Appendix C:

Negative Binomial Regression Models and Estimation Methods

Dominique Lord
Zachry Dept. of
Civil Engineering
Texas A & M University
College Station, TX

Byung-Jung Park
Korea Transport Institute
Goyang, South Korea

This appendix presents the characteristics of Negative Binomial regression models and discusses their estimating methods.

Probability Density and Likelihood Functions

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

Poisson-Gamma Model

The Poisson-Gamma model has properties that are very similar to the Poisson model discussed in Appendix B, 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." There are two major parameterizations that have been proposed and they are known as the NB1 and NB2, the latter one being the most commonly known and utilized. NB2 is therefore described first. Other parameterizations exist, but are not discussed here (see Maher and Summersgill, 1996; Hilbe, 2007).

NB2 Model

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

$$g\left(y_i;\lambda_i\right)=\frac{e^{-\lambda_i}\lambda_i}{y_i!} \tag{C.1}$$

where y_i is the observed number of counts for $i=1,2,\ldots,\text{n}$; 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, 1998):

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

where $\exp\left(\beta_0+\varepsilon_i\right)$ 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 the βs are the regression coefficients. As discussed in Appendix B, the relationship can also be formulated using vectors, such that $\mu_i=\exp(\mathbf{x}_i'\boldsymbol{\beta})$.

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

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

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

Assume that the variable ν_i follows a two-parameter Gamma distribution:

$$k(\nu_i;\psi,\delta)=\frac{\delta^{\psi}}{\Gamma(\psi)}\nu_i^{\psi-1}e^{-\nu_i\delta},\qquad \psi>0,\ \delta>0,\ \nu_i>0 \tag{C.4}$$

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 $VAR[\nu_i]=1/\psi$. We can transform the Gamma distribution as a function of the Poisson mean, which gives the following *probability density function* (PDF; Cameron and Trivedi, 1998):

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

Combining equations C-1 and C-5 with equation C-3 yields the marginal distribution of y_i:

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

Using the properties of the Gamma function, it can be shown that equation C-6 can be expressed as:

$$\begin{aligned}
f(y_i;\mu_i,\psi)&=\frac{(\psi/\mu_i)^{\psi}}{\Gamma(\psi)\Gamma(y_i+1)}\int_o^{\infty}\exp\left(-\lambda_i\left(1+\frac{\psi}{\mu_i}\right)\right)\lambda_i^{y+\psi-1}d\lambda_i\\
f(y_i;\mu_i,\psi)&=\frac{(\psi/\mu_i)^{\psi}\left(1+\frac{\psi}{\mu_i}\right)^{-(\psi+y_i)}\Gamma(\psi+y_i)}{\Gamma(\psi)\Gamma(y_i+1)}\\
f(y_i;\mu_i,\psi)&=\frac{\Gamma(y_i+\psi)}{\Gamma(y_i+1)\Gamma(\psi)}\left(\frac{\psi}{\mu_i+\psi}\right)^{\psi}\left(\frac{\mu_i}{\mu_i+\psi}\right)^{y_i}
\end{aligned} \tag{C-7}$$

The PDF of the NB2 model is therefore the last part of Equation C-7:

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

Note that the PDF has also been defined in the literature as:

$$f\left(y_i;\psi,\mu_i\right)=\begin{pmatrix} y_i+\psi-1 \\ \psi-1 \end{pmatrix}\left(\frac{\psi}{\mu_i+\psi}\right)^{\psi}\left(\frac{\mu_i}{\mu_i+\psi}\right)^{y_i} \tag{C.9}$$

The first two moments of the NB2 are the following:

$$E\left[y_i;\mu_i,\psi\right]=\mu_i \tag{C.10}$$

$$VAR\left[y_i;\mu_i,\psi\right]=\mu_i+\frac{\mu_i^2}{\psi} \tag{C.11}$$

The next step consists of defining the ***log-likelihood*** function of the NB2. It can be shown that:

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

By substituting equation C-12 into C-8, the log-likelihood can be computed using the following equation:

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

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

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

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

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

$$y_i \mid \lambda_i = Poisson(\lambda_i) \quad \text{i} = 1, 2, \ldots, \text{I} \tag{C.15}$$

where the mean of the Poisson is structured as:

$$\lambda_i = f(\mathbf{X};\boldsymbol{\beta})\exp(\varepsilon_i) = \mu_i \exp(\varepsilon_i) \tag{C.16}$$

and where, $f(.)$ is a function of the covariates, $\mathbf{X}$ (Miaou and Lord, 2003). As before, $\boldsymbol{\beta}$ is a vector of coefficients and ε_i is the model error independent of all the covariates with mean equal to 1 and a variance equal to $1/\psi$.

