Appendix A

Negative binomial regression models and estimation methods

This appendix presents the characteristics of negative binomial regression models and discusses their estimating methods. The material described below was originally written in Appendix C of the CrimeStat IV (version 4.02) program documentation (Lord and Park, 2013; Levine, 2015) and has been adapted for this textbook with permission.

Probability density and likelihood functions

The properties of the negative binomial models with and without spatial interaction are described in the next two sections.

Poisson-gamma model

The Poisson-gamma model has properties that are very similar to the Poisson model in which the dependent variable y_i is modeled as a Poisson variable with a mean λ_i where the model error is assumed to follow a gamma distribution. As its name implies, the Poisson-gamma is a mixture of two distributions and was first derived by Greenwood and Yule (1920). This mixture distribution was developed to account for over-dispersion that is commonly observed in discrete or count data (Lord et al., 2005). It became very popular because the conjugate distribution (same family of functions) has a closed form and leads to the negative binomial distribution. As discussed by Cook (2009), "the name of this distribution comes from applying the binomial theorem with a negative exponent." Two major parameterizations have been proposed and they are known as the NB-1 and NB-2, the latter one being the most commonly known and utilized. NB-2 is therefore described first. Other parameterizations exist but are not discussed here (see Maher and Summersgill, 1996; Hilbe, 2012).

NB-2 model

Suppose that we have a series of random counts that follows the Poisson distribution:

$$P(y_i|\lambda_i) = \frac{e^{-\lambda_i}\lambda_i}{y_i!} \tag{A.1}$$

where $P(y_i|\lambda_i)$ is the probability of roadway entity (or observation) i having y_i crashes per unit of time y_i; and, λ_i is the mean of the Poisson distribution. If the Poisson mean is assumed to have a random intercept term and this term enters the conditional mean function in a multiplicative manner, we get the following relationship (Cameron and Trivedi, 2013):

$$\begin{aligned}
\lambda_i &= \exp\left(\beta_0 + \sum\nolimits_{j=1}^{K} x'_{ij}\beta_j + \varepsilon_i\right) \\
\lambda_i &= e^{\sum_{j=1}^{K} x'_{ij}\beta_j} e^{(\beta_0+\varepsilon_i)} \\
\lambda_i &= e^{\left(\beta_0+\sum_{j=1}^{K} x'_{ij}\beta_j\right)} e^{\varepsilon_i} \\
\lambda_i &= \mu_i \nu_i
\end{aligned} \tag{A.2}$$

where $\exp(\beta_0 + \varepsilon_i)$ is defined as a random intercept; $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{K} x'_{ij}\beta_j\right)$ is the log-link between the Poisson mean and the covariates or independent variables xs; and, βs are the parameters or regression coefficients. The relationship can also be formulated using vectors, such that $\mu_i = \exp(\boldsymbol{x}'_i\boldsymbol{\beta})$.

The marginal distribution of y_i can be obtained by integrating the error term ν_i,

$$\begin{aligned}
f(y_i|\mu_i) &= \int_o^\infty g(y_i;\mu_i,\nu_i)h(\nu_i)d\nu_i \\
f(y_i|\mu_i) &= E_\nu[g(y_i;\mu_i,\nu_i)]
\end{aligned} \tag{A.3}$$

where $h(\nu_i)$ is a mixing distribution. In the case of the Poisson-gamma mixture, $g(y_i|\mu_i,\nu_i)$ is the Poisson distribution, and $h(\nu_i)$ is the gamma distribution. This distribution has a closed form and leads to the NB distribution.

