Responsible Land Governance:
Towards an Evidence Based Approach
ANNUAL WORLD BANK CONFERENCE ON LAND AND POVERTY
WASHINGTON DC, MARCH 20-24, 2017

Improving Land Valuation Models in Sparse Markets: A Comparison of Spatial Interpolation Techniques in Mass Appraisal **(Working Paper)**

Paul Bidanset (1); William McCluskey (2); Peadar Davis (1); Michael McCord (1)

Author contact: pbidanset@gmail.com

Organizations: 1: Ulster University, Newtownabbey, Northern Ireland; African Tax Institute, University of Pretoria, Pretoria, South Africa

Abstract

The research in this working paper has set up the groundwork for a national AVM of Malawi, Africa, using only secondary data collected by the 2010-2011 Integrated Household Survey. Model variables include physical characteristics of the property, economic variables, as well as location-specific distance and climate variables. This paper additionally helps bridge the current gap in development property tax literature by evaluating response surface analysis (RSA) at a national level, specifically with respect to technical standards of the International Association of Assessing Officers (IAAO). Our initial research shows that variables with positive effects on perception of value include agricultural plot size and estimated annual income (rental) potential. Plots situated further from agrimarkets and auction locations are perceived to be less valuable. Negative effects are associated with sandy soil, higher average annual rainfall, moderate to steep slopes, and plots situated in swamps or marshlands. The ordinary kriging-based predictions, while seemingly less likely to overestimate perceived value, are more regressive. Ordinary kriging achieves superior scores of vertical equity, but inferior scores of uniformity when compared to a current OLS model with location factor adjustment variables derived from RSA.

Introduction

Due to physical, legal, and other barriers, as well as cost-prohibitive reasons associated with data collection and storage, sparse data can be a common hurdle in the effectiveness of governments who depend on or are considering the implementation of a property tax regime. The ability to estimate–with some degree of confidence–property values for certain geographic areas is oftentimes a highly difficult task, particularly in areas with little or no sales transactions.

In developing countries with limited, inaccurate, or no cadastre or multiple listing service, the data needed to create reliable estimates of value is simply not available. Estimating the existing housing stock (a potential tax base) in such a country for research or implementation purposes would result in costly "boots on the ground" efforts, as well as heightened technical and data storage requirements. This research will be the first of its kind (to the authors' knowledge) in that it develops land valuation models across an entire country (Malawi, Africa) using data already collected from the Third Integrated Household Survey (Malawi, 2010-2011).

It should be noted that subjective, self-perceived value estimates introduce potential biases and errors likely not suitable for actual property tax regimes, but the benefits of understanding public perception and spatial variation of value extends implications across land governance. This research is concerned with the latter. This paper will add to the existing literature base by examining the feasibility of estimating agricultural land value in Malawi, Africa in the absence of sales.

Background

Property Tax Valuation Standards

The International Association of Assessing Officers (IAAO) maintains industry standards that guide government agencies in promoting equitability and uniformity in valuations. IAAO established statistical measurements allow these agencies to identify optimal valuation approaches and methodologies. Models and results in this paper will be compared and evaluated with respect to IAAO standards. The two main statistics that will be evaluated include the coefficient of dispersion (COD), which measures uniformity, and the price-related differential (PRD), which measures vertical equity. Higher PRD values are indicative or more regressive valuation approaches (IAAO 2003).

Mass Appraisal Modeling

Automated valuation models (AVMs) are computer algorithms used to predict values of large sets of properties at once. The use of such models (often referred to as computer-assisted mass appraisal or "CAMA" models) has been steadily increasing in the assessment community since the 1970's – primarily due to advancing technology with increased computational power, speed, and methodologies continuously making valuations more accurate and easier to execute (Moore 2009). Because property markets behave so differently over geographic space, and location plays such a large roll in value formation, conventional modeling techniques may not be able to accurately estimate value (Fotheringham et al. 2003; McMillen 2010).

Response Surface Analysis

Spatial interpolation of variables from known data locations provides a prediction surface or "response surface" from which non-measured locations can be assigned predictions. While this methodology is common across a variety of disciplines, in the property tax arena, it is most commonly referred to as "response surface analysis" (RSA). It can be used to predict property values or other variables and has demonstrated an ability to estimate sparse pockets of property markets, promoting equity and uniformity (O'Connor 1982; O'Connor & Eichenbaum 1988; Ward et al. 1999; McCluskey et al. 2000; González et al. 2005).

