2016

Spatio-Temporal Analysis Of The Occupational Fatal Victimization Of Law Enforcement Officers In The US

Xueyi Xing
University of South Carolina

SPATIO-TEMPORAL ANALYSIS OF THE OCCUPATIONAL FATAL VICTIMIZATION OF LAW ENFORCEMENT OFFICERS IN THE US

by

Xueyi Xing

Bachelor of Medcine
Sichuan University, 1996

Master of Medcine
Sichuan University, 1999

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Public Health in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2016

Accepted by:

Bo Cai, Director of Thesis

Robert Moran, Reader

Andrew Ortaglia, Reader

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the members of my thesis committee for providing their guidance and advice throughout this project. I would like to thank Dr. Bo Cai, for serving as my thesis Chair throughout my studies. As my thesis Chair, Dr. Cai has steered my thesis in a direction best suited for my area of study. I thank him for being there to answer any questions I had and for the random talks, especially when I would stop by his office just to say hi. I thank Dr. Robert Moran for being a great mentor, and providing convenience to support my study. I also want to thank Dr. Andrew Ortaglia for providing me with recommendations on advancing the front-end of my thesis; it truly develop the aspects behind the study’s purpose. I would like to thank you all for agreeing to be on my committee and for all your efforts that greatly improved the quality of this thesis.

I would also like to thank my family who have supported me throughout this journey. My wife Qihong and my children were by my side during this journey and always supported me with great encouragement. Without them, I would not be where I am today. I would like to thank my fellow graduate students, especially Brian Fulleihan, who was always willing to provide his assistance and friendship. Finally, I want to thank all of those people who, once upon a time, were my teachers and who helped me begin this journey. Their names are too numerous to mention, but many of them inspired me to continue learning and sharing with others

ABSTRACT

The models with constant coefficients of the covariates across space and time are commonly used in spatio-temporal analyses. However, the associations between risk factors and the outcome could have locally differential temporal trends in many cases. In this study, a Bayesian latent cluster modeling strategy is employed to identify potential spatial clusters in which locally specific sets of temporally varying coefficients of covariates are allowed. A state-level panel data of police officers occupational fatal victimization for the years 1979-2010 is used. To accommodate overdisperson and excess zeros, a negative binomial model and zero-inflated Poisson/negative binomial models are also utilized. A series of alternative models are also applied to this data. The model comparison shows that the proposed latent clusters Zero-Inflated Poisson model is superior to the other models. The analysis using the proposed model illustrates the heterogeneity in the associations between police fatal victimization outcome and specific risk factors across the latent spatial clusters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR .... Autoregressive

CAR .... Conditional Autoregressive

CPO .... Conditional Predictive Ordinate

DIC .... Deviance Information Criterion

FBI .... Federal Bureau of Investigation

GLMM .... Generalized Linear Mixed Model

LEOKA .... Law Enforcement Officers Killed and Assaulted

MCMC .... Markov Chain Monte Carlo

NLLK .... Negative Cross-validatory Predictive Log-Likelihood

NIOSH .... National Institute for Occupational Safety and Health

POFV .... Police Occupational Fatal Victimization

SMR .... Standardized Mortality Ratio

UCR .... Uniform Crime Report

ZIP .... Zero-Inflated Poisson

ZINB .... Zero-Inflated negative binomial

CHAPTER 1

INTRODUCTION

There has been growing interest in examining the distribution and the factors that affect the variations of incidents of health outcomes involving both spatial and temporal related information. The advantage of a spatial-temporal analysis over a pure geographical analysis is that it can illustrate the trend patterns over time and spatial pattern across regions simultaneously. It not only helps us accurately estimate the risks of outcome incidents and depict the clear relationship between the response variable and related risk factors, but also discern certain patterns from residuals due to unmeasured or unobservable covariates after taking into account the heterogeneity resulting from the variations in space and time. A common assumption used in the existing research is that the effects of risk factors are fixed over time and space. Such an assumption may be too restrictive in many cases. Motivated by a study exploring the relationship between occupational fatal violence victimization of police officers in the USA and related risk factors, I apply a space-time latent component model to identify potential spatial clusters in which locally specific sets of temporally varying coefficients of covariates are allowed.

Occupational violent victimization is a serious concern for law enforcement officers, because they are more likely to come into contact with the unstable elements of the population, face high levels of criminal violence, and work in more unpredictable situations than common people. Warchol (1998) and Duhart (2001) report that working as a police officer has the highest violence victimization risk in the work place among all occupations. Felonious murder is the most serious consequence of violence against police. National Institute for Occupational Safety and Health (NIOSH, 2002) research ranks law enforcement officers as second in terms of probability of being murdered during the work time, which is only lower than that of taxicab drivers. Besides the extremely traumatic experience it brought to victims' families, the loss of police officers has substantial adverse impacts to agencies and local communities. It could lower the morale of other colleagues or trigger unnecessary aggressive policing strategies, which, in turn, may damage the trust between police and the public. Therefore, the research on the related risk factors of fatal victimization of law enforcement officers draws scholars' attention considerably.

However, some methodological obstacles exist in incident level research. The measurement of police-citizen interactions and related contextual variables suffers from substantial reporting or recording biases. For example, there could be highly varied measurement errors regarding officers' dispositions in police-citizen contacts (Johnson, 2011). Also, the information regarding related situational factors during confrontations could be distorted, depending on whether it was gathered from officer-reported data or arrestee-reported data (Rojek, Alpert, & Smith, 2012). Employing observers to document police-citizen encounters could provide more detailed and accurate information, but because officers are aware of being observed, they may alter their routine practices and demonstrate more socially desirable behaviors than usual (reactivity bias, Spano, 2005).

Since the measurement issues make incident level research very difficult, most studies on the murder risk for officers were conducted at aggregate level. Extant research examined the relationship between victimization risks of police and a wide range of variables, i.e., negative social economic structures, mentally impaired population (people with mental illness and/or substance abuse), agency practice policy, firearm accessibility, first-aid availability, and political factors, etc. However, the findings of the research in this field are disparate. There has not been a single covariate identified as significant or with the same sign across existing studies (Kaminski, 2008). One reason for the lack of consistency is the heterogeneity of spatial and temporal variations of the data analyzed. Researchers agree that the felonious killings of officers are very rare incidents (Kaminski, et al, 2004; Kesic, et al., 2013). The rarity of observations of fatal victimization of police often makes the meaningful statistics analysis impractical. A common solution to this situation is to combine the rare event incidents across space and over time so as to yield enough observations, which also introduces substantial spatial and temporal variations into the data. The association between the victimization outcome and related risk factors may be influenced by such variations. For example, research shows that some negative social structural factors (i.e., poverty, unemployment, racial heterogeneity, etc.) could constitute a criminogenic environment in which police officers are close to the pool of potential offenders. Thus, it is expected that there are positive associations between these factors and the fatal victimization of officers. However, such associations may not be consistent across all the jurisdictions and the whole time period. The impacts of these factors on the outcome incidents could be subject to the local economic environments, the changes in social welfare policies, the evolution of policing ideology, and even the local cultures. These elements have striking variations in terms of space and time (Department of Justice, DOJ), but are usually difficult to measure or to observe.

Conventional modelling strategies often simply assume that the coefficients of the covariates are fixed. However, heterogeneity caused by massive space and time variations in the police occupational victimization study strongly challenges this assumption (Bailey, 1982; Bailey & Peterson, 1987, 1994; Fridell, et al., 2009; Jacobs & Carmichael, 2002; Kaminski, 2002, 2008; Kent, 2010; Moody, Marvell, & Kaminski, 2002; Mustard, 2001; Kaminski, 2002). Given this circumstance, a more practical assumption is that the effects of covariates on the fatal occupational violence victimization of police may vary across space and time units. In the proposed approach, I adopt a Bayesian space-time latent cluster model (Choi, et al. 2012) to explore the roles of the covariates in explaining occupational murders of police officers. This model allows detection of spatial clusters in which the associations between the outcome and covariates are homogeneous due to their similarity. It also enables the estimation of the varying temporal patterns within the clusters. To address overdispersion and large numbers of zero values in the respondent variable, two common problems in health data, negative binomial and zero-inflated Poisson/negative binomial models are applied. The model with the best fit is selected by using model comparison measures. This model is also compared to various models without considering spatial dependence and conventional spatio-temporal models to evaluate its performance.

This thesis proposal proceeds as follows. Chapter 2 provides a brief literature review of the Bayesian spatio-temporal analysis. Chapter 3 describes the proposed data collection and analytic strategy for this research. Chapter 4 explains the proposed Bayesian spatio-temporal latent model. Chapter 5 reports the results from the analysis. Chapter 6 ends the paper with discussions and conclusions.

CHAPTER 2

LITERATURE REVIEW

In the analysis of health data with spatial and temporal dimensions, some statistical approaches need to be applied to deal with the issues of dependences and the variations in space and time. The geographical or time-series dependences (i.e., autocorrelation) might exist because the values of outcome in any given region or time point could be impacted by its neighboring areas or time periods. One possible cause of such impacts could result from the similar characteristics, which are usually unobservable or unmeasurable, in these adjacent areas and time periods (Wakefield & Elliot, 1999). If an analysis fails to control for the heterogeneity introduced by spatial and temporal autocorrelations, it could yield misleading risk assessments and unstable coefficient estimates. Also, when the incident is rare, zero and low counts usually dominate the data. A raw mapping of incident risk rates is not an accurate reflection of the risk estimates, i.e., zero counts do not mean zero risks. Moreover, individual extreme values could distort the risk estimates markedly in rare event cases, especially for those areas with small sample sizes. Such “noise” covers the “true patterns of underlying risk” (Richardson, Abellan, & Best, 2006). Therefore, some type of smoothing, which means borrowing the information from the neighboring observed units, should be considered to take these issues into account (Knorr-Held & Besag, 1998).

Over the last few decades, hierarchical Bayesian modelling has been frequently applied in the analysis of spatio-temporal referenced epidemiological data. Briefly, while traditional frequentists view the parameters of the examined distributions as fixed, Bayesians believe these parameters also have their own distributions. The likelihood of observed data combined with researcher's prior beliefs can be used to compute the posterior distribution of these parameters (Greenland, 2006). In the analysis of space-time referenced data, the distinct advantage of Bayesian hierarchical modeling over the traditional frequentist approach is that it can easily combine certain spatial or temporal priors and hyperprior beliefs (distributions) in a hierarchical manner to incorporate geographical and time series information into the model (Waller, Carlin, Xia, & Gelfand, 1997, Carlin & Louis, 2000).

For a spatio-temporal referenced rare event data, a general Bayesian hierarchical model is given as below:

Let Y_{it} denote the observed counts of outcome for region *i* at year *t*. It is usually assumed that the outcome follows a Poisson distribution as Y_{it} ~Pois($E_{it}\theta_{it}$), where E_{it} is the expected count and θ_{it} is the relative risk in the *i*th region and tth year. Typically, the log relative risk can be written as

$$ln(\theta_{it}) = x'_{it}\beta + u_i + v_i + \delta_t \qquad (1)$$

where x'_{it} is the vector of covariates of state *i* at year *t*, β denotes the vector of the corresponding coefficients, u_i represents the unstructured spatial random effect for state *i*, v_i is a structured spatial element for state *i*, and δ_t denotes the excess variation coming from temporal autocorrelation. Therefore, this model captures the heterogeneity due to random sampling effects in states, the spatial dependence among adjacent regions, and the temporal autocorrelation between the values of outcome in the current and previous time period. Usually, a conditional autoregressive (CAR) distribution is assigned to v_i,

$$[v_i|v_j, j \neq i, \sigma_v^2] \sim \text{Normal}(\frac{1}{n_i}\sum_{j\neq i} w_{ij}v_j, \frac{1}{n_i}\sigma_v^2)$$

where n_i denotes the number of the neighbor states of state i, w_{ij}=1 if i and j are adjacent states and w_{ij}=0 otherwise, and σ_v^2 is the variation parameter. To consider the temporal autocorrelations between the outcome value in the current time period and the one in the previous time period, an autoregressive prior, i.e., a random walk process, is typically considered to δ_t (Cai, et al., 2013; Knorr-Held, 2000).

More complex models are developed when the effect of space-time interaction is considered. Bernardinelli (1995) suggests an approach that treats interaction effect as a linear time trend. Waller et al. (1997) develop a model in which spatial random effects are nested within a time period. Knorr-Held (2000) presents a set of models for four different types of inseparable space-time interactions. Richardson et al. (2006) focus on a model which can be used to analyze the risks of two related diseases, considering their shared and unique spatio-temporal components. Hossain and Lawson (2010) propose a spatial-temporal mixture model to detect clusters. The studies mentioned above mainly focus on the global space-time effect on the outcome. They assume the coefficients of the covariates are the same across the regions and over time, which may not be the case in practice. The roles of risk factors explaining the outcome can be influenced by the varying contexts across time and space, such as the changes in other unmeasurable confounders, newly emerging local health hazard or protective factors, or even the variations in the data collecting process. To examine the effect of covariates on the response variable, Gamerman et. al. (2003) and Gelfand, et. al. (2003) work on models considering geographical varied coefficients. Dreasii, et al. (2005) and Catelan, et al. (2005) propose models incorporating time dependent covariates. Some models have been developed to examine the space-time varied effect of risk factors on response variables. Lawson, et al. (2010) present a spatio-temporal mixture model to detect spatially varied temporal patterns. Cai, et al. (2012) apply a semiparametric approach to develop a non-linear time-space dependent coefficients model. Choi, et al. (2012), provide a latent model to discover spatial components each of which has a homogeneous temporal trend in the effects of covariates on the outcome. This model assumes that temporal patterns of the associations between outcomes and predictors are different across spatial clusters, while the patterns in the regions which belong to the same spatial cluster are alike. This model is flexible and convenient to use, but it does not consider two problems often encountered in health data. First, overdispersion, a situation in which the conditional variance of the response variable is greater than the mean, is very common in count data. Second, an overabundance of zero values in the outcome variable is a usual situation in the analysis. If either or both of these two problems exist, then using the Poisson distribution will not be appropriate and its use may result in lack of fit. Therefore, developing appropriate models to deal with the problems mentioned above is necessary. The proposed research is aimed to fill this gap in the literature by employing zero-inflated Poisson and Negative Binomial (ZIP and ZINB) models to improve the performance of Choi et al.'s spatio-temporal latent cluster model. As discussed in the previous section, a state level pooled time-series data for occupational violence victimization of police officers is an ideal data set for the purpose of our methodological evaluation, as the effects of the covariates on the fatal occupational violence victimization of police are logically believed to vary across space and time units.

CHAPTER 3

DATA DESCRIPTION

The data is comprised of state level aggregated summaries across 48 continental states for years 1979 to 2010.

3.1 Outcome measure

Data for the dependent variable, the number of occupational fatal victimization of law enforcement officers, is from Kaminski and Marvell (2002), and the Law Enforcement Officers Killed and Assaulted (LEOKA, 1997-2010) reports from the Federal Bureau of Investigation (FBI). The annual death tolls include all local, state, and federal law enforcement officers with arrest power murdered in each state. Officers who died during the September 11, 2001 terrorist attacks are excluded. Because District of Columbia is unique in terms of region size and agency structures, it is excluded from the analysis. Due to their distinctive spatial characteristics, the states of Alaska and Hawaii are excluded too. Since the data of total employed police officers in each state is not reliable (Kaminski, 2002), following Kaminski and Marvell, the total population in each state is used as a proxy to adjust for unequal exposure (offset). As estimates of relative risks (Lawson, et al., 2003, p. 4), the standardized mortality ratios (SMR) for each state at each time period are calculated. SMR is defined as the counts of murdered police within each state divided by the expected number of homicide of police. The expected number of homicide of police is calculated as the total police homicide numbers divided by total US population (excluding those in Alaska, Hawaii, and D.C.), and multiplied by this state's population at time period *t*.

The data is grouped into four-year periods to make it easy to illustrate spatial and temporal trends. Since four years is a relatively short time span, it is reasonable to assume homogeneity of the effects of the risk factors over this time period within each spatial unit. There are a total of 384 observations, the minimum value is 0, the maximum value is 33, the mean count 5.154, and the standard deviation 6.004. The standardized mortality ratios (SMRs) for police officers' fatal occupational violence victimization are mapped in Figure 3.1. The distribution of the number of law enforcement officers murdered is illustrated in Figure 3.2.

