Modeling the Spatial Interaction of Voting Participation vis-à-vis Poverty Using Spatial Regression: A Follow-up Study

Dennis P. Dizon
Alumnus[1], M.Sc. Cartography
Technische Universität München
dennisdizon@yahoo.com

Abstract

This study further builds on the findings of the previously conducted analysis on the “voting neighborhood effect”, where it was observed that cities and municipalities in the Philippines tend to exhibit spatially autocorrelated levels of voting participation, as well as voting participation with respect to socioeconomic variables such as poverty. In this follow-up analysis, we attempt to develop a statistical model that can estimate the degree of the neighborhood relationship exhibited in the said spatial phenomenon. Using official data on voters’ turnout and poverty incidence, three spatial regression models were tested: 1) a spatial lag model (SAR) for modeling the spatial interaction of voting participation among neighboring areas; and 2) a spatial Durbin model (SDM) for modeling the spatial interaction of neighboring areas’ poverty levels with their voting participation. Based on model performance measures and diagnostics for spatial dependence and homoscedasticity, a spatial Durbin model (SDM) containing spatially lagged voters’ turnout and spatially lagged poverty incidence as covariates sufficiently depicts how voting participation among cities and municipalities in the Philippines are affected by their mere geographic contiguity. In particular, high voting participation and high poverty levels are both found to induce a positive feedback on the voting participation of neighboring cities and municipalities.

Keywords: voting, poverty, spatial regression, spatial lag model, spatial Durbin model

Introduction

In a previously presented study, spatial autocorrelation and crossregression analyses have shown the presence of “voting neighborhood effect”, or the tendency of neighboring cities and municipalities to have similar levels of voting participation, measured in terms of voters’ turnout (Dizon, 2019). This spatial phenomenon particularly manifests through the formation of voting participation hotspots, i.e. positively autocorrelated spatial clusters; and cold spots, i.e. negatively autocorrelated spatial clusters; which can be partly attributed to the spatial variation of poverty among neighboring cities and municipalities (Dizon, 2019). Motivated by the insights derived from this analysis, this follow-up analysis explores another aspect in the spatiostatistical analysis of voting participation, this time by investigating the possible presence of spatial interaction, or the level of interdependence in voting participation and poverty among cities and municipalities of the Philippines (Flint, 2000).

This study again uses voters’ turnout data as basis for measuring voting participation, and poverty incidence as basis for measuring poverty.

---

[1] *graduated September 2020*

Methodology

Studies undertaken by Teng (2006) and Lay et al. (2007) exemplify the analysis of spatial interaction in election-related studies, a common ground of interest with this study. In these studies, which both focus on the case of election results in Taiwan, a spatiostatistical technique called spatial regression is commonly implemented, and is therefore being adopted in the analysis of spatial interaction of voting participation in the Philippines. In essence, spatial regression is used to develop a regression model where there is at least one spatially lagged covariate derived from either the dependent variable itself, or from any of the independent variables. In the context of this study, using spatial regression is meant to produce a model that can appropriately estimate the cities' and municipalities' voters' turnout based on the voters' turnout and poverty levels of their respective neighbors.

Figure 1 illustrates a framework in analyzing the spatial interaction of voting participation vis-à-vis poverty. An analogous framework was also previously applied in Dizon's (2019) spatial autocorrelation and crosscorrelation analysis of voting participation. This time however, the main relationship of interest is spatial interaction instead of spatial segregation.

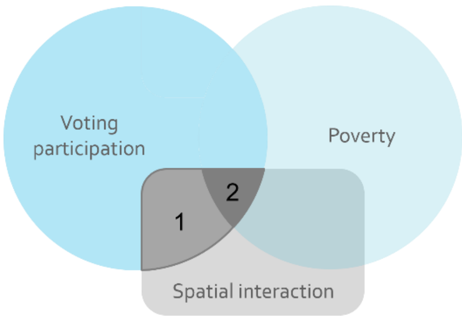

Figure 1. Framework for the analysis of spatial interaction of voting participation.

The analysis framework illustrated in Figure 1 distinguishes two aspects of spatial interaction that can be analyzed between voting participation and poverty. These are:

1. The spatial interaction of voting participation among neighboring cities and municipalities; and
2. The spatial interaction of voting participation and poverty levels among neighboring cities and municipalities

Taking on Teng's (2006) and Lay et al.'s (2007) use of spatial regression in their election-related studies, the two aspects of spatial interaction mentioned above are then respectively addressed using two types of spatial regression models.

