

PAPER • OPEN ACCESS

Relative Risk for Poverty in Kelantan –A Bayesian Approach

To cite this article: Siti Aisyah Nawawi *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **549** 012079

View the [article online](#) for updates and enhancements.

Relative Risk for Poverty in Kelantan –A Bayesian Approach

Siti Aisyah Nawawi^{1*}, Ibrahim Busu, Norashikin Fauzi¹, Mohamad Faiz Mohd Amin¹, Nik Raihan Nik Yusof¹

¹Faculty of Earth Science, Universiti Malaysia Kelantan, Jeli Campus, 17600 Jeli, Kelantan.

E-mail: aisyah.n@umk.edu.my

Abstract. Poverty eradication among poor household head becomes a significant concern. Previous research employed the traditional statistical method to model the poverty data. However, these traditional statistical methods do not consider the spatial elements of poverty data. This study compares the performance of Poisson log-linear Leroux Conditional Autoregressive (CAR) model with difference neighbourhood matrices. A Poisson Log-Linear Leroux Conditional Autoregressive model with different neighbourhood matrices was fitted to the poverty data for 66 districts in Kelantan for 2010. The results show that the performance of the model with the contiguity matrix was nearly similar to the Delaunay triangulation neighbourhood matrix in estimate poverty risk. The variables that are significantly associated with the poverty in Kelantan are the number of non-education, number of female household head and the average age of the household head.

1. Introduction

Many datasets are routinely available as summaries relating to non-overlapping areal units. Examples include disease counts, house prices, and household income. For example, in Scotland, there are Scottish statistics (<http://statistics.gov.scot>), which is an on-line government-run database of areal unit data on a wide variety of topics. This study used poverty data for each areal unit.

Poverty is one of the worst problems that the world faces today. It has received much attention due to its effects especially on academic achievement, economic recession, disease, alcohol-involved suicides, and suicide rates [1,2, 3]. Widely varying definitions of poverty have emerged. According to a definition provided by Santos & Villatoro [4], poverty is a multidimensional phenomenon that covers many aspects such as inability to work, inadequate housing and inadequate access to social health, education and utilities. This definition is close to those of Roganovi ´c & Stankov [5] who mention poverty exists in a society when one or more households do not meet the minimum standard of the economic well-being of the community. There are many definitions of the term poverty have been suggested. This research will use the definition of poverty as income insufficiency which is a household's income fails to meet a poverty line that differs across countries [6,7].

In Malaysia, Poverty Line Income (PLI) is different for peninsular Malaysia, Sabah and Sarawak. The level of PLI is RM930, RM1170 and RM990 for Peninsular Malaysia, Sabah and Sarawak, respectively [8]. Consequently, households with incomes below the poverty line are classified as living in poverty. Statistics have shown that the poverty rate has reduced from one year to another. However, the gap exists between the regions, states and rural-urban areas. In 2009, Melaka had the lowest poverty rate of 0.5%; in contrast, the highest poverty rate was Sabah with a value of 19.7% [9]. In another study,



Majid et al. [10] found that Klang valley had a low incidence of poverty while the highest poverty concentrations were in northeast Kelantan and Hulu Terengganu. According to the Department of Statistics Malaysia, Kelantan is known as the second-highest incidence of poverty after Sabah in 2016. Siwar et al. [11] in their study shows that the incidence of poverty in Kelantan was higher in urban areas with the value of 39.11% as compared to rural areas with the value of 30.18%.

Modelling data that relate to non-overlapping areal units is a common problem in several statistical applications. The response variable of poverty data which is the number of poor households counts for each district typically display spatial dependence, with observations from areal units close together tending to have similar values than further apart. Biased parameter estimates and overly optimistic standard errors will occur if this autocorrelation is ignored [12]. Apart from that, most studies overcome this problem by adding a set of autocorrelated random effects on the linear predictor of the regression model. These random effects are most often modelled by a Conditional AutoRegressive (CAR) prior as part of a hierarchical Bayesian model. Therefore this study will use the commonly used CAR model, which is Leroux CAR with different neighbourhood matrices \mathbf{W} to estimate poverty risk in Kelantan. The compare the different \mathbf{W} specifications using poverty data and quantify their adequacy regarding covariate effects and fitted value estimation. However, to our knowledge, no formal comparison has been made of the appropriateness of each of the neighbourhood matrices.