Let us assume the variable ν_i follows a two-parameter gamma distribution:

$$k(\nu_i|\psi,\delta) = \frac{\delta^\psi}{\Gamma(\psi)}\nu_i^{\psi-1}e^{-\nu_i\delta}, \quad \psi > 0, \delta > 0, \nu_i > 0 \tag{A.4}$$

where $E[\nu_i] = \psi/_\delta$ and $VAR[\nu_i] = \psi/_{\delta^2}$. Setting $\psi = \delta$ gives us the one-parameter gamma where $E[\nu_i] = 1$ and $VAR[\nu_i] = {}^1/_\psi$. (Note: in the main text $\psi = 1/\alpha = \phi$. The notation is changed in this appendix since more Greek letters are needed to describe spatial modeling.) We can transform the gamma distribution as a function of the Poisson mean, which gives the following *probability mass function* (PMF) (Cameron and Trivedi, 2013):

$$k(\lambda_i|\psi, \mu_i) = \frac{(\psi/\mu_i)^\psi}{\Gamma(\psi)} \lambda_i^{\psi-1} e^{-\frac{\lambda_i}{\mu_i}\delta} \tag{A.5}$$

Combining Eqs. (A.1) and (A.5) into Eq. (A.3) yields the marginal distribution of y_i:

$$f(y_i|\mu_i, \psi) = \int_o^\infty \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \frac{(\psi/\mu)^\psi}{\Gamma(\psi)} \lambda_i^{\psi-1} e^{-\frac{\lambda_i}{\mu_i}\delta} d\lambda_i \tag{A.6}$$

Using the properties of the gamma function, it can be shown that Eq. (A.6) can be defined as follows:

$$f(y_i|\mu_i, \psi) = \frac{(\psi/\mu_i)^\psi}{\Gamma(\psi)\Gamma(y_i+1)} \int_o^\infty \exp\left(-\lambda_i\left(1+\frac{\psi}{\mu_i}\right)\right) \lambda_i^{y+\psi-1} d\lambda_i$$

$$f(y_i|\mu_i, \psi) = \frac{(\psi/\mu_i)^\psi \left(1+\frac{\psi}{\mu_i}\right)^{-(\psi+y_i)} \Gamma(\psi+y_i)}{\Gamma(\psi)\Gamma(y_i+1)} \tag{A.7}$$

$$f(y_i|\mu_i, \psi) = \frac{\Gamma(y_i+\psi)}{\Gamma(y_i+1)\Gamma(\psi)} \left(\frac{\psi}{\mu_i+\psi}\right)^\psi \left(\frac{\mu_i}{\mu_i+\psi}\right)^{y_i}$$

The PMF of the NB-2 model is therefore (the last part of Eq. A.7):

$$f(y_i|\mu_i, \psi) = \frac{\Gamma(y_i+\psi)}{\Gamma(y_i+1)\Gamma(\psi)} \left(\frac{\psi}{\mu_i+\psi}\right)^\psi \left(\frac{\mu_i}{\mu_i+\psi}\right)^{y_i} \tag{A.8}$$

Note that the PMF has also been defined in the literature as follows:

$$f(y_i|\psi, \mu_i) = \binom{y_i+\psi-1}{\psi-1} \left(\frac{\psi}{\mu_i+\psi}\right)^\psi \left(\frac{\mu_i}{\mu_i+\psi}\right)^{y_i} \tag{A.9}$$

The first two moments of the NB-2 are the following:

$$E[y_i|\mu_i, \psi] = \mu_i \tag{A.10}$$

$$VAR[y_i|\mu_i, \psi] = \mu_i + \frac{\mu_i^2}{\psi} \tag{A.11}$$

The next steps consist of defining the *log-likelihood* (LL) function of the NB-2. It can be shown that

$$\ln\left(\frac{\Gamma(y_i+\psi)}{\Gamma(\psi)}\right)=\sum_{j=0}^{y-1}\ln(j+\psi) \tag{A.12}$$

By substituting Eq. (A.12) into (A.8), the log-likelihood can be computed using the following equation:

$$\ln L(\psi,\beta)=\sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{y-1}\ln(j+\psi)\right)-\ln y_i!-(y_i+\psi)\ln\left(1+\psi^{-1}\mu_i\right)+y_i\ln\psi^{-1}+y_i\ln\mu_i\right\} \tag{A.13}$$