According to Figure 3.1, the risks for police of being murdered indicate some spatio-temporal variations. For instance, it appears that the states on the west coast and northeast on average had the lower SMR for police murder than other states across time. The temporal profiles of the police homicide risk in some mid-west states varied dramatically from 1979-2010, making it hard to assess the general risk trend for police officers. This is probably due to the extreme values in these states at some individual time points. Take Montana, for an example: the murder risk for police in this state ranked very low during 1979-1982, but rapidly climbed to the top level during 1983-1990, then dropped again and remained low in the 1990s. The South tended to be more dangerous to law enforcement officers over the entire time interval from 1979-2010, where several southeast states kept staying in the high risk levels (SMR>1.5) most of the time. Although Florida is located in the South, it has a large variation in police homicide risks

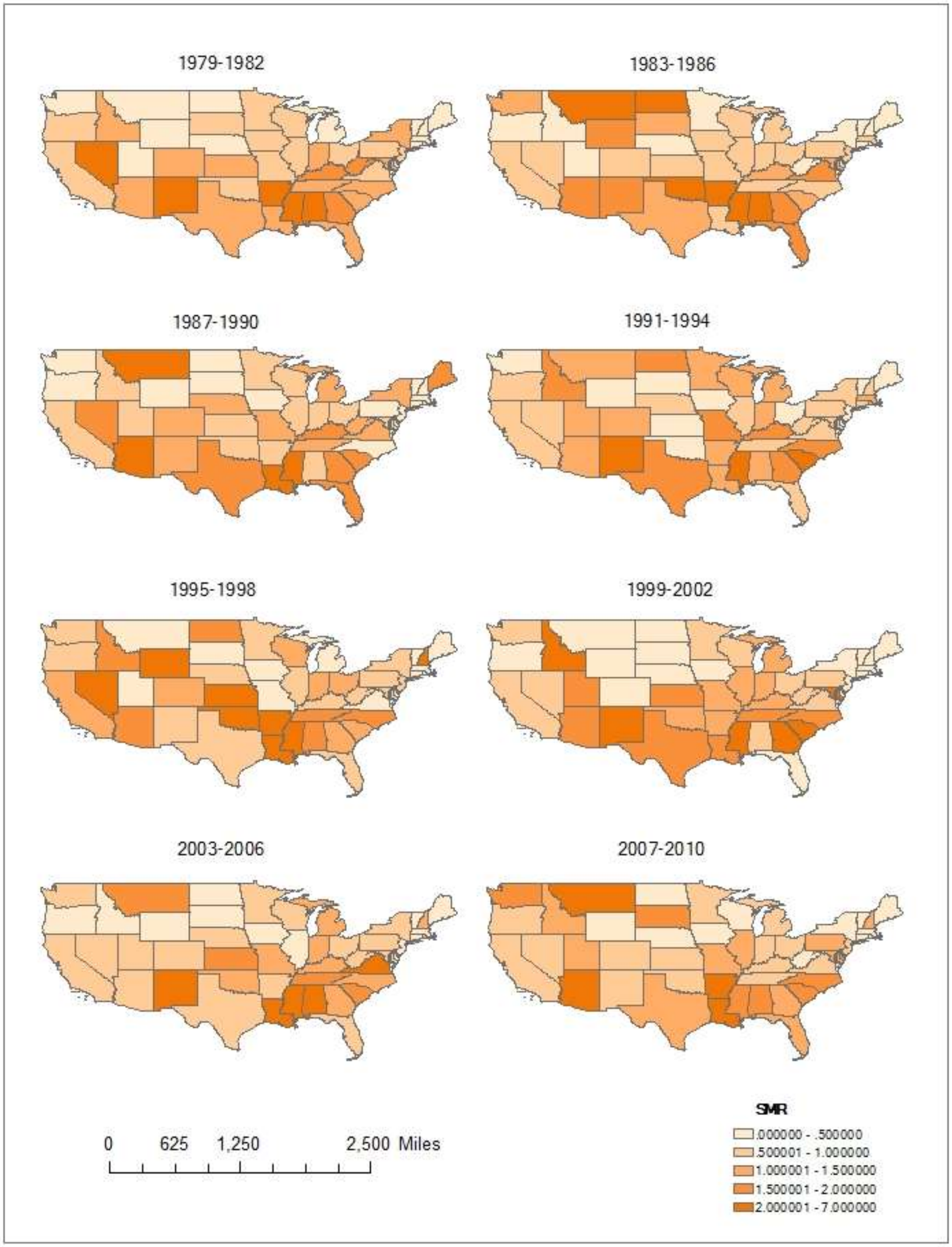

Figure 3.1 Standardized mortality ratios for police fatal occupational victimization in each time period

over time. In the period 1999-2002, Florida was actually one of the safest states for police officers.

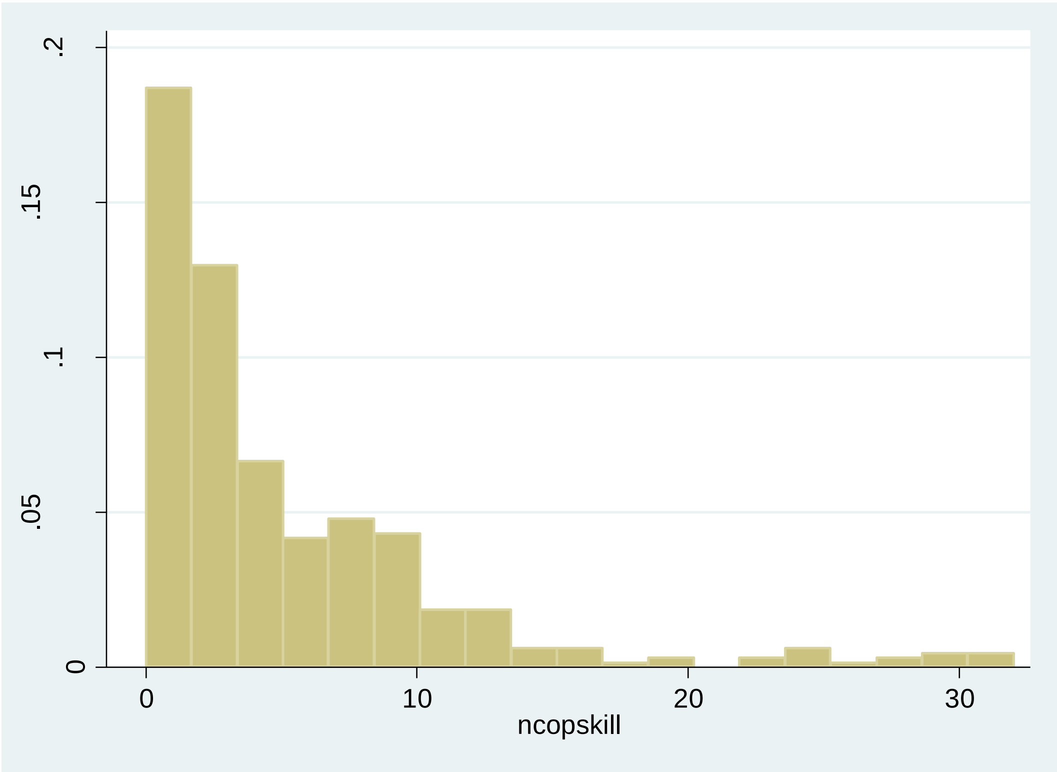

Figure 3.2. Distribution of the number of law enforcement officers murdered, 1979 - 2010

From Figure 3.2, we can see there are still many instances with 0 observed homicices even after grouping the data into four-year time periods. Moreover, the distribution shown in Figure 3.1 suggests two different types of zeroes exist. For certain states, there were no police murders at all across most time periods (i.e., Delaware, Rhode Island, Vermont, etc.). For other states, the zero observations can be only seen in few time units. As will be discussed later, zero-inflated models may help to explore the nature of these zeros and better explain the data observed.

3.2 Covariates

Following the convention used in law enforcement officer murders research, a number of eco-social structural factors including poverty rate, unemployment rate, the percentage of female-headed households with related children, and the percentage of African-Americans in the population are included in the model (Kaminski & Marvell, 2002; Kaminski, 2008; Fridell, et al., 2009).

Poverty rate and unemployment rate reflect the extent of economic disadvantages. The percentage of female householders living with related children represents the degree of family disintegration. The percentage of African-Americans in the population is chosen as the indicator of racial heterogeneity. Data source: Bureau of the Census.

The crime rate, which includes murder and non-negligent homicides, aggravated assaults, and robberies, is added in the model to control the level of risk exposure. This factor is assumed to be correlated to the proximity of officers to motivated offenders, thus being expected to have impacts on the felonious killings of officers. The data on crime rates is from the FBI's (1979-2010) Uniform Crime Reports (UCR).

Incarceration rate is included in the analysis as a confounder. Incarceration is assumed to be an indicator of formal social control and examined in many civilian homicide studies. It could affect the number of police killings due to the effect of incapacitation or deterrence. The data on incarceration is collected from Bureau of Justice Statistics annual report.

To make sure the estimation of the effect of violent crime rate and incarceration rate, are minimally affected by the collinearity between other covariates, the residuals of these two variables obtained by regressing them on other covariates are used in the model.

The residuals represent the variations in these variables unexplained by other predictors (Roncek, 1997).

All the annual data are grouped into four-year periods. The averages for each time period then are calculated. Summary statistics for the dependent and independent variables appear in Table 3.1. The unconditional variance of the outcome variable (36.048) is much larger than its unconditional mean (5.154), implying the possibility of overdispersion. An overdispersion test is then conducted through AER package in R after fitting the outcome with covariates in a generalized linear model (glm) regression. The result provides the evidence of overdispersion (alpha=3.315, p<.0001)[1].

Table 3.1 Summary statistics for variables used in the analysis (N=384)

Variables	Min	Max	Mean	SD
Officers murdered	0	33	5.154	6.004
Poverty rate	4.875	25.925	13.163	3.651
Unemployment rate	2.65	14.5	5.943	1.793
% of American Africa population	0.208	37.155	10.009	9.372
Population density	4.720	1186.337	177.172	243.863
Incarceration rates	30.859	863.975	292.865	158.95
Crime rates	40.674	1131.9	420.774	223.316

[1] When use "trafo=1" option in AER, a value of alpha much larger than zero (especially greater than 1) indicates overdispersion

CHAPTER 4

SPATIO-TEMPORAL LATENT CLUSTER MODELS

In the proposed study, I consider four models. They are spatio-temporal Poisson model, Negative Binomial model, zero-inflated Poisson model, and zero-inflated Negative Binomial model.

4.1 Spatio-temporal Poisson model with latent clusters (Model 1)

First, I consider a spatio-temporal Poisson model with latent clusters. As introduced in the Chapter 2, the occupational homicide of police is assumed to follow a Poisson distribution as Y_{it} ~Pois($E_{it}\theta_{it}$), where E_{it} is the expected count and θ_{it} is the relative risk in the i^{th} state and t^{th} year. θ_{it} is modeled by

$$ln(\theta_{it}) = \boldsymbol{x'}_{it}\boldsymbol{\beta}_{it}$$

where x'_{it} represents the vector of covariates (including intercept) of state *i* at year *t*, and β_{it} denotes the vector of the corresponding coefficients. In the proposed research, I assume that the temporal trends of the associations between outcomes and covariates are different across spatial domains. But such trends in the regions which belong to the same spatial cluster are alike. In other words, the time-dependent coefficients of the covariates are homogeneous within the domain of a spatial cluster. Thus, I specify the vector of covariates' coefficients, $\boldsymbol{\beta}_{it}$, as

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{S(m)t}$$

where *m* is the indicators of spatial clusters and $S(m)$ shows which spatial cluster the observed state belong to. I consider $S(m)$ follows a categorical distribution

$$S(m)\sim\text{Categorical}(q_{i1,\ldots,}\, q_{iM})$$

where q_{im} is the probability of state i belongs to cluster m. Hence, the q_{im} has two conditions: q_{im}>0 and $\sum_1^M q_{im} = 1$. Next, I model the q_{im} as

$$q_{im} = \frac{w_{im}}{\sum_{m=1}^{M} w_{im}}$$

where w_{im} is un-normalized weights. Following Choi, et al., I assume w_{im} , which are non-negative weights, has a lognormal distribution,

$$w_{im}\sim\text{LN}(\eta_{im}, \sigma_m^2)$$

where η_{im} is spatially dependent mean, and σ_m^2 is the variance of η_{im}.

To add a spatial dependency structure, I assign a conditional autoregressive (CAR) distribution to η_{im}

$$\eta_{im}|\eta_{jm,i\neq j}\sim N(\frac{1}{n_i}\sum_{j\neq i}\varphi_{ij}\eta_{jm}\,, \frac{\sigma_{\eta m}^2}{n_i}),$$

where n_i is the number of the neighboring states which are also in the same cluster of state *i*, φ_{ij}=1 if *i* and *j* are adjacent states and φ_{ij}=0 otherwise. Hence, assuming an CAR, the mean of state *i* is smoothed as the average of the means of its neighbor states of a same cluster, and the variance is the variance of $\sigma_{\eta_m}^2$ divided by n_i. This model is denoted as Model 1.

4.2 Spatio-temporal negative binomial model with latent clusters (Model 2)

As illustrated in the end of previous Chapter, there is an overdispersion issue in current data (See Table 3.1 and related overdispersion test). A space-time negative binomial latent cluster model is then considered to see if it improves the fitness in comparison to a Poisson distribution. Using this model, the distribution of an observed count Y_{it} is

$$Y_{it} \sim \text{NegBin}(E_{it}\theta_{it}, \alpha),$$

where E_{it} is the expected count, θ_{it} is the relative risk in the i^{th} state and t^{th} year, and α is the overdispersion parameter. The mean of the distribution is $E_{it}\theta_{it}$, and the variance is $E_{it}\theta_{it}\left(1+\frac{E_{it}\theta_{it}}{\alpha}\right)$. Note if $\alpha \to \infty$ then the distribution reduces to a regular Poisson. The parameter θ_{it} is modeled the same way as in the previous Poisson model. A Gamma prior is assigned to α. This model is denoted as Model 2.

4.3 Spatio-temporal Zero-Inflated models with latent clusters

As indicated in the histogram of the respondent variable (Figure 3.2), there were a large number of zero outcome observations in the data. The previous count models may not fit well in zero-dominated data. Therefore, it is worthy to consider zero-inflated count models to account for excess zeros issue. Instead of assuming all the zeros come from the same data-generating process in which the nonzero observations were produced, zero-inflated models assume that these zero counts could have been generated through two different processes: only part of zeros (sampling zeros) comes from the count model which also produced all other positive observations, and another process yields structural zeros (true zeros). Whether a zero observation belongs to structural zeros or sampling zeros is determined by a Bernoulli process (Lambert, 1992; Greene, 1994). Hence, a general structure of a zero-inflated count model is

$$\begin{cases} \Pr(y=0) = (1-p) + p * f(y=0), \text{if count is zero.} \\ \Pr(y=k) = p * f(y=k), \text{if count is any positive integer } k. \end{cases}$$

where p is the probability of the zero observation is a sampling zero, and $f(y)$ could be a Poisson or a Negative Binomial model.

4.3.1 Spatio-temporal Zero-Inflated Poisson model with latent clusters (Model 3)

For a zero-inflated spatio-temporal Poisson model, the model can be expressed as

$$\begin{cases} \Pr(Y_{it} = 0) = (1 - p_{it}) + p_{it} * e^{-E_{it}\theta_{it}} \\ \Pr(Y_{it} = k) = p_{it} * \dfrac{(E_{it}\theta_{it})^k e^{-E_{it}\theta_{it}}}{k!} \end{cases}$$

The relative risk, θ_{it}, can be modeled as introduced in the previous Poisson model

$$ln(\theta_{it}) = \boldsymbol{x'}_{it}\boldsymbol{\beta}_{it}$$

where $\boldsymbol{x'}_{it}$ represents the vector of covariates (including intercept) of state *i* at year *t*, and $\boldsymbol{\beta}_{it}$ denotes the vector of the corresponding coefficients. The probability of a zero counts belongs to sampling zeros can be modeled as

$$logit(p_{it}) = \boldsymbol{c'}_{it}\boldsymbol{\gamma}_{it}$$

where $\boldsymbol{c'}_{it}$ represents the vector of covariates (including intercept) of state *i* at year *t*, and $\boldsymbol{\gamma}_{it}$ denotes the vector of the corresponding coefficients. Theoretically, $\boldsymbol{c'}_{it}$ could use the same sets of predictors which are used in the Poisson model, but adding too many predictors into the model may cause difficulty in the model's convergence. Since the population density is probably the most influential factor for getting zero observations, the proposed ZIP model only includes the intercept and population density into the logistic model to predict whether a zero is a sampling zero.

