Journal of Tropical Resources and Sustainable Science

journal homepage: jtrss.org

Determinants of Poverty: A Spatial Analysis

Siti Aisyah Nawawi*, Ibrahim Busu, Norashikin Fauzi, Mohamad Faiz Mohd Amin

Faculty of Earth Science, Universiti Malaysia Kelantan, Jeli Campus, 17600 Jeli, Kelantan.

Received 13 May 2019
Accepted 19 May 2019
Online 31 December 2018

Keywords:
Poverty, Spatial Autocorrelation, Conditional Autoregressive Models

⌧*Corresponding author:
Siti Aisyah Nawawi,
Faculty of Earth Science,
Universiti Malaysia Kelantan,
17600 Jeli, Kelantan, Malaysia
Email: aisyah.n@umk.edu.my

Abstract

Eradicating poverty has become the main concern for Malaysian government since independence. Recognising the incidence of poverty through standard statistical data tables alone is no longer adequate. This study examines socio-demographic effects on poverty and measures spatial patterns in poverty risk looking for high risk of areas. The poverty data were counts of the numbers of poverty cases occurring in every ten districts of Kelantan. To model these data, a spatial autocorrelation was detected prior to a Poisson Log Linear Leroux Conditional Autoregressive was fitted to the data. The result shows the variables household members, number of non-education of household head and log number of female household head significantly associated with the number of poor households. Tumpat was found as the highest risk area of poverty.

1. INTRODUCTION

The most common criterion in defining poverty is income insufficiency that is a household's income fails to meet a poverty line (Braithwaite & Mont, 2009; Brandolini et al., 2010). In Malaysia, Poverty Line Income (PLI) is used as a poverty threshold for classifying poor households. Even though the statistics have shown a steady decrease in poverty from one year to another, the inequality gap between the regions, states and rural-urban areas remain wide. For examples, in 2009 Sabah has the highest poverty rate with a value of 19.7%, while Melaka has the lowest poverty rate of 0.5% (Mat Zin, 2011). The difference is almost 19%. Similarly, Majid et al. (2016) found that areas with highest poverty concentrations were in northeast Kelantan and Hulu Terengganu while Klang valley had a low incidence of poverty. Siwar et al. (2016) in their study found that Kelantan proportion of the poor households in urban areas was 39.11% and rural areas equal to 30.18%. The finding indicates that the incidence of poverty was higher in urban areas as compared to rural areas of the state. In 2012, the Department of Statistics Malaysia revealed that Kelantan is the second highest incidence of poverty after Sabah. Considering Kelantan is ruled by the opposition and not receiving oil royalties from the Federal Government compared with Sabah, Kelantan has potentially become the poorest state in the near future. Therefore, this study focuses on poverty in the state of Kelantan.

Recognising the incidence of poverty through standard statistical data tables alone is no longer adequate. The poverty data is in fact the number of poor household counts for each areal unit. This areal unit data typically exhibit spatial autocorrelation, with observations from areal units close together tending to have similar values. Ignoring this autocorrelation can result in biased parameter estimates and overly optimistic standard errors (Dormann, 2007). Therefore, this study was aimed to evaluate the determinant and pattern of poverty in Kelantan statistically. Research on poverty was carried out spatially so that a novel class of statistical models for estimating the spatial pattern in poverty rates, which has much greater flexibility than existing methods, was developed.

2. DATA AND METHODS

2.1. Data of study

Data for this study were obtained from the e-Kasih database from the Ministry of Women, Family and Community Development for the year 2010. Data in e-Kasih database was updated annually. The households that meet e-Kasih criteria such as income less than country's Poverty Line Income (PLI) or household's income less than Rm1000 per month at the rural area and RM1500 at the urban area were eligible to be inserted into e-Kasih. The state of Kelantan which comprises of 10 districts (see figure 1) was the study area. The poverty data will be the number of poor households including hardcore poor for each of the districts. The independent variables were the socio-demographic characteristics of the poor household head that comprises of the number of female household head, average age, number of non-education of household head, average income and the number of household members. These independent variables were chosen based

eISSN Number: 2462-2389

on studies from Crimmins et al. (2009), Chapoto et al. (2011), Anyanwu (2014), Majid et al. (2016), Lastrapes & Rajaram (2016) and others.

Figure 1: Map of 10 districts in Kelantan.