Note also that the log-likelihood has also been expressed as follows:

$$\ln L(\psi,\beta)=\sum_{i=1}^{n}\left\{y_i\ln\left(\frac{\psi\mu_i}{1+\psi\mu_i}\right)-\psi^{-1}\ln(1+\psi\mu_i)+\ln\Gamma\left(y_i+\psi^{-1}\right)-\ln\Gamma(y_i+1)-\ln\Gamma\left(\psi^{-1}\right)\right\} \tag{A.14}$$

Recall that $\mu_i=\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$.

In the statistical literature, the Poisson-gamma model has also been defined as follows:

$$P(y_i|\lambda_i)=\text{Poisson}(\lambda_i)\; i=1,2,\ldots,I \tag{A.15}$$

where the mean of the Poisson is structured as follows:

$$\lambda_i=f(\mathbf{x}_i|\boldsymbol{\beta})\exp(\varepsilon_i)=\mu_i\exp(\varepsilon_i) \tag{A.16}$$

and where $f()$ is a function of the covariates, $\mathbf{x}$ is a vector of explanatory variables; as before, $\boldsymbol{\beta}$ is a vector of estimable parameters; and, ε_i is the model error independent of all the covariates with mean equal to 1 and a variance equal to $1/\psi$.

NB-1 model

The NB-1 is very similar to the NB-2, but the parameterization of the variance (the second moment) is slightly different than in Eq. A.11.

$$E[y_i|\mu_i,\psi]=\mu_i \tag{A.17}$$

$$VAR[y_i|\mu_i, \psi] = \mu_i + \frac{\mu_i}{\psi} \tag{A.18}$$

The log-likelihood of the NB-1 is given by

$$\ln L(\psi, \beta) = \sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{y-1}\ln(j + \psi\mu_i)\right) - \ln y_i! - (y_i + \psi\mu_i)\ln\left(1 + \psi^{-1}\right) + y_i \ln \psi^{-1}\right\} \tag{A.19}$$

The NB-1 is usually less flexible in capturing the variance and is not used very often by analysts and statisticians. Interested readers are referred to Cameron and Trivedi (2013) for additional information about this parameterization.

Poisson-gamma model with spatial interaction

The Poisson-gamma (or NB model) can also incorporate data that are collected spatially. To capture this kind of data, a spatial autocorrelation term needs to be added to the model. Using the notation described in Eq. A.15, the NB-2 model with spatial interaction can be defined as follows:

$$P(y_i|\lambda_i) = \text{Poisson}(\lambda_i) \tag{A.20}$$

with the mean of Poisson-gamma organized as

$$\lambda_i = \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i + \phi_i\right) \tag{A.21}$$

The assumption on the uncorrelated error term ε_i is the same as in the Poisson-gamma model described earlier; same as before, $\mu_i = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$. The third term in the expression, φ_i, is a *spatial random effect*, one for each observation. Together, the spatial effects are distributed as a complex *multivariate normal* (or Gaussian) density function. In other words, the second model is a spatial regression model within a negative binomial model.

There are two common ways to express the spatial component, either as a *Conditional Autoregressive* (CAR) or as a *Simultaneous Autoregressive* (SAR) function (De Smith et al., 2007). The CAR model is expressed as

$$E\left(y_i\middle|\text{all}y_{j\neq i}\right) \underset{i\neq j}{=} \mu_i + \rho\sum_{ij}\left[w_{ij}\left(y_i - \mu_j\right)\right] \tag{A.22}$$

where μ_i is the expected value for observation i; w_{ij} is a spatial weight[1] between observation i and all other observations j (and for which all weights sum to 1.0); and, ρ is a spatial autocorrelation parameter that determines the size and nature of the spatial neighborhood effect. The summation of the spatial weights times the difference between the observed and predicted values is over all other observations ($i \neq j$). The reader is referred to Haining (1990) and LeSage (2001) for further details.