As introduced in the previous Poisson model, I assume that there are several latent clusters exist across all the spatial units, and the temporal pattern of the effect of predictors is unique in each cluster. Accordingly, I specify the vector of covariates' coefficients for the count model, $\boldsymbol{\beta}_{it}$, as $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{S(m)t}$, and the vector of covariates' coefficients for the probability of a zero being sampling zeros, $\boldsymbol{\gamma}_{it}$, as $\boldsymbol{\gamma}_{it} = \boldsymbol{\gamma}_{S(m)t}$, where *m* is the indicators of spatial clusters and $S(m)$ shows which spatial cluster the observed state belong to. Again, $S(\mathrm{m})$ follows a categorical distribution

$$S(m) \sim \text{Categorical}(q_{i1,\ldots,}\, q_{iM})$$

where q_{im} is the probability of state *i* belongs to cluster *m*. The modelling of q_{im} and following parameters (w_{im}) just follow the same steps as in the Model 1.

4.3.2 Spatio-temporal Zero-Inflated negative binomial model with latent clusters (Model 4)

The last model considered is a zero-inflated spatio-temporal negative binomial model with latent clusters,

$$\begin{cases} \Pr(\text{Yit} = 0) = (1 - p_{it}) + p_{it} * (\dfrac{\alpha}{E_{it}\theta_{it} + \alpha})^{\alpha} \\ \Pr(\text{Yit} = k) = p_{it} * \dfrac{\alpha^{\alpha}(E_{it}\theta_{it})^{k}\Gamma(k+\alpha)}{k!\,\Gamma(\alpha)(E_{it}\theta_{it} + \alpha)^{k+\alpha}} \end{cases}$$

where E_{it} is the expected count, θ_{it} is the relative risk in the ith state and tth year, and α is the overdispersion parameter. The parameter θ_{it} and p_{it} are modeled the same way as in the Model 3. A Gamma prior is assigned to α.

To detect the number of the latent clusters, Dirichlet process mixture model could be used. But this approach requires intensive computations. Since the range of the number of possible spatial clusters cannot be very large based on the limited numbers of observed spatial units (48 states), I adopt a simpler way: Several models with different numbers of spatial domains are estimated and the best one is chosen by using model diagnosis criterions, i.e., DIC and NLLK.

For the vector of coefficients within a spatial cluster $\boldsymbol{\beta}_{mtk}$=$(\beta_{mt0}, \beta_{mt1}, \dots, \beta_{mtk})$', a random walk process as $\boldsymbol{\beta}_{mtk} \sim Normal(\boldsymbol{\beta}_{m,t-1,k}, \sigma^2_{mk})$ is chosen to represent the temporal autocorrelations structure in spatial cluster *m*. By using this prior, I assume the current value of the outcome follows a normal distribution centered at the value of the coefficients at the previous time period and with variance σ^2_{mk}. For $\sigma^2_m, \sigma^2_{\eta_m}, and\ \sigma^2_{mk}$, since they have to be positive and there is no prior knowledge about them, a hyper-prior distribution Inverse-Gamma (0.025, 0.025), which allows wide variations in these parameters (Spiegelhalter, 2002), is used. The WinBUGS codes for these four models are attached in the Appendix A1-4.

4.4 Model computation and comparison

Four Bayesian hierarchical latent cluster models (Model 1-4) adopting Poisson or Negative Binomial distribution with or without zero-inflated structures are to be estimated respectively. As a comparison, a simple Bayesian Poisson model, a negative binomial model, a zero-inflated Poisson model, and a zero-inflated negative binomial model, which do not incorporate spatial and temporal structures, are fitted (Model 5-8) as baseline models. A generalized linear mixed model (GLMM) formulation (Laird & Ware, 1982), commonly employed in longitudinal studies, is used to estimate these four models as

$$\boldsymbol{\eta_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i},$$

where $\boldsymbol{\eta_i}$ is the linear predictor which combines the fixed effect ($\boldsymbol{X_i}\boldsymbol{\beta}$) and random effect ($\boldsymbol{Z_i}\boldsymbol{b_i}$). A log link function is used to relate the count outcomes to the linear predictor $\boldsymbol{\eta_i}$. Random slopes could be possible, but the proposed study just focuses on the random intercept model. In this case, the fixed-effect design matrix $\boldsymbol{X_i}$ includes columns for the globe intercept and other fixed effect covariates, $\boldsymbol{\beta}$ denotes the vector of the corresponding coefficients, the random effect design matrix $\boldsymbol{Z}$ only contains one column of 1s, and $\boldsymbol{b}$ only includes the intercept as a random effect. An equivalent expression in a composite multi-level model form is

$$\boldsymbol{\eta_i} = \boldsymbol{x'}_{it}\boldsymbol{\beta} + r_{0i}, \qquad r_{0i} \sim \mathrm{N}(0, \sigma_r^2)$$

where $\boldsymbol{x'}_{it}$ represents the vector of covariates of state *i* at year *t*, $\boldsymbol{\beta}$ denotes the vector of the corresponding coefficients, and r_{0i} is the subject specific intercept for state i.

Also, conventional spatio-temporal models (Eq. (1)) discussed in Chapter 2, using Poisson or Negative Binomial with or without zero-inflated structures (Model 7-10), are estimated. Posterior computation can be processed by WinBUGS software via the Markov Chain Monte Carlo (MCMC) algorithm. The labels of the spatial-temporal clusters can switch when using multiple chains to conduct MCMC simulations in Bayesian mixture modelling, which could cause the problem of identifiability (Stephens, 2000; Choi, et al., 2011). To avoid this issue, each model runs only one single chain to draw the samples. MCMC convergence can be diagnosed by trace plots, autocorrelations plots, and Geweke's z-test. A randomly scattered trace plot of the draws of a parameter surrounding a stable mean value indicates that the convergence point is reached. Upon convergence, the autocorrelation between the drawn samples should decrease rapidly. Additionally, the convergence can be diagnosed by the Geweke test. The Geweke diagnostic compares the means of the beginning part (i.e., the first 10%) and the last part (i.e., the second half) of the draws. The test statistic is a standard Z-score, with which a value between $|Z_a|$ (i.e., $|Z_{0.05}|$=1.96) shows convergence of the draws (2002).

After fitting these models, deviance information criterion (DIC), proposed by Spiegelhlter et al, (2002), can be used to evaluate which one best fits the data. DIC is defined as

$$\text{DIC}=\overline{D} + P_D,$$

which is a combination of $\overline{D}$, a summarized measure of the current model fit, and P_D, a penalty of the model complexity. Smaller values of DIC indicate better fittings of the model. P_D is calculated as

$$P_D = E_{\theta|y}(D) - D\left(E_{\theta|y}(\theta)\right) = \overline{D} - D(\bar{\theta}),$$

where $\overline{D} = E_{\theta|y}(D)$ and $\mathrm{D}(\theta) = -2\log p(y|\theta) + 2\log f(y)$. The DICs of ordinary models can be provided by WinBUGS. However, WinBUGS does not yield DICs for mixture models, such as zero-inflated Poisson and zero-inflated negative binomial models. In such cases, an approach suggested by Neelon, et al (2010) is adopted. Dbar ($\overline{D}$), the posterior mean of the deviance, is directly read from WinBUGS output. The average of the values of the parameters at stochastic parent nodes are calculated, and then are used to compute the deviance at the posterior means of these parameters, Dhat ($D(\bar{\theta})$). This step is completed by retrieving each draw from WinBUGS and processing in R afterward.

Also, the negative cross-validatory predictive log-likelihood (NLLK) based on the Conditional Predictive Ordinate (CPO) (Gelfand &Dey, 1994; Gesser, 1993; Dey, et al., 1997; Spiegelhalter et al., 1996) is considered to compare the prediction performance among these models. The CPO is the density of the posterior predictive distribution evaluated at an observation, given the data excluding the information of this observation.

Hence, the CPO is a cross-validation measure. The CPO for state *i* and time interval *t* is defined as

$$\text{CPO}_{it}=f(Y_{it}|\boldsymbol{Y}_{-it}) = \int f(Y_{it}|\boldsymbol{\theta},\ \boldsymbol{Y}_{-it})f(\boldsymbol{\theta}|\boldsymbol{Y}_{-it})d\boldsymbol{\theta}=\left(\int \frac{1}{P(Y_{it}|\boldsymbol{\theta})}P(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta}\right)^{-1}$$

where $\boldsymbol{Y}_{-it}$ denotes the vector of the police murder observations excluding Y_{it} and $\boldsymbol{\theta}$ is the vector of unknown parameters. The cross-validation likelihood as a summary measure is then calculated as

$$L_{cv}=\prod_{i=1}^{n}\prod_{t=1}^{T}\text{CPO}it.$$

A larger L_{cv} implies better fit. Usually, the values of L_{cv}s are very close to zero. Therefore, the negative cross-validatory log-likelihood can be used for model comparison:

$$\text{NLLK}_{cv}=-\sum_{i=1}^{n}\sum_{t=1}^{T}\text{logCPO}_{it}.$$

Thus, a less NLLK_{cv} indicates a better fit, which is consistent with other main model comparison criterions. The estimate of the CPO_{it} can be obtained by

$$\widehat{\text{CPO}}_{it} = \frac{1}{\frac{1}{T}\sum_{t=1}^{T}[P(Y_{it}|\theta^{(t)})]^{-1}}$$

where *T* is the number of samples drawn from the MCMC chain, and $\theta^{(t)}$ is the number *t* MCMC sample. $P(Y_{it}|\theta^{(t)})$ of each draw of the MCMC simulations is computed within WinBUGS, and then is exported to R to calculate the CPO and NLLK_{cv}. The related R codes for calculating DICs and NLLK_{cv} are attached in the Appendix.

CHAPTER 5

RESULTS

I apply the aforementioned models to the police occupational fatal victimization (POFV) data. For each model, 80,000 MCMC iterations are performed, and the first 30,000 samples are discarded as burn-in. To reduce the autocorrelations between the sampled parameters, only every 20th sample is kept. Hence, 2,500 final samples are collected to summarize the parameters of interest. Based on visual inspection of trace plots, autocorrelation plots, and the result of the Geweke test, convergence is reached for all the models. Some trace plots and autocorrelation plots of the parameters of interest are illustrated in Figure 5.1. In order to decide the best number of the spatial clusters in the models 1-4, the models with a range of the number of clusters are estimated. Because the total number of spatial units (48 states) is not large, a reasonable estimation of the number of clusters is between two to eight. The final number of the spatial clusters is determined by the model with the best comparison measures (DIC and NLLK) by estimating models with different numbers of clusters. The plots of DIC and NLLK values for model 1-4 fitted with different numbers of spatial groups are displayed in Figure 5.2. Table 5.1 provides detailed information about the DIC and NLLK values for each model.

From Figure 5.2, it can be found that there is no uniform pattern in terms of the changes of DICs and NLLKs with different numbers of clusters across four models. However, the changes of DICs are similar to that of NLLKs within each model, illustrating the consistence between these two comparison criterions. The results show that Model 3 (ZIP model) with four spatial clusters has the smallest DIC (1218.4) and NLLK (778.5). Model 1 (Poisson model) with four spatial clusters comes next with a DIC of 1226.5 and a NLLK of 783.6. Although the likelihood ratio test shows the sign of overdispersion in the POFV data, the negative binomial model (Model 2) does not perform better than the Poisson model with respect to DIC (1937.7) and NLLK (964.4). Theoretically, both unobserved heterogeneity among subjects and excess zeroes due to different zero generating mechanisms could produce an overdispersion in the raw data (Long, 1997). A negative binomial model should be more appropriate when the overdispersion is only a result of the subjects' heterogeneity, but may not fit well if the overdispersion is actually a reflection of the zero inflation. In this analysis, the ZIP model outperformed the negative binomial model, suggesting that the overdispersion may mainly come from the high proportion of zeroes. The ZINB model (Model 4) produces the highest DIC and NLLK (2711.3 and 2429.9 respectively), indicating that no extra overdispersion needs to be adjusted after applying a zero-inflated structure model (i.e., a ZIP model).

