Poisson Spatial Autoregression Modelling of Poverty Count Data in the Philippines

John Erwin Bañez
College of Social Work and Community Development
University of the Philippines Diliman

Count data with skewed distribution and possible spatial autoregression (SAR) often causes difficulty in modelling. Violations on the assumptions in ordinary least squares (OLS) may occur. While Poisson regression can offer some remedy in modelling count data, it still does not take into account the spatial dependencies of the data. This paper uses general linear estimation via backfitting algorithm in Poisson-SAR of poverty count in the Philippines for 2000. The model is assessed based on comparison from other models and the actual poverty count (MAPE and poverty map). MAPE was lowest in Poisson-SAR compared to other models.

Keywords: Spatial autoregression, backfitting algorithm, poisson regression

1. Introduction

Poverty is a worldwide concern. Although poverty across the globe differs in magnitude and form, poverty is not limited among developing countries alone. It remains the rule than the exception. In the Philippines, poverty remains to be the greatest concern of several sectors (Balisacan and Hill, 2003). And just like worldwide scenario, although poverty can be found elsewhere, the distribution of poverty seems to be spatially associated. That is, certain provinces of this country seem to have more poor than others (Barrios and Landagan, 2004; Ravallion and Jalan, 2002). In this study, we will examine the spatial distribution of poverty count in the Philippines. National data for 2000 will be used. Although this is an old data, we expect the data to illustrate modeling problems when the dependent variable is count. This paper addressed that modeling problem and asked: How is Poverty count in the Philippines represented in a space continuum using concepts behind generalized additive models?

The objectives of this paper are the following:

1. fit a spatial model using Poisson SAR and OLS.
2. compare the characteristics of Poisson SAR and OLS using selected poverty indicators.
3. determine significant poverty indicators that will explain the geographic distribution of poverty count.

This chapter describes how backfitting algorithm was implemented to fit Poisson-SAR models. There were five models fitted. To find the best model corresponding mean absolute percentage error (MAPE) were compared.

2. Methodology

2.1 Statistical models

In this paper, demographic profile, economic profile, living conditions and assets, and geographic attributes were postulated as indirect determinants of poverty. Inclusion of the variables and importance of each are justified by the results of previous researches on poverty and Philippine history (Balisacan and Hill, 2003; Ekelund and Herbert, 1997; Semple, 1996; UNDP, 2003; Henderson et al., 2000). The dependent variable was the number of poor household per province for the year 2000.

Five models were fitted in this study. First, we fit ordinary least squares (OLS) with regional dummy variables (*model 1*). Then we fit Poisson-SAR based models using Backfitting method. There are four Poisson-SAR models fitted, namely, Pure Poisson-SAR (*model 2*), Poisson-SAR with regional dummy variables (*model 3*), Poisson-SAR with regional and land type dummy variables (*model 4*), and Poisson-SAR with regional, landtype (topography) and Manila dummy variables (*model 5*).

2.2 Variables

The total number of families falling below the regional per capita poverty threshold of the census year was used as the dependent variable. As independent variables, this paper postulated demographic profile, economic profile, living conditions and assets, and geographic attributes as indirect determinants of poverty count. Below is the table of variables used in this study and their definition.

