Statistical Analysis of Vehicle Theft Crime in Peninsular Malaysia using Negative Binomial Regression Model
(Analisis Statistik Jenayah Kecurian Kenderaan di Semenanjung Malaysia menerusi Model Regresi Binomial Negatif)

MALINA ZULKIFLI, AHMAD MAHIR RAZALI*, NURULKAMAL MASSERAN & NORISZURA ISMAIL

ABSTRACT

The aim of this paper was to identify the determinants that influence vehicle theft by applying a negative binomial regression model. The identification of these determinants is very important to policy-makers, car-makers and car owners, as they can be used to establish practical steps for preventing or at least limiting vehicle thefts. In addition, this paper also proposed a crime mapping application that allows us to identify the most risky areas for vehicle theft. The results from this study can be utilized by local authorities as well as management of internal resource planning of insurance companies in planning effective strategies to reduce vehicle theft. Indirectly, this paper has built ingenuity by combining information obtained from the database of Jabatan Perangkaan Malaysia and insurance companies to pioneer the development of location map of vehicle theft in Malaysia.

Keywords: Crime; mapping; negative binomial; spatial analysis; vehicle theft

ABSTRAK

Tujuan penulisan kertas ini adalah untuk mengenal pasti penentu yang mempengaruhi kecurian kenderaan dengan menggunakan model regresi binomial negatif. Pengenalpastian penentu ini penting kepada pembuat dasar, pembuat kereta dan pemilik kereta kerana maklumat ini boleh digunakan untuk mewujudkan langkah-langkah praktikal dalam mencegah atau sekurang-kurangnya menghadkan kejadian kecurian kenderaan. Di samping itu, kertas ini juga mencadangkan suatu aplikasi pemetaan jenayah yang membolehkan kita mengenal pasti kawasan yang paling berisiko untuk berlakunya kecurian kenderaan. Hasil daripada kajian ini boleh digunakan oleh pihak berkuasa tempatan dan juga pihak pengurusan perancangan sumber dalaman syarikat insurans untuk merancang strategi yang berkesan bagi mengurangkan kecurian kenderaan. Secara tidak langsung, kertas kerja ini telah membina satu jalan pintar dengan menggabungkan maklumat yang diperoleh daripada pangkalan data Jabatan Perangkaan Malaysia dan syarikat-syarikat insurans untuk merintis kepada pembinaan peta lokasi kecurian kenderaan di Malaysia.

Kata kunci: Analisis reruang; binomial negatif; jenayah; kecurian kenderaan; pemetaan

INTRODUCTION

Vehicle theft is categorized as a property crime, also known as a highly organized crime and is one of the many global issues affecting the world, including Malaysia. Although most stolen vehicles are equipped with security systems or immobilizers, the number of vehicle theft cases is high and has increased over the years. In Malaysia, there are approximately 18 million registered vehicles and vehicle owners are required to purchase motor insurance. Statistics show that one vehicle was stolen every ten seconds during the first nine months of 2007 (Insurance Services Malaysia Bhd 2007). This insurance coverage will protect and pay the vehicle owner if their vehicle is stolen in which case the insurance covers the loss of the owner based on the market price of the stolen vehicle.

The crime index in Malaysia is defined as the index of crimes that are reported with sufficient regularity and significance and it is used to quantify crime and to evaluate the effectiveness of crime prevention measures.

The crime index in Malaysia can be divided into violent crimes and property crimes. Although violent crimes attract the greatest attention of both the public and the media in Malaysia, property crimes accounts for approximately 90% of all crimes reported from 1980 to 2004 and are closely linked to the total crime index. Approximately 49% of the total crime index is attributed to vehicle theft (Sidhu 2005).