2.2. Standardized Poverty Rate (SPR)

The measure of poverty risk was the standardised poverty ratio (SPR), which is the ratio of the observed to the expected numbers of poverty cases (Majid et al., 2016). The formula is shown in equation 1 and 2.

$$\text{SPR} = \frac{y_k}{E_k}, \tag{1}$$

$$E_k = \frac{\sum y_k}{\sum p_k} \times p_k. \tag{2}$$

Here, y_k for $k = 1, \ldots, n$ is the number of poor households in district k. While P_k is the number of living households and E_k is expected poverty rate for each district k. For example, if SPR for an area is equal to 1.20, then this means there is a 20% increased poverty risk relative to the expected cases.

2.3. Poisson Generalised linear model (GLM)

Generalised linear model (GLM) is an extension of the general linear model which allows for more flexibility in the modelling approach. The response variable, y is measured on a ratio scale or takes the form of a set of counts. Therefore, GLM allows the response variable, y to be one from a set of independent random variables from any exponential family distribution. A generalised linear model is given by

$$Y_k \sim f(y_k | \mu_k, \phi) \quad \text{for } k = 1, \ldots, n$$
$$g(\mu_k) = \eta_k = x_k^{\top} \beta \tag{3}$$

In the above model, $X = (x_1^{\top}, \ldots, x_n^{\top})$ is a matrix covariates where for observation k, $x_k^{\top} = (x_{k1}, \ldots, x_{kp})$. Here p is the number of covariates and regression parameters $\beta = (\beta_1, \ldots, \beta_p)$. The link function $g(.)$ must be monotone and differentiable. Example of the link function $g(.)$ is a log, square root and logit transformation. In this study, the poverty data used are counts data. Thus the Poisson GLM is employed. To ensures that the model fits non-negative values of a response variable, the log link function is used. This gives

$$Y_k \sim Poisson\ (\mu_k) \quad \text{for } k = 1, \ldots, n$$
$$\ln(\mu_k) = x_k^{\top} \beta. \tag{4}$$

2.4. Neighbourhood matrix W

In the spatial analysis, when undertaking a test for autocorrelation and modelling data at an area level, it is essential to identify the neighbourhood structure of the data being analysed. The neighbourhood structure is defined by a neighbourhood matrix W. Given a study region with n distinct areas $\{A_1, \ldots, A_n\}$, the neighbourhood matrix W is a $n \times n$ matrix whose element w_{kj} represents a measure of closeness between area A_k and area A_j. According to O'Sullivan & Unwin (2010), the elements w_{kj} of the neighbourhood matrix, W can be binary or non-binary. In this study, contiguity neighbour is considered, which is the most common specification in the literature, and is defined by whether two spatial units share a border or not.

$$w_{kj} \begin{cases} 1 & if\, A_k \text{ shares a common border with } A_j, \\ 0 & otherwise. \end{cases}$$

Tables 1 and 2 illustrate the spatial framework and W matrix for 6 areas. For area 1, it can be seen that it shares borders with area 2 and area 4. Therefore the matrix elements w_{12} and w_{14} are 1 which represent spatially close. Area 5 is not sharing a border with area 1 but only a single point.

Table 1: Spatial framework with $n = 6$ areas.

1	2	3
4	5	6

Table 2: Matrices for $n = 6$ areas.

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

2.5. Moran's I

The most common statistic that is used to measure the strength of spatial autocorrelation among areal units is Moran's I. It can be defined as

$$I = \frac{n \sum_{k=1}^{n} \sum_{j=1}^{n} w_{kj} (y_k - \bar{y})(y_j - \bar{y})}{\sum_{k=1}^{n} \sum_{j=1}^{n} w_{kj} \sum_{k}^{n} (y_k - \bar{y})^2}, \tag{5}$$

eISSN Number: 2462-2389

where y_k is the observed value in area k, $\bar{y}$ is the overall mean and w_{kj} come from the **W** matrix. The value of I is in the interval $[-1,1]$ where a positive value indicates a positive autocorrelation and a negative value a negative or inverse correlation.

The test statistic is Moran's I statistic. The p-value of Moran's I is computed using a Monte Carlo approach. Moran's I statistics are computed for K different random permutations of the data denoted $\{I_1, \dots, I_K\}$. The observed value of Moran's I is then compared to the simulated sample distribution, $\{I_1, \dots, I_K\}$. The null hypothesis is rejected if the probability of the observed value of Moran's I being bigger than the simulated sample distribution is less than 0.05. The hypotheses of this test are H_0 =no spatial autocorrelation and H_1 =some spatial autocorrelation.