Table 1 Table of Variables Used

Dependent Variable	Definition
s_poor	Number of poor

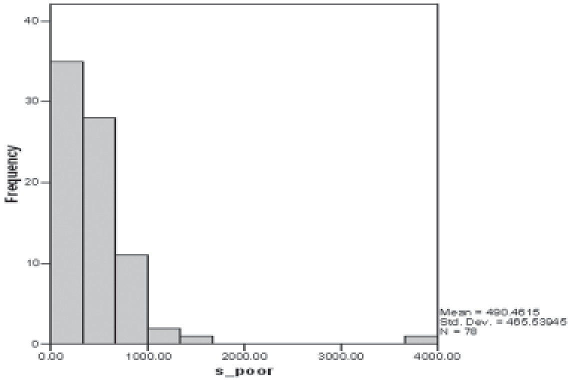
Independent Variables	Definition
mFsize	Average household size
mAge7_1	Proportion of households with children between 1 year and 6 years
mEmployed	Average number of employed persons per household
mgender1	Proportion of male-headed households
mMS1	Proportion of households headed by married persons
mjob1	Proportion of households with employed head
mAgind1	Proportion of households engaged in agriculture
mhhtype1	Proportion of nuclear families
mwemp1	Proportion of households where wife is employed
mbld_type1	Proportion of single detached houses
mroof1	Proportion of houses with strong roof
mwalls1	Proportion of houses with strong walls
melec1	Proportion of households with electricity
mwater1	Proportion of households with safe water
mtv1	Proportion of households with television
mref1	Proportion of households with refrigerator
mtoilet1	Proportion of households using water-sealed toilet
mRegion1	Dummy for Region 1
mRegion2	Dummy for Region 2
mRegion3	Dummy for Region 3
mRegion4a	Dummy for Mindoro, Marinduque, Romblon Palawan (Mimaropa)
mRegion4	Dummy for Region 4 Cavite, Laguna, Batangas, Rizal, Quezon (Calabarzon)
mRegion5	Dummy for Region 5
mRegion6	Dummy for Region 6
mRegion7	Dummy for Region 7
mRegion8	Dummy for Region 8
mRegion9	Dummy for Region 9
mRegion10	Dummy for Region 10
mRegion11	Dummy for Region 11
mRegion12	Dummy for Region 12
mRegion13	Dummy for National Capital Region (NCR)
mRegion15	Dummy for Region 15
mRegion14	Dummy for ARMM
mRegn16	Dummy for CAR
lt_is	Dummy for island land type
lt_il	Dummy for inland land type
lt_c	Dummy for coastal land type

2.3 Data

The main source of data for this study is the Family Income and Expenditure Survey (FIES) conducted by the National Statistics Office in 2000. The data was aggregated at provincial level (spatial units). The domain of FIES 2000 is rural-urban per province, with a total of 39, 615 sample households nationwide.

2.4 Modeling poverty count

Figure 1 Histogram of poverty count



Because of the skewed distribution and the suspected spatial autocorrelation of poverty count the spatial autoregressive model (SAR) introduced by Pace and Barry (2002) was modified to include Poisson terms given by:

$$Y_i = e^{\beta_0 + \beta_1 X_i} + \alpha D(Y_i - e^{\beta_0 + \beta_1 X_i}) + \varepsilon_i \tag{1}$$

where:

Y_i = number of poor families in province *i*
X_i = is a known constant, namely the value of the independent variable (e.g., those listed in Table 1 for province *i*.)
β_0 = is the intercept
β_1 = parameter for the province *i* for a specific variable (this can be easily generated to two or more variable.
α = spatial autocorrelation
D = spatial weight matrix defined by provinces belonging to the same region
ε_i = random error

The parameters will be estimated using the backfitting method.

2.5 General backfitting method

The general backfitting algorithm by Hastie and Tibshirani (1990) was modified by using a generalized linear model estimator in the smoothing part of the algorithm.

Step 1: The parameters β_0 and β_1 were estimated using Poisson regression model $Y_i=e^{\beta_0+\beta_1 X_i}+\varepsilon_i*$ where the component ε_i* still contains information on the spatial component. This step was implemented using generalized least squares estimation. Note that β estimated at this point does not necessarily inherit the optimal properties of OLS since the error component does not really match with the classical assumptions.

Define the resulting residuals as $e_i = Y_i - \hat{Y}_i$ where $\hat{Y}_i = e^{\hat{\beta}_0+\hat{\beta}_1 X_i}$. Note that these residuals contain information on both the error components ε_t and the spatial component w_t from the neighborhood system ignored initially.

Step 2: A final regression to estimate the spatial dimension was performed by regressing e_i* with the neighborhood variable to estimate the spatial parameter. This model can be defined by $e_i^* = \alpha D\left(Y_i - e^{\hat{\beta}_0+\hat{\beta}_1 X_i}\right) + a_i$ where a_i defines other systematic inadequacies of the predictor X, temporal dependencies, and other misspecification of the function link connecting the response variable to the independent variables which might still be present in the computed residuals.

Step 3: Results in steps 1 and 2 were combined to form our additive model.