2. Methodology

2.1. Data description

Data on poverty were obtained from the e-Kasih database from the Ministry of Women, Family and Community Development for the year 2010. The 2010 data was chosen since the census of household in Malaysia is performed for every ten years, where 2010 was the latest census organised. The state of Kelantan is split up into $n=66$ districts (see Figure 1). The response variable is the number of poor households, including hardcore poor for each of the 66 districts that comprise the study region. The covariates used in this study were the socio-demographic characteristics of the household head, including the number of female-headed household, average age and number of non-education of the household head in each district. These independent variables were chosen based on studies from Crimmins et al. [13], Chapoto et al. [14], Anyanwu [15], Majid et al. [10], Lastrapes & Rajaram [16] and others.

2.2. Standardised Poverty Ratio

The standardised poverty ratio (SPR) was used to measure poverty risk. It was calculated for district k as $SPR_k = y_k / E_k$. Here y_k denotes the number of poor household heads in district k . E_k is the expected number of cases which is calculated as $E_k = (\sum y_k / \sum y_k) \times p_k$ where p_k is the number of living households [10]. The value of SPR above 1 represents districts with elevated levels of poverty risk. However, elevated risks are likely to happen by chance if E_k is small, which can occur if the population at risk is small.

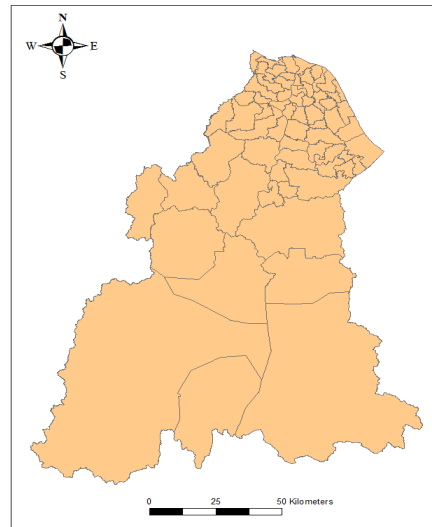


Figure 1. Map of Kelantan with 66 districts.

2.3. Poisson Generalised linear model (GLM)

The Poisson GLM was employed in this study since the observed of the poverty data used are counts variable. The log link function is used to ensures that the model fits the non-negative values of the response variable. The formula is given by;

$$Y_k \sim \text{Poisson}(\mu_k) \quad \text{for } k = 1, \dots, n$$

$$\ln(\mu_k) = \mathbf{x}_k^\top \boldsymbol{\beta}. \quad (1)$$

In the above model, \mathbf{x}_k^\top is a set of p covariates. The regression parameters are denoted by $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$.

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

A Bayesian Hierarchical model is adopted to estimates the poverty risks using covariates information and a set of random effect ϕ_k . The random effects are included to account for the spatial autocorrelation that remains in the data after adjusting for the covariate effects. The random effects in this study were modelled by Leroux prior [17]. The poverty data was fitted using Poisson log-linear Leroux CAR model, which is given by:

$$Y_k \sim \text{Poisson}(E_k R_k) \quad \text{for } k = 1, \dots, n,$$

$$\ln(R_k) = \mathbf{x}_k^\top \boldsymbol{\beta} + \phi_k, \quad (2) \quad \phi_k | \phi_{-k} \sim$$

$$N \left[\frac{\rho \sum_{j=1}^n w_{kj} \phi_j}{\rho \sum_{j=1}^n w_{kj} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^n w_{kj} + 1 - \rho} \right],$$

In the above model, R_k denotes the risk of poverty in district k . The parameter ρ takes values between 0 and 1. If ρ is close to 1, the strong spatial autocorrelation in the data exist. A uniform prior is specified for ρ where $\rho \sim U(0,1)$. Here, w_{kj} denotes a measure of closeness between areas. Inference for this model is based on Markov Chain Monte-Carlo (MCMC) simulation, using a combination of Gibbs sampling and Metropolis-Hasting steps. This study used *CARBayes* [18] software, which is an R package for Bayesian spatial modelling with conditional autoregressive priors.

2.5. Neighbourhood weight matrices

In the spatial analysis, identifying the neighbourhood structure of the data being analysed is important when conducting tests for autocorrelation data and modelling data at the areal level. The neighbourhood matrix W is a $n \times n$ matrix whose element w_{kj} represents a measure of closeness between area k and area j . These elements can be binary or non-binary. In this study, the most common specification of W in the literature, including contiguity neighbour, Delauney triangulation and binary distance neighbourhood matrices are considered. The contiguity matrix is defined by whether two spatial units share a border or not, where $w_{kj} = 1$ if area k and j share a common border and 0 otherwise. While the Delaunay triangulation is defined as $w_{kj} = 1$ if centroid of area k and j are connected by triangulation edge and 0 otherwise. Lastly, for the binary distance neighbourhood matrix, where the specification is $w_{kj} = 1$ if the Euclidean distance between the centroids of area k and j less than the threshold distance, which is the maximum of the minimum distance between each area and all the others and 0 otherwise.

