

## **Using spatial, hierarchical, and econometric models in urban data-poor areas to examine food security**

Anna Carla López-Carr<sup>\*</sup>, David López-Carr<sup>\*\*\*</sup>, Laura Grant<sup>\*\*</sup>, John Weeks<sup>\*</sup>

### **Abstract**

In this paper we explore three models of food security in Accra, Ghana. We use survey data from the 2003 Women's Health Study in Accra and satellite imagery from 2002 Quickbird satellite imagery to examine socio-economic, spatial, and environmental predictors of food insecurity (defined as poorly nourished households). The spatial model highlights areas of particular concern "hotspots" with statistically significant values. The hierarchical model separated the relative effects of household versus neighborhood level variables. The econometric model emphasized economic trends among household based on estimated values of household wealth. Together, results suggest that, while the data source is the same, outcomes differ, highlighting the caution researchers must use when determining an appropriate statistical approach. The choice of statistical model may point the researcher towards a path of ecological fallacy or delightful parsimony. Together, these three models allow us to draw a more complete

---

\* San Diego State University Department of Geography, San Diego, CA. USA

\*\* University of California, Department of Geography, Santa Barbara, CA. USA

\*\*\* University of California, Department of Economics, Santa Barbara, CA. USA

picture of food security patterns in Accra, and to draw important and more comprehensive conclusions for policy recommendations.

## Introduction

In data poor areas, the use of statistical models is often determined by the quantity and quality of the data. However, statistical modeling may also be limited by inherent biases in a researcher's native discipline. Geographers, for example, would prefer the use of spatial models, population scientists often use hierarchical models, and economists economic. Here, we explore the pros and cons of three model outcomes, which allow us to evaluate the range of predictions and how they would significantly influence our research conclusions.

Food security constraints can be ordered in the following categories: *Socio-economic*, *Home Environment*, *Neighborhood Environment*, and *Health and Wellness*. The combination of these variables determines a household's access and utilization, assuming constant food supply. Urban food supply converges with household (or individual) access and utilization given these endogenous and exogenous factors. When both supply and demand intersect at the desired inflection point, urban households are food secure. When, conversely, observed demand is less than desired demand, due to a combination of the identified constraints, households are food insecure. We address the relative merits of applying spatial, multi-level, and econometric models to the following research question: What socio-economic, environmental, and spatial variables predict urban food insecurity?

## **Data and methods**

Using survey data from the 2003 Women's Health Study in Accra and from 2002 Quickbird satellite imagery, this study determined socio-economic, spatial, and environmental predictors of food insecurity (defined as poorly nourished households) in Accra. The richest source of health data for Accra comes from Harvard University's 2003 Women's Health Study of Accra (WHSA), (Duda et al, 2007) and generously made available by Co-Principal Investigators Allan G. Hill and Rosemary Duda of Harvard University. Data were collected from 3,200 women aged 18 and older between April and July 2003 and provide self-report health data, data from a clinical examination and laboratory work, as well as data on the household's facilities matched to the census of 2000. The Demographic and Health Survey of 2003 was also used to calculate population data more detailed than that available in the census. A representative sample was taken of 6,600 households residing in private homes throughout the country, using Enumeration Areas (EAs) as an initial sampling frame. In the case of women residing in Accra, 930 households were sampled with a response rate of 95.8 (DHS, 2003).

### **Research site: Accra, Ghana**

The study is focused on the urban area known as the Accra Metropolitan Area (AMA) (Figure 1), the most densely populated and urbanized area of the Greater Accra Region. AMA is the capital city of Ghana, a West African country located on the Gulf of Guinea. According to the Ghana 2000 Census, AMA is currently home to circa two million people, with a growth rate of 3%. Therefore, the population density is close to 10,000 people per square

kilometer.

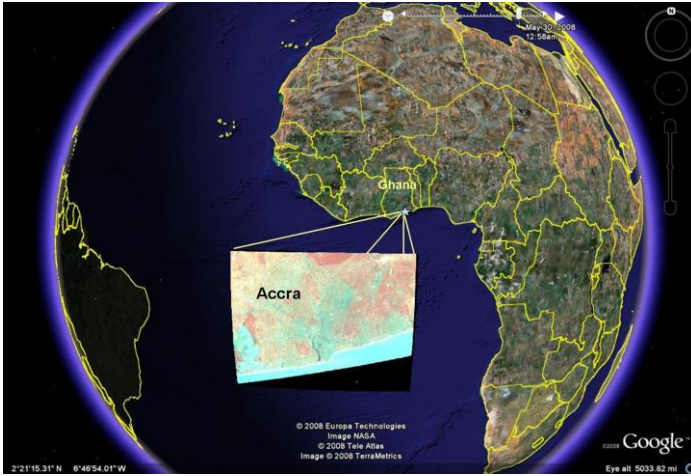


Figure 1 – Accra, Ghana (Sources, Google Earth, 2008; Digital Globe, 2002)

By the year 2020, over 3 million people are expected to live in Accra (UN Population Division, 2008). Though Accra is the heart of economic and political activity in Ghana, nearly 70 % of the city's population survives in slum-like conditions, without access to basic services such as water, sanitation, and health care (Weeks et al, 2007).