The model will be compared to OLS with regional dummy variables using mean absolute percentage error (MAPE).

Backfitting algorithm produces biased estimates. Hence the *p*-values produced by the software should not be interpreted as in linear regression where assumptions were met. However, the *p*-values can help us identify which variables are possibly strong determinants of poverty count. The interpretation of results should focus on the direction of the estimates.

2.5.1 Poisson Regression and rationale behind introduction of dummy variables in the model

According to Agresti (1996), in Poisson regression, the parameter (β) can be interpreted as follows: A unit increase in X has a multiplicative impact of e^{β} on μ. The mean of Y at $x+1$ equals the mean of Y at x multiplied by e^{β}. If $\beta=0$, then $e^{\beta}=e^{0}=1$ and the multiplicative factor is 1; that is the mean of Y does not change as X changes. If $\beta>0$ then, $e^{\beta}>1$, and the mean of Y increases as X increases. If $\beta < 0$, the mean decreases as X decreases.

Poisson Regression only takes non-negative response. In Backfitting algorithm, step three will result to a new dependent variable that can inevitably result to negative values. Because of this, our study is limited to a single iteration.

Poisson regression can become unstable if responses are clustered. To address this problem, we introduced dummy variables that would intuitively represent the clustering of responses. Regional dummy variables were considered. Dummy variables for type of land (inland, coastal, and island) were also considered since the islands may not be affected by the contiguity matrix since it is not connected to provinces of the same region. Also, the type of land may influence economic diversity. Intuitively, we can say that small islands are more vulnerable to poverty since most they depend solely on the sea for livelihood. They also have higher vulnerability to typhoon. The inland provinces are less vulnerable and more diverse in terms of economic resources. In these provinces, crop rotation can be practiced, and they can also reap some of the benefits of nearby coastal provinces. We can say that coastal provinces are in between islands and inlands in terms of economic diversity. Finally, dummy variable for NCR was also introduced.

2.5.2 Spatial autocorrelation

To account for spatial effect in the data, we tested the existence of spatial autocorrelation. Spatial autocorrelation is the association between the values of the same variables at different spatial locations.

The spatial effects in OLS were represented by regional dummies. The spatial weight matrix is the distinguishing characteristic of Poisson-SAR from OLS. The spatial linkages of the provinces were measured by defining a spatial weight matrix with dimension n*x*n. There is no standard specification of the elements of a spatial weight matrix. In this study, the spatial weight matrix was defined by a contiguity matrix that was based on the membership of a province to a region. As an illustration, Pangasinan, Ilocos Sur, Ilocos Norte, La Union is adjacent provinces, but Pangasinan is not contiguous to Manila. Pangasinan's row has the following weight: D = [0, 0.333, 0.333, 0.333, 0, ... 0]. Notice that the first entry, the intersection between Pangasinan and Pangasinan (in the weight matrix), is zero to prevent Pangasinan from predicting itself. The second, third and fourth entries are the intersection between Pangasinan, Ilocos Sur, Ilocos Norte, and La Union, which are equal to 0.333 since they are considered neighbors, The intersection of Pangasinan and Manila is also zero since they are not considered as neighbors. The row sums to 1.

The spatial weight matrix in SAR allows us to study neighborhood effect. Alpha represents the spatial pattern observed in the spatial weight matrix. Spatial autocorrelation is concern with the degree to which objects or activities are similar to other objects or activities located nearby. In contrast to other types of spatial statistical analysis, such as point pattern analysis for example, spatial autocorrelation deals simultaneously with both locational and attribute information. If objects which are similar in location also tend to be similar in attributes, then the pattern as a whole is said to show positive spatial autocorrelation. Conversely, negative spatial autocorrelation exists when objects which are close together in space tend to be more dissimilar in attributes than objects further apart. (Ding and Fotheringham, 1991). We also note that a negative association can only occur if units are heterogeneous.

2.6 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) measures the accuracy of prediction or performance of the model in forecasting. The lower MAPE suggests better prediction of poverty indicator. MAPE is computed as

$$\frac{\left|Y - \hat{Y}\right|}{Y} \text{ x } 100\% \; . \qquad (2)$$

2.7 Data Processing

Using FIES 2000 Data and aggregating household-level information at the provincial level, OLS and Poisson-SAR based models were run using *R* language and SAS.

