoDa 0.9 software (Anselin, 2003).

[8] We developed and assessed the GWR models using GWR 3.0 software (see Fotheringham et al., 2002, ch. 9).

Figure Legends

Fig. 1. Poverty mapping showing (A) poverty headcount (*p*0) estimate (%) and (B) standard error of *p*0 estimate for rural aggregated enumeration areas in Malawi.

Fig. 2. Local R^2 from the geographically weighted regression of the determinants of poverty prevalence for rural aggregated enumeration areas in Malawi.

Fig. 3. Maps of selected independent variables and *t*-statistics for each from geographically weighted regression analysis of the determinants of poverty prevalence for rural aggregated enumeration areas in Malawi.

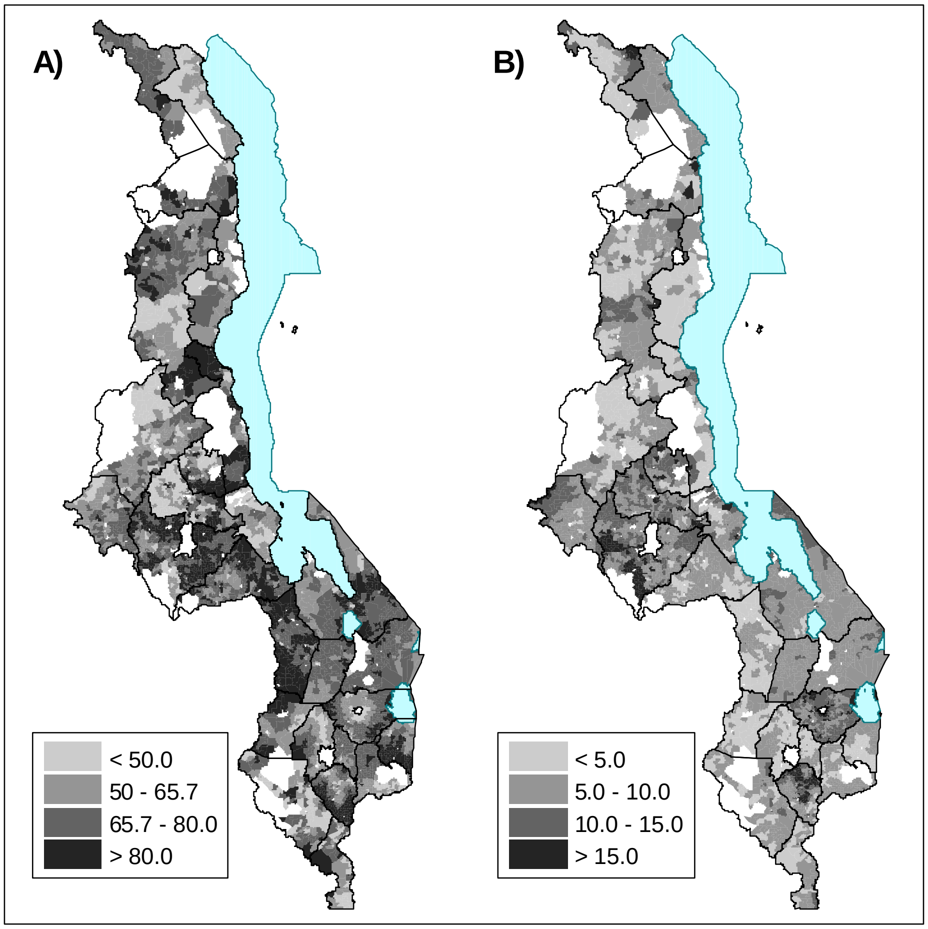
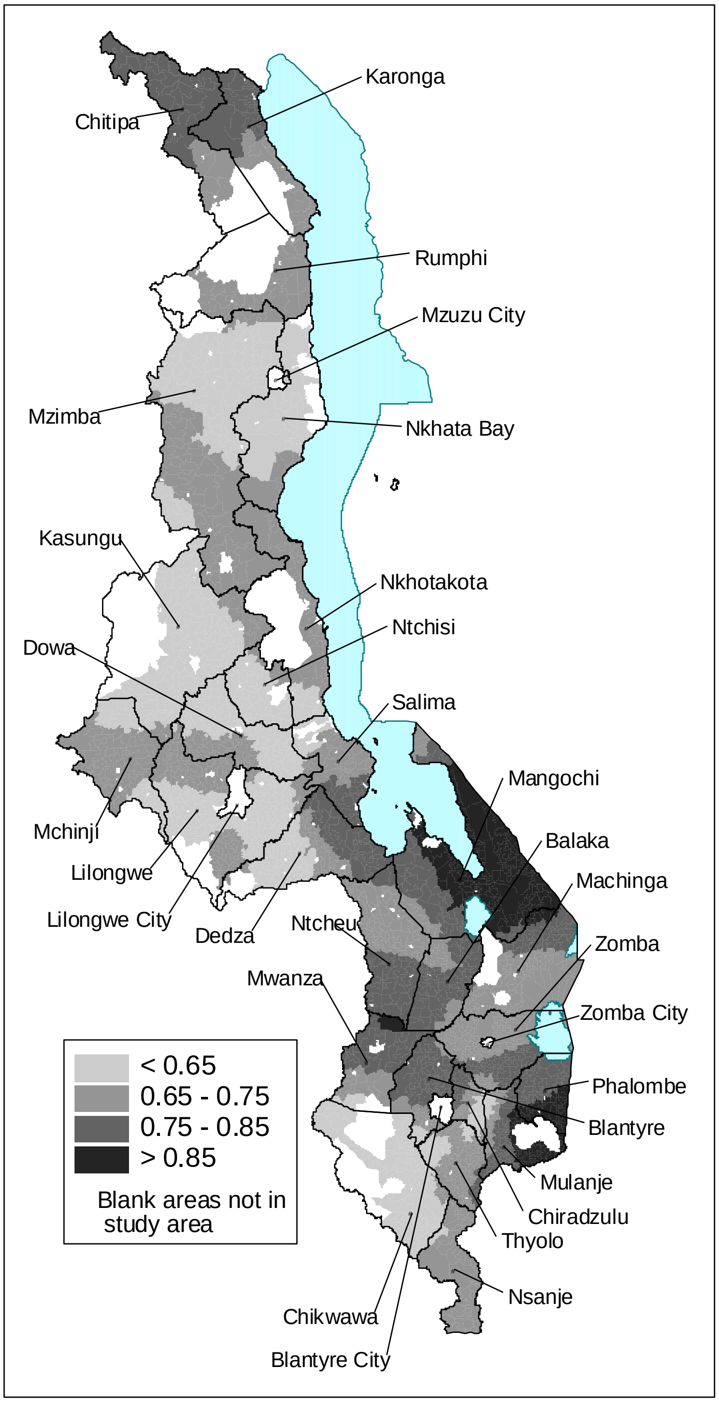
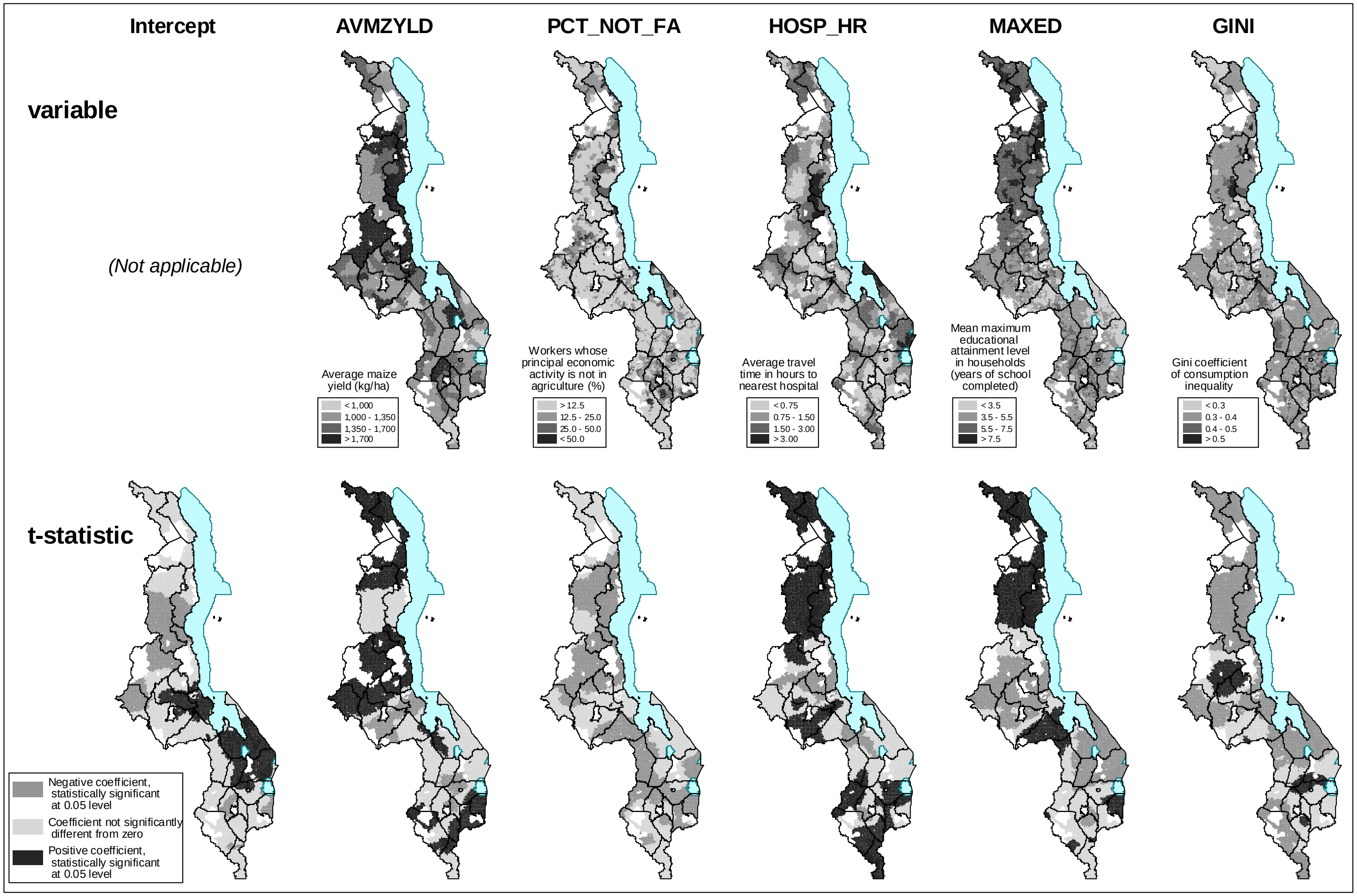