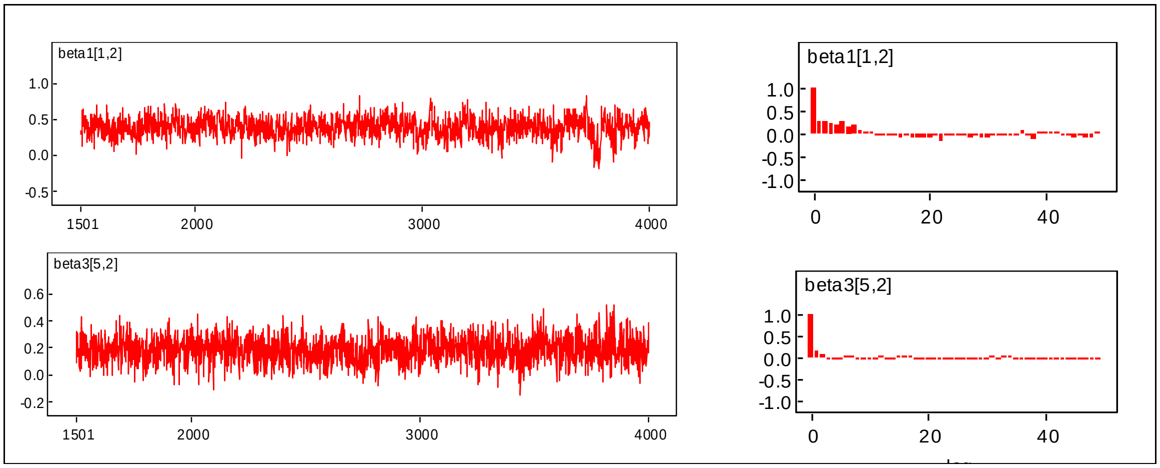

Figure 5.1. The trace plots and the autocorrelation plots of selected parameters

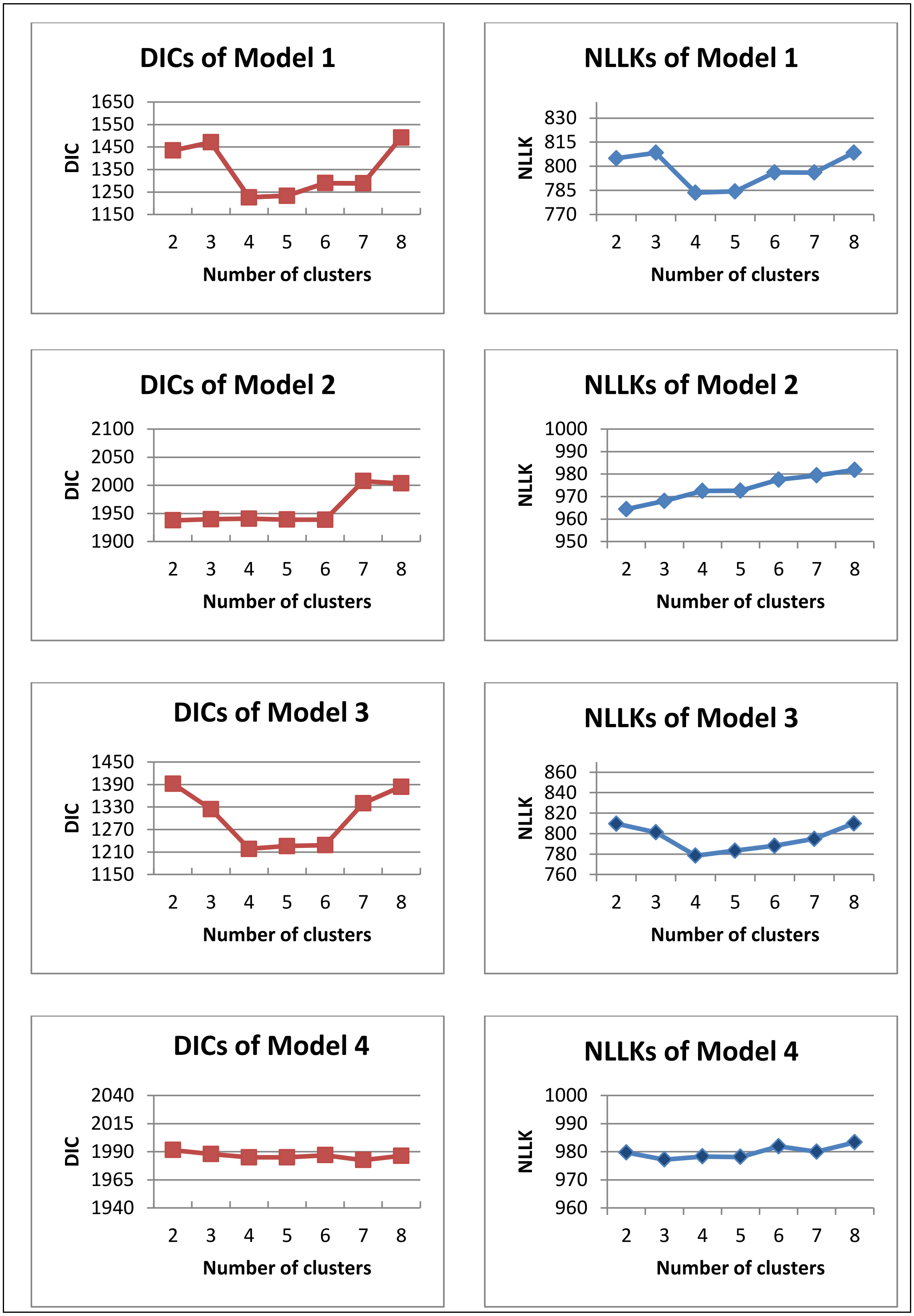

Figure 5.2 The DICs and NLLKs of Model 1-4 using different number of clusters

Table 5.1 Latent cluster models comparison using DIC and NLLK

Model	# Clusters	Dbar	Dhat	P_D	DIC	NLLK
Model 1	2	1468.4	1502.3	-33.9	1434.5	805.0
(cluster Poisson)	3	1482.7	1493.9	-11.2	1471.4	808.4
	4	1395.5	1564.5	-169.0	1226.5	783.6
	5	1404.4	1575.3	-170.9	1233.5	784.4
	6	1421.9	1553.7	-131.8	1290.0	796.2
	7	1474.8	1660.6	-185.9	1288.9	796.1
	8	1488.0	1484.0	4.0	1492.0	808.5
Model 2	2	1820.6	1703.4	117.2	1937.7	964.4
(cluster NB)	3	1821.7	1703.7	118.0	1939.7	968.0
	4	1822.4	1704.2	118.2	1940.6	972.5
	5	1821.4	1703.7	117.7	1939.1	972.6
	6	1821.0	1703.1	117.8	1938.8	977.5
	7	1841.2	1675.0	166.2	2007.4	979.4
	8	1841.5	1679.6	161.9	2003.4	981.8
Model 3	2	1504.4	1616.5	-112.1	1392.4	809.7
(cluster ZIP)	3	1482.6	1640.6	-158.0	1324.6	801.3
	4	1400.4	1582.5	-182.1	1218.4	778.5
	5	1403.8	1582.9	-179.1	1224.8	780.9
	6	1455.5	1690.5	-235.0	1220.5	785.3
	7	1457.2	1564.3	-107.1	1350.0	796.1
	8	1459.2	1534.3	-75.1	1384.1	809.2
Model 4	2	1833.9	1676.4	157.5	1991.4	979.7
(cluster ZINB)	3	1832.5	1677.1	155.43	1987.9	977.2
	4	1831.3	1677.7	153.6	1985.0	978.3
	5	1831.4	1677.8	153.6	1985.0	978.1
	6	1832.9	1678.9	154.0	1986.9	981.9
	7	1830.5	1678.4	152.1	1982.6	980.0
	8	1832.5	1678.6	153.9	1986.3	983.3

Next, a comparison is made between this model (Model 3) and the models without spatial structured variation (Model 5-8), as well as the models applying conventional spatio-temporal variation structures (Model 9-12). Table 5.2 reports the comparison measures of these models. In general, the negative binomial model and the ZINB model have higher DIC and NLLK values than the Poisson model and the ZIP model in each subgroup. This result, again, suggests that the overdispersion in the current data may mainly be due to excess zeroes rather than the heterogeneity among subjects. As expected, Models 5 and 7 have the largest DIC and NLLK values in comparison to their counterpart models considering spatio-temporal autocorrelations. This suggests that the models incorporating spatial and temporal autocorrelations are more appropriate for the analysis of this POFV data than the models which do not have space and time components. However, this finding does not appear in the negative binomial models and the ZINB models, implying that incorporating space and time information cannot improve the fitness if an improper model is chosen. When comparing the conventional spatio-temporal models with proposed latent cluster models, the latent cluster ZIP model (Model 3) with four spatial clusters still has the smallest DIC and NLLK. Therefore, the ZIP cluster model provides the best performance. All the ensuing analyses are based on the results of this model.

Table 5.2. Models comparison using DIC and NLLK

Model	Dbar	Dhat	P_D	DIC	NLLK
Model 3 (Latent cluster, ZIP)	1400.4	1582.5	-182.1	1218.4	778.5
Model 5 (Poisson, no spatial variation)	1493.6	1441.0	52.6	1546.2	816.4
Model 6 (NB, no spatial variation)	1809.0	1693.0	116.0	1925.0	937.0
Model 7 (ZIP, no spatial variation)	1544.1	1508.7	35.4	1579.5	810.9
Model 8 (ZINB, no spatial variation)	1816.8	1694.7	122.1	1938.9	942.4
Model 9 (Conventional Poisson)	1484.6	1430.0	54.6	1539.2	801.7
Model 10 (Conventional NB)	1822.4	1642.0	180.4	2002.8	980.3
Model 11 (Conventional ZIP)	1487.8	1434.5	53.3	1541.1	802.2
Model 12 (Conventional ZINB)	1818.0	1621.7	196.3	2014.3	987.3

The spatial clusters identified through Model 3 with four groups are mapped in Figure 5.3. The assignment of spatial clusters for each state is determined by the posterior mean of the conditional weight, q_{im}, which represents the probability of state i belonging to cluster m. A state is assigned to cluster k if q_{ik} is the largest one among all the q_{im}s ($k \in m$, $m=\{1,2,3,4\}$). The names of the states in each cluster are listed in Table 5.3. The numbers of states by spatial cluster are also presented. Cluster two has the largest number of member states (34 states in this group). This cluster includes the states from all the regions across the country, and does not show apparent spatial patterns. Most southern states and midwestern states are assigned into this cluster. Next comes cluster four with 11 states. Three southern states (Texas, Florida, and Virginia) and several northern states are allocated to this group. Also, Washington, Idaho, Utah, and Arizona in this cluster constitute a corridor in the south-north direction in the West. The remaining three states are classified into two clusters: Iowa and Vermont in cluster one and Maine in cluster three respectively. These states reported zero count observations in most time periods.

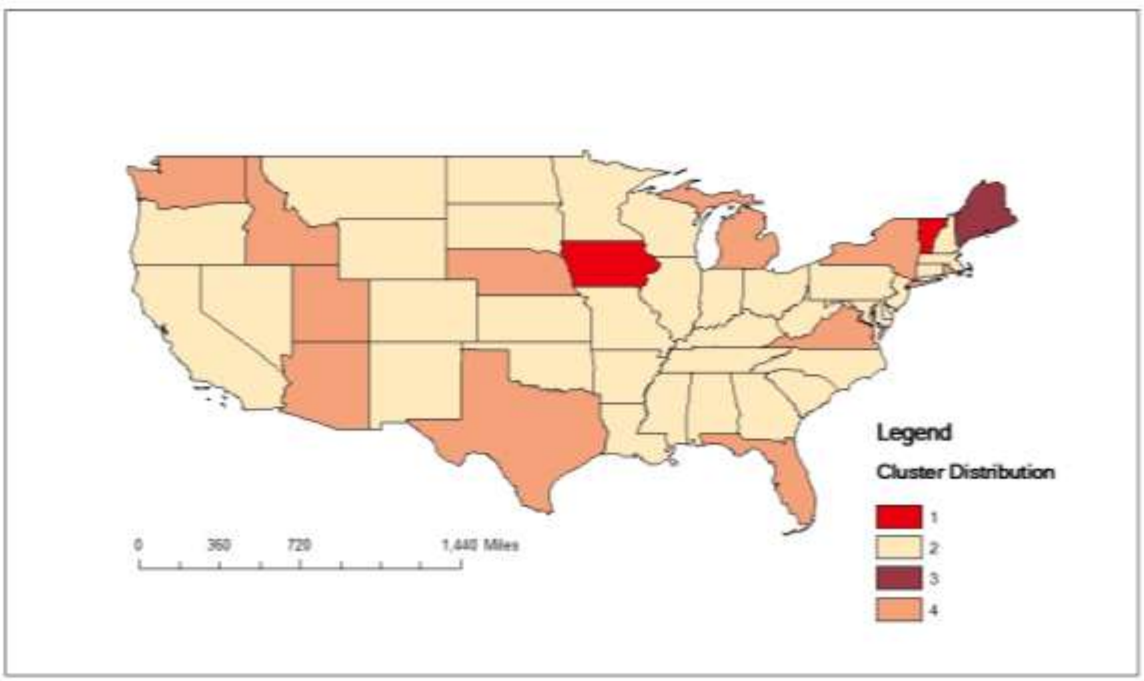

Figure 5.3 The distribution of the member states in each spatial cluster

Table 5.3 The distribution of the states in four spatial clusters

	Cluster1	Cluster2	Cluster3	Cluster4
States names	Iowa	Alabama	Maine	Arizona
	Vermont	Arkansas		Florida
		California		Idaho
		Colorado		Michigan
		Connecticut		Nebraska
		Delaware		New York
		Georgia		Rhode Island
		Illinois		Texas
		Indiana		Utah
		Kansas		Virginia
		Kentucky		Washington
		Louisiana		
		Maryland		
		Massachusetts		
		Minnesota		
		Mississippi		
		Missouri		
		Montana		
		Nevada		
		New Hampshire		
		New Jersey		
		New Mexico		
		North Carolina		
		North Dakota		
		Ohio		
		Oklahoma		
		Oregon		
		Pennsylvania		
		South Carolina		
		South Dakota		
		Tennessee		
		West Virginia		
		Wisconsin		
		Wyoming		
Number of states	2	34	1	11

The estimated risks of law enforcement officers being killed is mapped in Figure 5.4, after controlling for the effect of population density, unemployment rate, poverty rate, racial heterogeneity, crime level, and incarceration rate. Compared to the raw SMRs of police murder mapped in Figure 3.1, the estimated risks reflect more stable temporal and spatial patterns. Also, these patterns can be more easily detected than using raw SMRs. According to Figure 5.4, the lower risks of police killings in the Pacific states and the states in the Northeast region are more obvious. This pattern remained stable and did not suffer from the influence of extreme observations over time. In general, most states in the Midwest and the West stayed at the low and average risk level, with the exception of Montana, Idaho, and Nevada at several time points. It is more evident that most southern states plus several southwest states (i.e., New Mexico, Arizona) had heightened fatal threats to police over time. Several southeastern states constitute a core high risk region for police safety. Among these states, Mississippi and Louisiana remained the most dangerous places to law enforcement officers across almost all the time periods.

Because the number of the coefficients is large (288 coefficients), the estimated means and the 95% credible intervals of these parameters from the posterior samples are reported in the form of caterpillar plots (Figure 5.5). A rough idea from this figure is that the locations of the estimated means and the range of the credible intervals varied across four spatial clusters. To take a closer look at such variations, a further inspection of these estimates is needed. Therefore, Table 5.4 collects the detailed information of the parameters which have significant estimates. The results shows that poverty rates, the proportion of the black population, and incarceration rates had positive relationships with police murders in cluster two during 1979-1986, 1995-2010, and 1995-1998, respectively.

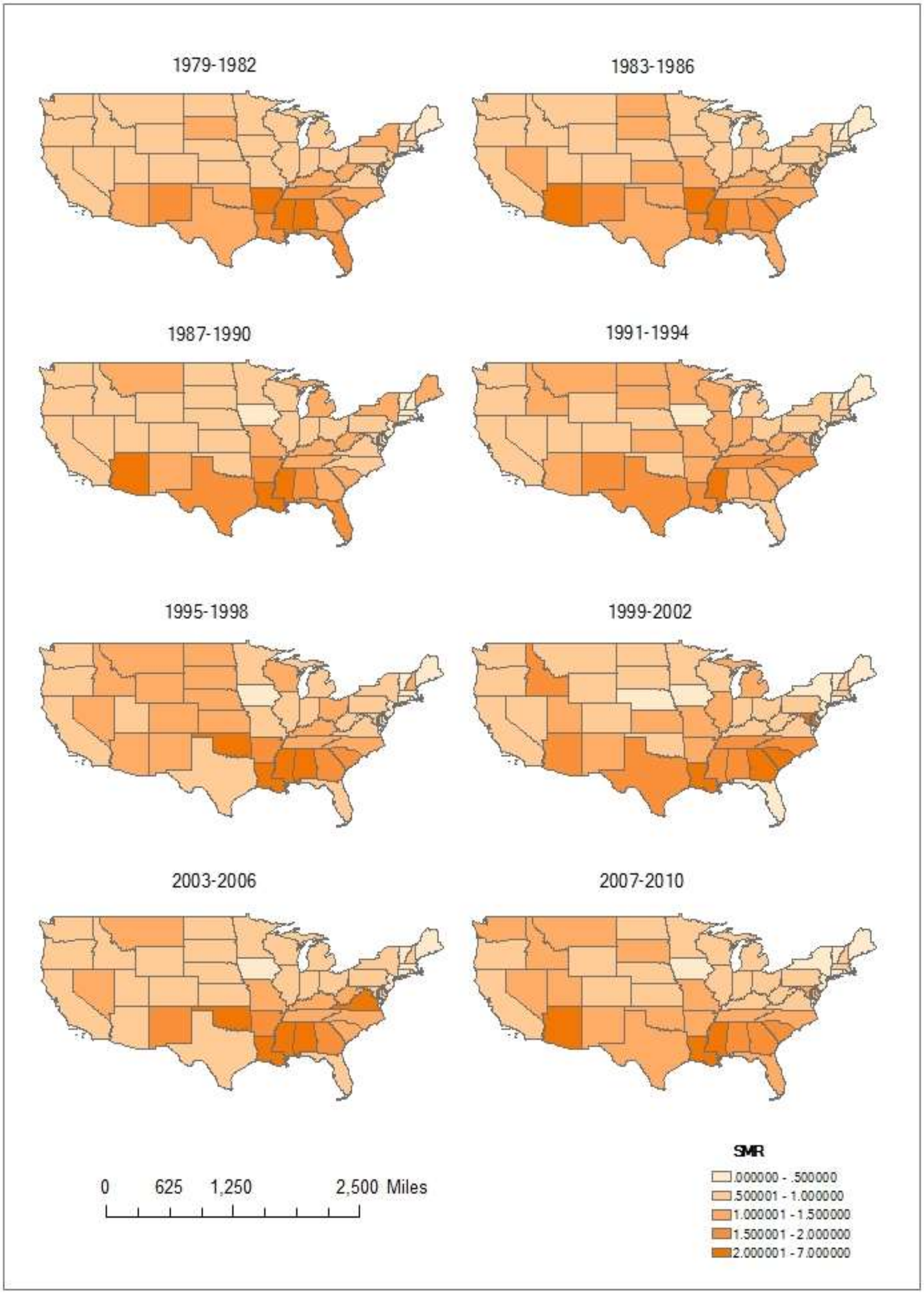

Figure 5.4 The estimated mortality risk for police in the U.S., 1979-2010

Contrary to common belief, unemployment rates were negatively associated with the police fatal victimization in cluster two during 1983-1986. The possible causes of this relationship deserve further investigation.