SAS was used to compute for the independent and dependent variables. Different poverty thresholds were used for different regions. This was done by first getting the per capita income by dividing the total income by family size. Then we define a new variable (poor) as 1 if the per capita income of trials is below the poverty line and 0 otherwise. Then we get the sum of poor (s_poor) per province. The National Statistics Coordinating Board (NSCB) annual per capita poverty thresholds were used. The Poisson regression was run in SAS. The residuals were computed then these residuals became the dependent variable in computing the spatial autocorrelation coefficient, α. The spatial weight matrix were done in excel then transferred into SPSS so that R could read it. R was used to compute for the spatial autocorrelation coefficient, α. In performing the sparse spatial autoregression the spdep package in R is essential. To compute for predicted values proc IML in SAS was used.

3. Results and Discussion

Different models were estimated. We started with OLS then continued with the Poisson-SAR based models. As earlier said, there are four Poisson-SAR based models.

3.2 OLS estimates

In OLS, only Proportion of single detached houses (mbld_typ1) is significant and its estimated coefficient is negative.

Table 2 Parameter Estimates of OLS with Dummy Variables

	OLS with dummy variables				
	estimate	**p-values**		**estimate**	**p-values**
alpha			mref1	2260.94837	0.0713
Intercept	4349.01058	0.1507	mtoilet1	-756.14478	0.2679
mFsize	16.15699	0.6992	mregion1	-80.55319	0.8296
mAge7_1	1158.27075	0.5383	mregion2	355.67434	0.3973
mEmployed	1341.68918	0.1326	mregion3	base	
mgender1	4411.74354	0.2102	mregion4	360.96895	0.3777
mMS1	-3815.45379	0.3029	mregion4a	-168.89114	0.5555
mjob1	-1514.58311	0.5203	mregion5	176.34976	0.6552
mAgind1	-31.81633	0.9716	mregion6	-20.56837	0.9554
mhhtype1	632.00905	0.5714	mregion7	429.09573	0.2703
mwemp1	-2853.84752	0.1208	mregion8	460.67442	0.2753
mbld_typ1	-6383.94163	0.0031	mregion9	261.35265	0.5407
mroof1	447.03797	0.5496	mregion10	-62.69286	0.8683
mwalls1	-1134.70705	0.1211	mregion11	-184.12383	0.6428
melec1	1037.26988	0.3031	mregion12	-11.13854	0.9786
mwater1	-92.2055	0.8673	mregion14	-208.22112	0.614
mtv1	-913.06081	0.4041	mregion15	306.5014	0.6201
			mregion16	256.32025	0.5103

3.4 Poisson-SAR with regional dummy variables

In this model, the estimates of the intercept, Proportion of male-headed households (*mgender1*), Proportion of households where wife is employed (*mwemp1*), Proportion of single detached houses (*mbld_typ1*), and Proportion of houses with strong walls (*mwalls1*) coefficients have p-values lower than 0.05. Among these only the estimated coefficient of Proportion of male-headed households (*mgender1*) is positive.

3.5 Poisson-SAR with regional and land type dummy variables

The estimates for the intercept, Proportion of male-headed households (mgender1), Proportion of households headed by married persons (mMS1), Proportion of single detached houses (mbld_typ1), and Proportion of houses with strong walls (mwalls1) coefficients in this model have p-values lower than 0.05. Again, among these, only the estimated coefficient of Proportion of male-headed households (mgender1) is positive.