The SAR model has a simpler form and can be expressed as

$$E\left(y_i \middle| \text{all} y_{j \neq i}\right) \underset{i \neq j}{=} \mu_i + \rho \sum_{ij} \left[w_{ij} y_j\right] \tag{A.23}$$

where the terms are as defined earlier. Note that in the CAR model the spatial weights are applied to the difference between the observed and expected values at all other locations whereas in the SAR model, the weights are applied directly to the observed value. In practice, the CAR and SAR models produce very similar results. Additional information about the Poisson-gamma-CAR is described in the following.

Estimation methods

This section describes two methods that can be used for estimating the coefficients of the regression NB models. The two methods are the maximum likelihood estimates (MLE) and the Bayesian method based on the Monte Carlo Markov Chain (MCMC).

Maximum likelihood estimation

The coefficients or parameters of the NB regression model are estimated by taking the first-order conditions and making them equal to zero. There are two first-order equations, one for the model's parameters and one for the dispersion parameter (Lawless, 1987). The two for the NB-2 are as follows:

$$\sum_{i=1}^{n} \frac{y_i - \mu_i}{1 + \psi^{-1} \mu_i} \mathbf{x_i} = \mathbf{0} \tag{A.24a}$$

[1]Note: there are different weight factors that have been proposed, such as the inverse distance weight function, exponential distance decay weight function and the Gaussian weighting function among others.

$$\sum_{i=1}^{n}\left\{\frac{1}{\left(\psi^{-1}\right)^{2}}\left(\ln\left(1+\psi^{-1}\mu_i\right)-\sum_{j=0}^{y_i-1}\frac{1}{(j+\psi)}\right)+\frac{y_i-\mu_i}{\psi^{-1}\left(1+\psi^{-1}\mu_i\right)}\right\}=0 \tag{A.24b}$$

where $\mathbf{x_i}$ is a vector of covariates.

The series of equations can be solved using the Newton-Raphson procedure or the scoring algorithm.

The confidence intervals on the βs and ψ^{-1} can be calculated using the covariance matrix that is assumed to be normally distributed:

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \psi^{-1}\end{bmatrix}\sim N\left(\begin{bmatrix}\boldsymbol{\beta}\\ \psi^{-1}\end{bmatrix},\begin{bmatrix}VAR[\boldsymbol{\beta}] & 0\\ 0 & VAR\left[\psi^{-1}\right]\end{bmatrix}\right) \tag{A.25}$$

where

$$VAR[\boldsymbol{\beta}]=\left(\sum_{i=1}^{n}\frac{\mu_i}{1+\psi^{-1}\mu_i}\mathbf{x_i}\mathbf{x}_i'\right)^{-1} \tag{A.26a}$$

$$VAR\left[\psi^{-1}\right]=\left(\sum_{i=1}^{n}\frac{i}{\left(\psi^{-1}\right)^{4}}\left(\ln\left(1+\psi^{-1}\mu_i\right)-\sum_{j=0}^{y_{j-1}}\frac{1}{(j+\psi)}\right)^{2}+\frac{\mu_i}{\left(\psi^{-1}\right)^{2}\left(1+\psi^{-1}\mu_i\right)}\right)^{-1} \tag{A.26b}$$

It should be pointed out that the NB-2 with spatial interaction model (Poisson-gamma-CAR) cannot be estimated using the MLE method. It needs to be estimated using the MCMC technique, which is described next.

Monte Carlo Markov Chain estimation

This section presents how to draw samples from the posterior distribution of the Poisson-gamma model and Poisson-gamma-conditional autoregressive (CAR) model using the MCMC technique.