Fig. 1



Karonga
Chitipa
Rumphi
Mzuzu City
Mzimba
Nkhata Bay
Kasungu
Nkhotakota
Ntchisi
Dowa
Salima
Mangochi
Mchinji
Balaka
Lilongwe
Machinga
Lilongwe City
Dedza
Ntcheu
Zomba
Mwanza
Zomba City
< 0.65
0.65 - 0.75
0.75 - 0.85
> 0.85
Blank areas not in study area
Phalombe
Blantyre
Mulanje
Chiradzulu
Thyolo
Nsanje
Chikwawa
Blantyre City

Fig. 2



Intercept
AVMZYLD
PCT_NOT_FA
HOSP_HR
MAXED
GINI
variable
(Not applicable)
Average maize yield (kg/ha)
< 1,000
1,000 - 1,350
1,350 - 1,700
> 1,700
Workers whose principal economic activity is not in agriculture (%)
> 12.5
12.5 - 25.0
25.0 - 50.0
< 50.0
Average travel time in hours to nearest hospital
< 0.75
0.75 - 1.50
1.50 - 3.00
> 3.00
Mean maximum educational attainment level in households (years of school completed)
< 3.5
3.5 - 5.5
5.5 - 7.5
> 7.5
Gini coefficient of consumption inequality
< 0.3
0.3 - 0.4
0.4 - 0.5
> 0.5
t-statistic
Negative coefficient, statistically significant at 0.05 level
Coefficient not significantly different from zero
Positive coefficient, statistically significant at 0.05 level

Fig. 3

Table 1
Variables selected for analysis of spatial determinants of poverty prevalence by rural aggregated enumeration area (EA)