Table 3 **Parameter Estimates of Pure Poisson-SAR**

	Pure Poisson-SAR				
	estimate	**p-values**		**estimate**	**p-values**
alpha	-0.16216	0.37096	mwemp1	-2.4936	0.1002
Intercept	10.1399	<.0001	mbld_typ1	-4.8334	0.0004
mFsize	-0.0123	0.7217	mroof1	0.6395	0.2966
mAge7_1	3.902	0.0695	mwalls1	-1.6911	0.0092
mEmployed	0.7001	0.3314	melec1	0.8108	0.468
mgender1	9.3172	0.0098	mwater1	-0.3714	0.4724
mMS1	-6.4744	0.0663	mtv1	-0.1575	0.88
mjob1	-3.0791	0.2123	mref1	2.2591	0.0381
mAgind1	-1.2432	0.1916	mtoilet1	-1.1143	0.1166
mhhtype1	0.0523	0.955			

Table 4 **Parameter Estimates of Poisson-SAR with Regional Dummy Variables**

	Poisson-SAR with regional dummy variables				
	estimate	**p-values**		**estimate**	**p-values**
alpha	-1	2.72E-06	mtoilet1	-1.0476	0.1848
Intercept	8.5259	0.0109	mregion1	0.3445	0.3747
mFsize	0.0196	0.6731	mregion2	1.0073	0.0211
mAge7_1	2.0552	0.3688	mregion3	base	
mEmployed	1.611	0.0862	mregion4	0.9125	0.0441
mgender1	10.5742	0.0074	mregion4a	0.0888	0.75
mMS1	-7.8554	0.056	mregion5	0.779	0.0789
mjob1	-2.5072	0.3621	mregion6	0.4936	0.2042
mAgind1	-0.4186	0.674	mregion7	0.9564	0.0129
mhhtype1	1.0985	0.3521	mregion8	1.3173	0.0036
mwemp1	-4.4364	0.0245	mregion9	1.1619	0.0097
mbld_typ1	-7.5525	0.0004	mregion10	0.4969	0.1849
mroof1	1.2127	0.1312	mregion11	0.2779	0.4876
mwalls1	-2.295	0.005	mregion12	0.6242	0.1465
melec1	1.2183	0.2888	mregion14	-0.1602	0.6842
mwater1	-0.3002	0.634	mregion15	1.0116	0.1337
mtv1	0.9285	0.403	mregion16	0.9675	0.0223
mref1	2.0923	0.1381			

Table 5 Parameter Estimates of Poisson SAR with Land-type Dummy Variables

	Poisson-SAR with land type dummy variable				
	estimate	**p-values**		**estimate**	**p-values**
alpha	-1	1.29E-06	mregion1	0.3398	0.4018
Intercept	9.8855	0.007	mregion2	0.9984	0.0247
mFsize	0.0423	0.4132	mregion3	base	
mAge7_1	0.7505	0.7731	mregion4	1.0075	0.0319
mEmployed	1.0504	0.3325	mregion4a	0.1487	0.6118
mgender1	9.9415	0.0143	mregion5	0.7867	0.0821
mMS1	-8.7662	0.0398	mregion6	0.4637	0.2686
mjob1	-1.9673	0.4906	mregion7	1.0215	0.0094
mAgind1	-0.469	0.6411	mregion8	1.2593	0.0082
mhhtype1	1.2518	0.2947	mregion9	1.2305	0.009
mwemp1	-3.427	0.1212	mregion10	0.5381	0.1683
mbld_typ1	-7.9992	0.0003	mregion11	0.3412	0.4214
mroof1	1.3592	0.097	mregion12	0.5698	0.2033
mwalls1	-2.5011	0.0029	mregion14	-0.4061	0.3797
melec1	1.802	0.167	mregion15	1.0279	0.1405
mwater1	-0.4884	0.4581	mregion16	1.0112	0.0201
mtv1	0.5514	0.6376	lt_is	-0.1442	0.49
mref1	1.6223	0.2893	lt_il	0.2023	0.3411
mtoilet1	-1.1183	0.1977			

3.6 Poisson-SAR with regional, land type, and NCR dummy variables

Only three estimates are significant in this model. These are the estimates for the coefficients of Proportion of male-headed households (*mgender1*), Proportion of houses with strong walls (*mwalls1*), and Proportion of households with television (*mtv1*). This time, among these, only the estimated coefficient of Proportion of houses with strong walls (*mwalls1*) is negative.

Note that examining all the Poisson-SAR based models, we see that the estimated coefficients for Proportion of male-headed households (*mgender1*) and Proportion of houses with strong walls (*mwalls1*) have p-values lower than 0.05. Also, the estimated coefficient of Proportion of male-headed households (*mgender1*) remained positive in all our models.