1. Spatial lag model (SAR) – a type of spatial regression model where one of its covariates is a spatially-lagged variable derived from the dependent variable (Golgher & Voss, 2016).
2. Spatial Durbin model (SDM) – a type of spatial regression model where, in addition to a spatially-lagged variable derived from the dependent variable, at least one covariate is a spatially-lagged variable derived from one of the independent variables (Golgher & Voss, 2016).

Consistent with the previously performed spatial autocorrelation and crosscorrelation analysis of voting participation by Dizon (2019), contiguity is still the primary basis for defining neighborhoods among the cities and municipalities. Furthermore, in the case of terrestrially isolated island cities and municipalities (51 out of the 1,634), neighborhood is established by manually assigning a neighbor to its nearest city and municipality based on Euclidean distance. Because of this neighborhood definition, a contiguity-based spatial weight matrix is to be used for the mathematical derivation of the spatially lagged variable coefficients.

The datasets used in this study are again sourced from official statistics published by the Philippine Commission of Elections (for the voters' turnout data), the Philippine Statistics Authority (for the poverty incidence data), and the National Mapping and Resource Information Authority (for the city and municipal boundaries data). The datasets (spatial and non-spatial) are compiled using ArcGIS, while the spatial regression is performed using Geoda (for constructing the spatial weight matrix) and R (for constructing the spatial regression models). Figure 2 shows a visualization of the spatial distribution of voters' turnout and poverty among the cities and municipalities.

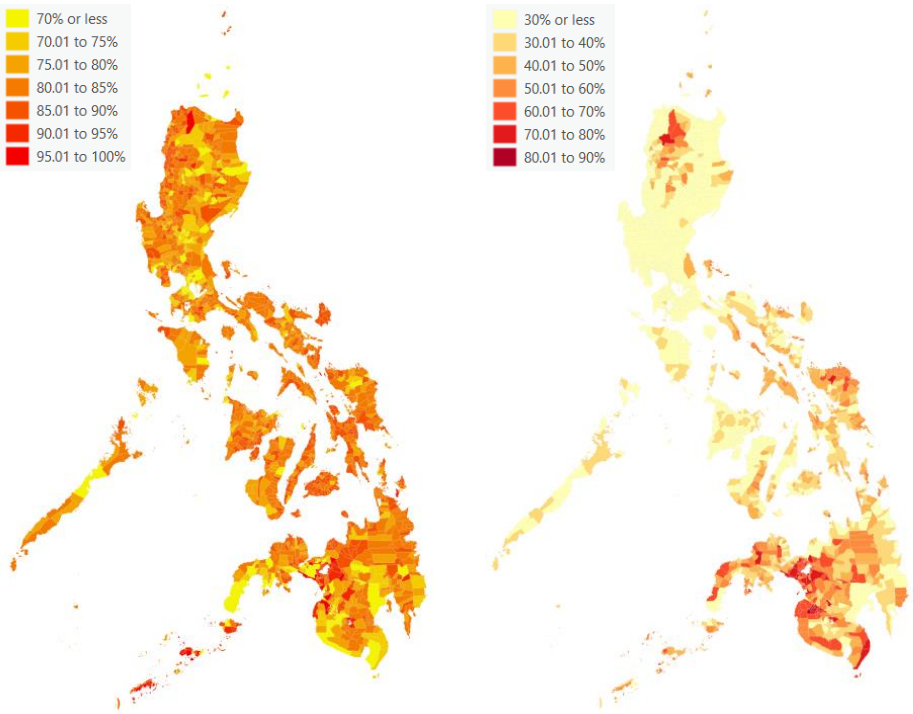

Figure 2. Voters' turnout (left) and poverty (right) of cities and municipalities.