NB1 Model

The NB1 is very similar to the NB2, but the parameterization of the variance (the second moment) is slightly different than in equation C-11.

$$E\left[y_i;\mu_i,\psi\right] = \mu_i \tag{C.17}$$

$$VAR\left[y_i;\mu_i,\psi\right] = \mu_i + \frac{\mu_i}{\psi} \tag{C.18}$$

The log-likelihood of the NB1 is given by:

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

The NB1 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 (1998) for additional information about this parameterization.

Poisson-Gamma Model with Spatial Interaction

The Poisson-Gamma (or negative binomial 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 Equation C-15, the NB2 model with spatial interaction can be defined as:

$$y_i \mid \lambda_i = Poisson(\lambda_i) \tag{C.20}$$

with the mean of Poisson-Gamma organized as:

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

The assumption on the uncorrelated error term ε_i is the same as in the Poisson-Gamma model described above; same as before, namely $\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. 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 Gausian) 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:

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

where μ_i is the expected value for observation *i*, w_{ij} is a spatial weight between the 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. Note that 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. 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:

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

where the terms are as defined above. 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 below.

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 Monte Carlo Markov Chain (MCMC).

Maximum Likelihood Estimation

The characteristics of the MLE method were described in Appendix B for the normal and Poisson regression. The same characteristics apply here. The coefficients 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 coefficients and one for the dispersion parameter (Lawson, 1987). The two for the NB2 are as follows:

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

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

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

Similar to the Poisson model, 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:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \alpha \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\beta} \\ \alpha \end{bmatrix}, \begin{bmatrix} VAR[\boldsymbol{\beta}] & 0 \\ 0 & VAR[\alpha] \end{bmatrix} \right) \tag{C.25}$$

where,

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

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

It should be pointed out that the NB2 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 discusses 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:

$$\text{(Likelihood)} \quad y_i \mid \lambda_i \sim Poisson\,(\lambda_i) \tag{C.27a}$$

$$\text{(First-stage)} \quad \lambda_i \mid \psi \sim \pi_\lambda(\psi)$$

(C.27b)

$$\text{(Second-stage)} \quad \psi \sim \pi_\psi(\cdot) \tag{C.27b}$$

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

In Equations C-27a and C-27b, 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 NB2 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 Equation C-5. Note that $\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ as described above.

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

$$\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \psi \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\lambda}) \cdot \pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \psi) \cdot \pi(\beta_0) \cdots \pi(\beta_J) \cdot \pi(\psi \mid a, b) \tag{C.28a}$$

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

The parameters of interest are $\boldsymbol{\lambda} = (\lambda_1, \cdots \lambda_n)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots \beta_J)$, 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 Equation C-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 Equation C-28b and are given as (Park, 2010):

$$\pi(\lambda_i \mid \boldsymbol{\beta}, \psi, y_i) \propto f(y_i \mid \lambda_i) \cdot \pi(\lambda_i \mid \boldsymbol{\beta}, \psi)$$
$$= Gamma\,(y_i + \psi, 1 + \psi e^{-\mathbf{x}_i'\boldsymbol{\beta}}), \text{ for } i = 1, 2, \cdots, n \qquad \text{(C.29a)}$$

$$\pi(\beta_j \mid \boldsymbol{\lambda}, \boldsymbol{\beta}_{-j}, \psi) \propto \pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}_{-j}, \psi) \cdot \pi(\beta_j)$$
$$= \exp\left\{-\psi\left[\left(\sum_{i=1}^{n} x_{ij}\right)\beta_j + \sum_{i=1}^{n} \lambda_i e^{-\mathbf{x}_i'\boldsymbol{\beta}}\right]\right\}, \text{ for } j = 0, 1, \cdots, J \qquad \text{(C.29b)}$$

$$\pi(\psi \mid \boldsymbol{\lambda}, \boldsymbol{\beta}, a, b) \propto \pi(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \psi) \cdot \pi(\psi \mid a, b) \qquad \text{(C.29c)}$$
$$= \exp\left\{-n \ln(\Gamma(\psi)) + \psi\left(n \ln(\psi) - \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta} + \ln(\lambda_i) + \lambda_i e^{-\mathbf{x}_i'\boldsymbol{\beta}}\right)\right) + (a-1)\ln(\psi) - b\psi\right\}$$