3. Results

In this study, The SPR values in Kelantan range between 0.27 and 6.40. Values above 1 represent areas with elevated levels of poverty risk, while values below 1 correspond to comparatively non-poverty areas. The map of their spatial pattern is displayed in Figure 2. It shows that Kursial has the highest SPR (the darker area) with the SPR 6.40. Followed by Pengkalan Kubor and Perupok with the values of 3.82 and 3.57, respectively. The Variance Inflation Factor (VIF) was used to measure collinearity among the variables in a regression model. It was found that there was no collinearity were observed between the covariates except for the number of female-headed households and the number of non-education of household head, where the value of $VIF > 10$. Apart from that, to reduce the collinearity between variables, the number of the female-headed household variable was transformed into a natural log. Thus, VIF becomes less than 10 and multicollinearity did not exist between covariates.

The areal unit's data tend to have spatial autocorrelation even after covariates were included, which was due to unmeasured covariates [19,20]. A Poisson log-linear model (Model 1) without any spatial structure was fitted to the poverty data to measure the existence of spatial autocorrelation. Then, the residuals from the model were tested for the presence of spatial autocorrelation using Moran's I statistic. The results of Moran's I p-value was less than 0.05, which indicated that the residuals of the model contain a strong spatial autocorrelation structure. Thus, the assumption of independence in the model was omitted. Since the model does not allow for residual spatial autocorrelation, a Poisson log-linear Leroux CAR model (Model (2)) with three different neighbourhood matrices was fitted to the poverty data. 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.

The deviance information criterion (DIC) [21] and RMSE were also measured to determine how well a model fits a set of data. The DIC and RMSE indicate a better fitting model if it has a low value. Table 1 shows the results of DIC and RMSE among the W matrices. The results across the three models are very similar. The model with a binary distance matrix appears to be the worst fit, with DIC and RMSE values of 563.52 and 0.88, respectively. The remaining two models have similar DIC values ranging between 548.00 and 550.00, but different RMSE. The RMSE value for the model with contiguity and Delaunay triangulation are 0.14 and 0.72, respectively. Therefore both neighbourhood matrices performed the best in estimating poverty risk.

The effects of the covariates are presented in Table 2, 3, and Table 4. These tables display the medians as well as 95% credible intervals of posterior poverty relative risk for all 66 districts in Kelantan. Estimates are obtained using Poisson log-linear Leroux CAR model with different neighbourhood weight matrices W and average age, number of non-education and number of the poor female head of household as covariates. The tables show that the choice of neighbourhood matrices has little effect on the results. Both the relative risks and credible intervals were consistent across the three different W . The household head age appears to be associated with poverty reduction, with a relative

risk of 0.93 for 2 years increase in the average age of the household head. There is convincing evidence that an increase in the number of non-education of the headed household by 62, is associated with an increased poverty risk by 21%, indicating that this covariate is a very informative covariate. Areas with a higher number of the poor female head of household were at high risk of poverty, with a relative risk of 1.12 for an increase of 0.16 in the poor female head of household.

Figure 3 and 4 show the estimated poverty risks from the Poisson log-linear Leroux CAR model with contiguity and Delaunay triangulation, respectively. The estimated risk maps are smoother than the raw SPR values and are also less extreme. For example, the SPR for poverty ranges between 0.27 and 6.40, while the corresponding model estimates range between 0.28 and 6.32. However, the estimated risk surface exhibits a similar spatial pattern to the SPR map, with the highest risk is Kusial, followed by Pengkalan Kubor and Perupok. In order to determine the appropriateness of Poisson log-linear Leroux CAR model for the data, the residuals from the models were tested again 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.

Table 1. DIC and RMSE values for the fitted models using neighbourhood weight matrices W .

Model	DIC	RMSE
Contiguity	549.03	.14
Delaunay triangulation	548.64	.72
Binary distance	563.52	0.88

Table 2. Median and 95% credible interval of posterior poverty relative risk for all 66 districts in Kelantan. Estimates obtained using Poisson log-linear Leroux CAR model with different neighbourhood weight matrices W and age as a covariate.