It is also found that the temporal trends of the coefficients for some covariates varied across four spatial groups. For example, Figure 5.6 illustrates the different temporal profiles of the effects of the percentage of the population that was African-American on the fatal victimization risk for police across four spatial clusters. While cluster one showed a near flat trend over time, cluster three indicated a slightly increasing temporal pattern of the coefficients of the covariate. The temporal changes in the effect of the proportion of black population on murders of police in cluster two and cluster four displayed totally different profiles. Overall, the corresponding coefficient in cluster two increased over time and became significant after 1995. In contrast, although the estimates of this coefficient in cluster four were not significant, they showed an apparent downward pattern after 1983 and dropped to zero gradually. This indicates that the positive association between the percentages of the black population (which represents racial heterogeneity) and the killings of officers became stronger over time in cluster two, while such an association diminished in cluster four. Compared to cluster two and cluster four, the effect of the percentage of the black population did not show any obvious temporal variations in cluster one and cluster three. Since cluster two has the largest number of member states, this positive association between the proportion of the black population and killings of police may have a heavier influence on the general association estimate than that of any other spatial group. It is possible that a positive relationship between these two variables is found if no latent cluster analysis is considered, while in fact such relationships do not hold in all the states. The temporal profiles of the coefficients corresponding to other covariates are displayed in Appendix B.

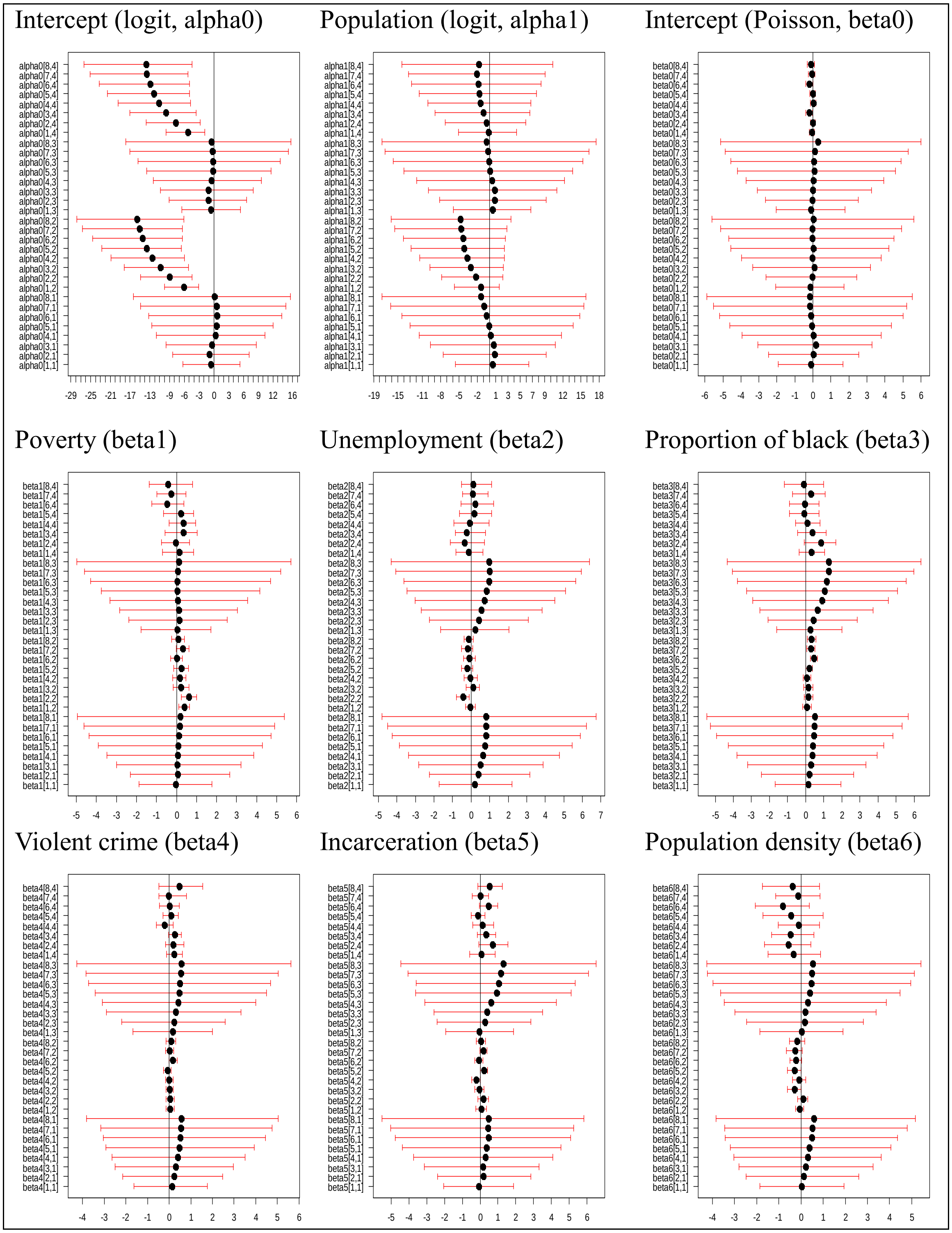

Figure 5.5 The caterpillar plots of the coefficient estimates from the posterior samples in Model 3 (The first number in the square bracket after the coefficient name denotes the order of the time periods, and the second number denotes the spatial cluster to which the coefficient belongs.)

Table 5.4 Parameter estimates from the posterior distribution in Model 3 (Posterior mean, standard deviation and 95% credible interval are shown. The first number in the square bracket after the coefficient name denotes the order of the time periods, and the second number denotes the spatial cluster to which the coefficient belongs.)

	Time period	**median**	**sd**	**MC error**	**2.50%**	**mean**	**97.50%**
(logit)Intercept[1,2]	1979-1982	-5.86	1.86	0.03	-10.04	-6.05	-3.02
(logit)Intercept[2,2]	1983-1986	-8.68	2.75	0.06	-14.85	-8.95	-4.40
(logit)Intercept[3,2]	1987-1990	-10.59	3.35	0.08	-18.19	-10.85	-5.14
(logit)Intercept[4,2]	1991-1994	-12.18	3.76	0.08	-20.82	-12.48	-5.94
(logit)Intercept[5,2]	1995-1998	-13.34	4.16	0.09	-22.76	-13.64	-6.56
(logit)Intercept[6,2]	1999-2002	-14.06	4.67	0.12	-24.55	-14.47	-6.42
(logit)Intercept[7,2]	2003-2006	-14.6	5.21	0.11	-26.64	-15.09	-6.34
(logit)Intercept[8,2]	2007-2010	-14.99	5.64	0.13	-27.8	-15.55	-6.08
(logit)Intercept[1,4]	1979-1982	-4.96	1.98	0.04	-9.70	-5.20	-1.88
(logit)Intercept[2,4]	1983-1986	-7.54	2.88	0.06	-13.72	-7.73	-2.72
(logit)Intercept[3,4]	1987-1990	-9.47	3.40	0.07	-16.99	-9.69	-3.63
(logit)Intercept[4,4]	1991-1994	-10.82	3.77	0.08	-19.39	-11.14	-4.70
(logit)Intercept[5,4]	1995-1998	-11.81	4.25	0.10	-21.57	-12.18	-4.92
(logit)Intercept[6,4]	1999-2002	-12.41	4.67	0.11	-23.24	-12.92	-4.89
(logit)Intercept[7,4]	2003-2006	-12.96	5.15	0.12	-25.1	-13.57	-5.26
(logit)Intercept[8,4]	2007-2010	-13.18	5.69	0.12	-26.31	-13.75	-4.41
(ln(mu))Poverty[1,2]	1979-1982	0.39	0.14	0.01	0.11	0.39	0.64
(ln(mu))Poverty[2,2]	1983-1986	0.63	0.19	0.01	0.24	0.63	1.01
(ln(mu))Unemployment[2,2]	1983-1986	-0.45	0.18	0.01	-0.80	-0.45	-0.09
(ln(mu))BlackPct[5,2]	1995-1998	0.19	0.09	0.01	0.00	0.19	0.37
(ln(mu))BlackPct[6,2]	1999-2002	0.45	0.10	0.01	0.27	0.46	0.66
(ln(mu))BlackPct [7,2]	2003-2006	0.28	0.11	0.01	0.06	0.28	0.52
(ln(mu))BlackPct [8,2]	2007-2010	0.32	0.12	0.01	0.08	0.32	0.56
(ln(mu))Incarceration[5,2]	1995-1998	0.22	0.10	0.01	0.01	0.22	0.41

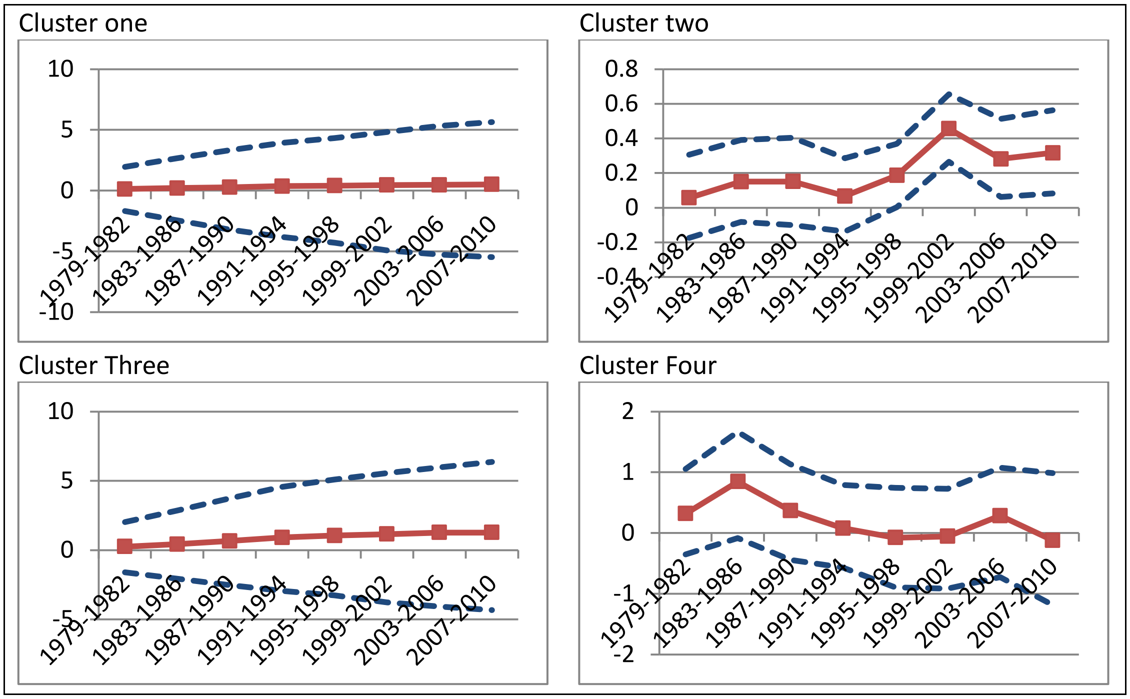

Figure 5.6 Temporal profiles of the effects of the proportion of blacks (beta3) on the fatal victimization risk for police in four spatial clusters (The solid lines denote the posterior means of the parameter and the dotted lines mark the 95% confidence intervals)

CHAPTER 6

DISCUSSION

This paper applies a series of latent cluster models for the spatio-temporal analysis of the POFV data. Overall, the latent cluster Poisson and ZIP models provide better fits than their conventional spatio-temporal analysis counterparts, proving their values in analyzing the data in which the effects of the covariates on the outcomes are believed to vary spatially and temporally. This paper illustrates the heterogeneity in the associations between police fatal victimization outcome and specific risk factors (i.e., the proportion of the black population) across the latent clusters. If such heterogeneity is not fully considered, it could lead to inaccurate and misleading conclusions. The proposed latent cluster models are superior to conventional models in that they incorporate spatial and temporal information to estimate the coefficients of the covariates instead of only considering the space and time variations in the global effects. On the other hand, unlike estimating the coefficients of covariates for each time or space unit, this approach only focuses on detecting spatial groups in which the association between the outcome and exposures have unique temporal trend, thus providing a relative parsimonious model.

The present study reveals that the positive association between the percentages of the black population and the killings of officers became stronger over time in the states belonging to cluster two. Such a increasing trend should gain the attention of policy makers. Although there are different theories explaining the positive association between percent black and police homicides (i.e., social disorgnization theory and racial threat theory)[2], racial inequality is generally recognized as the fundamental cause of such an association (Jacobs & Carmichael, 2002; Kaminski, 2002).Therefore efforts should be made to identify the factors aggravating racial inequality in these states. Resources should be allocated to eliminate these hazard factors.

Note that although overdispersion exists in the POFV data, the model fit does not benefit from the negative binomial model. However, the zero-inflated Poisson (ZIP) model does help improve the fitness of the estimations. This result suggests that the overdispersion in the raw data could come from the same process that also leads to zero-inflations. A negative binomial model may not fit well in such a situation. In contrast, a ZIP model could account for the overdispersion if it is actually a result of excess zeroes. If a ZIP model has fully addressed the overdisperson issue by modeling excess zeroes, a more complicated ZINB model may not be necessary, which is the case in this analysis. This finding suggests that the choice of the analysis model should be carefully considered in terms of specific fitness criterions, rather than being determined only by the existences of conditional dispersion and/or excess zeroes.

This study provides a flexible latent cluster model for analyzing spatial-temporal health data with excess zeros and overdispersion issues. However, some limitations have to be pointed out. First, the performance of other models handling excess zeros and overdispersions, such as hurdle models or zero-altered models, is not examined. These models have different explanations for the generation of excess zeros. Further research on these models may improve our understanding of the nature of excess zero observations.

---

[2] The social capital/collective efficacy framework rooted in social disorganization theory argues that high racial heterogeneity could impede a community to nurture mutual trust and support, thus resulting in reduced the willingness and capability to control disorder behaviors (Parker, McCall, and Land, 1999; Sampson and Groves, 1989; Sampson and Laub, 1993; 2005), which in turn increase the proximity of police to potential offenders. In contrast, the political explanation based on conflict theory and racial threat theory (Blalock, 1967; Eitle, D'Alessio, & Stolzenberg, 2002; Jackson, 1989) posits that the elevated violence against police has reflected suppressed minority groups' inarticulate protest or primitive rebellion toward the state's control force (Jacobs & Carmichael, 2002).

Second, the proposed approach does not consider the modelling of multivariate outcomes, which may improve inference. Extant studies suggest that general homicides and police homicides are correlated and have similar sets of predictors. Modeling general homicides and police homicides simultaneously, employing a multivariate CAR structure, could provide a clearer picture of the effects of related covariates across time and space.

REFERENCES

Besag, J., York, J., & Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics, 43,* 1–59.

Bernardinelli. L. , Clayton, D., Pascutto, C. , Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space-time variation in disease risk . *Statistics in Medicine*, *14*, 2433-2443.

Cai B., Lawson, A. B., Hossain, M. M., & Choi, J. (2012). Bayesian Latent Structure Models with Space-Time Dependent Covariates. *Statistical Modeling. 12,* 145–164.

Cai, B., Lawson, A. B., Hossain, M. M., Choi, J., Kirby, R. & Liu, J. (2013). Bayesian semiparametric model with spatially-temporally varying coefficients selection. *Statistics in Medicine, 32,* 3670-3685.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis:* Second Edition, Chapman & Hall, London.

Catelan, D., Biggeri, A., Dreassi, E., Lagazio, C. (2006). Space-cohort Bayesian models in ecological studies, *Statistical Modelling, 6,* 159-173.

Choi, J., Lawson, A. B., Cai, B., Hossain, M. M., Kirby, R. & Liu, J. (2012). A Bayesian latent model with spatio-temporally varying coefficients in low birth weight incidence data. *Statistical Methods in Medical Research*, *21*, 445-456.

Cooper, N. J., Sutton, A. J., Mugford, M., & Abrams, K. R. (2003). Use of Bayesian Markov Chain Monte Carlo methods to model cost-of-illness data. *Medical Decision Making, 23,* 38-53.