2.6. Poisson log-linear Leroux Conditional Autoregressive (CAR) model

A Bayesian Hierarchical model is adopted to model the data y_k using covariates information $x_k = (x_{1k}, \dots, x_{pk})$ and a random effect ϕ_k, the latter representing the unmeasured spatial structure in the poverty cases. The random effects $\phi = (\phi_1, \dots, \phi_n)$ are included to model any spatial autocorrelation in the data, that persist after adjusting for the available covariate information. The random effects are modelled by a CAR prior distribution, which is a type of Gaussian Markov Random Field (GMRF) model. The model is determined by a set of n univariate full conditional distributions $f(\phi_k|\ \phi_{-k})$, where $\phi_{-k} = (\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n)$ for $k = 1, \dots, n$. In this study, the random effects are given the Leroux CAR prior (Leroux et al. 2000). Therefore, the formulation of the Poisson log-linear Leroux CAR model used in this analysis is shown below:

$$
\begin{aligned}
&Y_k \sim Poisson\ (E_k R_k) \quad \text{for } k = 1, \dots, n, \\
&\ln(R_k) = x_k^\top \beta + \phi_k, \qquad\qquad (6) \\
&\phi_k|\ \phi_{-k} \sim N\left[\frac{\rho \sum_{j=1}^{n} w_{kj}\phi_k}{\rho \sum_{j=1}^{n} w_{kj} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^{n} w_{kj} + 1 - \rho}\right], \\
&\beta \sim N\ (\mu_\beta, V_\beta), \\
&\rho \sim U\ (0,1), \\
&\tau^2 \sim Inverse - gamma\ (0.001, 0.001).
\end{aligned}
$$

In the above equation, R_k which is the poverty risk in area k will be estimated. If $R_k = 1$, then $\mathbb{E}(y_k) = E_k$ which is thus the average risk. While if $R_k = 1.2$, then $\mathbb{E}(y_k) = 1.2E_k$ which means 20% more cases than expected. Besides, the value of the regression parameters β also will be estimated, which quantify the effects of the covariates on poverty risk. The covariates $x_k^\top$ are the number of female household head, average age, number of non-education of household head, average income and the number of household members. Here ρ is the level of spatial autocorrelation in the random effects, where $\rho = 1$ shows strong spatial autocorrelation between random effect and corresponds to the intrinsic model and $\rho = 0$ corresponds to independence ($\phi_k \sim N\ (0, \tau^2)$). Finally, τ^2 the conditional variance of $\phi_k|\ \phi_{-k}$. Inference for this type of model is typically based on Markov Chain Monte-Carlo (MCMC) simulation, using a combination of Gibbs sampling and Metropolis-Hasting steps. The software used for this study is $CARBayes$ (Lee, 2013), which is an R package for Bayesian spatial modelling with conditional autoregressive priors.

3. RESULTS AND DISCUSSION

3.1 Exploratory analysis for poverty data

In this study, the SPR values for poor household risk in Kelantan range between 0.59 and 1.31, and a map of their spatial pattern was displayed in Figure 2. The figure shows that the highest SPR was found in Tumpat (the darker area in the upper part of the state) followed by Gua Musang, Tanah Merah and Bachok. While the areas of the lowest SPR were Machang, Kota Bharu and Pasir Puteh. Nevertheless, SPR was an unstable estimator of poor household risk especially when the expected counts E were small, which can occur when the population at risk was small or the cases in the location was rare. For instance if $E_k = 1$, and the observe $y_k = 1$ or $y_k = 2$, then the SPR for area k doubles from 1 to 2. To overcome the instability of the SPR, a Bayesian modelling approach was typically adopted to estimate poor household risk, using both covariate information and a set of random effects.

The autocorrelations between the covariates were measured to assess collinearity. It was found that there was no clear relationship were observed between the covariates except for number of non-education of household head and the number of female household head. Note that the relationship between these two variables was highly autocorrelated, with an autocorrelation coefficient of (0.84). Hence, a natural log transformation was applied to the number of female household head to reduce the collinearity with the number of non-education of the household head (down to 0.64).

3.2 Residuals spatial autocorrelation

The areal unit's data tend to have spatial autocorrelation even after covariates were included, which was due to unmeasured covariates (Lee et al., 2014; Lee & Shaddick, 2010). Therefore, to measure the existence of spatial autocorrelation, a Poisson log-linear model (Model 4) without any random effects or any spatial structure was fitted to the poverty data. Residuals from Model 4 were tested for the presence of spatial autocorrelation. Moran's I statistic with the commonly used neighbourhood matrix based on geographical contiguity was used. The result of Moran's I p-value was less than 0.05 indicating that the

eISSN Number: 2462-2389

residuals of the non-spatial model contain a strong spatial autocorrelation structure. Thus, the null hypothesis of no spatial autocorrelation was rejected. Therefore, the assumption of independence in the model was not satisfied overall. Since the model was inappropriate for the data as it does not allow for residual spatial autocorrelation. Then, Poisson log-linear Leroux CAR model was applied to the poverty data described above, to find the best fitting model.