Results and discussion

Before proceeding with the spatial models, a linear regression model is first constructed to test and justify the necessity of using spatial regression for modeling voters' turnout vis-à-vis poverty incidence. Figure 3 shows the output of the linear regression modeling, generated using the Generalized Linear Regression tool in ArcGIS. The output shows that despite the significance of the covariates (p-value < 0.05), the linear regression model is not able to satisfy the assumptions for residual normality and homoscedasticity, as indicated by the Jarque-Bera and the Koenker (BP) statistics having p-value < 0.05. This confirms the need to apply spatial regression for the model.

```
--------------------------------------------------------------------------------------------
 Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE   Robust_t Robust_Pr [b]
Intercept       80.028026 0.341731  234.184413       0.000000*  0.398522 200.811893      0.000000*
  POVERTY        0.038367 0.009823    3.905667       0.000107*  0.013912   2.757840      0.005884*
--------------------------------------------------------------------------------------------
--------------------------------------------- GLR Diagnostics --------------------------------------------
----------------------------------------------------------------------------------------------------------
Input Features:              citmun_withlags.shp   Dependent Variable:                            TURNOUT_CR
Number of Observations:                     1634   Akaike's Information Criterion (AICc) [d]:   10983.047684
Multiple R-Squared [d]:                 0.009260   Adjusted R-Squared [d]:                          0.008653
Joint F-Statistic [e]:                 15.254238   Prob(>F), (1,1632) degrees of freedom:           0.000505*
Joint Wald Statistic [e]:               7.605683   Prob(>chi-squared), (1) degrees of freedom:      0.005818*
Koenker (BP) Statistic [f]:            55.830456   Prob(>chi-squared), (1) degrees of freedom:      0.000000*
Jarque-Bera Statistic [g]:           1531.344510   Prob(>chi-squared), (2) degrees of freedom:      0.000000*
----------------------------------------------------------------------------------------------------------
```

Figure 3. Linear regression results.

A necessary step prior to formulating the spatial regression models is to assess if introducing a spatially lagged dependent variable results in improved spatial independence (Bivand et al., 2005). A Lagrange multiplier test for spatial dependence is performed using R for this assessment (Anselin, 1988). Figure 4 shows the result of the test for the spatial lag model (SAR). With the test yielding a p-value < 0.05, it is confirmed that having a spatially lagged dependent variable gives model stability in terms of having spatially independent residuals. This again suggests the suitability of spatial regression modeling.

```
        Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = turnout_cr ~ poverty, data = shapecitmun)
weights: nbcq2

LMlag = 85.668, df = 1, p-value < 2.2e-16
```

Figure 4. Lagrange multiplier test for spatial dependence results.

The first candidate spatial model is spatial lag model (SAR) with the “poverty” covariate and a spatially lagged “voters’ turnout” covariate. In equation form, this spatial lag model (SAR) is specified as follows:

$$Y = \alpha + \rho WY + \beta X + \varepsilon \tag{1}$$

where Y is the estimated voters’ turnout (in percent) of a city or municipality; α is the constant (intercept) term; ρ is the spatial lag coefficient associated with voters’ turnout; WY is the spatially lagged voters’ turnout within a city or municipality’s neighborhood; β is the coefficient for the “poverty” covariate; X is a city or municipality’s poverty incidence (in percent); and ε is the residual. Figure 5 shows the output of the constructed spatial lag model (SAR) using R.

```
Type: lag
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value  Pr(>|z|)
(Intercept) 57.643804   2.488753 23.1617 < 2.2e-16
poverty      0.029173   0.009520  3.0644  0.002181

Rho: 0.27888, LR test value: 77.023, p-value: < 2.22e-16
Asymptotic standard error: 0.030623
    z-value: 9.1071, p-value: < 2.22e-16
Wald statistic: 82.94, p-value: < 2.22e-16

Log likelihood: -5450.005 for lag model
ML residual variance (sigma squared): 45.342, (sigma: 6.7336)
Number of observations: 1634
Number of parameters estimated: 4
AIC: 10908, (AIC for lm: 10983)
LM test for residual autocorrelation
test value: 5.2132, p-value: 0.022416
```

Figure 5. Spatial lag model (SAR) results.

It can be seen that the coefficients corresponding to the constant (intercept) term and the "poverty" covariate are simultaneously significant components of this model. In addition, the ρ coefficient (labeled Rho in the output) is also significant based on the p-values (< 0.05) associated with the likelihood ratio (LR), z, and Wald statistics, further strengthening the usefulness of this model. However, one issue with this spatial lag model (SAR) is the presence of spatially autocorrelated residuals as reflected in the significant p-value (0.0224) of the Lagrange multiplier (LM) residual autocorrelation test. This means that the resulting spatial lag model (SAR) is still not able to completely capture the spatial interaction in voters' turnout among the neighboring cities and municipalities.