Survey data indicate that nearly 4% of the population in the Greater Accra region is underweight, and that 11.5%, 7.5%, and 11.4% of children in the same area are underweight, stunted, and wasted respectively (DHS, 2003). Nearly 65% percent of children consume an inadequate amount of Vitamin A, and another 60% percent are lacking sufficient quantities of iodine.

The average life expectancy for Ghanaians at birth is 59 years (UNICEF, 2008), a level not seen in the United States since the 1930s.

## Results

### *Econometric Model*

Income affects food security measures in many dimensions. Directly, wealth allows greater flexibility of food choices and stability through lean times. Indirectly, wealth may be correlated with other variables that also reflect food choices: location, education, marital status, and culinary amenities. The strategies for identifying the wealth effects build up from a basic model. Wealth is split into four income ranges. The first model reports the effects of the four incomes. In the second model, we add binary indicators for localities. While there is not a spatial network built into this regression design, the location indicators allow food security to vary in different places. Finally, we add personal and household characteristics: age, marital status, ownership of a refrigerator, and access to cooking fuel.

We use an aggregated measure of food access as the dependent variable, a binary indicator of food security. Typically limited dependent variables require a non-linear form, such as a logistic regression model. However, here we use ordinary least squares regression. The signs of the coefficients will conform to the alternative model. The limitation is predicted values that lie out of the bounds zero and one. The advantages are easily interpretable and comparable coefficients, particularly those of binary independent variables. In all cases, standard errors are clustered on location to control for common variation by place. The specification is

$$\begin{aligned}
 food\_secure_i = & \beta^K * income_i^K + \beta * married_i + \beta * age_i + \\
 & \beta * fridge_i + \beta * cook\_oil_i + \beta_i + \beta_i \quad (X)
 \end{aligned}$$

where  $i$  denotes a household.

This regression controls for the different levels of income through a series of indicators,  $income^K$ , where  $K = 1,2,3,4$  (and 1 = less than 300,000 cedís, 2 = 300,000 – 500,000 cedís, 3 = 500,000 – 1million cedís, and 4 = more than 1 million cedís). The vector of  $\beta^K$  is of primary interest – each coefficient is interpreted as the weighted average effect on food security of having income in range  $K$  relative to the lowest income group. Food secure indicates a positive attribute and we expect that increased income is related to higher probability of access to food. We are particularly interested in how these coefficients change as other covariates are included. Also, the relative weights of income and the other factors are important. The covariates include personal and household characteristics. The series of location fixed effects,  $\beta_i$ , measures characteristics of each place remaining constant for all households in that region.

Table 1 – Regression results

	-1	-2	-3	-4
<i>Income, mid-low</i>	0.006	0	-0.003	-0.011
	-0.871	-0.994	-0.94	-0.763
<i>Income, mid-high</i>	0.152	0.121	0.107	0.066
	(0.004)**	(0.005)**	(0.026)*	-0.15
<i>Income, high</i>	0.278	0.224	0.213	0.155
	(0.000)**	(0.000)**	(0.001)**	(0.016)*
<i>Location 401</i> (max)		0.51	0.355	0.278
		(0.000)**	(0.000)**	(0.000)**
<i>Location 410</i> (min)		-0.39	-0.427	-0.475
		(0.000)**	(0.000)**	(0.000)**
<i>Married</i>			0.043	0.032
			(0.068)+	-0.172
<i>Age</i>			0	0
			-0.649	-0.621
<i>Cookfuel</i>				0.074
				(0.033)*
<i>Refrigerator</i>				0.1
				(0.004)**
<i>Constant</i>	0.366			
	(0.000)**			
<i>Observations</i>	2310	2310	2104	2100
<i>R-squared</i>	0.037	0.17	0.183	0.196