RSA is a multi-step process. First, the modeler obtains a location adjustment. One common method to achieve this is by calculating a Z-score for each observation in a dataset (for example, on the sale price, price-per-square foot, or error term) (McCluskey et al. 2000, Ward 2006, D'Amato 2011). The next step is employing a spatial interpolation process that "smooths" the Z-score points into a surface. When mapped, this allows continuous variations in value to be visually identified. In order to obtain a location adjustment for an unobserved location that is covered by this response surface, a spatial join is executed using GIS-capable software (e.g. R, ArcGIS). This process can be used to create location adjustment variables able to be used in AVMs or to estimate an overall value independently.

Data

Produced by the National Statistical Office (NSO) and the Ministry of Economic Planning and Development (MoEPD), several questions in this large-scale survey ask Malawi residents about their perceived selling value of their land, in addition to the current physical characteristics of their respective property (size, soil condition, distance to agricultural market, etc.). These variables can be incorporated into AVMs for study and analysis.

4,682 respondents received their respective agricultural plot via an inheritance. This subset of respondents is used for the analysis because it was considerably larger than any other, and due to the inherent familial element, these respondents are arguably more likely to possess a stronger understanding of and familiarity with the properties in question. Examples of unused subsets of respondents include respondents who moved in without permission (*n=40*), purchased with a title (*n=70*), or were granted access by local leaders (*n=564*).

Due to the dependency upon location for spatial modeling, observations with missing or incomplete coordinates (latitude & longitude) are omitted from the analysis (n= 172). To protect against missing or incorrect data, observations with irrational values (e.g. zero or negative land size) are also omitted. The final dataset consists of 4,398 survey responses. The data are then divided into two sets–80% and 20%–for model training and validation, respectively.

Variables are tested for multicollinearity and those with a variance inflation factor (VIF) of five or greater are omitted. Table 1 shows descriptions and descriptive statistics of the final variables used in this study.

Table 1. Model Variables and Descriptions

Dependent Variable	Description	Mean	St. Dev.	Min.	Max.
logESP	Estimated Sale Price (Nat. Log.)	11.08	1.11	6.91	18.42
Continuous Independent Variables					
logEARI	Estimated Annual Rental Income (Nat. Log.)	8.49	0.82	4.61	13.71
logSize	Plot Area (in Acres) (Nat. Log.)	-0.47	0.94	-4.61	6.28
dist_auction	HH Distance in (KMs) to Nearest Tobacco Auction Floor	78.82	49.12	1.00	236.00
dist_agmrkt	HH Distance in (KMs) to Nearest Agricultural Market	25.05	13.85	0.00	72.00
AvgRain	Avg. 12-month Total Rainfall (mm) for July-June (Nat. Log.)	6.73	0.09	6.54	7.10
Discrete Independent Variables		**n**			
Soil_Sand	Predominant Soil Type = Sand (Default is Clay)	874			
Slope_Mod	Slope of Plot = Moderate (Default is Flat)	326			
Slope_Steep	Slope of Plot = Steep or Hilly (Default is Flat)	128			
Swamp	Marsh or Wetland (Default is Dryland)	717			

Research Methodology

Spatial interpolation utilizes measured locational values to calculate predictions for non-measured locations. During estimation calculations, a higher weight is given to nearer observations. The purpose of such a weighting scheme is to yield estimates more congruent with the geographic realities of a region.

One such spatial interpolation technique is ordinary kriging. Its estimation procedure is given by the following formula:

$$\hat{Z}_0 = \sum_{i=1}^{N} \lambda_i Z_i$$

where:

$\hat{Z}_0$ = estimated value at location 0
λ_i = weight at location *i*
Z_i = observed value at location *i*

Inverse-distance weighting (IDW) is another common spatial interpolation technique, and is calculated as follows:

$$\hat{Z}_0 = \frac{\sum_{i=1}^{N}\left(\frac{Z_i}{d_i}\right)}{\sum_{i=1}^{N}\left(\frac{1}{d_i}\right)}$$

where:

$\hat{Z}_0$ = estimated value at location 0
Z_i = observed value at location *i*
d_i = distance between *n* locations and location 0

This paper will compare various RSA approaches with a spatially unaware OLS AVM (model 1). The RSA models include model 2: *logESP* as predicted by IDW; model 3: *logESP* as predicted by ordinary kriging; model 4: *logESP* as estimated by the addition of a location factor adjustment variable (as predicted by IDW) to the baseline OLS model; and model 5: *logESP* as estimated by the addition of a location factor adjustment variable (as predicted by ordinary kriging interpolation) to the baseline OLS model.

Results

Table 2 (below) shows the regression results of a spatially unaware regression model applied to the holdout test dataset. *logEARI* and *logSize* are the only two variables with positive coefficients.