The identification of the determinants of vehicle theft is important to policymakers, car manufacturers and car owners because they can indicate possible practical steps for preventing or at least limiting, vehicle theft. When predicting the risk of vehicle theft per vehicle, the exposure is the number of vehicles owned and the possible factors are vehicle use, vehicle model, vehicle age, vehicle location such as residential or business and vehicle geographical location. When relating vehicle theft with social and economic factors per census block size or district, the exposure is the size of the population which is in term of census block size or district size and the possible

factors are population density, average number of persons per household, proportion of the population aged 15-24, proportion of certain races and unemployment rates.

Kelly (2000) studied the relationship between inequality and crime using a quasi-Poisson regression model. Osgood (2000) analyzed juvenile arrest rates for robbery by implementing the Poisson and negative binomial regression models. Kleck and Chiricos (2002) studied the effect of unemployment on property crime by fitting the Poisson and negative binomial regression models. Di Tella and Schargrodsky (2004) estimated the effect of police presence on car theft by implementing the least squares regression model. Demombynes and Ozler (2005) examined the effects of local inequality on property and violent crime in South Africa using the negative binomial regression model.

Aitkin et al. (1990) and Renshaw (1994) used the Poisson regression model for fitting claim count data. Brockman and Wright (1992), Ismail and Jemain (2007) and McCullagh and Nelder (1989) suggested a quasi-Poisson regression model to accommodate overdispersion in claim count and count data in other areas. The negative binomial regression model has been fitted for overdispersed claim count and count data in other areas by Ismail and Jemain (2007), Lawless (1987), McCullagh and Nelder (1989) and Zulkifli et al. (2013). Generalized Poisson regression models have been applied to overdispersed and underdispersed count data by Consul (1989), Consul and Famoye (1992), Wang and Famoye (1997), Zamani and Ismail (2012) and Zamani and Ismail (2014).

Vehicle theft, like most other crime, is spatially concentrated. Crime mapping has become an important tools in crime and justice. Advances in information technology and software development provide new opportunities for digital mapping to assist crime and prevention programs. Crime mapping is also useful in studying the environmental and geographical aspects of crime. Maps allow the areas of unusually high or low concentration to be identified visually. However, the map is only a pictorial representation of the results of complex spatial data analysis.

Crime mapping is often thought of as the simple displaying and querying of crime data. However, it is a general term that encompasses the technical aspects of visualization and statistical techniques, as well as the practical aspects of geographic principles and criminological theories. In fact, the idea of crime mapping is not new, as it dated back to the early 1800's in France. A review of the literature from that period to the present shows several epochs during which interest in crime mapping was high, but then faded dramatically (Weisburd & McEwen 1997). The first map of crime was created by Adriano Balbi and Andre-Michel Guerry in 1829 (Beirne 1993; Kenwitz 1987). Using criminal statistics from 1825 to 1827 and demographic data from France's census, they developed maps of crimes against property, crimes against persons and levels of education. In recent years, research by Di Tella and Schargrodsky (2004) found a large, negative

and highly local effect of police presence on car theft. Blocks that receive police protection experience 0.081 fewer car thefts per month than blocks that do not.

In this paper, we proposed an application of the negative binomial regression model for identifying the determinants influencing vehicle theft and an application of crime mapping that identifies the riskiest areas for vehicle theft. In the next section, we discuss the material and methods used in this study, followed by the results and discussion sections. The last section is the conclusion.

MATERIALS AND METHODS

SPATIAL ANALYSIS

The Moran's I test is a test measures the spatial autocorrelation in a random field. Spatial autocorrelation in a random field is the correlation between an observation at region i, y_i and an observation at region j, y_j. Spatial autocorrelation measures the correlation between the same attributes at two locations. A positive autocorrelation implies that observations in close spatial proximity are expected to be more similar than observations that are more spatially separated. Conversely, a negative spatial autocorrelation implies that, proximity in space should not provide similar attribute values. The Moran's I test is calculated as the ratio of the product of the variable of interest and its spatial lag, to the cross product of the variable of interest, adjusted for the spatial weights used. Its formula is given by:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \overline{y})(y_i - \overline{y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} (y_i - \overline{y})^2}, \quad (1)$$

where y_i is the observation or attribute value in region i, $\overline{y}$ is the mean of the variable of interest and w_{ij} is the spatial weight of the link between region i and region j. The weight w_{ij} assigned to region j, if it is a neighbour of region i, can be written as:

$$w_{ij} = \begin{cases} 1 & \text{if region } i \text{ and } j \text{ are conntected} \\ 0 & \text{if region } i \text{ and } j \text{ are not conntected} \end{cases}. \quad (2)$$

