Presented at the DLSU Research Congress 2019
De La Salle University, Manila, Philippines
June 19 to 21, 2019

Poisson Simultaneous Autoregressive Modelling of National Household Targeting System for Poverty Reduction (NHTS-PR) Data

Ellaine Krishel SD. Go[1*], Paula Margarita J. Tse Wing[1], and Frumencio F. Co[1]
[1]*Mathematics and Statistics Department, De La Salle University*
**Corresponding Author: ellaine.sdgo@gmail.com*

Abstract: Creating the most efficient solution to diminish the prevalence of poverty in the Philippines remains as one of the country's major struggles. This paper formulates a spatial model that could aid in poverty reduction using the NHTS-PR 2015 data to identify which indicator variables have significant relationships with poverty count. Given the use of count data for modeling, the Simultaneous Autoregressive (SAR) model was modified to include the Poisson regression approach in the estimation of parameters. The Poisson-SAR model was generated using the backfitting algorithm and were compared with the Ordinary Least Squares (OLS) model for model accuracy. It was found that the Poisson-SARerr model with regional and class dummy variables had the lowest Mean Absolute Percentage Error (MAPE) value and provides the most accurate poverty map.

Key Words: poverty; spatial autocorrelation; simultaneous autoregressive models; Poisson regression; backfitting algorithm

1. INTRODUCTION

Finding an efficient way to solve poverty has been a constant struggle in the Philippines. For the past four decades, the number of people living below the poverty line in the country has declined considerably slow compared to other Association of Southeast Asian Nations (ASEAN) countries (Asian Development Bank, 2009). The measurement of poverty in most practices is primarily scaled in terms of income and/or expenditure. However, there are also a number of indicators such as life expectancy and safe water access that are being utilized in studies that establish poverty as a multidimensional phenomenon (Balisacan, 2011).

The NHTS-PR was initiated by DSWD in 2006 to classify households as either poor or nonpoor based on their living conditions. By 2012, NHTS-PR has identified a total of 5.2 million poor households out of the 11 million families found in 1,647 municipalities and cities nationwide (Fernandez, 2012).

In response to the problems involving poverty, the researchers intended to use spatial analysis in generating a model for poverty count in the Philippines using multidimensional indicators. Specifically, the study aimed to: (1) create spatial models using Poisson-SAR analysis and OLS regression to observe spatial patterns of poverty in the country by municipality and city; (2) measure and identify the underlying association between poverty count and the significant indicator variables; (3) compare the results and prediction accuracy of these models; (4) develop poverty maps representing actual, OLS, and Poisson-SAR generated poverty counts; and (5) propose some recommendations that will possibly aid concerned agencies focused on poverty alleviation

2. METHODOLOGY

2.1 Data

In this paper we use the National Household Targeting System for Poverty Reduction (NHTS-PR) data with coverage from 2014 until 2015. The total number of poor households per municipality or city based on the NHTS-PR classification was considered as the dependent variable for the study. As for the independent variables, datasets for the non-socioeconomic factors, mainly focusing on the living conditions of households in the municipality and city were considered. To represent the clustering of responses by region and by municipality or city classes, dummy variables were included in the study.

The NHTS-PR data, a set of observations for barangays, was aggregated to their corresponding municipalities or cities with PSGC as reference. To check the assumptions, compute for parameter estimates, and generate the poverty maps, the following analyses and software were used: OLS assumptions using SAS, 'spdep' package for R, poisson regression in R, and 'tmap' package for R to generate the maps.

2.2 Statistical Models

This study fitted two poverty models, namely the ordinary least squares (OLS) model with regional dummy variables and the Poisson-SAR model. The Poisson-SAR with municipality/city class and regional dummy variables treating NCR and 1st class municipality as reference categories was fitted using the backfitting algorithm.