However, unlike Equation C-29a, the full conditional distributions for the βs and ψ (Equations C-29b and C-29c) do not belong to any standard distribution family so it is not easy to draw samples directly from their full conditional distributions. While there are several approaches to sampling from such a complex distribution, the particular sampling algorithm used in *CrimeStat* is a 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 $\boldsymbol{\lambda}^{(0)}$, $\boldsymbol{\beta}^{(0)}$ and $\psi^{(0)}$. Repeat the following steps for $t = 1, \cdots, T_0, \cdots, T_0 + T$.
2. *Step 1*: Conditional on knowing $\boldsymbol{\beta}^{(t\text{-}1)}$ and $\psi^{(t\text{-}1)}$, draw $\boldsymbol{\lambda}^{(t)}$ from Equation C-29a independently for $i = 1, 2, \cdots, n$.
3. *Step 2*: Conditional on knowing $\boldsymbol{\lambda}^{(t)}$ and $\psi^{(t\text{-}1)}$, draw $\boldsymbol{\beta}^{(t)}$ from Equation C-29b independently for $j = 0, 1, \cdots, J$ using the slice sampling method.
4. *Step 3*: Conditional on knowing $\boldsymbol{\lambda}^{(t)}$ and $\boldsymbol{\beta}^{(t)}$, draw $\psi^{(t)}$ from Equation C-29c using the slice sampling method.

5. *Step 4*: Store the values of all parameters (i.e., $\boldsymbol{\lambda}^{(t)}$, $\boldsymbol{\beta}^{(t)}$ and $\psi^{(t)}$). Increase t by one and return to Step 1.
6. *Step 5*: Discard the first *k* draws as a *burn-in* period, where *k* is defined by the user.

After equilibrium is reached at the k^{th} iteration, sampled values are averaged to provide the consistent estimates of the parameters:

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

where θ denotes any parameter of interest 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:

$$\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i + \phi_i) \tag{C.31}$$

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

$$p(\phi_i \mid \boldsymbol{\Phi}_{-i}) \propto \exp\left(-\frac{w_{i+}}{2\sigma_\phi^2}\left[\phi_i - \rho \sum_{j \neq i} \frac{w_{ij}}{w_{i+}} \phi_j \right]^2 \right) \tag{C.32}$$

where $p(\phi_i \mid \boldsymbol{\Phi}_{-i})$ 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, j (i.e., all other zones) except for the record of interest, i. This formulation gives a conditional normal density with mean $= \rho \sum_{j \neq i} \frac{w_{ij}}{w_{i+}} \phi_j$ and variance $= \frac{\sigma_\phi^2}{w_{i+}}$. 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 $\sigma_\phi^2 = 1/\tau_\phi$ 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, Tsutakawa, and Speckman (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 $\mathbf{\Phi}$ has a multivariate normal distribution with mean **0** and nonsingular covariance matrix $\sigma_\phi^2(\mathbf{D} - \rho\mathbf{W})^{-1}$.

$$\mathbf{\Phi} = (\phi_1, \cdots \phi_n)' = MNV_n\left(\mathbf{0}, \sigma_\phi^2 \mathbf{A}^{-1}\right) = \frac{|\mathbf{A}|^{1/2}}{(2\pi\sigma_\phi^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_\phi^2}\mathbf{\Phi}'\mathbf{A}\mathbf{\Phi}\right) \tag{C.33}$$

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
$\beta_j \ (j = 0, 1, \cdots, J)$	$Uniform(-\infty, \infty)$
ψ	$Gamma\,(a_\psi, b_\psi)$
$\tau_\phi (= \sigma_\phi^{-2})$	$Gamma\,(a_\phi, b_\phi)$
ρ	$Uniform(\kappa_{\min}^{-1}, \kappa_{\max}^{-1})$

The parameters in the Poisson-Gamma-CAR model are $\boldsymbol{\lambda} = (\lambda_1, \cdots \lambda_n)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots \beta_J)$, ψ, $\mathbf{\Phi} = (\phi_1, \cdots \phi_n)$, 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:

$$\pi(\lambda_i \mid others) \sim Gamma\,(y_i + \psi, 1 + \psi e^{-\mathbf{x}_i'\boldsymbol{\beta} - \phi_i}), \text{ for } i = 1, 2, \cdots, n \tag{C.34a}$$

$$\pi(\beta_j \mid others) \propto \exp\left\{-\psi\left[\left(\sum_{i=1}^{n} x_{ij}\right)\beta_j + \sum_{i=1}^{n} \lambda_i e^{-\mathbf{x}_i'\boldsymbol{\beta} - \phi_i}\right]\right\}, \text{ for } j = 0, 1, \cdots, J \tag{C.34b}$$