Inference for I can be derived by using permutation tests, Monte-Carlo tests or approximate tests based on the asymptotic distribution of I. Here, the mean and variance of I can be derived via the assumption of randomizing the attribute values to the lattice regions. The mean for I via the assumption of randomizing is given as:

$$E_r[I] = -\frac{1}{n-1}, \quad (3)$$

The Moran's I statistic is as follows. If $I > E_r[I]$, the spatial autocorrelation is positive and increases in strength with $|I - E_r[I]|$ and the connected regions tend to have

similar attribute values. In contrast, if $I < E_r[I]$, the attribute value of connected regions tends to be dissimilar, which implies a negative spatial autocorrelation in a random field (Schabenberger & Gotway 2005).

INVERSE DISTANCE WEIGHTING METHOD (IDW)

IDW is a simple method for the spatial estimation of a random field. IDW is a weighted average interpolator, which can either be an exact or a smoothing interpolator. In IDW, data are weighted during interpolation such that the influence of one point relative to another declines as the distance increases. The value $p(Y; s_0)$ at the pivot point s_o can be estimated by using a weighted mean of the available measurements through the expression.

$$p(y; s_0) = \frac{\sum_{i=1}^{n} W(s_i, s_0) y_i}{\sum_{i=1}^{n} W(s_i, s_0)}, \qquad (4)$$

where y_i is the observed data for region; i, (s_i, s_0) is the distance between the centre region i and the pivot point s_0, and $W(s_i, s_0)$ is the weighting factor that decreases as the distance increases. Its value decreases with distance following a quadratic or exponential law.

The IDW has been recognized as a good method for a spatial interpolation that is based on the knowledge of distance. For the vehicle theft crime data, we believe that the adjacent region will have similar characteristics in term of the vehicle theft crime. Thus, it is reasonable to give a large weighted for the adjacent region and a smaller weighted for the regions that have a long distance between each other. The interpolation values for the number of vehicle theft crime at the unobservable point are determined using the observable pair of data accompanied with the weighted assigned. Apart from that, although there are many other methods such as based on the Kriging, triangulation and trend surface, all of these methods have its own inherent assumptions and strict requirement about the spatial in order to apply it to a real data set. While the IDW does not require any strong assumption, it can always be a good alternative method in spatial analysis. For example, the Kriging method requires knowledge of spatial correlation in the random field that is translated using the semivariogram model. In our pre-determined analysis (not shown in this paper), we have found that the spatial correlation for the number of vehicle theft crime in Peninsular Malaysia cannot be fitted to any of the available semivariogram. Consequently, a spatial interpolation based on Kriging will not provide a good result. Thus, according to Hengl (2007) and Masseran et al. (2012a, 2012b), the IDW is a reliable method to overcome this problem.

NEGATIVE BINOMIAL REGRESSION MODELS

Let $(Y_1, Y_2, \ldots, Y_n)^T$ be the vector of count random variables where Y_i and Y_j are independent for any $i \neq j$ and n is the sample size. The probability mass function (pmf) of a Poisson regression model is given by:

$$\Pr(Y_i = y_i | \mu_i) = \frac{\mu_i^{y_i}}{y_i!} \exp(-\mu_i), \quad y_i = 0, 1, 2, \ldots, \qquad (5)$$

where the mean and the variance are equal, $E(Y_i) = Var(Y_i) = \mu_i$.