Name	Definition[a]	Assumed risk chain link position	Assumed relationship to poverty prevalence[b]	Descriptive statistics	
				Mean	SD
Dependent variable					
FGT_0	Poverty prevalence (as a proportion) in rural aggregated EAs	Outcome	n/a	0.661	0.168
Agroclimatological					
CLIOPT5PRE	Avg. rainfall (mm) in 5 mo. following precipitation to potential evapotranspiration ratio triggered plant date	Risk	Negative	913	272
CVRAIN	Avg. rainfall coefficient of variation during rainy season (Dec.-Mar.), percentage, (100 * [*s* / mean])	Risk	Positive	24.5	2.8
HIRAIN9798	In highest quintile of rainfall deviation from long-term mean in 1997-98 season (0/1) – much higher rainfall than avg.	Risk	Unknown or negative	0.200	0.400
LORAIN9798	In lowest quintile of rainfall deviation from long-term mean in 1997-98 season (0/1) – much lower rainfall than avg.	Risk	Positive	0.200	0.400
Natural hazards					
FLOOD	Dominant soils subject to flooding (0/1)	Risk	Positive	0.046	0.210
STEEP	Steep slopes common (0/1)	Risk	Positive	0.204	0.403
Agriculture and livelihoods					
SOLGOODD	Dominant soils have relatively good agricultural potential, based on FAO soil classification (0/1)	Risk/coping	Negative	0.527	0.499
AVMZYLD	Mean maize yield (kg/ha), 1995-96 to 1999-2000	Risk/coping	Negative	1381	333
CVMAIZE	Maize yield coefficient of variation, 1995-96 to 1999-2000, (100 * [*s* / mean])	Risk	Positive	24.9	10.6
CROPDIVERS	Cropped area not in staple crop (%)	Risk/coping	Negative	0.443	0.127
PCT_NOT_FA	Workers whose principal economic activity not in agriculture (%)	Risk/coping	Negative	16.3	18.0
Access to services					
HOSP_HR	Avg. travel time (h) to nearest hospital – district-level services proxy	Coping	Positive	0.90	0.65
GAZ_AREA_H	Avg. travel time (h) to nearest major forest reserve or national park – access to common property resources proxy	Coping	Positive	1.57	0.92
MKT_ALL_HR	Avg. travel time (h) to nearest subdistrict market centre	Coping	Positive	0.77	0.61
MKT_1_HR	Avg. travel time (h) to nearest of six major regional markets – Blantyre, Lilongwe, Mzuzu, Zomba, Kasungu and Karonga	Coping	Positive	1.94	1.07
RD_WT_PAV	Avg. weighted road density (m/km^2), weighted by potential speed on different qualities of road	Coping	Negative	3286	1989
Demography					
MSXRT20_49	Sex ratio (modified), 20-49 y ([no. men per 100 women] – 100)	Coping	Negative	-10.9	15.4
DEPRATIO	Dependency ratio (total aged under 15 and > 65 y / total pop.)	Coping	Positive	0.484	0.028
FEMHHH	Households (HH) headed by women (%)	Coping	Positive	32.8	12.2
POPDENS	Population density (persons/km^2)	Risk/coping	Unknown	256	521
Education					
SEXDIFF_LI	Literacy rates differences between adult men and women (%)	Coping	Positive	21.6	8.3
MAXED	Mean max. educational attainment in HH (y. school completed)	Coping	Negative	5.1	1.5
Other					
ORPH_PREV	Those aged < 15 y having at least one parent dead (%) – proxy for general health status, adult mortality, level of care	Risk/coping	Positive	7.5	3.4
GINI	Gini coefficient of consumption inequality	Risk/coping	Unknown	0.352	0.055
CHEWA_YAO	Population with Chichewa, Chinyanja, or Chiyao as mother tongue (%) – proxy for matrilineality	Coping	Unknown	81.7	31.6
OLDPARTY	Parliamentarian from historical ruling party, Malawi Congress Party, elected from area in 1999 (0/1)	Coping	Negative	0.354	0.478

[a]Notation (0/1) in variable definition indicates that variable is a binary, dummy variable.
[b]Negative relationship to poverty prevalence indicates expectation of increases in determinant's value leading to poverty reduction.

Table 2
Results of spatial error maximum-likelihood estimation model on the determinants of poverty prevalence for rural aggregated enumeration areas in Malawi[a]

Variable	Coefficient	Std. Error	z-statistic[b]
Constant	0.37336	0.09399	3.97219**
λ – LAMBDA	0.79898	0.01240	64.45346**
CLIOPT5PRE	0.00005	0.00004	1.47527
CVRAIN	0.00228	0.00262	0.86813
HIRAIN9798	-0.02429	0.01231	-1.97327*
LORAIN9798	-0.01531	0.01155	-1.32620
FLOOD	-0.00297	0.01026	-0.28968
STEEP	0.00257	0.00601	0.42705
SOLGOODD	0.00171	0.00552	0.30937
AVMZYLD	0.00003	0.00001	2.34613*
CVMAIZE	0.00006	0.00042	0.14645
CROPDIVERS	-0.13085	0.03977	-3.29023**
PCT_NOT_FA	-0.00171	0.00022	-7.83898**
HOSP_HR	0.02158	0.01526	1.41423
GAZ_AREA_H	-0.00739	0.00898	-0.82355
MKT_ALL_HR	0.00906	0.01430	0.63321
MKT_1_HR	-0.00063	0.00999	-0.06270
RD_WT_PAV	0.00000	0.00000	-0.11647
MSXRT20_49	0.00015	0.00022	0.66903
DEPRATIO	0.64136	0.09686	6.62166**
FEMHHH	0.00040	0.00022	1.78551
POPDENS	-0.00001	0.00000	-1.71496
SEXDIFF_LI	0.00011	0.00028	0.40414
MAXED	-0.00720	0.00260	-2.77261**
ORPH_PREV	0.00128	0.00071	1.79468
GINI	-0.34611	0.04953	-6.98783**
CHEWA_YAO	0.00054	0.00017	3.16258**
OLDPARTY	-0.01505	0.01325	-1.13621