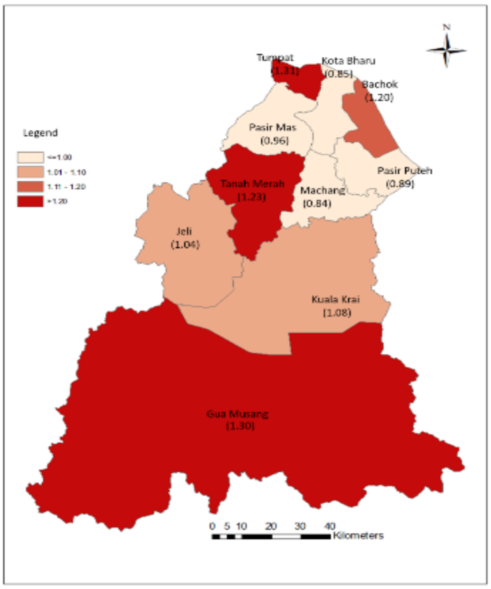

Figure 2: Spatial map of SPR for poor household in Kelantan.

3.3 Poisson log-linear Leroux CAR model

The previous section has shown the existence of a spatial autocorrelation structure of the residuals for the Poisson log-linear model. Therefore, the data set was modelled using Model (5) with the neighbourhood matrix W was binary. Inference for each model was based on 50,000 MCMC samples with a burn-in until convergence of the first 10,000 samples and the rest of the samples were thinned by 10, to reduce their autocorrelation resulting in 4,000 samples.

Table 3 presents the 95% credible intervals for model coefficients of Poisson log-linear Leroux CAR model. The result shows that the variables household members, number of non-education of household head and log number of female household head significantly associated with the number of poor households. The effects of the covariates were illustrated in Table 4, which shows the estimates and 95% credible intervals on the relative risk scale for a standard deviation increase in each covariates value. There was convincing evidence that an increase in the natural log of the number of the female head by 0.64, was related to an increased poor household risk of 63%, indicating that the number of the female head was a very informative covariate. Besides, an area with a high number of non-education of household head was at risk of being poor with a relative risk of 1.114 (11.4%) increase. Similarly, the increasing house member size by 0.2, the poor household risk increase by 13.9%. While, age and the income of the household head exhibit no relationship to poverty risk, as the 95% credible intervals for the estimates contain the null risk of 1.

Figure 3 shows the estimated poverty risks from the Leroux model, where the scales were the same as those used for the SPRs in Figure 2. The estimated risk maps were slightly different with the raw SPR values. For example, the SPR ranges between 0.84 and 1.31, while the corresponding model estimates range between 0.91 and 1.30. The estimated risk surface exhibits a similar spatial pattern to the SPR map, with the highest risks being observed in Tumpat followed by Gua musang and Tanah Merah.

In order to determine the appropriateness of Poisson log-linear Leroux CAR model for the data, the residuals from the model were tested for the presence of spatial autocorrelation. The p-value of Moran's I statistic was greater than 0.05, indicates that there was no spatial autocorrelation as the model remove the spatial autocorrelation present in the data. Poisson log linear Leroux CAR model fit the data, with RMSE, DIC and p.d values were 0.126, 99.78 and 0.88 respectively.

Table 3: Posterior median and 95% credible intervals of the regression from fitting a Poisson log linear Leroux CAR model.

Variable	Median	2.5%	97.5%
Household members	0.6491	0.3821	0.9522
Age	0.0341	-0.0023	0.0571
No Education	0.0003	0.0002	0.0005
Log female head	0.7731	0.6683	0.8463
Income	0.0014	-0.0023	0.0034

Table 4: Estimates and 95% credible interval for the regression parameters. The results were presented on the relative risk scale for a standard deviation increase in each covariates value.

Variable	Median	2.5%	97.5%
Household members	1.139	1.080	1.211
Age	1.074	0.996	1.127
No Education	1.114	1.075	1.197
Log female head	1.632	1.527	1.709
Income	1.031	0.941	1.096