Table 6 Parameter estimates of Poisson-SAR with NCR Dummy Variable

	Poisson-SAR with NCR dummy variable				
	estimate	**p-values**		**estimate**	**p-values**
alpha	-1	1.01E-07	mregion1	0.502	0.1525
Intercept	4.3343	0.1987	mregion2	1.2644	0.0012
mFsize	-0.0081	0.8555	mregion3	0.1261	0.0063
mAge7_1	1.8357	0.3945	mregion4	1.0974	<.0001
mEmployed	0.9336	0.3007	mregion4a	0.4882	0.0631
mgender1	9.3456	0.0059	mregion5	1.1137	0.0051
mMS1	-5.024	0.1757	mregion6	0.8332	0.0264
mjob1	-3.816	0.1207	mregion7	1.203	0.0005
mAgind1	-0.9067	0.287	mregion8	1.6126	0.0001
mhhtype1	0.2787	0.7893	mregion9	1.6291	<.0001
mwemp1	-2.8696	0.1272	mregion10	1.0786	0.0028
mbld_typ1	-1.1345	0.6334	mregion11	0.9054	0.0209
mroof1	0.9654	0.1616	mregion12	1.1967	0.0036
mwalls1	-1.6474	0.0227	mregion13	base	
melec1	0.3865	0.7313	mregion14	0.874	0.0756
mwater1	-0.3871	0.4879	mregion15	1.4567	0.0152
mtv1	3.0455	0.0088	mregion16	1.413	0.0003
mref1	0.1217	0.926	lt_is	-0.0116	0.9483
mtoilet1	-1.0374	0.1477	lt_il	-0.0224	0.903

3.7 Spatial effects

Table 7 below summarizes the spatial effects as estimated by different models. It is noted that no spatial effects were estimated via OLS.

Referring to table 7, we can see that for Pure Poisson-SAR the spatial coefficient is -0.16. Although negative, this is not indicative of a significant spatial correlation (p-value = 0.37). For Poisson-SAR with regional dummy variables the spatial coefficient is -1 suggesting a perfect and significant spatial correlation (p-value = 2.75E-06). For Poisson-SAR with regional and land type dummy variable the spatial coefficient is also -1 and is also highly significant (p-value = 1.29E-06). The Poisson-SAR with regional, land type and NCR dummy variable has similar results.

Table 7 Estimated spatial coefficients (alpha)

	alpha
with regional dummy	-1
p-value	2.72E-06
with land type	-1
p-value	1.29E-06
pure Poisson-SAR	-0.16216
p-value	0.37096
ols	-
p-value	-
with Manila dummy variable	-1
p-value	1.01E-07

4. Examining the Significant Variables

Although p-values from backfitting are not real, they can give us an idea of which variables are possibly strong determinants of poverty count. In this section we examine the p-values of the parameter estimates. We still use 0.05 in examining the p-values.

For all the Poisson-SAR based model Proportion of male-headed households (*mgender1*) and Proportion of houses with strong walls (*mwalls1*) have p-values lower than 0.05. Proportion of male-headed households (*mgender1*) showed a positive estimated coefficient, meaning, for every increase in the proportion of male-headed household an increase in the poverty count is expected holding other predictors constant. The significance of Proportion of houses with strong walls (*mwalls1*) indicates the importance of household characteristics in poverty reduction.

5. Assessment of Model Performance

5.1 Reduction in MAPE

The mean absolute percentage error (MAPE) measures the accuracy of prediction or performance of the model in forecasting. The lower MAPE suggests better prediction of poverty indicator.

Poisson-SAR (MAPE=36.16) outperforms OLS (MAPE=51.46) substantially. This indicates that the correction for spatial dependence improved the predictive capability of the model. Poisson-SAR with regional dummy variables is also a considerable improvement (MAPE = 22.53) while Poisson-SAR with regional and land type dummy variables outperforms them all (MAPE = 22.3). Poisson-SAR with regional, land type, and NCR dummy variable has the best predictive performance (MAPE = 18.47).