MCMC Poisson-gamma model

The Poisson-gamma model can be formulated from a two-stage hierarchical Poisson model:

$$\text{(Likelihood) } y_i|\lambda_i\sim\text{Poisson }(\lambda_i) \tag{A.27a}$$

$$\text{(First}-\text{stage) } \lambda_i|\psi\sim\pi_\lambda(\psi) \tag{A.27b}$$

$$\text{(Second}-\text{stage) } \psi\sim\pi_\psi(\,\cdot\,) \tag{A.27c}$$

where $\pi_\lambda(\psi)$ is the *prior distribution* imposed on the Poisson mean, λ_i, with a prior parameter ψ, and $\pi_\psi(\bullet)$ is the *hyper-prior* on ψ with known *hyper-parameters* (a, b, for example).

In Eqs. (A.27b) and (A.27c), if we specify $\lambda_i = \nu_i \mu_i$ (where $\nu_i(=e^{\varepsilon_i})\sim$Gamma $(\psi,\ \psi)$ in the first stage and $\psi\sim$Gamma $(a,\ b)$ in the second stage), these result in exactly the NB-2 regression model described in the previous section. With this specification, it is also easy to show that λ_i in the first stage follows Gamma $(\psi,\ \psi/\mu_i)$ as shown in Eq. (A.5). Note that $\mu_i = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$ as described earlier.

For simplicity, if a *flat uniform prior* is assumed for each β_j $(j = 0, 1, \ldots, J)$ and the parameters βs and ψ are mutually independent, the joint posterior distribution for the Poisson-gamma model is defined as

$$\pi(\lambda, \beta, \psi|y) \propto f(y|\lambda)\cdot\pi(\lambda|\beta,\psi)\cdot\pi(\beta_0)\ldots\pi(\beta_J)\cdot\pi(\psi|a,b) \tag{A.28a}$$

$$= \left(\prod_{i=1}^{n}\frac{e^{-\lambda_i}(\lambda_i)^{y_i}}{y_i!}\right)\times\left(\prod_{i=1}^{n}\frac{\left(\psi e^{-\mathrm{x}_i'\beta}\right)^{\psi}}{\Gamma(\psi)}\lambda_i^{\psi-1}e^{-\left(\psi e^{-\mathrm{x}_i'\beta}\right)\lambda_i}\right)\times\left(\psi^{(a-1)}e^{-b\psi}\right) \tag{A.28b}$$

The parameters of interest are $\boldsymbol{\lambda} = (\lambda_1, \ldots\lambda_n)$, $\boldsymbol{\beta} = (\beta_0,\ \beta_1, \ldots\beta_J)$, and the inverse dispersion parameter ψ (or the *dispersion parameter* $\gamma = 1/\psi$). Ideally, samples need to be drawn of each parameter from the joint posterior distribution. However, the form in Eq. (A.28b) is very complex and it is difficult to draw samples from such a distribution. Consequently, samples are drawn from the full conditional distribution *sequentially* (that is, one at a time). This iterative process is called the Gibbs sampling method.

Therefore, once the full conditionals are known for each parameter, Gibbs sampling can be implemented by drawing samples of each parameter sequentially. The full conditional distributions for each parameter for the Poisson-Gamma model can be easily derived from Eq. (A.28b) and are given as (Park, 2010)

$$\begin{aligned}&\pi(\lambda_i|\beta,\psi,y_i)\propto f(y_i|\lambda_i)\cdot\pi(\lambda_i|\beta,\psi)\\&=\text{Gamma}\left(y_i+\psi, 1+\psi e^{-x_i'\beta}\right),\quad\text{for } i=1,2,\ldots,n\end{aligned} \tag{A.29a}$$

$$\begin{aligned}&\pi\left(\beta_j|\lambda,\beta_{-j},\psi\right)\propto\pi\left(\lambda|\beta_{-j},\psi\right)\cdot\pi\left(\beta_j\right)\\&=\exp\left\{-\psi\left[\left(\sum_{i=1}^{n}x_{ij}\right)\beta_j+\sum_{i=1}^{n}\lambda_i e^{-x_i'\beta}\right]\right\},\quad\text{for } j=0,1,\ldots J\end{aligned} \tag{A.29b}$$