Bañez (2012) modified the original SAR model to include Poisson terms and dummy variables. In the case of data containing clustered responses, the model may become unstable. The inclusion of dummy variables could resolve this by representing the clustering of observations. For i =1, 2, …, n and j =1, 2, …, m, the modified SAR_{err} (eq. 1) equation with Poisson terms based on Bañez were written as follows.

$$Y_i = e^{\beta_0 + \sum_{j=1}^{m} \beta_{ij} X_{ij}} + \lambda D(Y_i - e^{\beta_0 + \sum_{j=1}^{m} \beta_{ij} X_{ij}}) + \varepsilon_i \text{, (Eq. 1)}$$

where:

Y_i = number of poor families in municipality/city *i*

X_{ij} = the value of the jth explanatory variable for the municipality/city *i*

β_0 = parameter for the intercept

β_{ij} = parameter for the jth independent variable for the municipality/city *i*

λ = spatial autocorrelation coefficient for SAR_{err}

D = contiguity matrix defined by municipalities/ cities located in the same region

ε_i = random disturbance for the municipality/city *i*

The additive model (AM) from Bañez (2012) utilized the backfitting algorithm introduced by Hastie and Tibshirani (1986) as an estimation technique. He modified the backfitting algorithm by using a generalized linear model (GLM) estimator, where the dependent variable Y_i is assumed to follow a Poisson distribution. The estimation of the modified SAR_{err} model was composed of three main steps:

(1) The β parameters were estimated by applying generalized least squares estimation on the Poisson regression model, $Y_i = exp(\beta_0 + \sum_{j=1}^{m} \beta_{ij} X_{ij}) + \varepsilon_i^*$. It was noted that ε_i^* holds information on the spatial component. The partial residual, ei, computed by subtracting $\hat{Y}_i = exp(\hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_{ij} X_{ij})$ value from Y_i, contains information on the error component, *t* and spatial component, d_t.

(2) To generate spatial parameter estimates, regression analysis was applied on the model $e_i^* = \lambda D(Y_i - exp(\hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_{ij} X_{ij})) + a_i$ for SAR_{err} model, where e_i^* was the partial residual for the i^{th} municipality or city computed in Step 1 and a_i represented temporal influences, systematic inadequacies, and function link misspecification (Bañez, 2012).

(3) The AM was constructed by combining the two major steps formerly discussed.

The process is repeated continuously until the smoothing component estimate remains constant, however Bañez only performed this once since more than one iteration yielded a new dependent variable that gives off either negative or non-integer estimates that are unacceptable for Poisson regression.

After the models' parameter estimates are computed, the corresponding MAPE values for the models are calculated and compared. Finally, mapping was done to generate maps of the *npoor* counts of the actual observation, the OLS model, and for the Poisson-SAR model.

3. RESULTS AND DISCUSSION

3.1 Model Fitting

The OLS assumptions were checked and were satisfied. For the OLS model, 14 non-income indicator variables and eight regional dummy variables were significant at $\alpha = 0.05$. Independent variables such as electricity, subsidized housing, refrigerators, and PhilHealth memberships, among others, were highly significant (p-value < 0.01) and had negative coefficients.

Afterwards, the computed Moran's I statistic ($(MI)^3 = 0.01909$) was found to be significant (p-value < 0.0001), indicating the existence of spatial autocorrelation in the residuals. This justifies the use of spatial regression models. All Lagrange Multiplier tests were significant with equal magnitude. Therefore, indicating that both Poisson simultaneous autoregressive error (Poisson-SAR_{err}) and lag (Poisson-SAR_{lag}) models are applicable. However, only the Poisson-SAR_{err} model's spatial coefficient, ρ, was significant.

All parameter estimates for the Poisson-SAR_{err} model have p-values less than 0.01. Therefore, all non-income indicators and both sets of dummy variables representing regions and classes were found to be strong determinants of poverty count. Consistent among all the parameter estimates for regional dummy variables were their negative coefficients which imply the significance of including regional classifications in poverty reduction. It was also found that municipality classes had negative valued coefficients, while all city classes had positive estimated coefficients except for the 6th class city classification. Hence, considering municipality or city class may improve strategies for poverty reduction. The spatial autocorrelation coefficient, λ, having a negative value, signifies that the municipalities and cities vary in poverty count relative to their neighbors within a certain region, or are more heterogeneous in nature.