eISSN Number: 2462-2389

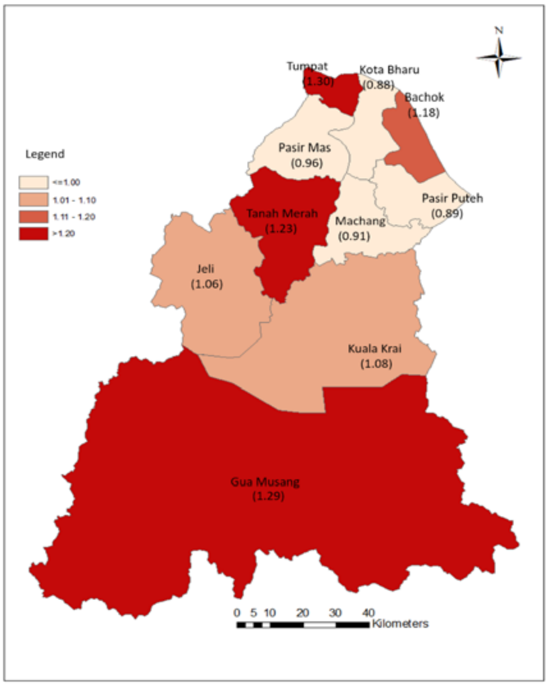

Figure 3: Estimated poverty risks from the Poisson Log Linear Leroux CAR model.

4. CONCLUSION

This study explored the most commonly used conditional autoregressive prior distributions, which was Leroux. The results relate to the wealth of the overall population rather than applying directly to individuals. Overall, Tumpat appears to have a higher risk of poverty with mean risks of 1.31 than the other districts considered in this study. In contrast, Machang, Kota Bharu and Pasir Mas have much lower risks of poverty. These differences in risk appear to be partly due to the covariates. Decreased numbers of household members, non-education and female household head appear to decrease the poverty risk. The model shows that numbers of household members, non-education and female household head significantly influenced poverty. Based on the utilisation of Poisson log linear Leroux CAR model in this study, it conjures up the obvious pattern of poverty as compared to the traditional statistics.

ACKNOWLEDGEMENT

The authors would like to thank University Malaysia Kelantan for the approval given to conduct this research.

REFERENCES

Anyanwu, J. C. (2014), ‘Marital status, household size and poverty in nigeria: evidence from the 2009/2010 survey data’, African Development Review 26(1), 118–137.

Braithwaite, J. & Mont, D. (2009), ‘Disability and poverty: a survey of world bank poverty assessments and implications’, ALTER-European Journal of Disability Research/Revue Europ´eenne de Recherche sur le Handicap 3(3), 219–232.

Brandolini, A., Magri, S. & Smeeding, T. M. (2010), ‘Asset-based measurement of poverty’, Journal of Policy Analysis and Management 29(2), 267–284.

Chapoto, A., Banda, D. J., Haggblade, S., Hamukwala, P. et al. (2011), Factors affecting poverty dynamics in rural zambia, Technical report, Michigan State University, Department of Agricultural, Food, and Resource Economics.

Crimmins, E. M., Kim, J. K. & Seeman, T. E. (2009), ‘Poverty and biological risk: the earlier “aging” of the poor’, Journals of Gerontology Series A: Biomedical Sciences andMedical Sciences 64(2), 286–292.

Dormann, C. F. (2007), ‘Effects of incorporating spatial autocorrelation into the analysis of species distribution data’, Global Ecology and Biogeography 16(2), 129–138.

Lastrapes, W. D. & Rajaram, R. (2016), ‘Gender, caste and poverty in india: evidence from the national family health survey’, Eurasian Economic Review 6(2), 153–171.

Lee, D. (2013), ‘CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors’, Journal of Statistical Software 55(13).

Lee, D. & Mitchell, R. (2014), ‘Controlling for localised spatio-temporal autocorrelation in long-term air pollution and health studies.’, Statistical
methods in medical research 0(0), 1–19.

Lee, D. & Shaddick, G. (2010), ‘Spatial modeling of air pollution in studies of its short-term health effects’, Biometrics 66(4), 1238–1246.

Leroux, B. G., Lei, X. & Breslow, N. (2000), ‘Estimation of disease rates in small areas: A new mixed model for spatial dependence’, Institute for Mathematics and Its Applications 116, 179.

Majid, M. R., Jaffar, A. R., Man, N. C., Vaziri, M. & Sulemana, M. (2016), ‘Mapping poverty hot spots in peninsular malaysia using spatial autocorrelation analysis’, PLANNING MALAYSIA JOURNAL 14(4).

Mat Zin, R. (2011), ‘Measuring and monitoring poverty and inequality: The Malaysian experience’, Singapore: Institute of Southeast Asian Studies .

O’Sullivan, D. & Unwin, D. J. (2010), Geographic information analysis.

Siwar, C., Ahmed, F., Zahari, S. Z., Idris, N. D. M., Mia, M. S. & Bashawir, A. (2016), ‘Poverty mapping and assessing the relationship between poverty and socio demographic characteristics of households: A study in kelantan, malaysia’, Journal of Environmental Science and Technology 9(6), 407–416.