Department of Justice, Officer Safety and Wellness Group. (2016). *Understanding Firearms Assaults against Law Enforcement Officers.*

Dey, D.; Chen, M.; Change, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics , 53,* 1239–1252.

Dreassi, E., Biggeri, A., Catelan, D. (2005). Space-time models with time dependent covariates for the analysis of the temporal lag between socio-economic factors and mortality. *Statistics in Medicine, 24*, 1919–32.

Duhart, D. T. (2001, December). *Violence in the workplace: 1993-99* (Bureau of Justice Statistics Special Report). Washington, DC: U.S. Department of Justice.

Fridell, L., Faggiani, D., Taylor, B., Brito, C., & Kubu, B. (2009). The impact of agency context, policies, and practices on violence against police. *Journal of Criminal Justice, 37*, 542-552.

Gamerman, D., Moreira, A.R.B., & Rue, H.(2003). Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis, 42,* 513–33.

Geisser. S. (1993). *Predictive Inference: An Introduction*. New York: Chapman & Hall.

Gelfand, A., Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B, 56*, 501–514.

Gelfand, A.E., Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics, 4*, 11–25.

Geweke, J. (1992). Evaluating the accuracy of sampling - based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), 169 - 193. Oxford: Oxford University Press.

Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology, 35,*765–775.

Harcourt, B. E. (2011). An Institutionalization Effect: The Impact of Mental Hospitalization and Imprisonment on Homicide in the United States. *The Journal of Legal Studies, 40,*. 39-83.

Hossain, M.M., & Lawson, A.B. (2010). Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environmental and Ecological Statistics, 17,* 73–95.

Jacobs, D., & Carmichael, J. T. (2002). Subordination and violence against state control agents: Testing political explanations for lethal assaults against the police. *Social Forces, 80*, 1223−1251.

Kaminski, R. J. (2002). An opportunity model of police homicide victimization. *Dissertation Abstracts International*, 1-216 (UMI No. 3053970).

Kaminski, R. J. (2007). The Impact of deinstitutionalization on police homicide victimization (unpublished).

Kaminski, R. J. (2008). Assessing the county-level structural covariates of police homicides. *Homicide Studies, 12*, 350−380.

Kaminski, R. J., Digiovanni, C., & Downs, R. (2004). The use of force between the police andpersons with impaired judgment. *Police Quarterly*, *7,* 311–338.

Kaminski, R. J., & Marvell, T. B. (2002). A comparison of changes in police and general homicides, 1930–1998. *Criminology*, *40*, 171−191.

Kent, S. L. (2010). Killings of police in U.S. cities since 1980: An examination of environmental and political explanations. *Homicide Studies*, *14,* 3023.

Kesic, D., Thomas, S. DM. & Ogloff, J. R. P. (2013). Use of nonfatal force on and by persons with apparent mental disorder in encounters with police. *Criminal Justice and Behavior, 40,* 321-337.

Knorr-Held, L. (2000). Bayesian modeling of inseparable space-time variation in disease risk. *Statistics in Medicine, 19,* 2555–2567.

Knorr-Held, L. & Besag, J. (1998), Modelling risk from a disease in time and space. *Statistics in Medicine*, *17*, 2045–2060.

Lagazio, C., Dreassi, E. & Biggeri, A. (2001). A hierarchical Bayesian model for space–time variationof disease risk. *Statistical Modelling*, *1,* 17–29.

Lagazio, C., Biggeri, A., & Dreassi, E. (2003). Age-period-cohort models and disease mapping. *Environmetrics,14,* 475–490.

Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics, 38*: 963–974

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34,* 1-14.

Law enforcement officers killed and assaulted. (1997-2010). Washington, DC: Federal Bureau of Investigation.

Lawson, A. B., Biggeri, A. , Böhning, D., Lesaffre, E., Viel, J. F., Clark, A., Schlattmann, P., Divino, F., & Group), (Disease Mapping Collaborative (2000). Disease Mapping Models: An Empirical Evaluation. *Statistics in Medicine, 19,* 2217-2242.

Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). *Disease Mapping with WinBUGS and MLwiN.* John Wiley, Chichester.

Lawson, A. B., Song, H., Cai, B., Hossain, M.M. & Huang, K. (2010). Space-time latent component modeling for geo-referenced health data. *Statistics in Medicine, 29,* 2012-2027.

National Institute for Occupational Safety and Health. (2002). *Worker health chartbook, 2000.* Washington, DC: U.S. Department of Health and Human Services.

Raphael, S. (2000). *The deinstitutionalization of the mentally ill and growth in the U.S. prison population: 1971 to 1996*. Unpublished manuscript. University of California, Goldman School of Public Policy, Berkeley.

Richardson, S., Abellan, J.J., & Best, N. (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). *Statistical Methods in Medical Research, 15,* 385-407.

Salzer, M., Kaplan, K., & Atay, J. (2006). State psychiatric hospital census after the 1999 Olmstead decision. Evidence of decelerating deinstitutionalization. *Psychiatric Services*, *57*, 1501–1504.

Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996). *BUGS 0.5: Bayesian inference Using Gibbs Sampling - Manual (version ii),* Medical Research Council Biostatistics Unit, Cambridge.

Spiegelhalter, D.J., Best, N.G., Carlin, B.R. & van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 64,* 583-616.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society Series B, 62,* 795-809.

Wall, M. (2004). A close look at the spatial structure implied by the 462 CAR and SAR models. *Journal of Statistical Planning and Inference, 121*, 311–324

Waller, L., Carlin, B., Xia, H., & Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of American Statistical Association, 92,* 607–617.

Wakefield, J, Elliott, P. (1999). Issues in the statistical analysis of small area health data. *Statistics in Medicine*, *18*, 2377–2399.

Warchol, G. (1998). *Workplace violence, 1992–96* (Bureau of Justice Statistics Special Report). Washington, DC: Department of Justice, Bureau of Justice Statistics.

Yu, Q., Scribner, R., Carlin, B., Theall, K., Simonsen, N., Ghosh-Dastidar, B., Cohen, D., & Mason, K. (2008). Multilevel spatio-temporal dual change point models for relating alcohol outlet destruction and changes in neighborhood rates of assaultive violence. *Geospatial Health*, *2*, 161–172.

APPENDIX A –THE WINBUGS CODES FOR SELECTED MODELS

APPENDIX A.1The WinBUGS codes for the spatio-temporal latent cluster Possion model (Model 2)

```
# OBSERVED[i,j], is the observed murdered cop count in i th time interval and state j.

# EXPECTED[i,j], is the expected murdered cop count in i th time interval and state j.

model {

        for (i in 1:T){                    # T= 8 time periods.

                for (j in 1: N){                    # T= 8 time intervals.

                 OBSERVED[i,j]  ~ dpois(muf[i,j])

                log(mu[i,j])<-log(EXPECTED[i,j])+beta0[i,cluster[j]]+beta1[i,cluster[j]]*(POV[i,j]-
mean(POV[i,]))/sd(POV[i,])+beta2[i,cluster[j]]*(UNEMPLY[i,j]-
mean(UNEMPLY[i,]))/sd(UNEMPLY[i,])+beta3[i,cluster[j]]*(BLACKPCT[i,j]-
mean(BLACKPCT[i,]))/sd(BLACKPCT[i,])+beta4[i,cluster[j]]*(CRIMErs[i,j]-
mean(CRIMErs[i,]))/sd(CRIMErs[i,])+beta5[i,cluster[j]]*(INCARrs[i,j]-
mean(INCARrs[i,]))/sd(INCARrs[i,])+beta6[i,cluster[j]]*(POPDENSITY[i,j]-mean(POPDENSITY[i,]))/sd(POPDENSITY[i,])


                        muf[i,j]<-min(2.0E+3,max(1.0E-20,mu[i,j])) # Prevent error message

                        # Likelihood

                        L[i,j]<-exp(-muf[i,j])*pow(muf[i,j],OBSERVED[i,j])/exp(loggam(OBSERVED[i,j]+1))

                        Lf[i,j]<-max(1.0E-20,L[i,j]) # Prevent error message.

                        LogL[i,j]<-log(Lf[i,j])

                        D[i,j]<--2*LogL[i,j]    #Deviance

                        ci[i,j]<-1/exp(LogL[i,j])   #  will be used to calculate CPO. The average of ci[i,j]s will be the
posterior mean of the inverse of CPO[i,j]

                }

                        DevI[i]<-sum(D[i,])

                }

                        Dev<-sum(DevI[ ])

                #Prior distributions

                for (j in 1:N){

                   for (c in 1: NumCluster){
```

```
  w[j,c]~dnorm(spatial[j], taum)     # w[j,c] is the unstandardized weight, a log norm prior is assigned.
                    wb[j,c]<-max(-80,min(80,w[j,c]))  # Put a limit to prevent error message.
                    we[j,c]<-exp(wb[j,c])
}
for (c in 1: NumCluster){
              q[j,c]<-(we[j,c])/sum(we[j,])  # q[j,c] is the standardized weight,
}
cluster[j]~dcat(q[j,]) # Assign each state into one cluster
          }
          #Random walk for betas
          for (c in 1: NumCluster) {
            beta0[1,c]~dnorm(0, taub)
            beta1[1,c]~dnorm(0, taub)
            beta2[1,c]~dnorm(0, taub)
            beta3[1,c]~dnorm(0, taub)
            beta4[1,c]~dnorm(0, taub)
            beta5[1,c]~dnorm(0, taub)
            beta6[1,c]~dnorm(0, taub)
            for (t in 2 : T) {
              beta0[t, c]~dnorm(beta1[t-1, c], taub)
              beta1[t, c]~dnorm(beta1[t-1, c], taub)
              beta2[t, c]~dnorm(beta2[t-1, c], taub)
              beta3[t, c]~dnorm(beta3[t-1, c], taub)
              beta4[t, c]~dnorm(beta4[t-1, c], taub)
                    beta5[t, c]~dnorm(beta3[t-1, c], taub)
              beta6[t, c]~dnorm(beta4[t-1, c], taub)
                                   }
          }
```

```
#CAR

spatial[1:N] ~ car.normal(adj[], weights[], num[], tau)

for(k in 1:sumNumNeigh) {weights[k] <- 1}

#Other Priors

taum~ dgamma(0.025,0.025)  # prior on precision for the unstandardized weights

tau  ~ dgamma(0.01,0.01)          # prior on precision for CAR

taub<-1/pow(tem1,2.0)             # prior on precision for betas.

tem1~dunif(1,6)

}
```

APPENDIX A.2 The WinBUGS codes for the spatio-temporal latent cluster negative binomial model (Model 2)

```
model {

for (i in 1:T){                     # T= 8 time periods.

        for (j in 1: N){            # N=48 states.

                OBSERVED[i,j]~dnegbin(p[i,j],r1[i,j])

                p[i,j]<-r1[i,j]/(r1[i,j]+muf[i,j])

                r1[i,j]~dgamma(alpha[i,cluster[j]], alpha[i,cluster[j]])                    mu[i,j]<-
exp(log(EXPECTED[i,j])+beta0[i,cluster[j]]+beta1[i,cluster[j]]*(POV[i,j]-
mean(POV[i,]))/sd(POV[i,])+beta2[i,cluster[j]]*(UNEMPLY[i,j]-
mean(UNEMPLY[i,]))/sd(UNEMPLY[i,])+beta3[i,cluster[j]]*(BLACKPCT[i,j]-
mean(BLACKPCT[i,]))/sd(BLACKPCT[i,])+beta4[i,cluster[j]]*(CRIMErs[i,j]-
mean(CRIMErs[i,]))/sd(CRIMErs[i,])+beta5[i,cluster[j]]*(INCARrs[i,j]-
mean(INCARrs[i,]))/sd(INCARrs[i,])+beta6[i,cluster[j]]*(POPDENSITY[i,j]-mean(POPDENSITY[i,]))/sd(POPDENSITY[i,]))


                muf[i,j]<-min(2.0E+3,max(1.0E-20,mu[i,j]))  # Prevent error message

                #Log of Likelihood

                LogL[i,j]<- loggam( OBSERVED[i,j]+r1[i,j] ) - loggam( r1[i,j] ) - loggam( OBSERVED[i,j]+1 ) +
r1[i,j]*log( p[i,j] ) + OBSERVED[i,j]*log( 1-p[i,j] )

                        D[i,j]<--2*LogL[i,j]   #Deviance

                        ci[i,j]<-1/exp(LogL[i,j])   #  will be used to calculate CPO. The average of c[i,j]s will be the
posterior mean of the inverse of CPO[i,j]

                        }
```

```
                DevI[i] <-sum(D[i,])

}

                Dev<-sum(DevI[ ])

            for (j in 1:N){

              for (c in 1: NumCluster){

                w[j,c]~dnorm(spatial[j], taum)      # w[j,c] is the unstandardized weight, a log norm prior is assigned.

                wb[j,c]<-max(-80,min(80,w[j,c]))    # Put a limit to prevent error message.

                we[j,c]<-exp(wb[j,c])

                 }

    for (c in 1: NumCluster){

                q[j,c]<-(we[j,c])/sum(we[j,])   # q[j,c] is the standardized weight,

    }

                    cluster[j]~dcat(q[j,]) # Assign each state into one cluster (four clusters)


            }



                for (i in 1:T){

                    for (c in 1: NumCluster){

                        alpha[i,c]~dunif(0.25,1) # too small low bounary will cause error message.


                    }

                    }

           #The other priors are the same as in the Model 1.
```

APPENDIX A.3 The WinBUGS codes for the spatio-temporal latent cluster ZIP model (Model 3)

```
model {

    for (i in 1:T){                     # T= 8 time periods.

        for (j in 1: N){            # N=48 states

        OBSERVED[i,j]~dpois(muzip[i,j])
```

```
u[i,j]~dbern(p0[i,j])   #p0 is the prob. of the zero is a structural zero.

muzip[i,j]<-(1-u[i,j])*muf[i,j]  # if not an excess zero, follows a Pois(mu),

p0[i,j]<-exp(alpha0[i,cluster[j]]+alpha1[i,cluster[j]]*(POPDENSITY[i,j]-
mean(POPDENSITY[i,]))/sd(POPDENSITY[i,]))/(1+exp(alpha0[i,cluster[j]]+alpha1[i,cluster[j]]*(POPDENSITY[i,j]-
mean(POPDENSITY[i,]))/sd(POPDENSITY[i,])))  #logit part

log(mu[i,j])<-log(EXPECTED[i,j])+beta0[i,cluster[j]]+beta1[i,cluster[j]]*(POV[i,j]-
mean(POV[i,]))/sd(POV[i,])+beta2[i,cluster[j]]*(UNEMPLY[i,j]-
mean(UNEMPLY[i,]))/sd(UNEMPLY[i,])+beta3[i,cluster[j]]*(BLACKPCT[i,j]-
mean(BLACKPCT[i,]))/sd(BLACKPCT[i,])+beta4[i,cluster[j]]*(CRIMErs[i,j]-
mean(CRIMErs[i,]))/sd(CRIMErs[i,])+beta5[i,cluster[j]]*(INCARrs[i,j]-
mean(INCARrs[i,]))/sd(INCARrs[i,])+beta6[i,cluster[j]]*(POPDENSITY[i,j]-mean(POPDENSITY[i,]))/sd(POPDENSITY[i,])
        #Poisson part

# likelihood of the Poisson part

muf[i,j]<-max(1.0E-20,min(2.0E+3,mu[i,j])) # to prevent WinBUGS errors.

fd[i,j]<-exp(-muf[i,j]+OBSERVED[i,j]*log(muf[i,j])-(loggam(OBSERVED[i,j]+1)) )

# Likelihood

        L[i,j]<-p0[i,j]*equals(OBSERVED[i,j],0)+(1-p0[i,j])*fd[i,j]

        Lf[i,j]<-max(1.0E-20,L[i,j]) # to prevent WinBUGS errors.

        LogL[i,j]<-log(Lf[i,j])

        D[i,j]<--2*LogL[i,j]    #Deviance

        ci[i,j]<-1/exp(LogL[i,j])   #  will be used to calculate CPO. The average of c[i,j]s will be the
posterior mean of the inverse of CPO[i,j]

}

DevI[i]<-sum(D[i,])

}

Dev<-sum(DevI[ ])

#Prior distributions

for (z in 1:N){

  for (c in 1: NumCluster){

w[z,c]~dnorm(spatial[z], taum)    # w[j,c] is the unstandardized weight, a log norm prior is assigned.

                wb[z,c]<-max(-80,min(80,w[z,c])) # put a limit to prevent error messages

                we[z,c]<-exp(wb[z,c])
```

```
}
for (c in 1: NumCluster){
q[z,c]<-(we[z,c])/sum(we[z,]) # q[j,c] is the standardized weight,
}
cluster[z]~dcat(q[z,]) # Assign each state into one cluster
}
#Random walk for betas
for (c in 1: NumCluster) {
beta0[1,c]~dnorm(0, taub)
beta1[1,c]~dnorm(0, taub)
beta2[1,c]~dnorm(0, taub)
beta3[1,c]~dnorm(0, taub)
beta4[1,c]~dnorm(0, taub)
beta5[1,c]~dnorm(0, taub)
beta6[1,c]~dnorm(0, taub)
alpha0[1,c]~dnorm(0, 1.0E-1)
alpha1[1,c]~dnorm(0, 1.0E-1)
for (t in 2 : T) {
beta0[t, c]~dnorm(beta1[t-1, c], taub)
beta1[t, c]~dnorm(beta1[t-1, c], taub)
beta2[t, c]~dnorm(beta2[t-1, c], taub)
beta3[t, c]~dnorm(beta3[t-1, c], taub)
beta4[t, c]~dnorm(beta4[t-1, c], taub)
beta5[t, c]~dnorm(beta3[t-1, c], taub)
beta6[t, c]~dnorm(beta4[t-1, c], taub)
alpha0[t, c]~dnorm(alpha0[t-1, c], 1.0E-1)
alpha1[t, c]~dnorm(alpha1[t-1, c], 1.0E-1)
}
```

```
}

#The other priors are the same as in the Model 1.
```