$$\pi(\psi|\lambda,\boldsymbol{\beta},a,b)\propto\pi(\boldsymbol{\lambda}|\boldsymbol{\beta},\psi)\cdot\pi(\psi|a,b)$$

$$=\exp\left\{-n\ln(\Gamma(\psi))+\psi\left(n\ln(\psi)-\sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}+\ln(\lambda_i)+\lambda_i e^{-\mathbf{x}_i'\boldsymbol{\beta}}\right)\right)\right.$$
$$\left.+(a-1)\ln(\psi)-b\psi\right\} \tag{A.29c}$$

However, unlike Eq. (A.29a), the full conditional distributions for the βs and ψ (Eqs. (A.29b) and (A.29c)) do not belong to any standard distribution family so it is not easy to draw samples directly from their full conditional distributions. Although there are several approaches to sampling from such a complex distribution, the popular sampling algorithm used in practice is the Metropolis-Hastings (or MH) algorithm with *slice sampling* of individual parameters.

The MCMC sampling procedure using the slice sampling algorithm within Gibbs sampling, therefore, can be summarized as follows:

1. Start with initial values $\lambda^{(0)}$, $\boldsymbol{\beta}^{(0)}$ and $\psi^{(0)}$. Repeat the following steps for $t=1,\ldots,T_0,\ldots,T_0+T$.

2. *Step 1*: Conditional on knowing $\boldsymbol{\beta}^{(t-1)}$ and $\psi^{(t-1)}$, draw $\lambda^{(t)}$ from Eq. A.29a independently for $i=1,2,\ldots,n$.

3. *Step 2*: Conditional on knowing $\lambda^{(t)}$ and $\psi^{(t-1)}$, draw $\boldsymbol{\beta}^{(t)}$ from Eq. A.29b independently for $j=0,1,\ldots,J$ using the slice sampling method.

4. *Step 3*: Conditional on knowing $\lambda^{(t)}$ and $\boldsymbol{\beta}^{(t)}$, draw $\psi^{(t)}$ from Eq. A.29c using the slice sampling method.

5. *Step 4*: Store the values of all parameters (i.e., $\lambda^{(t)}$, $\boldsymbol{\beta}^{(t)}$ and $\psi^{(t)}$). Increase t by one and return to Step 1.

6. *Step 5*: Discard the first T_0 draws as a *burn-in* period.

After equilibrium is reached at the T_0 iteration, sampled values are averaged to provide the consistent estimates of the parameters:

$$\widehat{E}[h(\theta)]=\frac{\sum_{t=T_0+1}^{T}h(\theta)^{(t)}}{T} \tag{A.30}$$

where θ denotes any interest parameter in the model.

MCMC Poisson-gamma-CAR model

For the Poisson-gamma-CAR model, the only difference from the Poisson-gamma model is the way λ_i is structured. The mean of Poisson-gamma-CAR is organized as

$$\lambda_i = \exp\left(x_i'\boldsymbol{\beta} + \varepsilon_i + \varphi_i\right) \tag{A.31}$$

where φ_i is a spatial random effect, one for each observation. As in the Poisson-gamma model, we specify $e^{\varepsilon_i} \sim \text{Gamma}(\psi, \psi)$ to model the independent error term. To model the spatial effect, φ_i, we assume the following:

$$p(\varphi_i|\boldsymbol{\Phi}_{-i}) \propto \exp\left(-\frac{w_{i+}}{2\sigma_\varphi^2}\left[\varphi_i - \rho\sum_{j\neq i}\frac{w_{ij}}{w_{i+}}\varphi_j\right]^2\right) \tag{A.32}$$