To incorporate covariates and to ensure non-negativity, the mean is included in the Poisson model through a log link function, $\mu_i = e_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ or $\ln(\mu_i) = \ln(e_i) + \sum_k \beta_k x_{ik}$, where x_{ik} denotes the explanatory variables, β_k are the regression parameters, e_i is the exposure, $\mathbf{x}_i$ is the vector of the explanatory variables and $\boldsymbol{\beta}$ is the vector of the regression parameters.

Under the Poisson distribution, the mean is assumed to be constant or homogeneous within case i, or level i, or cell i. By defining a specific distribution for the mean, heterogeneity within cases is allowed. Assuming $Y_i | \lambda_i$ is distributed as a Poisson with conditional mean $E(Y_i | \lambda_i) = \lambda_i$, λ_i is distributed as a gamma with mean $E(\lambda_i) = \mu_i$ and variance is $Var(\lambda_i) = \mu_i^2 v_i^{-1}$, the marginal distribution of Y_i follows a negative binomial with pmf;

$$\Pr(Y_i = y_i) = \int \Pr(Y_i = y_i | \lambda_i) f(\lambda_i) d\lambda_i$$
$$= \frac{\Gamma(y_i + v_i)}{\Gamma(y_i + 1)\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i}, \qquad (6)$$

where the mean is $E(Y_i) = \mu_i$ and the variance is $Var(Y_i) = \mu_i(1 + \mu_i v_i^{-1})$. When $v_i = a^{-1}$, the negative binomial distribution is produced, with mean $E(Y_i) = \mu_i$ and variance $Var(Y_i) = \mu_i(1 + a\mu_i)$. The pmf (Cameron & Trivedi 1986; Lawless 1987) is

$$\Pr(Y_i = y_i | \mu_i, a) = \frac{\Gamma(y_i + a^{-1})}{y_i! \Gamma(a^{-1})} \left(\frac{a^{-1}}{a^{-1} + \mu_i}\right)^{a^{-1}} \left(\frac{\mu_i}{a^{-1} + \mu_i}\right)^{y_i},$$
$$y_1 = 0, 1, \ldots \qquad (7)$$

where a denotes the dispersion parameter. If a equals to zero, the mean and the variance are equal; if $a > 0$, the variance exceeds the mean and the NB allows for overdispersion.

If the mean is assumed to follow a log link function, $\mu_i = e_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, the log likelihood for the NB regression model can be written as

$$\ln L(\boldsymbol{\beta}, a) = \sum_i \ln\left(\Gamma\left(y_i + a^{-1}\right)\right) - \ln\left(\Gamma\left(a^{-1}\right)\right) \qquad (8)$$
$$- \ln(y_i!) + y_i \ln(a\mu_i)$$
$$- \left(y_i + a^{-1}\right) \ln(1 + a\mu_i).$$

The maximum likelihood estimates for the negative binomial regression model can also be obtained by maximizing $\ln L(\boldsymbol{\beta}, a)$ with respect to $\boldsymbol{\beta}$ and a.

LIKELIHOOD RATIO TEST, WALD TEST, AIC AND BIC

The test of overdispersion in the NB regression models can be assessed using the likelihood ratio test because NB regression reduces to Poisson regression in the limit when $a \rightarrow 0$. The hypothesis are $H_0 : a = 0$ against $H_1 : a > 0$ which is a one-sided test. The likelihood ratio is $T = 2(\ln L_1 - \ln L_0)$, where $\ln L_1$ and $\ln L_0$ are the model's log likelihood under the respective hypothesis. Since the null hypothesis is on the boundary of parameter space, T has an asymptotic distribution of probability mass of 0.5 at zero and 0.5 of chi-square distribution with one degree of freedom. In other words, to test the null hypothesis at significance level α, the critical value of chi-square distribution with significance level 2α is used, or reject H_0 if $T > \chi^2_{1-2\alpha,1}$. As an example, for 0.05 significance level, the critical value is $\chi^2_{0.90,1} = 2.705$ instead of $\chi^2_{0.95,1} = 3.8415$.

