Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models: A Bayesian Perspective

Dominique Lord
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX
77843-3136
Tel (979) 458-3949
Fax (979) 845-6481
e-mail: d-lord@tamu.edu

Luis F. Miranda-Moreno
Department of Civil Engineering
University of Waterloo
Waterloo, ON, Canada
&
Instituto Mexicano del Transporte,
Qro., Mexico
Phone: (519) 888 4567, ext. 3872
e-mail: lfmirand@uwaterloo.ca

Draft

Paper to be submitted for publication

May 8th, 2006

ABSTRACT

There has been considerable research conducted on the development of statistical models for predicting motor vehicle crashes on highway facilities. Many of these developments were performed for the likelihood-based or frequentist modeling approach. Over the last few years, there has been a significant increase in the application hierarchical Bayes method for modeling motor vehicle crashes. Whether the inferences are estimates using the likelihood-based or the Bayesian method, the most common probabilistic structure used for modeling this type of data remains the traditional Poisson-gamma (or Negative Binomial) distribution. Crash data collected for highway safety studies often have the unusual attributes of being characterized by low sample mean values and, due to the prohibitive costs of collecting data, small sample sizes. Previous studies have shown that the dispersion parameter of Poisson-gamma models can be seriously mis-estimated when the coefficients are estimated using the maximum likelihood method (MLE) for these extreme conditions. Despite important work done on this topic for the MLE, nobody has so far examined how low sample mean values and small sample sizes affect the posterior mean of the dispersion parameter of Poisson-gamma models estimated using the hierarchical Bayes method. The inverse dispersion parameter (posterior mean) plays an important role in various types of highway safety studies, such as building confidence intervals for comparing the safety performance of different highway design alternatives and the application of the empirical Bayes (EB) method for refining the long-term mean of a highway entity, and is particularly important for practitioners who are not familiar with Bayesian methods. It is therefore vital to determine the conditions in which the inverse dispersion parameter may be mis-estimated for this category of models.

To accomplish the objectives of this study, a series of Poisson-gamma distributions are simulated using different values describing the mean, the dispersion parameter, the sample size, and the prior specification. Non-informative and informative prior specifications are tested for determining the magnitude of the biases introduced by low sample mean values and small sample sizes. A series of Poisson-lognormal distributions are also simulated, in the light of recent work done by statisticians on this mixed distribution. The study shows that a dataset characterized by a low sample mean combined with a small sample size can seriously affect the estimation of the posterior mean of the dispersion parameter when a non-informative prior specification is used to characterize the gamma hyper-parameter. The risk of a mis-specified posterior mean can be greatly minimized when an appropriate informative prior distribution is used. Finally, the study shows that Poisson-lognormal models are recommended over Poisson-gamma models whenever crash data characterized by low sample mean values are used for developing crash prediction models.

Keywords: Poisson-gamma, Poisson-lognormal, highway safety, small sample size, low sample mean

INTRODUCTION

There has been considerable research conducted on the development of statistical models for predicting motor vehicle crashes on highway facilities (Abbess et al., 1981; Hauer et al., 1988; Persaud and Dzbik, 1993; Kulmala, 1995; Poch and Mannering, 1996; Lord, 2000; Ivan et al., 2000; Lyon et al., 2003; Miaou and Lord, 2003; Oh et al, 2003; Lord et al., 2005a; Miaou and Song, 2005). Many of these developments were performed for the likelihood-based (referred to as MLE) modeling approach. Over the last few years, there has been a significant increase in the application hierarchical Bayes method for modeling motor vehicle crashes (Qin et al, 2003; Miaou and Lord, 2003; Miaou and Song, 2005). This increase can be attributed to various reasons, including the rediscovery of the Markov Chain Monte Carlo (MCMC) methods by statisticians in the last 15 years (Besag et al., 1995; Gilks et al., 1996; Roberts and Rosenthal, 1998; Robert and Casella, 1999; Carlin and Louis, 2000), the enhancements in desktop computing power, and the availability of Bayesian software programs, such as WinBUGS (Spiegelhalter et al., 2003) and MLwiN (Yang et al., 1999). Whether the inferences are estimated using the likelihood-based or the Bayesian method, the most common probabilistic structure of the models developed for modeling motor vehicle crashes remains the traditional Poisson and Poisson-gamma (or Negative Binomial) distribution.