where $p(\varphi_i|\boldsymbol{\Phi}_{-i})$ is the probability of a spatial effect given a lagged spatial effect, $w_{i+} = \sum_{i\neq j} w_{ij}$ which sums all over all records, and j (i.e., all other zones) except for the record of interest, i. This formulation gives a conditional normal density with mean $=\rho\sum_{j\neq i}\frac{w_{ij}}{w_{i+}}\varphi_j$ and variance $=\frac{\sigma_\varphi^2}{w_{i+}}$. The parameter ρ determines the direction and overall magnitude of the spatial effects. The term w_{ij} is a spatial weight function between zones i and j. In the algorithm, the term for the variance is $\sigma_\varphi^2 = 1/\tau_\varphi$ and the same variance is used for all observations.

We define the spatial weight matrix $\mathbf{W}$ with the entries w_{ij} and the diagonal entries $w_{ii} = 0$. The matrix $\mathbf{D}$ is defined as a diagonal matrix with the diagonal entries, w_{i+}. Sun et al. (1999) show that if $\kappa_{\min}^{-1} < \rho < \kappa_{\max}^{-1}$ where $\kappa_{\min}$ and $\kappa_{\max}$ are the smallest and largest eigenvalues of $\mathbf{WD}^{-1}$, respectively, then, $\boldsymbol{\Phi}$ has a multivariate normal distribution with mean $\mathbf{0}$ and nonsingular covariance matrix $\sigma_\varphi^2(\mathbf{D} - \rho\mathbf{W})^{-1}$.

$$\boldsymbol{\Phi} = (\varphi_1, \ldots \varphi_n)' = MNV_n\left(\mathbf{0}, \sigma_\varphi^2\mathbf{A}^{-1}\right) = \frac{|\mathbf{A}|^{1/2}}{\left(2\pi\sigma_\varphi^2\right)^{n/2}}\exp\left(-\frac{1}{2\sigma_\varphi^2}\boldsymbol{\Phi}'\mathbf{A}\boldsymbol{\Phi}\right) \tag{A.33}$$

where $\mathbf{A} = (\mathbf{D} - \rho\mathbf{W})$ and $\kappa_{\min}^{-1} < \rho < \kappa_{\max}^{-1}$.

Prior distributions for MCMC Poisson-gamma-CAR

For the prior distributions, we assume the following distributions for each parameter:

Parameter	Prior distribution
β_j $(j = 0, 1, \ldots, J)$	Uniform$(-\infty, \infty)$
ψ	Gamma(a_ψ, b_ψ)
$\tau_\varphi\left(=\sigma_\varphi^{-2}\right)$	Gamma(a_φ, b_φ)
ρ	Uniform$(\kappa_{\min}^{-1}, \kappa_{\max}^{-1})$

The parameters in the Poisson-gamma-CAR model are $\boldsymbol{\lambda} = (\lambda_1, \cdots \lambda_n)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots \beta_J)$, ψ, $\boldsymbol{\Phi} = (\varphi_1, \ldots \varphi_n)$, τ_φ and ρ. Then, the random samples can be drawn from the full conditional distributions of each parameter. It can be shown that the full conditional distributions for each parameter are given as follows:

$$\pi(\lambda_i|\text{others}) \sim \text{Gamma}\left(y_i + \psi, 1 + \psi e^{-x_i'\beta - \varphi_i}\right), \quad \text{for } i = 1, 2, \ldots, n \tag{A.34a}$$

$$\pi\left(\beta_j|\text{others}\right) \propto \exp\left\{-\psi\left[\left(\sum_{i=1}^{n} x_{ij}\right)\beta_j + \sum_{i=1}^{n} \lambda_i e^{-x_i'\beta - \varphi_i}\right]\right\}, \quad \text{for } j = 0, 1, \ldots, J \tag{A.34b}$$