The test of overdispersion in the negative binomial regression model can also be performed by the Wald statistic, which is defined as the ratio of the estimated overdispersion parameter to its standard error, $\frac{\hat{a}}{\sqrt{Var(\hat{a})}}$, where asymptotically, the statistic follows a standard normal distribution.

One can also compare the performance of alternative models based on several likelihood measures, such as the Akaike Information Criteria (AIC) and the Bayesian Schwartz Information Criteria (BIC). The AIC penalises a model with a larger number of parameters and is defined as $AIC = -2 \ln L + 2p$, where $\ln L$ denotes the fitted log likelihood and p is the number of parameters. The BIC penalises a model with a larger number of parameters and a larger sample size and is defined as $BIC = -2 \ln L + p \ln(n)$, where n is the sample size. The best model is indicated by the smallest AIC and BIC.

RESULTS

DATA

Crime data for private car theft were obtained and compiled from insurance companies in Malaysia. In particular, automobile theft insurance indemnifies the insured against the loss of a motor vehicle through theft. The data are from 2001 to 2003 and were supplied by Insurance Services Malaysia (ISM).

Population data on the social and economic background of each district were obtained and compiled from the Malaysian 2000 census data, supplied by Jabatan Perangkaan Malaysia. The data provide information on 81 districts of all 11 states in Peninsular Malaysia. The district population ranges from 11,183 to 1,305,792 with an average of 163,856.

Table 1 shows the demographic and sociological factors considered for the independent (explanatory) variables. Further explanations on the demographic and social factors in Table 1 are as follow:

TABLE 1. Demographic and sociological factors for vehicle theft data

Factors	Classes
Population density	1. Low
	2. Medium
	3. High
Gender	1. Men
	2. Women
Bumiputra	1. Less
	2. Moderate
	3. High
Chinese	1. Less
	2. Moderate
	3. High
India	1. Less
	2. Moderate
	3. High
Others :	1. Less
	2. Moderate
	3. High
Foreign migrants	1. Less
	2. Moderate
	3. High
Professional :	1. Less
	2. Moderate
	3. High
Work :	1. Less
	2. Moderate
	3. High
Productivity	1. Not productive
	2. Productive
	3. Very productive
Police station	1. Less
	2. Moderate
	3. High

Population density Population density is defined as number of population per km². Population density is divided into three quantiles; low, medium and high

Gender Gender is divided into two proportions; male and female

Bumiputra Bumiputra is defined as number of bumiputra per 1000 population. This factor is divided into three quantiles; less, moderate and high

Chinese Chinese is defined as number of chinese per 1000 population. This factor is divided into three quantiles; less, moderate and high

Indian Indian is defined as number of indian per 1000 population. This factor is divided into three quantiles; less, moderate and high

Others   Others is defined as the number of other races (besides bumiputra, chinese and indian) per 1000 population. This factor is divided into three quantiles; less, moderate and high

Foreign migrant   Foreign migrant is defined as number of foreign migrant per 1000 population. Foreign migrant is divided into three quantiles; less, moderate and high

Professional   Professional is defined as number of professional workers per 1000 population. Professional is divided into three quantiles; less, moderate and high

Work   Work is defined as number of working employee per 1000 population. Work is divided into three quantiles; less, moderate and high

Productivity   Productivitiy is defined as productivity per 1000 population. Productivity is divided into three quantiles; not productive, productive and very productive

Police station   Police station is defined as number of police station per 1000 populatio. Police station is divided into three quantiles; less, moderate and high

RESULTS ON NEGATIVE BINOMIAL REGRESSION ANALYSIS

In determining the factors that influence the occurrence of vehicle theft in certain areas, independent variables (explanatory variables) for each district are analysed. In this study, negative binomial regression is used to model the relationship between the explanatory variables with the number of burglaries in the district, where population is used as offset.

Table 2 shows the parameters, log likelihood, AIC and BIC for the fitted regression model. Based on the *p*-value, the results indicated that high population density is a significant factor in the occurrence of vehicle theft. Thus, the district with higher population density tends to have more cases of car theft. It also appears that the number of police stations in a district has a significant effect on the rate of car thefts in the area. However, racial composition, foreign migration, productivity, number of workers and number of professional employees do not play an important role in influencing the rate of car theft in any district.