Table 8 MAPE for the Different Models

	MAPE
OLS	51.4585427
Pure Poisson-SAR	36.1630374
Poisson-SAR with regional dummy variables	22.5342472
Poisson-SAR with regional and landtype dummy variables	22.3015031
Poisson-SAR with Manila dummy variable	18.4702081

5.2 Poverty maps

The results above indicates that Poisson-SAR with regional, landtype, and NCR dummy variable (*model 5*) has the best predictive power (MAPE = 18.47). Thus it is much worthwhile examining this model. Below, are poverty maps plotting their estimated poverty counts. The distribution of actual poverty count and estimated poverty count were divided into three, High Poverty Count (those above the 0.75 percentile), Medium Poverty Count (those between 0.25 and 0.75 percentile), and Low Poverty Count (those below 0.25 percentile). The table below shows the different cut-offs for the different maps.

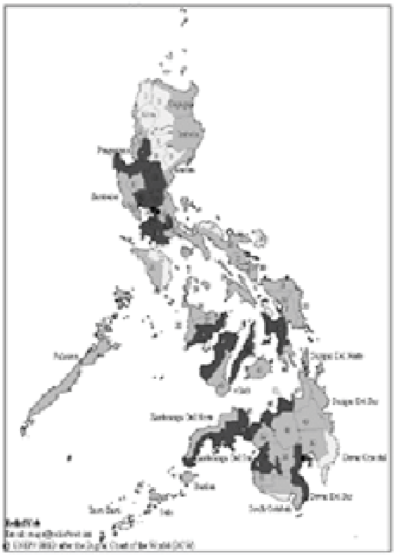
Table 9 Cut-offs used in Constructing Poverty Maps

	0.25 percentile	**0.75 percentile**
Actual Poverty Count Map	275	562
OLS Poverty Count Map	245.4755111	631.7843889
Poisson-SAR with NCR Dummy Variable Poverty Count Map	282.74847	582.13579

Since Poisson-SAR with NCR dummy variable has the lowest MAPE (18.47), it is expected that Figure 3 is more similar to Figure 2 than Figure 4. An obvious difference is, in Figure 4 the clustering of Luzon provinces with high poverty count leaned towards the south-west part of Luzon rather than being in the central part of Luzon as in Figure 2 and Figure 3. Difference in clustering in Mindanao can also be seen when Figure 2 and Figure 4 are compared.

Our best model shows only three possibly strong determinants of poverty count, these are Proportion of male-headed households (*mgender1*), Proportion of houses with strong walls (*mwalls1*), and Proportion of households with television (*mtv1*).

Figure 2. Actual Poverty Map



Legend

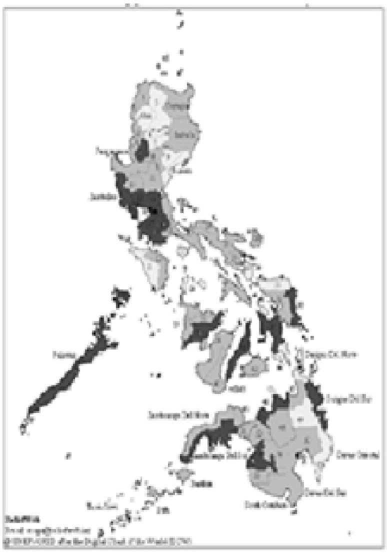
■ **High Poverty Count**
(> 0.75 percentile of estimated count)

▨ **Moderate Poverty Count**
(>= 0.25 and <= 0.75 percentile of estimated poverty count)

□ **Low Poverty Count**
(< 0.25 percentile of estimated count)

Figure 3. Poisson-SAP
with NCR dummy variable Poverty Map

Figure 4. OLS POverty Map



As earlier said, the implication of the direction of parameter estimates in the Poisson part of our additive model is this; A unit increase in X has a multiplicative impact of eβ on μ. The mean of Y at x + 1 equals the mean of Y at x multiplied by e^{β}. If $\beta = 0$, then $e^{\beta} = e^{0} = 1$ and the multiplicative factor is 1; that is the mean of Y does not change as X changes. If $\beta > 0$ then, $e^{\beta} > 1$, and the mean of Y increases as X increases. If $\beta < 0$, the mean decreases as X increases (Agresti, 1996).