Model	Median	2.5%	97.5%
Contiguity	0.933	0.907	0.965
Delaunay triangulation	0.931	0.904	0.963
Binary distance	0.933	0.908	0.961

Table 3. Median and 95% credible interval of posterior poverty relative risk for all 66 districts in Kelantan. Estimates obtained using Poisson log-linear Leroux CAR model with different neighbourhood weight matrices W and number of non-education of household head as a covariate.

Model	Median	2.5%	97.5%
Contiguity	1.207	1.133	1.284
Delaunay triangulation	1.206	1.133	1.284
Binary distance	1.207	1.133	1.284

Table 4. Median and 95% credible interval of posterior poverty relative risk for all 66 districts in Kelantan. Estimates obtained using Poisson log-linear Leroux CAR model with different neighbourhood weight matrices W and the number of female household head as a covariate.

Model	Median	2.5%	97.5%
Contiguity	1.116	1.097	1.132
Delaunay triangulation	1.117	1.100	1.133
Binary distance	1.116	1.100	1.132

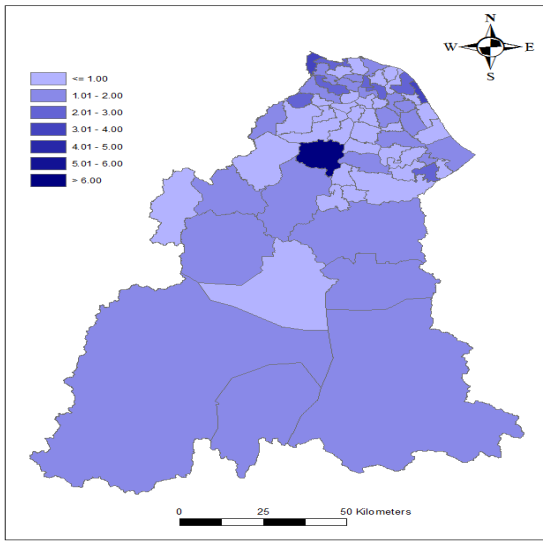


Figure 2. Spatial map of SPR for poor households in Kelantan.

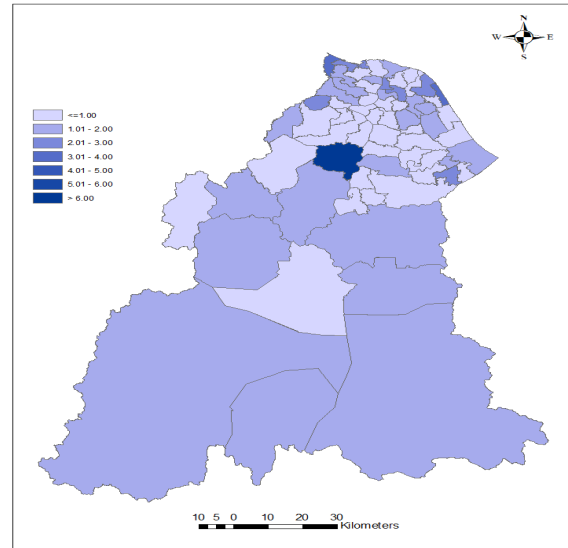


Figure 3. Estimated poverty risks from the Poisson Log Linear Leroux CAR model with contiguity matrix.

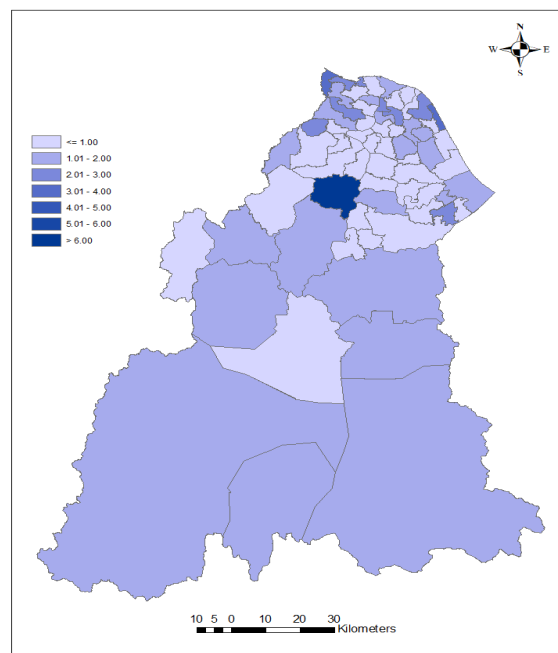


Figure 4. Estimated poverty risks from the Poisson Log-Linear Leroux CAR model with Delaunay triangulation matrix.