Crash data have been shown to exhibit over-dispersion, meaning that the variance is greater than the mean. The over-dispersion can be caused by various factors, such as data clustering, unaccounted temporal correlation, model mis-specification, but it has been shown to be mainly attributed to the actual nature of the crash process, namely the fact that crash data are the product of Bernoulli trials with unequal probability of events (this is also known as Poisson trials). Lord et al. (2005b) have reported that as the number of trials increases and becomes very large, the distribution may be approximated by a Poisson process, where the magnitude of the over-dispersion is dependent on the characteristics of the Poisson trials. (Note: the over-dispersion can be minimized using appropriate mean structures of statistical models, as discussed in Miaou and Song, 2005).

Although different mixed-Poisson distributions have been developed to accommodate the over-dispersion (e.g., Poisson-lognormal, Poisson-Inverse gaussian, etc.), the most common distribution used for modeling crash data remains the Poisson-gamma or NB distribution. The Poisson-gamma distribution offers a simple way to accommodate the over-dispersion, especially since its marginal distribution has a close form and this mixture results in a conjugate model (Hauer, 1997). In addition, most statistical software programs, such as SAS (SAS, 2002) or Genstat (Payne, 2000), accommodate NB regression models.

Poisson-gamma models in highway safety applications have been shown to have the following probabilistic structure: the number of crashes at the i-th entity (road section, intersections, etc.) and t-th time period, Y_{it}, when conditional on its mean m_{it}, is assumed to be Poisson distributed and independent over all entities and time periods as:

$$Y_{it} \mid m_{it} \sim Poisson(m_{it}) \qquad i = 1, 2, \ldots, I \text{ and } t = 1, 2, \ldots, T \qquad (1)$$

The mean of the Poisson is structured as:

$$m_{it} = \mu_{it}\, exp(e_{it}) . \tag{2}$$

where,

m_{it} = is the mean for site i and time period t;

$\mu_{it} = f(x_{it}^{'}, \boldsymbol{\beta})$ is a function of the covariate vector ($x_{it}^{'}$);

$\boldsymbol{\beta}$ is a vector of unknown coefficients; and,

e_{it} is a the model error independent of all the covariates.

It is usually assumed that exp(e_{it}) is independent and gamma distributed with a mean equal to 1 and a variance $1/\phi$ for all i and t (with $\phi > 0$). With this characteristic, it can be shown that Y_{it}, conditional on $f(.)$ and ϕ, is distributed as a Poisson-gamma random variable with a mean $f(.)$ and a variance $f(.)(1+f(.)/\phi)$ respectively. (Note: other variance functions exist for Poisson-gamma models, but they are not covered here since they are seldom used in highway safety studies. The reader is referred to Cameron and Trevidi (1998) and Maher and Summersgill (1996) for a description of alternative variance functions.) The probability density function (PDF) of the Poisson-gamma structure described above is given by the following equation:

$$f\left(y_{it};\phi,\mu_{it}\right)=\frac{\Gamma\left(y_{it}+\phi\right)}{\Gamma\left(\phi\right)y_{it}!}\left(\frac{\phi}{\mu_{it}+\phi}\right)^{\phi}\left(\frac{\mu_{it}}{\mu_{it}+\phi}\right)^{y_{it}} \tag{3}$$

Where,

y_{it} = response variable for observation i and time period t;

μ_{it}= mean response for observation i and time period t; and,

ϕ= inverse dispersion parameter of the Poisson-gamma distribution.