Proportion of male-headed households (*mgender1*) and Proportion of households with television (*mtv1*) both have positive estimated coefficients. This means that as Proportion of male-headed households (*mgender1*) increases poverty count is expected to increase holding all the other predictors constant. The same goes for Proportion of households with television (*mtv1*). Proportion of houses with strong walls (*mwalls1*) has negative estimated coefficient. This means as Proportion of houses with strong walls (*mwalls1*) increases poverty count is expected to decrease holding other predictors constant.

Based on model 5, we can also say, an increase in family size (*mFsize*) would mean a decrease in poverty count holding other predictors constant. The same goes for proportion of household headed by a married couple (*mMS1*), proportion of household with employed head (*mjob1*), proportion of household engaged in agriculture, proportion of household where wife is employed (*mWemp1*), proportion of single detached household (*mbld_typ1*), proportion of household with safe water (*mwater1*), proportion of household using water sealed toilets (*mtoilet1*).

Our study yields a negative alpha (spatial coefficient). This suggests that the provinces within a region are heterogeneous in terms of poverty count distribution. That there is a clustering of high and low poverty count within a region can be statistically supported. This suggests that there are indeed pockets of poverty in the regions.

6. Conclusions and Recommendations

There is reason to believe that there is spatial clustering among provinces in the Philippines in terms of poverty count. Our model, where *Y* is discrete, yielded similar results in terms of improvement in MAPE to the model of Barrios and Suratos (2008) and Suratos (2005), where *Y* is continuous. Poisson-SAR outperforms OLS. Introduction of dummy variables to Poisson-SAR produces lower MAPE and highly significant spatial coefficient. The backfitting algorithm can resolve the estimation problem associated with a Poisson-SAR model where the response variable is assumed to be discrete (count).

We also note that the p-values are not real since backfitting method produces biased estimates. However those with very low p-values can still give us an idea of which variables are possibly strong determinants of poverty count.

Perhaps incorporating temporal effects can reduce the MAPE further. Smaller geographic units such as municipality may be more effective. Inclusion of variables other than socio-eco variables can also be explored. Perhaps variables on governance, environment, peace and order, and temporal effect should be considered. Different contiguity matrix is also worthy to explore.

Since the p-values in this model are not real we cannot really test whether the parameter estimate is zero or not. Perhaps, the use of bootstrap can be used in order to have a better hypothesis testing on the parameter estimates.

References

AGRESTI, A., 1996, *An Introduction to Categorical Data Analysis*, New York: John Wiley & Sons, Inc.

BALISACAN, A. and HILL, H., 2003, *The Philippine Economy*, New York: Oxford University Press, Inc.

BARRIOS, E. and LANDAGAN, O., 2004, Geographic Distribution of the Poor: Is Poverty Contaminating?, 9th National Convention of Statistics, Manila.

BARRIOS, E. and SURATOS, G., 2008, Regional clustering of poverty in the Philippines, UPSS Working Paper No. 2008-06.

DING, Y. and FOTHERINGHAM, S., 1992, The Integration of Spatial Analysis and GIS, *Computers, Environment, and Urban Systems* 16:3-19.

EKELUND, R. and HEBERT, R., 1997, *A History of Economic Theory and Method*, McGraw-Hill.

HASTIE, T.J. and TIBSHIRANI, R.J., 1990, Generalized Additive Models, Chapman & Hall.

PACE, R. K. and BARRY, R., 1997, Sparse spatial autoregressions, *Statistics and Probability Letters* 33: 291-287.

SEMPLE, E.C., 1996, *Influences of Geographic Environment*,. New York: Henry Holt and Company (text converted in Spring 1996 by Lance Christian).

RAVALLION, M. and JALAN, J., Geographic Poverty Traps? A Micro Model Consumption Growth in Rural China, *Journal of Applied Econometric*, 17(4):329-346.

United Nations Development Program, 2003, Left Behind: Domestic Inequalities and Fate of the Poor, Human Development Report 2003: Millenium Development Goals: A compact among nations to end human poverty.

HENDERSON, J.V., SHALIZI, Z., VENABLES, A.J., 2000, Geography and Development. Available at: http://time.dufe.edu.cn/jingjiwencong/waiwenziliao/37877.pdf