Robust p-values in parentheses

+ significant at 10%; \* significant at 5%; \*\* significant at 1%

Table 1 gives the results in four columns. The first column shows the relative role wealth plays in food access without accounting for other attributes. Relative to low-income households, mid-low income does not significantly correspond to increased food security. However, once the threshold of mid-high income is reached, probability of being food secure increases by 15%. High income households are nearly 30% more food secure than low income ones. When locations are added, each coefficient of wealth decreases. We expect this result because neighborhoods are certainly correlated with income. However, the general results on wealth still hold. The coefficients for all 41 locations are not reported. Instead the table displays the range of maximum and minimum relative to Location 101. The range is considerable. The location with maximum food security is 51% more than Location 101; the place with lowest food security is almost 40% worse off. Column three adds personal characteristics of marital status and age to the preceding specification. Again, this addition diminishes the income effects but both mid-high and high income are significant at 11% and 21%, respectively. The coefficients on locations also change slightly, indicating differing patterns of personal characteristics over neighborhoods. Age is not statistically significant, but being married does increase the probability of being food secure by a small amount, 4.3% – married couples typically create more stable income and health. Column four includes households' appliances and cooking fuel use; both are binary indicator variables. These attributes swamp the effects of having mid-high income and reduce the effect of high income to 16%. Being married is also no longer significant. The effect of non-raw material cooking fuel (gas, kerosene, or electricity versus coconut husks or wood) improves food security by 7%, having a refrigerator by 10%. However, distributing

refrigerators is not a passable solution to food insecurity – the appliance requires consistent electricity, an added cost, and adequate space in the home. Similarly, cooking fuel is a reflection of owning a cooking appliance which requires electricity, gas, or kerosene as an input to function. Using wood or coconut husks for fuel implies outdoor cooking area or fire pit.

## Spatial Model

The local Moran test (Anselin, 1995), detects local spatial autocorrelation. It can be used to identify local clusters (regions where adjacent areas have similar values) or spatial outliers (areas distinct from their neighbors). Because the data consist of a sample of EAs and not the entirety of the units which comprise the map of Accra, Bayesian smoothing was applied to fill the “empty spaces” on the map. Bayesian. The algorithm works by predicting the probability of an event occurring in a non-sampled area by manipulating the observed rates of that event in neighboring areas (Cressie, 1995).

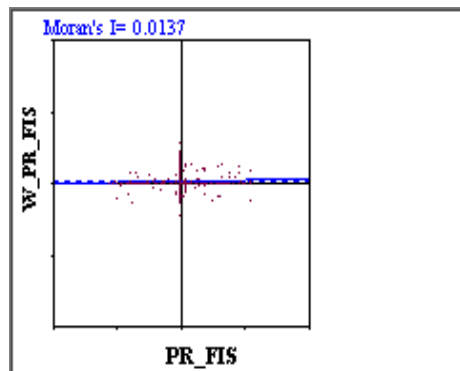


Figure 2 - Moran's I scatter plot for food insecurity per EA in Accra



This normal or Gaussian model is heteroskedastic and spatial, with parameters estimated using restricted maximum likelihood. Moran's I spatial autocorrelation statistic is visualized as the slope in the scatter plot with the spatially lagged variable on the vertical axis and the original variable on the horizontal axis. The slope of the regression line is Moran's I statistic, indicated at the top of the window. The four quadrants in the scatter plot correspond to different types of spatial correlation. Spatial clusters are found in the upper right (high-high) and lower left (low-low) quadrants, and spatial outliers in the lower right (high-low) and upper left (low-high) quadrants. Note that the magnitude of Moran's I as such does not indicate significance, nor are the statistics directly comparable across weights and variables. The scatter plot shows rates for food insecure households. Here, the data are located around the center intersect of the graph some points in both the upper right and lower-left quadrants. Because the upper-right and lower-left quadrants correspond to clustered data, we can conclude that there is some clustering in the data. To verify the local spatial pattern for food insecurity,  $G_i^*$  was employed where the value of the target feature was included. Z-scores for the statistic were mapped (Figure 2). Statistically significant polygons (greater than 1.96 or less than -1.96) indicated clusters of high and low values respectively. In this instance, only high valued clusters are manifest. Low rates of food insecurity have no significant clusters with like values. Instead, high rates of food insecurity have some clustering on the map.

## Multi-level Model

A two-level random intercept logistic multi-level model was developed in MLWin 2.01. The random intercept model allowed the overall probability of food security to vary

across EAs. The binary response was  $y_{ij}$  which equaled 1 if a household  $i$  in EA  $j$  was food secure, and 0 if it was not. Similarly, a  $j$  subscript was added to the proportion so that  $\pi_{ij} = \Pr(y_{ij} = 1)$ . If there is a single explanatory variable,  $x_{ij}$ , measured at the household level, then a two-level random intercept model will resemble the following:

$$\begin{aligned}\text{logit}(\pi_{ij}) &= \beta_{0j} + \beta_1 x_{ij} \\ \beta_{0j} &= \beta_0 + u_{0j}\end{aligned}$$

Here, the intercept consists of two terms: a fixed component  $\beta_0$  and an EA-specific component, the random effect  $u_{0j}$ . It is assumed that the  $u_{0j}$  follows a Normal distribution with mean zero and variance  $\sigma^2_{u_0}$ . All variables except for *vegetation* were entered at the household level. *Vegetation* was entered at the EA level.