The term ϕ is usually defined as the "inverse dispersion parameter" of the Poisson-gamma distribution (note: in the statistical and econometric literature, $\alpha = 1/\phi$ is usually defined as the dispersion parameter; in some published documents, the variable α has also been defined as the “over-dispersion parameter”). Usually the dispersion parameter or its inverse is assumed to be fixed, but recent research in highway safety has shown that the variance structure can potentially be dependent on the covariates (Heydecker and Wu, 2001; Miaou and Lord, 2003; Lord et al., 2005a; Miranda-Moreno et al., 2005; El-Basyouny and Sayed, 2006).

As opposed to data collected in other fields of research, crash data have the uncommon attribute to frequently exhibit a distribution with a low sample mean. Similarly, it is not unusual for researchers and practitioners to develop statistical models using a limited number of observations (or sites) where the data can be collected (see e.g., Lord, 2000;

Oh et al., 2003; Kumara et al., 2003). Small sample sizes are attributed to the prohibitive costs of collecting crash data and other relevant variables (Lord and Bonneson, 2005).

Data characterized by low sample mean values have been sporadically studied in the traffic safety literature. As such, Maycock and Hall (1984) first raised the issue related to the low sample mean. Fridstrøm et al. (1995) further discussed this issue, while Maher and Summersgill (1996) showed how the goodness-of-fit of statistical models could be affected by a low sample mean. They defined this issue as the “low mean problem” (LMP). Subsequent to the identification and its effects on the development of statistical models, Wood (2002) proposed a method to test the fit of MLE models estimated using data characterized with low sample mean values. Lord (2006) reported that the dispersion parameter of Poisson-gamma models estimated from data characterized by low sample mean values and small sample size can be significantly biased (the value is likely to be mis-estimated) and negatively affect analyses commonly performed in highway safety. They include using confidence intervals for comparing the safety performance of different highway design alternatives (Agrawal and Lord, 2006) and the application of the empirical Bayes (EB) method (Hauer, 1997). Despite important work done on this topic, nobody has so far examined how small sample mean values and small sample size affect the dispersion parameter of a Poisson-gamma model estimated using the hierarchical Bayes method. In highway safety studies, the posterior value of the dispersion parameter has many applications as listed above and is particularly important for practitioners who are not familiar with Bayesian methods.

This study is a follow-up to a research project carried out last year on the effects of the LMP and small sample sizes on the estimation of the dispersion parameter of Poisson-gamma models using the MLE method (Lord, 2006). The purpose of this study is two-fold. The first objective seeks to evaluate whether the LMP affects the estimation of the posterior value of the dispersion parameter when the hierarchical Bayes method is used for developing crash prediction models, and if so, to determine the conditions, including the assumptions used for describing the prior, affecting the estimation of Poisson-gamma models. The second objective consists of determining whether better prior specifications or mixed-Poisson models could be used for minimizing biases caused by low sample mean values and small sample sizes for hierarchical Bayes models.

To accomplish the objectives of this study, a series of Poisson-gamma distributions are simulated using different values describing the mean, the dispersion parameter, the sample size, and the prior specification. Non-informative and informative prior specifications are tested for determining the magnitude of the biases introduced by low sample mean values and small sample sizes. A series of Poisson-lognormal distributions are also simulated, in the light of recent work done by statisticians on this mixed distribution. The study shows that a dataset characterized by a low sample mean combined with a small sample size can seriously affect the estimation of the posterior mean of the dispersion parameter when non-informative prior specification is used to characterize the gamma hyper-parameter. The risk of a mis-specified posterior mean can be greatly minimized when an appropriate informative prior specification is used. Finally, the study demonstrates that Poisson-lognormal models are recommended over Poisson-gamma models whenever crash data characterized by low sample mean values are used for developing crash prediction models.