3.2 Poverty Maps

For the actual and model estimates, *npoor* counts were divided into three classifications: Low (below the 25th percentile), Moderate (from the 25th to 75th percentile), and High npoor count (above the 75th percentile). Table 1 gives the respective boundary values used in mapping the poverty counts.

Table 1. *npoor* Boundary Values used in Poverty Maps

Poverty Count	25^{th} Percentile	75^{th} Percentile
Actual *npoor* Count	1325	3982
OLS *npoor* Count	1552.186	3418.481
Poisson-SAR_{err} *npoor* Count	1496.099	4051.083

It could be seen from the figures that the Poisson-SAR map closer resembles the actual, compared to that of the OLS model. An apparent difference between Figure 1 and both Figures 2 and 3, are some of the municipalities and cities in the actual map that were failed to be classified to have moderate poverty count which may be due to their higher boundary values for the 25th percentile. Some evidences of similarities between Figures 1 and 2 are more observable in larger islands, e.g., Panay and Palawan. Whereas for Figure 3, due to its lower cutoff value for High npoor count, has classified more municipalities and cities higher than 75th percentile in the eastern side of Mindoro, the island of Negros, and the regions Mindanao.

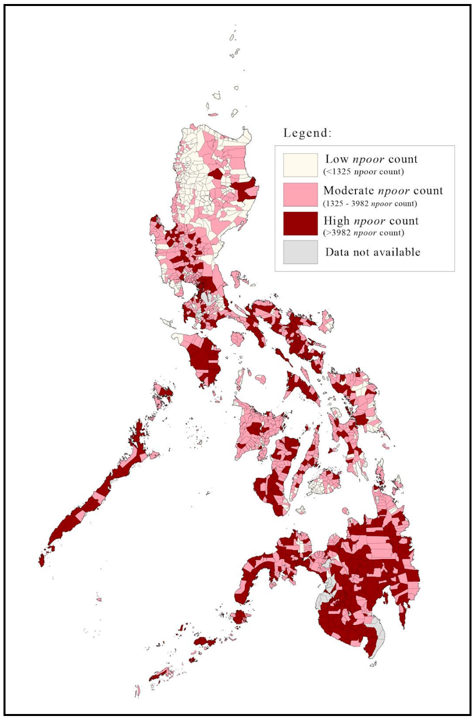

Figure 1. Actual *npoor* Count Map

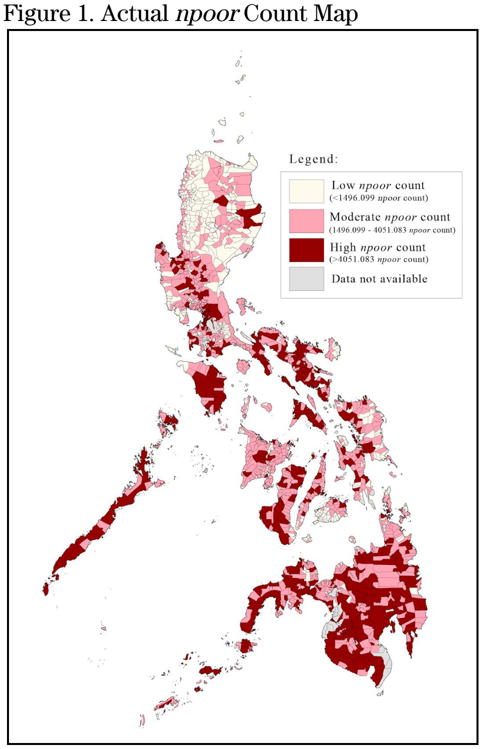

Figure 2. Poisson-SAR *npoor* Count

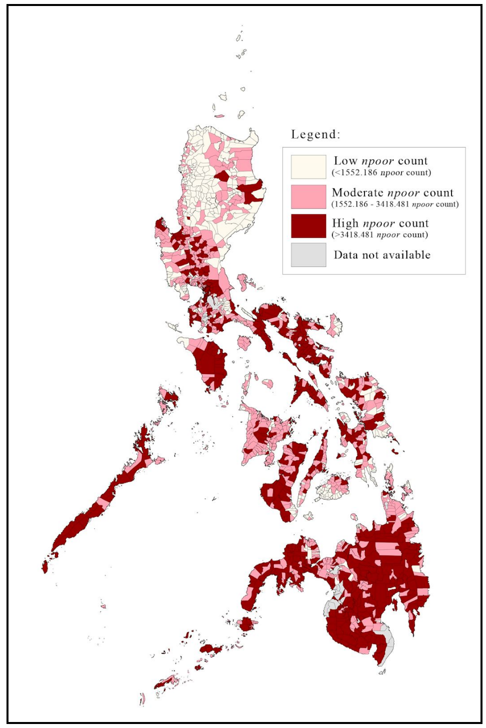

Figure 3. OLS *npoor* Count

4. CONCLUSIONS

This study utilized spatial analysis to generate a Poisson-SAR$_{err}$ model with non-socioeconomic variables that would estimate poverty count in the Philippines. Only a single iteration of the backfitting algorithm was performed for the estimation process to yield positive integer values of the response variable. Otherwise, either negative or non-integer values are obtained. Among smaller geographic units, there exists spatial clustering related to the number of poor households. The Poisson-SAR$_{err}$ model had yielded better poverty count estimates over the OLS model. Consistent with initial variable screening using Pearson's r, all the considered non-income indicators were found to be highly significant and possible strong determinants of poverty count, which may also correspond to the notion that poverty in the country is a multidimensional phenomenon that is not limited to a certain characteristic such as income measures.

Further, due to the heterogeneity of units within a region, it could be said that there are pockets of poverty in the country such as the Mindanao regions and the province of Palawan.