$$\pi(\psi \mid others) \tag{C.34c}$$

$$\propto \exp\left\{-n\ln(\Gamma(\psi))+\psi\left(n\ln(\psi)-\sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}-\phi_i+\ln(\lambda_i)-\lambda_i e^{-\mathbf{x}_i'\boldsymbol{\beta}-\phi_i}\right)\right)+(a_\psi-1)\ln(\psi)-b_\psi\psi\right\}$$

$$\pi(\phi_i \mid others) \propto \exp\left\{-\psi\phi_i-\psi\lambda_i e^{-\mathbf{x}_i'\boldsymbol{\beta}-\phi_i}-\frac{\tau_\phi}{2}(\boldsymbol{\Phi}^T\mathbf{A}\boldsymbol{\Phi})\right\},\ \text{for } i=1,2,\cdots,n \qquad \text{(C.34d)}$$

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

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

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

Since the full conditional distributions were specified, the Gibbs sampling method can be applied sequentially. It is easy to generate random samples from Equations C-34a and C-34e. The other full conditional distributions are not of closed form, so the slice sampling method should be applied.

Likelihood Statistics

There are many measures that can be used for estimating how well the model fits the data. Some of them have already been discussed in Appendix B but are also included here for the sake of completeness. They fall into three groups. First, there are statistics for indicating the likelihood level of a model, that is, how well the model maximizes the likelihood function. Among these statistics are:

Akaike Information Criterion (AIC)

The AIC is another measure of fit that can be used to assess models. This measure also uses the log-likelihood, but add a penalizing term associated with the number of variables. It is well known that by adding variables, one can improve the fit of models. Thus, the AIC tries to balance the goodness-of-fit versus the inclusion of variables in the model. The AIC is computed as:

$$AIC = -2\ln L + 2p \qquad \text{(C.37)}$$

where p is the number of unknown parameters included in the model (this also includes the inverse dispersion parameter ψ and random spatial effect f_i) and $\ln L$ is the log-likelihood described in Equations C-13 or C-14. Smaller values are better.

Bayes Information Criterion (BIC)

Similar to the AIC, the BIC also employs a penalty term associated with the number of parameters (p) and the sample size (n). This measure is also known as the Schwarz Information Criterion. It is computed the following way:

$$AIC = -2\ln L + p\ln n \tag{C.38}$$

Again, smaller values are better.

Deviance Information Criterion (DIC)

When the Bayesian estimation method is used, the DIC is often used as a goodness-of-fit (GOF) measure instead of the AIC or BIC. The latter ones are generally used for the maximum likelihood method. The DIC is defined as follows:

$$DIC = \hat{D} + 2(\overline{D} - \hat{D}) \tag{C.39}$$

where $\overline{D}$ is the average of the deviance ($-2\ln L$) over the posterior distribution, and $\hat{D}$ is the deviance calculated at the posterior mean parameters. As with the AIC and BIC, the DIC uses $p_D = \overline{D} - \hat{D}$ (effective number of parameters) as a penalty term on the goodness of fit. Differences in DIC from 5-10 indicate that one model is clearly better (Spiegelhalter et al., 2002).

Deviance

The deviance is a measure of goodness of fit that can be used to assess models. It is defined as twice the difference between the maximum likelihood achievable ($y_i = \hat{\mu}_i$) and the likelihood of the fitted model (the ^ refers to the estimate of the variable that is based on the data):

$$D(\mathbf{y}, \mathbf{u}) = 2\left\{L(\mathbf{y}) - L(\hat{\boldsymbol{\mu}})\right\} \tag{C.35}$$

For the NB2 model, the deviance can be computed as:

$$D = 2\sum_{i=1}^{n}\left\{y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - \left(y_i + \psi^{-1}\right)\ln\left[\frac{y_i + \psi^{-1}}{\hat{\mu}_i + \psi^{-1}}\right]\right\} \tag{C.36}$$

Smaller values mean that the model fits the data better.

Pearson Chi-Square

Another useful likelihood statistic is the *Pearson Chi*-square and is defined as

$$Pearson - \chi^2 = \sum_{i=1}^{N}\frac{(y_i - \hat{\mu}_i)^2}{VAR(y_i)} \tag{C.37}$$

If the mean and the variance are properly specified, then $E\left[\sum_{i=1}^{n}\left(y_i-\mu_i\right)^2/VAR\left(y_i\right)\right]=n$ (Cameron and Trivedi, 1998). Values closer to n (the sample size) show a better fit. Recall that the variance for the NB2 model is $VAR\left(y_i\right)=\hat{\mu}_i+\hat{\mu}_i^2/\psi$.