PREVIOUS WORK

The estimation of the dispersion parameter using the MLE approach has been evaluated extensively in various fields, including statistics, econometrics, and biology. It is generally agreed that the first estimator for calculating the dispersion parameter was initially proposed by Fisher (1941). Fisher discussed how the MLE method could be used for estimating the parameter. Following the publication of Fisher, several researchers expanded on his work by either refining the estimation method (Anscombe, 1950; Shenton and Wallington, 1962; Fletcher, 1970; Walsh, 1975; Gourieroux et al., 1984a & 1984b; Cameron and Trivedi, 1986 & 1990; Lawless, 1987) or describing potential biases, such as unstable variance and issues related to small sample sizes (Pieters et al., 1977; Wilson et al., 1984 & 1986; Davidian and Caroll, 1988; Clark and Perry, 1989; Piegorsh, 1990; Dean, 1994). The studies listed above are only a fraction of the published studies done on this topic and, consequently, the reader is referred to Piegorsh (1990), Dean (1994) and Cameron and Trivedi (1998) for additional information on different estimation techniques.

Up until the late 80's, researchers who have worked on small sample size (n) estimation of ϕ or α have usually focused their effort on determining whether ϕ should be estimated directly or, indirectly through its reciprocal $\alpha = 1/\phi$ (Clark and Perry, 1999). Given the outcome of these studies, it is the generally agreed that the dispersion parameter α should be estimated directly rather than its inverse ϕ (via the PDF of a Poisson-gamma function). According to the literature, the MLE estimator of ϕ does not have any formal distribution, since there exists a finite probability that ϕ may not be calculable (Piegorcsh, 1990); this usually occurs for data characterized with under-dispersion. It has also been shown that confidence intervals build for α are continuous and usually more symmetric than for ϕ (Ross and Preece, 1985).

Most recent studies on small sample size estimation techniques have usually focused on estimating potential biases small sample sizes exert on different estimators. The studies have shown that different estimators (among them the MM and MLE) perform well, except when the sample mean is low and the sample size becomes small. When this occurs, many estimators provide a biased estimate of the dispersion parameter (the distribution becomes highly skewed) and has a high probability of being mis-specified.

Among the most significant studies on this topic, Clark and Perry (1989) compared two estimators, the MM and Maximum Quasi-Likelihood method, for different sample sizes (n=10, 20, 30, 50) and sample means (λ=1, 3, 5, 10, 15, 20). Using simulated data, they reported that both estimators become biased when $\lambda \leq 3.0$ and $n < 20$. In addition, under these conditions, the bias becomes more important as $\phi \to \infty$ (if the true value of the inverse of the dispersion parameter is known).

In a follow up study, Piegorsch (1990) examined the MLE estimator and compared the results to the ones of Clark and Perry. Again, using a simulation experiment, the author noted that the ML estimator performed as well as the Quasi-likelihood for large sample sizes. However, Piegorsch reported that the MLE estimator was slightly less accurate for small sample sizes than the Quasi-likelihood.

Dean (1994) evaluated the effects of small sample sizes on the estimation of the dispersion parameter for seven different estimators, including the MLE and MM. She reported that the MLE estimator produced a biased estimate as the sample size decreased and ϕ increased (even for a sample mean equal to 6). In fact, the biased results influenced the standard errors of the coefficients of the models.

Toft et al. (2006) studied the stability of the parameters of the Poisson-gamma model when it is used for modeling micro-organisms that are randomly distributed in a food matrix. These authors also used simulation to test the stability using the MLE method. However, in this case, they examined an alternative parameterization of the Poisson-gamma model commonly used in biology, where the mean and variance functions are defined as $\mu = \tau\nu$ and $\sigma^2 = \tau\nu^2$, respectively. Toft et al. (2006) reported that the parameter estimation becomes unstable when the parameter $\nu \rightarrow 0$ and $n \rightarrow 0$. Indeed, even for $\mu = 10$ and $n = 100$, the MLE method did not provide a reliable estimate of the parameters.

Lord (2006) evaluated the same effects, but used more extreme conditions commonly observed with crash data (i.e, sample means below 1.0). Using simulation, he reported that no matter which estimator, MM, MLE and Weighted Regression (Cameron and Trivedi, 1998), is used, the dispersion parameter is very likely to be mis-estimated when the data are characterized by extreme conditions. In addition, the mis-estimated may be unbeknownst to the analyst performing the regression analysis (based on the output of the statistical software program) and may even affect common analyses performed in highway safety, such as the application of the EB method. Consequently, he suggested minimum sample size requirements for developing crash prediction models as function of the sample mean values for minimizing the risk of mis-estimating the dispersion parameter.