[a] Dependent variable: FGT_0; no. of observations: 3004; no. of variables: 27 + spatial error lag, which takes λ coefficient; R^2: 0.6777; Akaike information criterion: -5014.12.
[b] ** significant at $p \leq 0.01$ level, * at $p \leq 0.05$ level.

Table 3
Descriptive statistics of the coefficients for each independent variable for the geographically weighted regression models of the determinants of poverty prevalence for rural aggregated enumeration areas (EAs) in Malawi (n = 3004)

Variable	Minimum	Lower quartile	Median	Upper quartile	Maximum	Rural aggregated EAs with significant coefficient (%)		Spatial non-stationarity test sig. level[a]
						Negative	Positive	
Constant	-1.94981	-0.33394	0.10816	0.77342	2.89514	15.3	22.5	**
CLIOPT5PRE	-0.00063	-0.00006	0.00010	0.00024	0.00284	9.0	34.8	**
CVRAIN	-0.04102	-0.00633	0.00575	0.01658	0.04435	11.0	31.4	**
HIRAIN9798	-0.32874	-0.06148	-0.01496	0.00000	0.32003	25.0	4.1	**
LORAIN9798	-0.32573	-0.08776	-0.01043	0.01650	0.47181	30.7	11.8	**
FLOOD	-0.17801	-0.02738	0.00000	0.02224	0.45353	11.2	4.8	ns
STEEP	-0.24000	-0.00776	0.00730	0.02696	0.32526	0.2	15.8	**
SOLGOODD	-0.52372	-0.02072	0.00116	0.01985	0.30624	10.9	12.7	**
AVMZYLD	-0.00033	-0.00003	0.00006	0.00019	0.00057	15.7	42.1	**
CVMAIZE	-0.00922	-0.00224	-0.00008	0.00379	0.01634	24.5	26.8	**
CROPDIVERS	-1.74229	-0.27547	-0.04368	0.12295	1.46350	24.8	16.0	**
PCT_NOT_FA	-0.00587	-0.00249	-0.00153	-0.00103	0.00232	50.7	0.1	**
HOSP_HR	-0.33748	-0.01792	0.05296	0.10709	0.34218	10.9	43.3	**
GAZ_AREA_H	-0.33622	-0.03974	-0.00204	0.03655	0.14040	27.6	18.7	**
MKT_ALL_HR	-0.31533	-0.07275	-0.02747	0.01997	0.40941	26.5	12.4	**
MKT_1_HR	-0.36103	-0.04106	-0.00706	0.03964	0.15979	18.5	19.8	**
RD_WT_PAV	-0.00002	-0.00001	0.00000	0.00000	0.00003	14.1	4.4	ns
MSXRT20_49	-0.00248	-0.00070	-0.00001	0.00067	0.00296	2.5	7.6	ns
DEPRATIO	-1.01700	0.20705	0.50681	0.84924	1.90329	0.4	33.6	ns
FEMHHH	-0.00334	-0.00055	0.00026	0.00103	0.00438	7.6	11.1	ns
POPDENS	-0.00018	-0.00006	-0.00002	0.00000	0.00014	17.6	3.5	ns
SEXDIFF_LI	-0.00486	-0.00044	0.00037	0.00118	0.00343	1.7	6.3	ns
MAXED	-0.10897	-0.03926	-0.01032	0.01797	0.06743	41.1	23.3	**
ORPH_PREV	-0.01149	-0.00107	0.00136	0.00277	0.01066	1.2	4.6	ns
GINI	-1.48925	-0.85225	-0.28509	0.10171	1.19286	44.8	10.8	**
CHEWA_YAO	-0.01535	-0.00069	0.00029	0.00118	0.00913	5.8	8.8	**
OLDPARTY	-0.90323	-0.03388	0.00000	0.01500	0.39417	14.7	11.4	**

[a] For spatial non-stationarity test: 100 Monte Carlo simulations run; * significant at $p \leq 0.05$ level, ** at $p \leq 0.01$ level, ns = not significant.