Table 2. Baseline OLS Regression Output (Model 1)

	Estimate	St. Error	t-value	Pr(>|t|)	
(Intercept)	9.22	1.28	7.19	7.93E-13	***
logEARI	0.67	0.02	32.58	< 2e-16	***
logSize	0.11	0.02	6.09	1.24E-09	***
dist_auction	-0.00	0.00	-4.95	7.80E-07	***
dist_agmrkt	-0.01	0.00	-8.11	6.82E-16	***
Soil_Sand	-0.10	0.04	-2.44	0.014945	*
AvgRain	-0.50	0.20	-2.53	0.011428	*
Slope_Mod	-0.21	0.06	-3.41	0.000647	***
Slope_Steep	-0.29	0.09	-3.07	0.002192	**
Swamp	-0.13	0.04	-3.01	0.00263	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared: 0.3086
Adjusted R-squared: 0.3069

As evidenced by the lowest COD in table 3 (below), model 4 achieves the most uniform results, though models 1 and 5 are only slightly higher. Model 3 has the highest PRD (6.57), followed by model 2 (5.59), and model 5 (5.00).

Estimated values are divided by their observed values to produce “predicted-to-actual” ratios. Table 3 (below) shows the ratio distribution of each model. The maximum ratio of model 2 (319.57) is considerably higher than any other model, with the second highest maximum ratio only reaching 63.67 (model 5). The model yielding the lowest maximum ratio is model 3 (27.34).

Table 3. Test Sample Performance

	COD	PRD	Min. Ratio	1st Quart. Ratio	Median Ratio	3rd Quart. Ratio	Max. Ratio
Model 1 (Baseline OLS)	78.54%	4.97	0.01	0.59	0.97	1.79	61.34
Model 2 (IDW Prediction)	110.85%	5.59	0.00	0.49	0.97	2.15	319.57
Model 3 (Kriging Prediction)	86.70%	6.57	0.00	0.56	0.95	1.87	27.34
Model 4 (IDW Loc. Fact.)	77.58%	4.77	0.01	0.57	0.96	1.77	57.44
Model 5 (Kriging Loc. Fact.)	79.02%	5.00	0.01	0.58	0.96	1.77	63.67

Conclusions

In its current stage, this paper has set up the groundwork for a national AVM in Malawi, Africa, using only secondary data collected by the 2010-2011 Integrated Household Survey. Model variables include physical characteristics of the property, economic variables, as well as location-specific distance and climate variables. Variables with positive value perception effects include those pertaining to plot size and estimated annual income (rental) potential. Plots situated further from agrimarkets and auctions are perceived to be less valuable. Negative effects are associated with sandy soil, higher average annual rainfall, moderate to steep slopes, and plots situated in swamps or marshlands.

The kriging estimates, while more regressive in nature, are less offensive than their IDW counterparts. However, as evidenced by a lower COD. The OLS model that included an IDW-estimated location factor adjustment is the most uniform and least regressive model (though not by a drastic amount). The next stages of this paper will include more advanced spatial interpolation techniques including universal kriging with explanatory variables, co-kriging, and block kriging; experimentation with additional survey variables; as well as the inclusion of other spatial valuation models (e.g. geographically weighted regression, spatial lag model, etc.).

References

Aurélio Stumpf González, M., Soibelman, L., & Torres Formoso, C. (2005). A new approach to spatial analysis in CAMA. Property Management, 23(5), 312-327.

d'Amato, M. (2010). A location value response surface model for mass appraising: An "iterative" location adjustment factor in Bari, Italy. *International Journal of Strategic Property Management*, (3), 231-244.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.

IAAO. 2003. Standard on automated valuation models (AVMs). Assessment Journal 10, no. 4 (Fall): 109–154

Kim, K. R., Kim, G., Kim, J. Y., Park, K., & Kim, K. W. (2014). Kriging interpolation method for laser induced breakdown spectroscopy (LIBS) analysis of Zn in various soils. Journal of Analytical Atomic Spectrometry, 29(1), 76-84.

McCluskey, W. J., Deddis, W. G., Lamont, I. G., & Borst, R. A. (2000). The application of surface generated interpolation models for the prediction of residential property values. Journal of Property Investment & Finance, 18(2), 162-176.

McMillen, D. P. (2010). Issues in spatial data analysis. Journal of Regional Science, 50(1), 119-141.

Moore, J. W. (2009). A history of appraisal theory and practice: Looking back from IAAO's 75th year. Journal of Property Tax Assessment & Administration, 6(3), 23-50.

O'Connor, P. M., & Eichenbaum, J. (1988). Location value response surfaces: The geometry of advanced mass appraisal. Property Tax Journal, 7(3), 277-296.

Ward, R. D., Weaver, J. R., & German, J. C. (1999). Improving CAMA models using geographic information systems/response surface analysis location factors. Assessment Journal, 6, 30-39.