The effects of a small sample size in model development for hierarchical Bayes models have been less extensively investigated. In fact, no documents have been found that specific treated its effects on the dispersion parameter for hierarchical Negative Binomial (HNB) models. Among the few studies related to the issue small sample size, Natarajan and McCulloch (1998) showed using a simulation study that the use of the Gibbs sampler with non-informative proper priors can lead to inaccurate posterior estimates. They found that such inaccuracies are not necessarily limited to small samples, but in fact may occur even for moderate sample sizes. In their study, they consider a probit-normal hierarchy model with a single normally distributed random-effect and a single fixed-effect. In their simulation runs, the number of observations (sample size) ranged from a small (n = 25), to a moderate (n = 50) and a large dataset (n = 100).

Kass and Wassermann (1996) have reported that, in a hierarchical Bayes setting, the differences between alternative prior assumptions can be negligible for large data samples. Unfortunately, they have not examined the differences for small datasets. It is expected that the differences can become important for small sample sizes, where the treatment of prior knowledge may have a large effect on the outcome of the analysis, as discussed below.

In order to explore the effects of small sample and low sample mean values on the posterior value of the dispersion parameter, a sensitivity analysis using different prior specifications was performed in this study. But first, the characteristics of the hierarchical mixed-Poisson models that were used in the simulation study are described in the next section.

HIERARCHICAL MIXED-POISSON MODELS

As described above, the number of crashes occurring at site i during a given time period t (Y_{it}) is assumed to be Poisson distributed with mean m_{it}. A hierarchical modeling framework can be then defined as follows:

$$\begin{aligned}
&\text{(i)} \quad && Y_{it} \mid m_{it} \sim \text{Poisson}(m_{it}) \\
& && \quad \sim \text{Poisson}(\mu_{it} e^{\varepsilon_{it}}) \\
&\text{(i)} \quad && e^{\varepsilon_{it}} \mid \eta \sim \pi_{\varepsilon}(\eta), \\
&\text{(ii)} \quad && \eta \sim \pi_{\eta}(.) \qquad (4)
\end{aligned}$$

This means that a prior distribution π_ε is assumed on the unobserved model error (e^{ε_i}), which depends on hyper-parameter η, with hyper-prior π_η. The third level of randomness assumed on η is one of the differences with the Poisson-gamma model, previously discussed. Moreover, parameters $\mu_{it} = f(x'_{it}, \beta)$ and η are assumed to be mutually independent (Rao, 2003).

Various prior choices can be considered for modeling the parameters $e^{\varepsilon_{it}}$ and η. Depending on the specification of the priors $\pi_\varepsilon(\cdot)$ and $\pi_\eta(\cdot)$, different alternative hierarchical models can be defined. By specifying a gamma prior on $e^{\varepsilon_{it}}$ with shape (a) and scale (b) parameters to be equal, the HNB model is derived with the following assumptions (Miaou and Song, 2005):

$$e^{\varepsilon_{ti}} \mid \phi \sim gamma(\phi, \phi) \text{ and } \phi \sim gamma(a, b). \qquad (5)$$

As in the classical NB model, $e^{\varepsilon_{it}}$ follows a gamma distribution with E[$e^{\varepsilon_{it}}$] = 1 and Var[$e^{\varepsilon_{it}}$] = 1/ϕ. Instead of assuming a gamma distribution as a prior distribution for e^{ε_i}, the lognormal distribution can be used as an alternative PDF. With this prior choice, the hierarchical Poisson-lognormal (HPL) model is derived by assuming a proper hyper-prior for the parameter σ^2 such that (Rao, 2003):

$$\varepsilon_{it} = \log(e^{\varepsilon_{it}}) \mid \sigma^2 \sim Normal(0,\sigma^2) \text{ and } \sigma^{-2} \sim gamma(a,b). \qquad (6)$$

The choice of prior for the parameter σ^2 relies on the fact that a conjugate distribution of the Normal distribution is the Inverse-gamma. Convenient priors are conjugate distributions that produce full conditional posteriors of the same form. Furthermore, the hyper-prior parameters a and b have fixed values and must be specified by statistical modelers. In this paper, whether the choice of prior on e^{ε_i} and these hyper-parameters (a and b) affect the outcome of the analysis are investigated.