The inclusion of other socioeconomic variables which are not probed in this study (*e.g.*, average household income and household size) could improve the model. Assumptions for Poisson regression such as the property of equidispersion could also be checked to verify its applicability on the count data. Additionally, despite the absence of high multicollinearity in the data, dimension reduction could be explored by applying either Principal Component Analysis (PCA) or Factor Analysis (FA) for a more parsimonious model. FA could also be used to discover the underlying factors that could further explain pockets of poverty. Given that the backfitting algorithm produces biased parameter estimates, the use of other estimation methods such as bootstrapping could be a possible option. Higher-order spatial models such as the Kelejian-Prucha model (SAC) and Spatial Durbin model (SDM) could also be explored. Further studies could also focus on smaller geographical sections and utilize temporal data to have a more detailed analysis.

For poverty alleviation, policy makers are advised to concentrate on pockets of poverty and explore its prevalence. Such concentrated studies may delve into possible poverty indicators that are specific to regions with high poverty rates. Concerned agencies are also advised to examine and provide or improve the predictor variables with negative parameter estimates, especially those involving government-provided social services.

5. REFERENCES

Anselin, L. (2003). Spatial Econometrics in Companion to Theoretical Econometrics. Hoboken, NJ: Blackwell Publishing Ltd. DOI:10.1002/9780470996249

Balisacan, A. (2011). What has really happened to poverty in the Philippines? new measures, evidence, and policy implications. (Discussion Paper, School of Economics, University of the Philippines, No. 2011-14). Retrieved from http://hdl.handle.net/10419/93555

Bañez, J.E. (2012). Poisson spatial autoregression modelling of poverty count data in the Philippines. The Philippine Statistician. Retrieved from http://www.psai.ph/docs/publications/tps/tps_2012_61_2_5.pdf

Buja, A., Hastie, T., Tibshirani, R. (1989). Linear smoothers and additive models. The Annals of Statistics. DOI:10.1214/aos/1176347115

Hastie, T., Tibshirani, R. (1986). Generalized Additive Models. Statistical Science. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.ss/1177013604

Kissling, W.D., Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive model. Global Ecology and Biogeography. DOI: 10.1111/j.14668238.2007.00334.x

Mapa, D.S., Albis, M.L.F. (2013). New proxy means test (PMT) models: improving targeting of the poor for social protection. 12th National Convention on Statistics (NCS). Retrieved from http://nap.psa.gov.ph

Philippine Standard Geographic Code. (2017, December). Retrieved from https://psa.gov.ph/classification/psgc/downloads