In an attempt to improve the spatial regression model, a second candidate model based on the spatial Durbin model (SDM) is then constructed using R. The spatial Durbin model (SDM)'s equation is specified as follows:

$$Y = \alpha + \rho WY + \beta X + \theta WX + \varepsilon \tag{2}$$

where all notations are interpreted similarly with the spatial lag model (SAR), with the addition of θ, which is the spatial lag coefficient associated with poverty incidence; and WX, the spatially lagged poverty incidence within a city or municipality's neighborhood. Figure 6 shows the output of the constructed spatial Durbin model (SDM) using R.

```
Type: mixed
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 57.538645   2.479830 23.2027  < 2e-16
poverty      0.012529   0.012149  1.0313  0.30239
lag.poverty  0.034832   0.015742  2.2126  0.02692

Rho: 0.27338, LR test value: 73.543, p-value: < 2.22e-16
Asymptotic standard error: 0.030792
    z-value: 8.8782, p-value: < 2.22e-16
Wald statistic: 78.822, p-value: < 2.22e-16

Log likelihood: -5447.558 for mixed model
ML residual variance (sigma squared): 45.241, (sigma: 6.7261)
Number of observations: 1634
Number of parameters estimated: 5
AIC: 10905, (AIC for lm: 10977)
LM test for residual autocorrelation
test value: 1.8954, p-value: 0.1686
```

Figure 6. Spatial Durbin model (SDM) results.

Again, it can be seen from that the constant term and ρ coefficients remain to be statistically significant components in the model. In addition, the θ coefficient (labeled lag.poverty in the output) was also found to be statistically significant, with an attained p-value = 0.0269. However, with the introduction of the spatially lagged "poverty" variable, the "poverty" covariate consequently lost its significance (p-value = 0.3024), which means that the original form of the variable does not anymore have a bearing in the spatial Durbin model (SDM). Despite this tradeoff however, it can be noted that the spatial Durbin model (SDM) already passes the Lagrange multiplier (LM) residual autocorrelation test (p-value = 0.1686), indicating that the residuals are not anymore spatially autocorrelated and that the model closely approximates the spatial interaction of neighboring cities and municipalities voters' turnout vis-à-vis poverty. This is backed by running the Moran's test for spatial autocorrelation, which indicates that the residuals are indeed not anymore spatially autocorrelated (p-value = 0.5661). The output for the Moran's test for spatial autocorrelation is shown in Figure 7.

```
        Moran I test under randomisation

data:  spatdurb$residuals
weights: nbcq2

Moran I statistic standard deviate = -0.16639, p-value = 0.5661
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
    -0.0035169101     -0.0006123699      0.0003047150
```

Figure 7. Moran's test for spatial autocorrelation results.

Figure 8 shows the spatial distribution of residuals from the spatial Durbin model (SDM) expressed in standard deviation units.

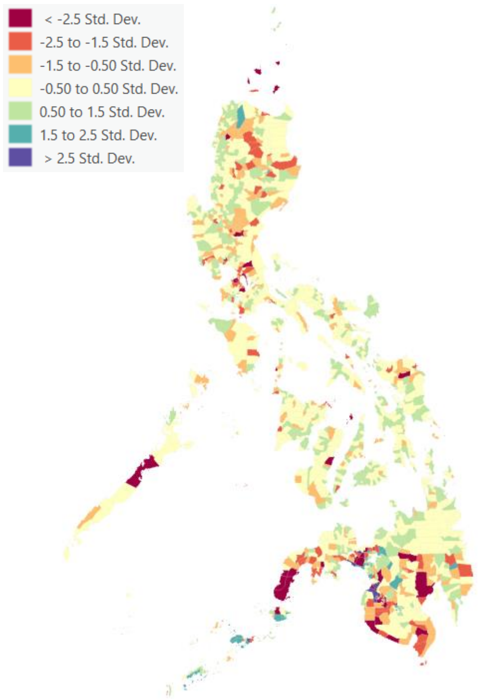

Figure 8. Residuals from the spatial Durbin model (SDM).