Model Error Estimates

Second, there are statistics for estimating how well the model fit the data and the converse, how much error was in the model. Two error statistics are particularly useful.

Mean Absolute Deviation (MAD)

This criterion has been proposed by Oh et al. (2003) to evaluate the fit of models. The Mean Absolute Deviance (MAD) calculates the absolute difference between the estimated and observed values

$$MAD=\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\mu}_i-y_i\right| \tag{C.38}$$

Mean Squared Prediction Error (MSPE)

The Mean Squared Prediction Error (MSPE) is a traditional indicator of error and calculates the difference between the estimated and observed values squared.

$$MPSE=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mu}_i-y_i\right)^2 \tag{C.39}$$

A value closer to 1 means the model fits the data better.

Over-dispersion Tests

Third, there are statistics for indicating the degree of over-dispersion in the model, including:

Adjusted Deviance

The *adjusted deviance* is defined as the deviance divided by the degrees of freedom (N-K-1). A value closer to 1 indicates a satisfactory GOF. Usually, values greater than 1 indicate signs of over-dispersion, while values below 1 show signs of under-dispersion.

Adjusted Pearson Chi-Square

The *adjusted Pearson Chi-square* is defined as the Pearson Chi-square divided by the degrees of freedom. A value closer to 1 indicates a satisfactory goodness-of-fit.

Dispersion Multiplier

The *dispersion* multiplier, γ, measures the extent to which the conditional variance exceeds the conditional mean (conditional on the independent variables and the intercept term) and is defined by $Var(y_i) = \mu_i + \gamma\mu_i^2$

Inverse Dispersion Multiplier

The *inverse dispersion multiplier* (ψ) is simply the reciprocal of the dispersion multiplier $(\psi = 1/\gamma)$; some users are more familiar with it in this form.

It should be pointed out that many GOF measures are not useful when a single model is evaluated. The measures are therefore relevant when several models are compared with each other (i.e., different functional forms or when different variables are included in the models).

There are other measures that can be used for estimating the goodness-of-fit and the amount of error in models, but are not presented here. Readers can find additional measures in Mitra and Washington (2007) and Lord and Park (2008).

References

Cameron, A.C. & Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press. Cambridge, UK.

Cook, J.D. (2009). Notes on the Negative Binomial Distribution. www.johndcook.com/negative_binomial.pdf (accessed on Aug. 15th, 2010).

De Smith, M., Goodchild, M.F., & Longley, P.A. (2007). *Geospatial Analysis* (second edition). Matador: Leicester, U.K.

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. Journal of the Royal Statistical Society A, Vol. 83, pp. 255-279.

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

Hilbe, J.M. (2007) *Negative Binomial Regression*. Cambridge University Press, Cambridge, U.K.

Lawless, J.F. (1987). Negative binomial and mixed Poisson regression. The Canadian Journal of Statistics, Vol. 15, No. 3, pp. 209-225.

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

Lord, D., & Park, P.Y.J. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-Gamma models on empirical Bayes estimates. Accident Analysis & Prevention, Vol. 40, No. 4, pp. 1441-1457.

References (continued)

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. Accident Analysis & Prevention. Vol. 37, No. 1, pp. 35-46.

Maher M.J., & Summersgill, I. (1996) A Comprehensive Methodology for the Fitting Predictive Accident Models. Accident Analysis & Prevention, Vol. 28, No. 3, pp.281-296.

Miaou, S.-P., & Lord, D. (2003). Modeling TraffiC.Flow Relationships at Signalized Intersections: Dispersion Parameter, Functional Form and Bayes vs Empirical Bayes. Transportation Research Record 1840, pp. 31-40.

Mitra, S., & Washington, S. (2007). On the nature of over-dispersion in motor vehicle crash prediction models. Accident Analysis & Prevention, Vol. 39, No. 3, pp. 459-468.

Oh, J., Lyon, C., Washington, S., Persaud, B., & Bared, J. (2003). Validation of FHWA crash models for rural intersections: lessons learned. Transportation Research Record 1840, pp. 41-49.

Park, B.-J. (2010). Application of the Finite Mixture Models for Vehicle Crash Data Analysis. PhD Dissertation, Texas A&M University, College Station, TX.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B, Vol. 64, No. 4, pp. 583-639.

Sun, D., Tsutakawa, R. K., & Speckman, P. L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions. Biometrika, Vol. 86, No. 2, pp. 341-350.