The multi-level model results accounts appropriately for vegetation, a neighborhood level variable. In the equation below, the intercept ( $\beta_{0j}$ ) for EA  $j$  is  $-1.372 + u_{0j}$  where  $-1.372$  is the average rate of food security for each EA and the variance among EAs ( $u_{0j}$ ) is estimated as 0.382 (SE = 0.151). In other words, when all coefficients equal zero, the average food security rate in each EA is  $-1.372$  plus or minus 0.382. Because the intercept is below the x-axis and the variance is small, the likelihood of an EA containing a majority of food insecure households is greater than that of it having a majority of food secure households.

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_i$$

In the multi-level model, chances of household food security increase with ownership of a fridge, adequate bathing facilities, access to better solid waste collection services, better overall health, and infant breast feeding practices. Again, the variable *tenure* is ordered so the least secure situations have a higher ranking, resulting in a

negative output in the model. Therefore, chances of being food secure increase with better housing tenure. Consistent with the household-level model, in the multi-level model vegetation resulted in a positive outcome, indicating that chances of a household being food secure increase with greater proportions of vegetation at the EA level.

For discrete response models, the likelihood ratio test is

$$\begin{aligned} \text{foodsec}_{ij} &\sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \beta_{0j} \text{cons} + 0.581(0.116) \text{fridge}_{ij} + -0.205(0.072) \text{tenure}_{ij} + \\ &\quad 0.626(0.133) \text{bath}_{ij} + 0.918(0.158) \text{solid}_{ij} + 0.607(0.437) \text{veg}_{ij} + \\ &\quad 0.408(0.187) \text{brstfd}_{ij} + 0.497(0.125) \text{overhlth}_{ij} \\ \beta_{0j} &= -1.372(0.276) + \mu_{0j} \\ \left[ \mu_{0j} \right] &\sim N(0, \Omega_u) : \Omega_u = \left[ 0.382(0.151) \right] \end{aligned}$$

$$\text{var}(\text{foodsec}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

unavailable and a Wald test is a viable alternative. A Wald test was therefore carried out in MLWin although this test is approximate, as variance parameters are not normally distributed. The test statistic was 24.775, which we compare to a chi-squared distribution on (n-1) or 6 d.f. The returned p-value was 0.0003, a value of high statistical significance. Therefore we can conclude that there are significant differences between EAs in terms of food security, and that a multi-level approach to analysis was useful in extracting data that provides information for policy and in confirming the results of the household-level model.

## Conclusion

While the data source is the same, the outcomes are different, highlighting the caution researchers must use

when determining an appropriate statistical approach. The spatial model delivered vital information on the geographic distribution of food security across the urban landscape, highlight areas of particular concern “hotspots” with statistically significant values. Our use of the hierarchical model, or multi-level model, separated the effects of household versus neighborhood level variables, allowing us to distinguish the level at which variables were most influential. Lastly, our econometric model emphasized the economic trends among household based on estimated values of household wealth. The choice of statistical model may point the researcher towards a path of ecological fallacy or delightful parsimony. Together, these three models allow us to draw a more complete picture of food security patterns in Accra, and to draw important and more comprehensive conclusions for policy recommendations.

## References

- Anselin L. (1995), Local indicators of spatial association: LISA, *Geographical Analysis*, 27, pp. 93–115.
- Cressie N. (1995), Bayesian smoothing of rates in small geographic areas, *Journal of Regional Science*, 35, pp. 659–673
- Duda R., Kim M., Darko R., Adanu R., Seffah J., Anarfi J., Hill . (2007), Results of the women's health study of Accra: Assessment of blood pressure in urban women. *International Journal of Cardiology*, 117, pp. 115–122.
- United Nations (2001), *The State of the World's Cities 2001*. Kenya, UN HABITAT.
- United Nations (2003), *Global Report on Human Settlements*, Kenya, UN HABITAT.
- United Nations (2004), *The State of the World's Cities 2004/2005*, Kenya, UN HABITAT.
- United Nations (2006), *Statistics*. [www.unhabitat.org]
- United Nations (2007), *Population Information Network – POPIN* [www.un.org/popin/data.html]
- United States Census Bureau (2008) [www.census.gov]
- Weeks, J.R., Hill A., Stow D., Getis A., Fugate D. (2007), Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana. *GeoJournal* 69(9)