Table 3 shows the parameters, log likelihood, AIC and BIC for the negative binomial regression model with covariates that are significant at 5% level. It appears that population density, Chinese composition and the number

TABLE 2. Estimated parameters for negative binomial regression model

Parameter		Est.	*t*-ratio	*p*-value	
Intercept		-9.5634	-34.455	0.000	***
Density :	medium	0.2209	0.596	0.551	
	high	1.1256	2.916	0.004	**
Gender :	female	-0.3576	-1.115	0.265	
Bumiputra :	moderate	0.2628	0.542	0.588	
	high	-0.5479	-0.533	0.594	
Chinese :	moderate	0.4955	0.455	0.649	
	high	1.7038	1.517	0.129	
India :	moderate	-0.5474	-1.153	0.249	
	high	-0.5535	-0.982	0.326	
Others :	moderate	-0.4496	-1.204	0.229	
	high	0.2807	0.805	0.421	
Foreign migrants :	moderate	-0.0438	-0.115	0.908	
	high	-0.3593	-0.815	0.415	
Professional :	moderate	0.1946	0.417	0.677	
	high	0.6486	1.157	0.247	
Work :	moderate	-0.4115	-0.939	0.348	
	high	-0.8040	-1.437	0.151	
Productivity :	productive	-0.2349	-0.559	0.576	
	very productive	-0.1983	-0.429	0.668	
Police station :	less	-1.0616	-2.760	0.006	*
	moderate	-0.8509	-2.041	0.041	**
Log likelihood			-269.943		
AIC			559.286		
BIC			583.830		

***, ** and * indicates that the parameters are significant at 0.01, 0.05 and 0.10 levels, respectively

of police stations is significant factors contributing to the car theft rate in the districts of Peninsular Malaysia.

RESULTS ON SPATIAL ANALYSIS

Figure 1 shows the population density of the districts of Peninsular Malaysia. The darkest area indicates the district with the highest population density.

    Figure 2 shows the map of vehicle theft in districts of Peninsular Malaysia based on the negative binomial regression model shown in Table 3. The darkest area on the map has the highest rate of car theft. As shown, areas with high rate of car theft are Klang Valley, Johor Bahru and Penang, which are areas of high population density,

with many industrial activities and locations of tourism attraction.

DISCUSSION

This paper builds an initiative by combining the information from Jabatan Perangkaan Malaysia and insurance companies to pioneer the creation of map of vehicle theft location in Peninsular Malaysia. The map and the accompanying analyses will help policymakers to classify districts and identify primary factors regarding vehicle theft. These information can also be supplied to local authorities and management of resource planning department of insurance companies for planning strategies of reducing car theft.

TABLE 3. Negative binomial regression model with significant covariates

Parameter		Est.	t-ratio	p-value	
Intercept		2.4364	4.729	0.000	***
Density :	moderate	0.4883	1.248	0.212	***
	high	1.7285	3.830	0.000	
Chinese :	moderate	-0.3523	-0.833	0.405	**
	high	1.0400	1.983	0.047	
Police station :	less	-1.8930	-4.379	0.000	***
	moderate	-1.4454	-3.642	0.000	***
Log likelihood			-265.275		
AIC			552.57		
BIC			578.89		