$$\begin{aligned}\pi(\psi|\text{others}) \propto \exp\Bigg\{ & -n\ln(\Gamma(\psi)) + \psi\left(n\ln(\psi) - \sum_{i=1}^{n}\left(x_i'\beta - \varphi_i + \ln(\lambda_i)\right.\right. \\ & \left.\left. - \lambda_i e^{-x_i'\beta - \varphi_i}\right)\right) + (a_\psi - 1)\ln(\psi) - b_\psi\psi\Bigg\}\end{aligned} \tag{A.34c}$$

$$\pi(\varphi_i|\text{others}) \propto \exp\left\{-\psi\varphi_i - \psi\lambda_i e^{-x_i'\beta - \varphi_i} - \frac{\tau_\varphi}{2}\left(\boldsymbol{\Phi}^T\mathbf{A}\boldsymbol{\Phi}\right)\right\}, \quad \text{for } i = 1, 2, \ldots, n \tag{A.34d}$$

$$\pi\left(\tau_\varphi|\text{others}\right) \propto \text{Gamma}\left(a_\varphi + \frac{n}{2}, b_\varphi + \frac{1}{2}\boldsymbol{\Phi}^T\mathbf{A}\boldsymbol{\Phi}\right) \tag{A.34e}$$

$$\pi(\rho|\text{others}) \propto \exp\left\{\frac{1}{2}\sum_{i=1}^{n}\ln(1 - \rho\kappa_i) - \frac{\tau_\varphi}{2}\left(\boldsymbol{\Phi}^T\mathbf{A}\boldsymbol{\Phi}\right)\right\} \tag{A.34f}$$

where $\kappa_1, \ldots, \kappa_n$ are the eigenvalues of $\mathbf{WD}^{-1}$.

As the full conditional distributions were specified, the Gibbs sampling method can be applied sequentially. It is easy to generate random samples from conditionals from Eqs. (A.34a) and (A.34e). The other full conditional distributions are not of closed form, and so the slice sampling method should be applied.

References

Cameron, A.C., Trivedi, P.K., 2013. Regression Analysis of Count Data, second ed. Cambridge University Press, Cambridge, UK.

Cook, J.D., 2009. Notes on the Negative Binomial Distribution. www.johndcook.com/negative_binomial.pdf. (Accessed 1 August 2020).

De Smith, M., Goodchild, M.F., Longley, P.A., 2007. Geospatial Analysis, second ed. Leicester, U.K, Matador.

Greenwood, M., Yule, G.U., 1920. An inquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. J. R. Stat. Soc. A 83, 255−279.

Haining, R., 1990. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press, Cambridge, U.K.

Hilbe, J.M., 2012. Negative Binomial Regression, second ed. Cambridge University Press, Cambridge, U.K.

Lawless, J.F., 1987. Negative binomial and mixed Poisson regression. Can. J. Stat. 15 (3), 209−225.

LeSage, J.P., 2001. A Family of Geographically Weighted Regression Models. Working Paper. Department of Economics, University of Toledo, Toledo, OH.

Levine, N., 2015. CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 4.02. Ned Levine & Associates, Houston, Texas, and the National Institute of Justice, Washington, D.C. August.

Lord, D., Park, B.-J., 2013. CrimeStat IV: Appendix C − Negative Binomial Regression Models and Estimation Methods. Ned Levine & Associates and the National Institute of Justice, Washington, D.C. https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/CrimeStat%2520IV%2520Appendix%2520C.pdf.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-Gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Acc. Anal. Prev. 37 (1), 35−46.

Maher, M.J., Summersgill, I., 1996. A Comprehensive methodology for the fitting predictive accident models. Acc. Anal. Prev. 28 (3), 281−296.

Park, B.-J., 2010. Application of the Finite Mixture Models for Vehicle Crash Data Analysis. Ph.D. dissertation. Texas A&M University, College Station, TX.

Sun, D., Tsutakawa, R.K., Speckman, P.L., 1999. Posterior distribution of hierarchical models using CAR(1) distributions. Biometrika 86 (2), 341−350.