APPENDIX A.4 The WinBUGS codes for the spatio-temporal latent cluster ZINB model (Model 4)

```
model {

Con<-10000 #set a large constant

  for (i in 1:T){                         # T= 8 time intervals.

          for (j in 1: N){          # N=48 states

          zeros[i,j]<-0  #zeros trick

          zeros[i,j]~dpois(zeros.means[i,j])

          zeros.means[i,j]<--LogL[i,j]+Con

          LogL[i,j]<-log(p0[i,j]*equals(OBSERVED[i,j],0)+(1-p0[i,j])*fd[i,j] )

          r1[i,j]~dgamma(alpha[i,cluster[j]], alpha[i,cluster[j]])

          p0[i,j]<-exp(alpha0[i,cluster[j]]+alpha1[i,cluster[j]]*(POPDENSITY[i,j]-
mean(POPDENSITY[i,]))/sd(POPDENSITY[i,]))/(1+exp(alpha0[i,cluster[j]]+alpha1[i,cluster[j]]*(POPDENSITY[i,j]-
mean(POPDENSITY[i,]))/sd(POPDENSITY[i,])))                  #logit part

          log(mu[i,j])<-log(EXPECTED[i,j])+beta0[i,cluster[j]]+beta1[i,cluster[j]]*(POV[i,j]-
mean(POV[i,]))/sd(POV[i,])+beta2[i,cluster[j]]*(UNEMPLY[i,j]-
mean(UNEMPLY[i,]))/sd(UNEMPLY[i,])+beta3[i,cluster[j]]*(BLACKPCT[i,j]-
mean(BLACKPCT[i,]))/sd(BLACKPCT[i,])+beta4[i,cluster[j]]*(CRIMErs[i,j]-
mean(CRIMErs[i,]))/sd(CRIMErs[i,])+beta5[i,cluster[j]]*(INCARrs[i,j]-
mean(INCARrs[i,]))/sd(INCARrs[i,])+beta6[i,cluster[j]]*(POPDENSITY[i,j]-
mean(POPDENSITY[i,]))/sd(POPDENSITY[i,])  #NB part

          muf[i,j]<-max(1.0E-20,min(2.0E+3,mu[i,j])) # in case mu too large or too small.

          # likelihood of the NB part

          lfd[i,j]<-loggam( OBSERVED[i,j]+r1[i,j] ) - loggam( r1[i,j] ) - loggam( OBSERVED[i,j]+1 ) +
r1[i,j]*log( p1[i,j] ) + OBSERVED[i,j]*log( 1-p1[i,j] )


   fd[i,j] <- exp( lfd[i,j] )


          p1[i,j]<-r1[i,j]/(r1[i,j]+muf[i,j])
```

```
                D[i,j]<--2*LogL[i,j]   #Deviance

                ci[i,j]<-1/exp(LogL[i,j])   #  will be used to calculate CPO. The average of c[i,j]s will be the
posterior mean of the inverse of CPO[i,j]

    }

                DevI[i]<-sum(D[i,])

                }

                Dev<-sum(DevI[ ])



    #Prior distributions

                for (j in 1:N){

                  for (c in 1: NumCluster){

          w[j,c]~dnorm(spatial[j], taum)     # w[j,c] is the unstandardized weight, a log norm prior is assigned.

                         wb[j,c]<-max(-60,min(60,w[j,c]))  # put limit to prevent error messages

                         we[j,c]<-exp(wb[j,c])

                        #w[j,c]~dlnorm(spatial[j], taum)

            }

            for (c in 1: NumCluster){

                        q[j,c]<-(we[j,c])/sum(we[j,])  # q[j,c] is the standardized weight,

            }

                                cluster[j]~dcat(q[j,]) # Assign each state into one cluster (four clusters)

                }



                        for (i in 1:T){

                                for (c in 1: NumCluster){

                                                alpha[i,c]~dunif(0.25,1) # too small low bounary will
cause error message.

                                        }

                                }
```

}#The other priors are the same as in the Model 1 and Model 3.

APPENDIX A.5 The WinBUGS codes for the mixed effect Poisson model (Model 5)

```
model {
	# Likelihood
	for (i in 1:N){			# N= 384
		ncopskill[i] ~dpois(mu[i])
		log(mu[i])<-log(EXPECTEDs[i])
+beta0+beta1*POVERTYs[i]+beta2*UNEMPLYs[i]+beta3*BLACKPCTs[i]+beta4*CRIMEs[i]+beta5*INCARs[i]+beta6*POPDEN
SITYs[i]+alpha0[stateid[i]]
		muf[i]<-min(2.0E+3, max(1.0E-20,mu[i])) #Prevent WinBUGS error message
		# alpha0[] within state random effect
		L[i]<-exp(-muf[i]+ncopskill[i]*log(muf[i])-loggam(ncopskill[i]+1))   #Likelihood
			LogL[i]<-log(L[i])
			D[i]<--2*LogL[i]	#Deviance
			c[i]<-1/exp(LogL[i])  #  will be used to calculate CPO. The average of c[i]s will be the posterior
mean of the inverse of CPO[i]
		}
			Dev<-sum(D[])
	#Prior distributions
		#Priors for alpha
				for (i in 1: 48){
					alpha0[i]~dnorm(0, 1.0E-2)
				}
		#CAR
		spatial[1:48] ~ car.normal(adj[], weights[], num[], tau)
		for(k in 1:sumNumNeigh) {
		weights[k] <- 1
		}
		#Other Priors
```

```
beta0 ~ dflat()

beta1~dnorm(0, taum)

        beta2~dnorm(0, taum)

        beta3~dnorm(0, taum)

        beta4~dnorm(0, taum)

        beta5~dnorm(0, taum)

        beta6~dnorm(0, taum)

        tau  ~ dgamma(0.01,0.01)        # prior on precision for CAR

        taum<-1/pow(tem1,2.0)           # prior on precision for betas.

        tem1~dunif(1,6)

}
```

APPENDIX A.6 The WinBUGS codes for the conventional spatio-temporal Poisson model (Model 9)

```
model {

        # Likelihood

        for (i in 1:T){                        # T= 8

                for (j in 1: N){               # N=48 states

                OBSERVED[i,j]  ~ dpois(muf[i,j])                        log(mu[i,j])<-
log(EXPECTED[i,j])+beta0+beta1*(POV[i,j]-mean(POV[i,]))/sd(POV[i,])+beta2*(UNEMPLY[i,j]-
mean(UNEMPLY[i,]))/sd(UNEMPLY[i,])+beta3*(BLACKPCT[i,j]-
mean(BLACKPCT[i,]))/sd(BLACKPCT[i,])+beta4*(CRIMErs[i,j]-mean(CRIMErs[i,]))/sd(CRIMErs[i,])+beta5*(INCARrs[i,j]-
mean(INCARrs[i,]))/sd(INCARrs[i,])+beta6*(POPDENSITY[i,j]-
mean(POPDENSITY[i,]))/sd(POPDENSITY[i,])+alpha0[j]+spatial[j]+temp[i,j]

                muf[i,j]<-min(2.0E+3, max(1.0E-20,mu[i,j])) # prevent WinBUGS error message

                # spatial[] represent spatial structured variation (CAR). temp[] represent random walk temporal autocorrelation.

                # alpha0[] unstructured spatial random effect.

                        L[i,j]<-exp(-muf[i,j]+OBSERVED[i,j]*log(muf[i,j])-loggam(OBSERVED[i,j]+1))   #Likelihood

                        LogL[i,j]<-log(L[i,j])

                        D[i,j]<--2*LogL[i,j]   #Deviance
```

```
			Lf[i,j]<-max(1.0E-20,L[i,j])
			c[i,j]<-1/exp(LogL[i,j]) #  will be used to calculate CPO. The average of c[i,j]s will be the posterior mean of the inverse of CPO[i,j]
			cf[i,j]<-1/Lf[i,j]  #in case L is zero
		}
			DevI[i]<-sum(D[i,])
		}
			Dev<-sum(DevI[ ])
	#Prior distributions
			#Random walk
		for (p in 1: N) {
		temp[1,p]~dnorm(0, taum)
		for (t in 2 : T) {
		temp[t, p]~dnorm(temp[t-1, p], taum)
		}
				}
		#CAR
		spatial[1:N] ~ car.normal(adj[], weights[], num[], tau)
		for(k in 1:sumNumNeigh) {
		weights[k] <- 1
		}
		#Other Priors
		for (s in 1:N){
		alpha0[s] ~ dnorm(0.0, 1.0E-3)
		}
	beta0 ~ dflat()
	beta1~dnorm(0, taum)
		 beta2~dnorm(0, taum)
		 beta3~dnorm(0, taum)
```

```
beta4~dnorm(0, taum)

beta5~dnorm(0, taum)

beta6~dnorm(0, taum)

taum ~dgamma(0.025,0.025)

tau  ~ dgamma(0.01,0.01)          # prior on precision for CAR

}
```

APPENDIX A.7 The R codes for calculating DIC and NLLK for Model 3.

```
Q=4 # number of clusters. Can be changed if the number of clusters changed.

#read raw data

data <- readRDS("newdata.rds")

N = 48

T=8

attach(data)

# The following are the dependent variable and covariates.

OBSERVED<-(array(ncopskill,dim=c(8,48)))

EXPECTED<-(array(exp4yearstatecopskill,dim=c(8,48)))

BLACKPCT<-(array(BLACKPCT,dim=c(8,48)))

POPDENSITY<-(array(POPDENSITY,dim=c(8,48)))

POV<-(array(POVERTY,dim=c(8,48)))

UNEMPLY<-(array(UNEMPLY,dim=c(8,48)))

INCARrs<-(array(residincarrbs,dim=c(8,48)))

CRIMErs<-(array(residcrimebs,dim=c(8,48)))

# attach BUGS object

library(R2WinBUGS)

attach.bugs(res.sim)

# Calculate Dhat and NLLK

# Preparing work

r<-rep(NA,384)
```

```
rhohat<-(array(r,dim=c(8,48)))
muhat<-(array(r,dim=c(8,48)))
spatialhat<-rep(NA, 48)
alpha0hat<-rep(NA, 48)
temphat<-(array(r,dim=c(8,48)))
mustarhat<-(array(r,dim=c(8,48)))
fdhat<-(array(r,dim=c(8,48)))
p0hat<-(array(r,dim=c(8,Q)))
L.hat<-(array(r,dim=c(8,48)))
LogL.hat<-(array(r,dim=c(8,48)))
LogL1.hat<-(array(r,dim=c(8,48)))
D.hat<-(array(r,dim=c(8,48)))
DevI.hat<-rep(NA,8)
Dev.hat<-NULL
ic<-(array(r,dim=c(8,48)))
cpo<-(array(r,dim=c(8,48)))
lgcpo<-(array(r,dim=c(8,48)))
DvI.hat<-rep(NA,8)
lgcpoi<-rep(NA,8)
Dv.hat<-NULL
NLLK<-NULL
beta0hat<-(array(r,dim=c(8,Q)))
beta1hat<-(array(r,dim=c(8,Q)))
beta2hat<-(array(r,dim=c(8,Q)))
beta3hat<-(array(r,dim=c(8,Q)))
beta4hat<-(array(r,dim=c(8,Q)))
beta5hat<-(array(r,dim=c(8,Q)))
beta6hat<-(array(r,dim=c(8,Q)))
```

```
ind<-(array(r,dim=c(8,48)))

#obtain posterior means of the parameters

for (i in 1:T){

        for (j in 1:Q){

                        beta0hat[i,j]<-mean(beta0[,i,j])

                        beta1hat[i,j]<-mean(beta1[,i,j])

                        beta2hat[i,j]<-mean(beta2[,i,j])

                        beta3hat[i,j]<-mean(beta3[,i,j])

                        beta4hat[i,j]<-mean(beta4[,i,j])

                        beta5hat[i,j]<-mean(beta5[,i,j])

                        beta6hat[i,j]<-mean(beta6[,i,j])

}}

        for (i in 1:T){

        for (j in 1:N){

                        #NLLK

                        ic[i,j]<-mean(ci[,i,j]) # posterior mean of the inverse of CPO_it

                        cpo[i,j]<-1/ic[i,j]

                        lgcpo[i,j]<-log(cpo[i,j])

                        #Dhat

                        # compute the mu in the Poisson part by the posterior means of the parameters

                        muhat[i,j]<-exp(log(EXPECTED[i,j])+beta0hat[i,cluster[j]]+

                        beta1hat[i,cluster[j]]*(POV[i,j]-mean(POV[i,]))/sd(POV[i,])+

                        beta2hat[i,cluster[j]]*(UNEMPLY[i,j]-mean(UNEMPLY[i,]))/sd(UNEMPLY[i,])+

                        beta3hat[i,cluster[j]]*(BLACKPCT[i,j]-mean(BLACKPCT[i,]))/sd(BLACKPCT[i,])+

                        beta4hat[i,cluster[j]]*(CRIMErs[i,j]-mean(CRIMErs[i,]))/sd(CRIMErs[i,])+

                        beta5hat[i,cluster[j]]*(INCARrs[i,j]-mean(INCARrs[i,]))/sd(INCARrs[i,])+

                        beta6hat[i,cluster[j]]*(POPDENSITY[i,j]-mean(POPDENSITY[i,]))/sd(POPDENSITY[i,]))

                        if (OBSERVED[i,j]==0) {ind[i,j]=1 } else {ind[i,j]=0}
```

```
            p0hat[i,cluster[j]]<-mean(p0[,i,cluster[j]])

            #likelihood of Poisson part

            fdhat[i,j]<-exp(-muhat[i,j]+OBSERVED[i,j]*log(muhat[i,j])-log(factorial(OBSERVED[i,j])))

            #likelihood

            LogL.hat[i,j]<-log(p0hat[i,cluster[j]]*ind[i,j]+(1-p0hat[i,cluster[j]])*fdhat[i,j])


            }

            DvI.hat[i]<-sum(LogL.hat[i,])

            lgcpoi[i]<-sum(lgcpo[i,])

        }

        Dv.hat<--2*sum(DvI.hat[])

        Dhat<-Dv.hat

        NLLK<--sum(lgcpoi[])

DIC<-2*mean(Dev)-Dv.hat

pD<-mean(Dev)-Dv.hat

Dhat<-Dv.hat

Dbar<-mean(Dev)

Dbar

Dhat

pD

DIC

NLLK
```