***, ** and * indicates that the parameters are significant at 0.01, 0.05 and 0.10 levels, respectively
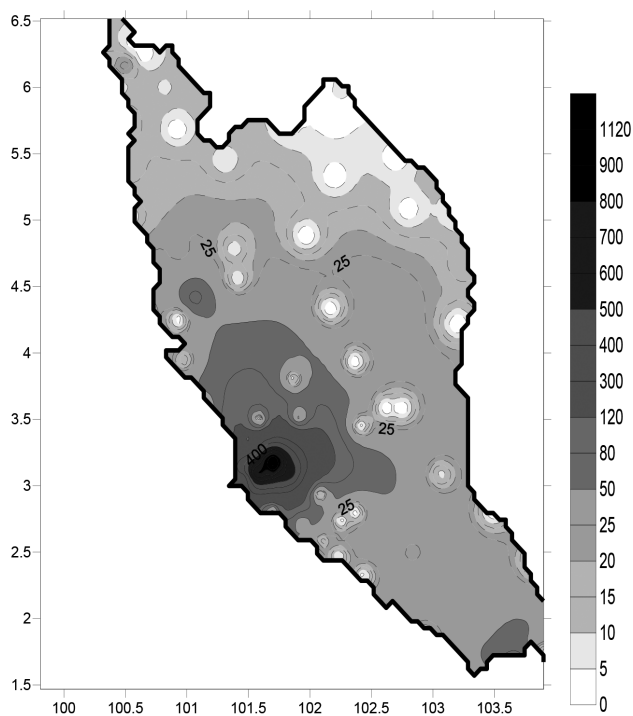

FIGURE 1. Map of population density in districts of Peninsular Malaysia
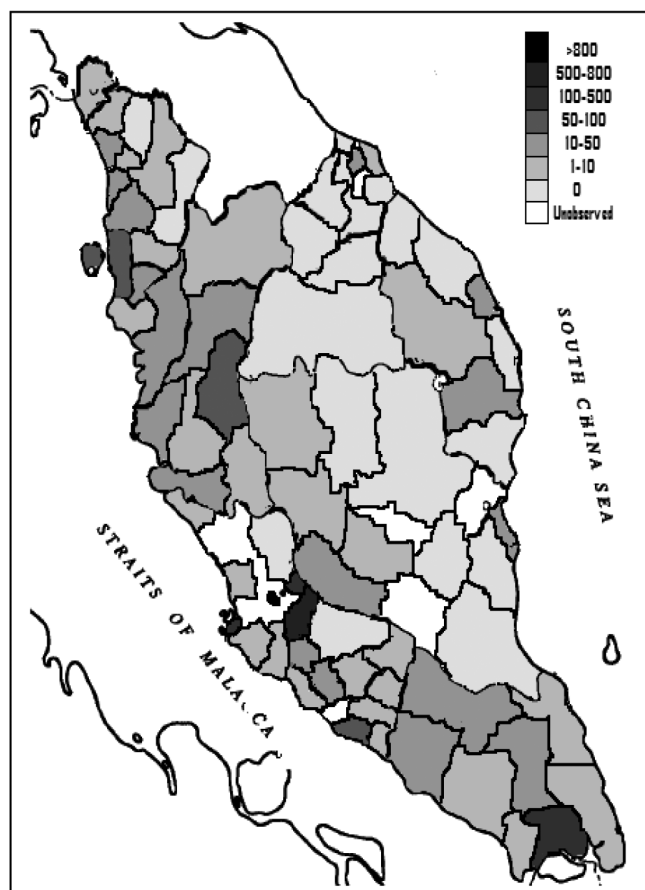
FIGURE 2. Map of vehicle theft in districts of Peninsular Malaysia
based on negative binomial regression model

There are several preventive and corrective measures that can be implemented by manufacturers to reduce the number of stolen vehicles such as anti-theft alarms and anti-theft key locks. Property managers may also assist by providing protected and secured parking, as well as hiring security guards.

The Royal Malaysian Police may also strategize in increasing the number of police stations and locating them in strategic and dangerous areas, especially in the areas that show large numbers of reported vehicle thefts. A study on the existence of anti-theft devices and secured parking, as well as the number and the location of the police stations, can be carried out if information on these factors is provided in police reports (for the number of stolen vehicles), insurance claim reports (for the number of vehicle theft claims) or insurance policies (for the number of vehicle exposures). That information can be embedded in a regression model by using the contributing factors as covariates.

REFERENCES

Aitkin, M., Anderson, D., Francis, B. & Hinde, J. 1990. *Statistical Modelling in GLIM*. New York: Oxford University Press.