4. Conclusion

This study explored the most commonly used conditional autoregressive prior distributions, which was Poisson log-linear Leroux CAR model with different W . Previous studies commonly used contiguity matrix as neighbourhood matrix. However, this study showed that the performance of Poisson log-linear Leroux CAR with Delaunay triangulation was almost similar to contiguity matrix. Therefore, both neighbourhood matrices fit the poverty data best. Overall, Kursial appears to have a higher risk of poverty than the other districts considered in this study. These differences in risk appear to be partly due

to the covariates. Decreased numbers of non-education and female household head appear to decrease the poverty risk. In contrast, increase the age of poor household head appears to decrease the poverty risk.

This study found considerable differences in the smoothing properties of the Poisson log-linear Leroux CAR model, depending on the type of neighbours matrices W . This, in turn, had an effect on the ability of the models to predict the observed risk in an area. These results have significant implications for all researchers using Poisson log-linear Leroux CAR models, since the neighbourhood weight matrices chosen may markedly influence the findings of the study.

Acknowledgement

The authors would like to thank Universiti Malaysia Kelantan to give approval to conduct this research.

References

- [1] Claro, S., Paunesku, D. & Dweck, C. S. (2016), *Proceedings of the National Academy of Sciences* 113(31), 8664–8668.
- [2] Brody, G. H., Gray, J. C., Yu, T., Barton, A. W., Beach, S. R., Galv'an, A., MacKillop, J., Windle, M., Chen, E., Miller, G. E. et al. (2017), *JAMA pediatrics* 171(1), 46–52.
- [3] Kerr, W. C., Kaplan, M. S., Huguet, N., Caetano, R., Giesbrecht, N. & McFarland, B. H. (2017), *American journal of preventive medicine* 52(4), 469–475.
- [4] Santos, M. E. & Villatoro, P. (2018), *Review of Income and Wealth* 64(1), 52–82.
- [5] Roganovi'c, M. & Stankov, B. (2018), *Skola biznisa* (1), 21–38.
- [6] Braithwaite, J. & Mont, D. (2009), *ALTER-European Journal of Disability Research/Revue Europ'eenne de Recherche sur le Handicap* 3(3), 219–232.
- [7] Brandolini, A., Magri, S. & Smeeding, T. M. (2010), *Journal of Policy Analysis and Management* 29(2), 267–284.
- [8] EPU (2017), 'Malaysia-sustainable development goals voluntary national review 2017: high-level political forum'.
- [9] Mat Zin, R. (2011), *Singapore: Institute of Southeast Asian Studies*.
- [10] Majid, M. R., Jaffar, A. R., Man, N. C., Vaziri, M. & Sulemana, M. (2016), *Planning Malaysia Journal* 14(4).
- [11] Siwar, C., Ahmed, F., Zahari, S. Z., Idris, N. D. M., Mia, M. S. & Bashawir, A. (2016), *Journal of Environmental Science and Technology* 9(6), 407–416.
- [12] Dormann, C. F. (2007), *Global Ecology and Biogeography* 16(2), 129–138
- [13] Crimmins, E. M., Kim, J. K. & Seeman, T. E. (2009), *Biomedical Sciences and Medical Sciences* 64(2), 286–292.
- [14] Chapoto, A., Banda, D. J., Haggblade, S., Hamukwala, P. et al. (2011) *Technical report, Michigan State University, Department of Agricultural, Food, and Resource Economics*.
- [15] Anyanwu, J. C. (2014) *African Development Review* 26(1), 118–137.
- [16] Lastrapes, W. D. & Rajaram, R. (2016), *Eurasian Economic Review* 6(2), 153–171.
- [17] Leroux, B. G., Lei, X. & Breslow, N. (2000), *Institute for Mathematics and Its Applications* 116, 179.
- [18] Lee, D. (2013), *Journal of Statistical Software* 55(13).
- [19] Lee, D. & Mitchell, R. (2014), *Statistical methods in medical research* 0(0), 1–19.
- [20] Lee, D. & Shaddick, G. (2010), *Biometrics* 66(4), 1238–1246.
- [21] Spiegelhalter D, Best N, Carlin B, Van der Linde A. (2002) *J R Stat Soc Ser B* 2002;64:583–639.