When the resulting diagnostics for the two candidate spatial models are juxtaposed, it can easily be determined that the spatial Durbin model (SDM) is better for modeling the spatial interaction of voting participation and poverty than the spatial lag model (SAR). Another indication of this is the log likelihoods of the two spatial models, where the spatial Durbin model (SDM) garnered a lower value of -5447.558 compared to -5450.005 for the spatial lag model (SAR). However, since the difference between the two models' log likelihoods (2.447) is relatively small, a likelihood ratio test is run using R to determine if this small difference is significant enough to conclude that the spatial Durbin model (SDM) is indeed better than the spatial lag model (SAR). Figure 9 shows the result of the test, and the attained p-value = 002696 leads to the conclusion of better preference for the spatial Durbin model (SDM).

```
        Likelihood ratio for spatial linear models

data:
Likelihood ratio = -4.8933, df = 1, p-value = 0.02696
sample estimates:
    Log likelihood of spat Log likelihood of spatdurb
                 -5450.005                 -5447.558
```

Figure 9. Likelihood ratio test results.

Based on the above results, the final equation representing the spatial regression model for voting participation vis-à-vis poverty of neighboring cities and municipalities of the Philippines is as follows:

$$Y = 57.5386 + 0.2734WY + 0.0348WX + \varepsilon \tag{3}$$

From this spatial regression model, it can be roughly interpreted that a city or municipality's voters' turnout may increase by 27.34% of the average voters' turnout of its neighbors; and 3.48% of the average poverty levels of its neighbors.

Conclusion

Based on the aspect of spatial interaction, this study has again demonstrated another manifestation of voting neighborhood effect among the cities and municipalities of the Philippines. Through the use of spatial regression modeling, the following specific conclusions can be established:

1. There is a positive relationship between an area's voting participation and the voting participation of its neighbors. When neighboring areas vote more, the tendency for nearby areas is to vote more as well as a consequence of voting neighborhood effect; and
2. Voting neighborhood effect is spatially influenced by poverty. The tendency to vote is seen to increase with increasing levels of poverty. In other words, the richer towns and cities get, the less they tend to vote.

References

Anselin, L. (1988). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20(1), 1-17.

Anselin, L. (2003). *An Introduction to Spatial Regression Analysis in R*. Urbana-Champaign: University of Illinois.

Anselin, L. (2004). *Exploring Spatial Data with Geodata: A Workbook*. Urbana, Illinois: Center for Spatially Integrated Social Science, University of Illinois, Urbana-Champaign.

Anselin, L. (2017, November 28). GeoDa 1.12.1.59.

Anselin, L. (2018, March 16). *GeoDa Documentation*. Retrieved April 22, 2018

Bivand, R., & Piras, G. (2019). spatialreg: Spatial regression analysis. *R package version,* 1-1.

Dizon, D. (2019). *Spatial Autocorrelation and Crosscorrelation Analysis of Voting Particiation in the Philippines*. Geomundus Conference 2020.

Esri. (2017). ArcGIS Desktop 10.5.1.7333. Redlands, California.

Esri, Inc. (2018, February). *Modeling spatial relationships*, 2.1.2. (Esri, Inc.) Retrieved April 2018, from ArcGIS Pro Tool Reference: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/modeling-spatial-relationships.htm

Flint, C. (2000). Electoral Geography and the Social Construction of Space: The Example of the Nazi Party in Baden. *GeoJournal*, 51(3), 145-156.

Golgher, A.B., Voss, P.R. How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects. *Spatial Demography 4,* 175–205 (2016). https://doi.org/10.1007/s40980-015-0016-y

Lay, J.-g., Yap, K.-h., & Chang, C.-c. (2007). Spatial Perspectives and Analysis on Voting Behavior: A Case Study of the 2004 Taiwan Presidential Election. (60, Ed.) *Journal of Electoral Studies*, 14(1), 33.

National Statistical Coordination Board (NSCB). (2013). *2012 First Semester Official Provincial Poverty Statistics of the Philippines*. Makati City: National Statistical Coordination Board.

NIST Statistical Engineering Division. (2015, October 19). *Jarque Bera Test*. Retrieved from NIST Information Technology Laboratory: https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/jarqbera.htm

Philippine Statistics Authority (PSA). (2016). *2012 Municipal and City Level Poverty Estimates*. Quezon City: Poverty and Human Development Statistics Division, PSA.

Teng, C.-S. (2006, December). The Spatial Factors in Electoral Studies: The Example of Presidential Elections in Taiwan. *Journal of National Development Studies*, 6, 89-144.

Wartenberg, D. (1985, May). Multivariate Spatial Correlation: A Method for Explanatory Geographical Analysis. *Geographical Analysis*, 36(2), 90-104.

Williams, R. (2015, January 30). *Heteroskedasticity*. Retrieved from University of Notre Dame Web site: https://www3.nd.edu/~rwilliam/stats2/l25.pdf