Beime, P. 1993. *Inventing Criminology*. Albany, NY: State University of New York Press.

Brockmann, M.J. & Wright, T.S. 1992. Statistical motor rating: making effective use of your data. *Journal of the Institute of Actuaries* 119(3): 457-543.

Cameron, A.C. & Trivedi, P.K. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1: 29-53.

Consul, P.C. 1989. *Generalized Poisson Distribution: Properties and Application*. New York: Marcel Dekker.

Consul, P.C. & Famoye, F. 1992. Generalized Poisson regression model. *Communications in Statistics (Theory & Methodology)* 2(1): 89-109.

Demombynes, G. & Ozler, B. 2005. Crime and local inequality in South Africa. *Journal of Development Economics* 76: 265-292.

Di Tella, R. & Schargrodsky, E. 2004. Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *The American Economic Review* 94(1): 115-133.

Hengl, T. 2007. *A Practical Guide to Geostatistical Mapping of Environmental Variables*. Italy: European Communities.

Insurance Services Malaysia Bhd. 2007. Insurance industry statistics on stolen vehicles. http://www.piam.org.my/news/piamnews/p014.htm. Accessed 22 June 2011.

Ismail, N. & Jemain, A.A. 2007. Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuarial Society Forum Winter*. pp. 103-158.

Kelly, M. 2000. Inequality and crime. *The Review of Economics and Statistics* 82(4): 530-539.

Kenwitz, J.W. 1987. *Cartography in France: 1660-1848*. Chicago, IL: University of Chicago Press.

Kleck, G. & Chiricos, T, 2002. Unemployment and property crime: A target-specific assessment of opportunity and motivation as mediating factors. *Criminology* 40(3): 649-680.

Lawless, J.F. 1987. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics* 15(3): 209-225.

Masseran, N., Razali, A.M. & Ibrahim, K. 2012a. An analysis of wind power density derived from several wind speed density functions: The regional assessment on wind power in Malaysia. *Renewable and Sustainable Energy Reviews* 16(8): 6476-6487.

Masseran, N., Razali, A.M., Ibrahim, K., Zin, W.Z.W. & Zaharim, A. 2012b. On spatial analysis of wind energy potential in Malaysia. *WSEAS Transactions on Mathematics* 11(6): 467-477.

McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman and Hall.

Osgood, W. 2000. Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology* 16: 21-43.

Renshaw, A.E. 1994. Modelling the claims process in the presence of covariates. *ASTIN Bulletin* 24(2): 265-285.

Schabenberger, O. & Gotway, C.A. 2005. *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall/CRC Press.

Sidhu, A.S. 2005. The rise of crime in Malaysia: An academic and statistical analysis. *Journal of the Kuala Lumpur Royal Malaysia Police College* 4: 1-28.

Wang, W. & Famoye, F. 1997. Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics* 10: 273-283.

Weisburd, D. & McEwen, T. 1997. *Introduction: Crime Mapping and Crime Prevention*. Monsey, NY: Criminal Justice Press.

Zamani, H. & Ismail, N. 2012. Functional form for the generalized Poisson regression model. *Communications in Statistics - Theory and Methods* 41(20): 3666-3675.

Zamani, H. & Ismail, N. 2014. Functional form for the zero-inflated generalized Poisson regression model. *Communications in Statistics - Theory and Methods* 43(3): 515-529.

Zulkifli, M., Ismail, N. & Razali, A.M. 2013. Analysis of vehicle theft: A case study in Malaysia using functional forms of negative binomial regression models. *Applied Mathematics and Information Sciences* 7(2L): 389-395.

Malina Zulkifli
School of Quantitative Sciences, College of Arts and Science
Universiti Utara Malaysia
06010 Sintok, Kedah Darul Aman
Malaysia

Ahmad Mahir Razali* , Nurulkamal Masseran & Noriszura Ismail
School of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: mahir@ukm.edu.my