APPENDIX B TEMPORAL PROFILES OF THE COEFFICIENTS IN

MODEL 3

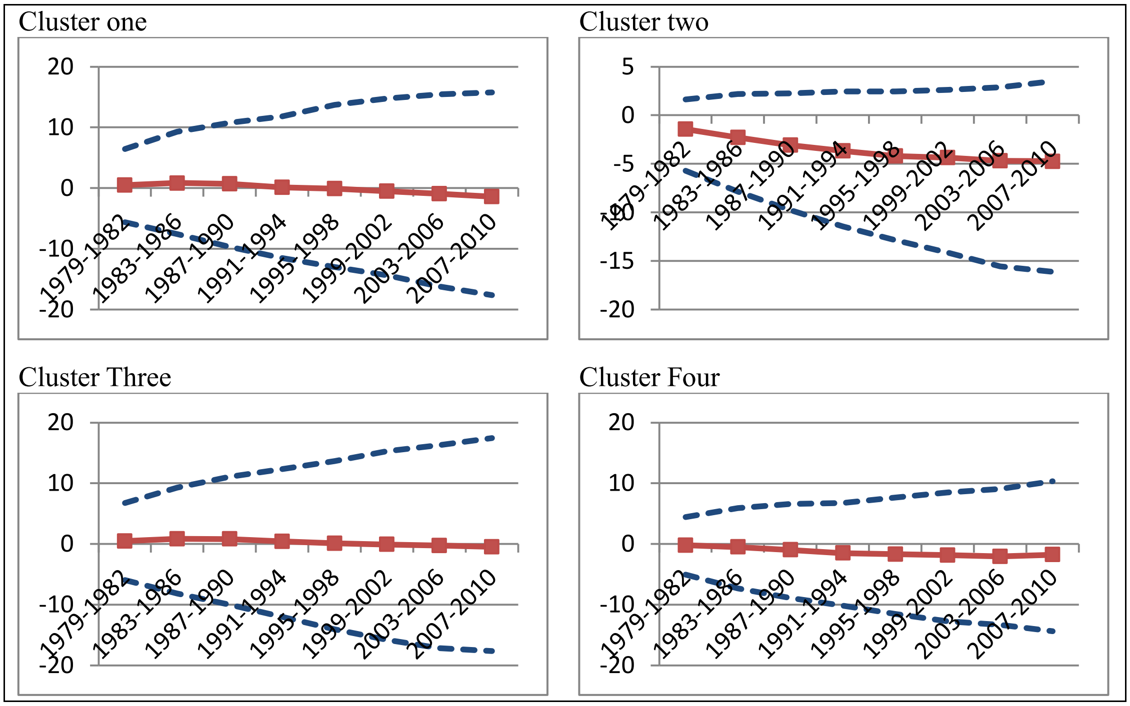

Figure B.1 Temporal profiles of the effects of population density (alpha1)

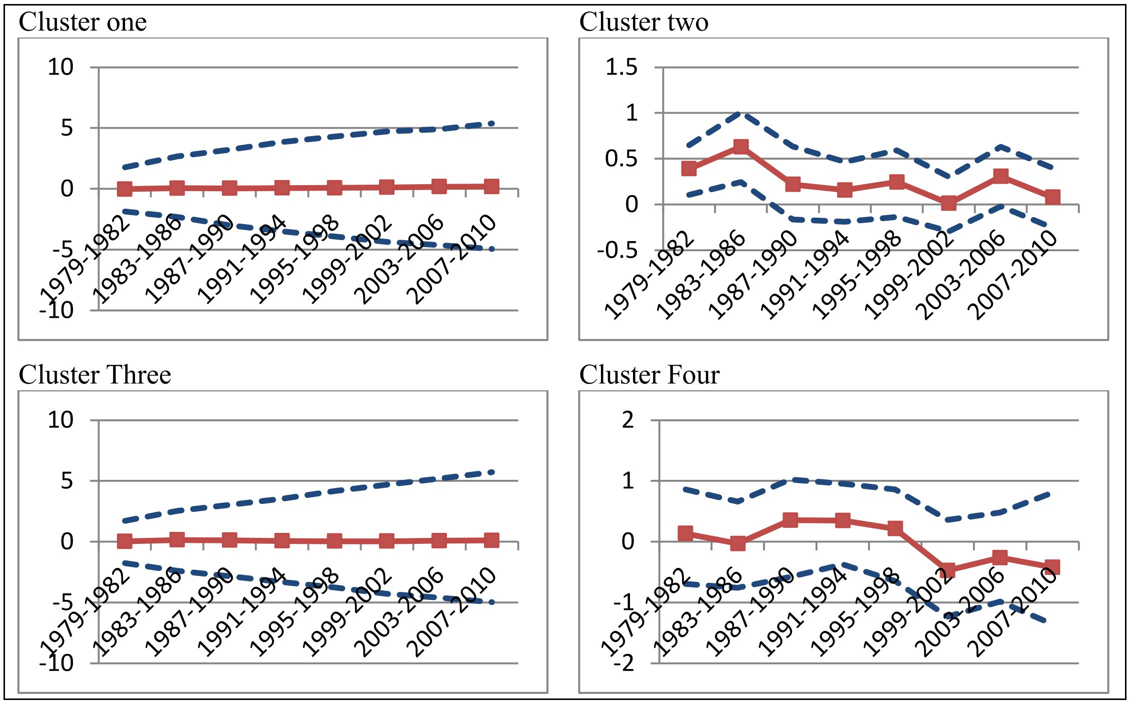

Figure B.2 Temporal profiles of the effects of poverty (beta1)

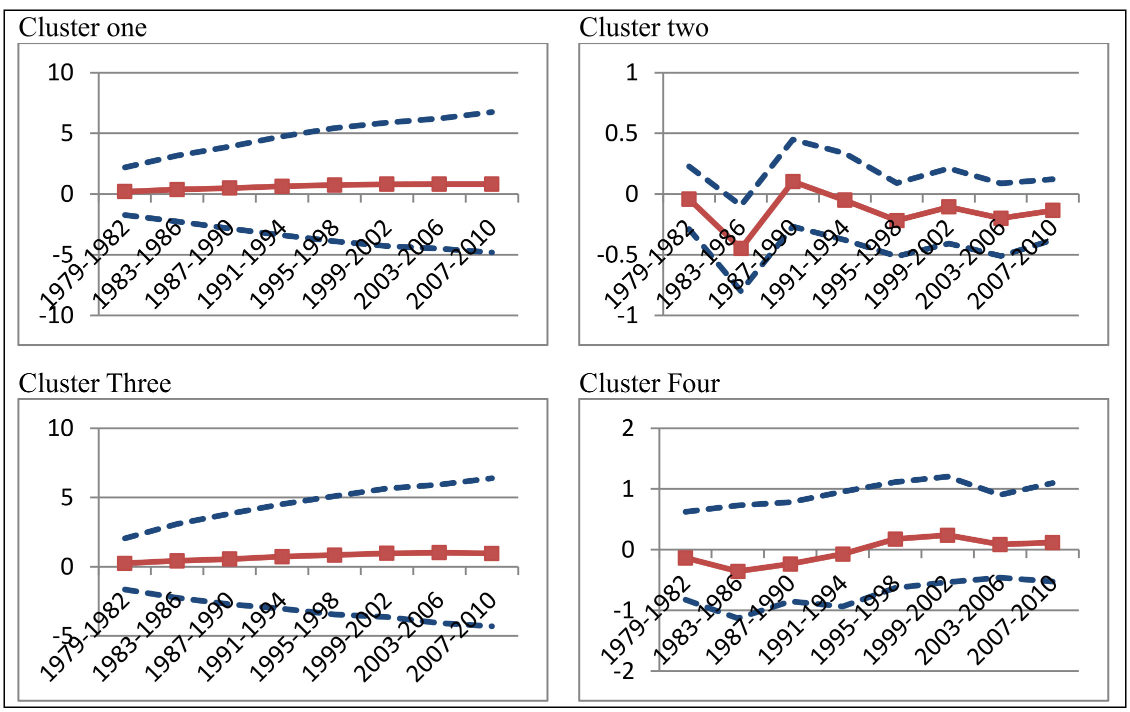

Figure B.3 Temporal profiles of the effects of unemployment rates (beta2)

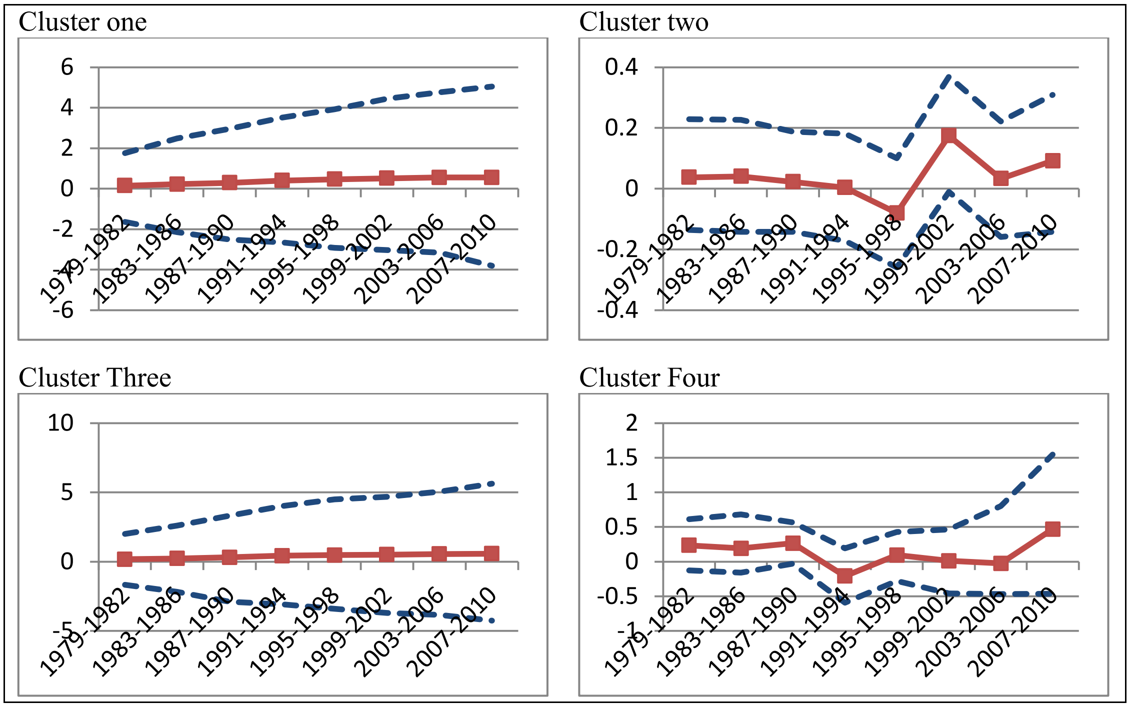

Figure B.4 Temporal profiles of the effects of violent crime rates (beta4)

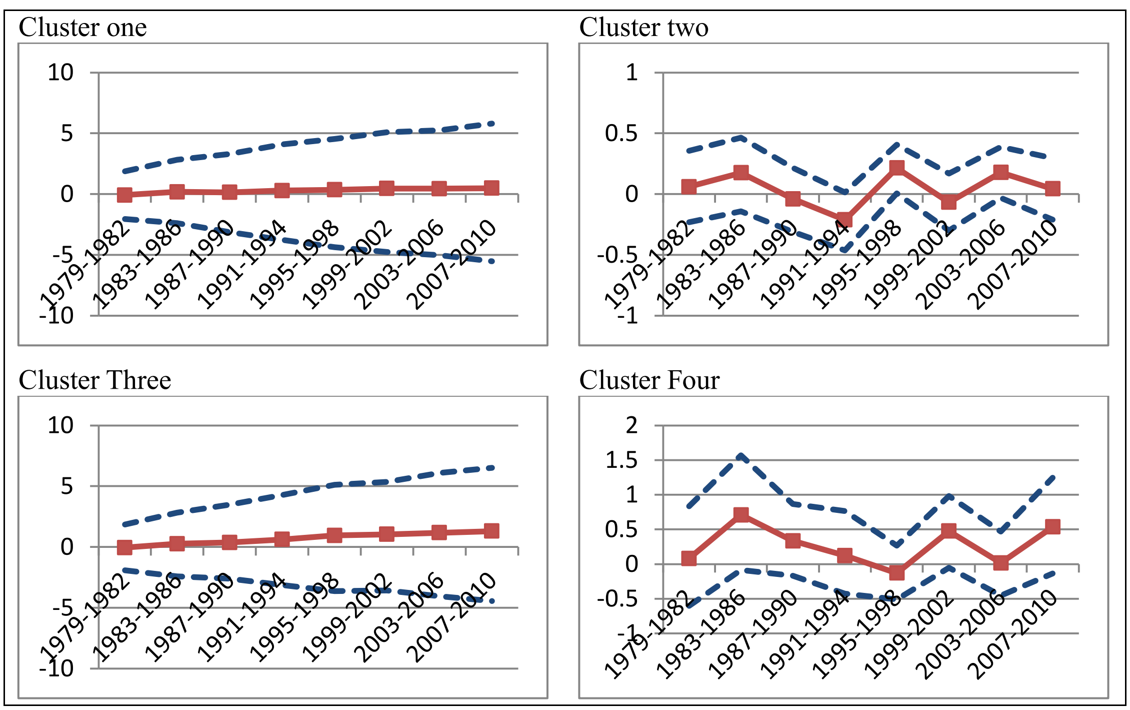

Figure B.5 Temporal profiles of the effects of incarceration rates (beta5)

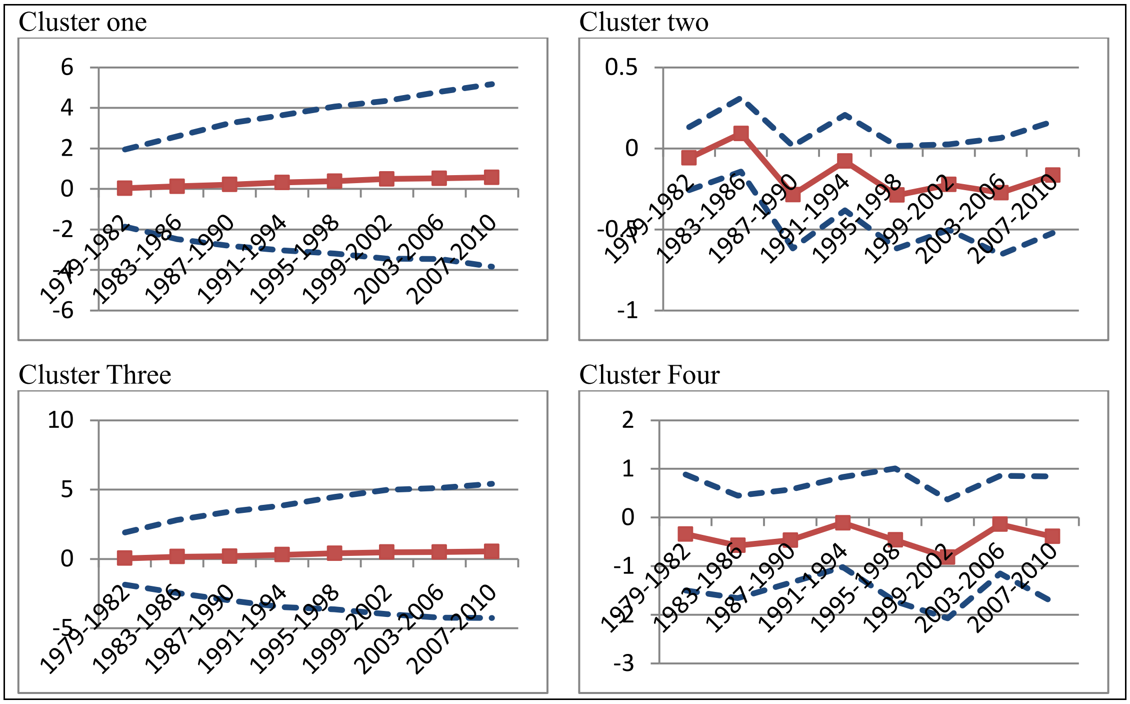

Figure B.6 Temporal profiles of the effects of population density (beta6)