POSTERIOR INFERENCE

To give an introduction to posterior inference in the present context, $\mathbf{y} = (y_{1t},\ldots,y_{nt})$ (i=1,…n and t=1,…T) is defined as the vector of observed motor vehicle collisions for n locations for each time t, and y_{it} represents the observed value of Y_{it}. Here, $\mathbf{y}$ is assumed to be generated according to a PDF $f(\mathbf{y} \mid \mathbf{m})$, i.e., Poisson distribution with the mean given by $\mathbf{m} = (m_{1t},\ldots,m_{nt})$ (i=1,…n and t=1,…T). In Bayesian analysis, a prior distribution on the parameters $\mathbf{m}$ needs to be assumed and is denoted as $\pi(\mathbf{m} \mid \eta)$, where η is a vector of hyper-parameters (e.g., ϕ and $\boldsymbol{\beta}$ in the NB model introduced previously). In hierarchical Poisson models, a hyper-prior $\pi(\eta)$ is unspecified for each hyper-parameters η. In other words, the parameters ϕ and $\boldsymbol{\beta}$ in the Poisson-gamma model are assumed to be random. As discussed above, the parameters for the NB model are assumed fixed and estimated using a MLE method. In the present context, the problem with the MLE method is that it does not allow for uncertainty in the estimation process.

In general, if $\pi(\mathbf{m} \mid \eta)$ denotes the chosen prior distribution involving a vector of hyper-parameters η and if $\pi(\eta)$ is the specified hyper-prior, then the joint posterior distribution of all of the parameters given the data $\mathbf{y}$ is derived from the relationship:

$$p(\mathbf{m},\eta \mid \mathbf{y}) \propto f(\mathbf{y} \mid \mathbf{m})\pi(\mathbf{m} \mid \eta)\pi(\eta) \qquad (7)$$

where, $f(\mathbf{y} \mid \mathbf{m})$ is the likelihood of the crash data. When using hierarchical Poisson models, the parameter estimation is often carried out on the basis of the posterior distribution. Given $p(\mathbf{m},\eta \mid \mathbf{y})$, the parameters of interest, θ or η, can be estimated from this posterior distribution. Unfortunately, direct mathematical derivation of $p(\mathbf{m},\eta \mid \mathbf{y})$ usually involves a high-dimensional integration to obtain the constant of proportionality (the normalizing constant) and is not mathematically tractable. In many cases, posterior distributions will not have a known closed form, but rather complicated high dimensional densities, which makes direct inference almost impossible. This problem can be solved by generating a large number of samples from the posterior distribution using MCMC algorithms (Gilks et al., 1996). From these samples, posterior quantities of interest are computed for the model parameters. MCMC algorithms are sophisticated techniques and the reader is referred to Carlin and Louis (2001) and Gelman et al. (2003) for additional details. From a practical point of view, the implementation of MCMC methods proceeds by building the joint posterior density of the model parameters, from which the full conditional distribution for each parameter is derived. This is the conditional distribution of a parameter of interest given the data and all the other parameters in the model. If the full conditional is proportional to a known distribution, the Gibbs–sampling algorithm can be applied. Otherwise, Metropolis–Hastings (M–H) sampling algorithm is implemented.

The implementation of a HNB model is straightforward, however one needs to specify priors on $\pi(\eta)$, which may be informative or "diffuse". Informative priors can be specified based on substantial past experiences, such as rigorous previous studies in similar contexts. However, they are not always available in traffic safety applications and they can be very subjective. Diffuse priors, also called non-informative or flat priors, are designed to reflect lack of information on η. These priors are flat relative to the likelihood and are usually constructed by using proper priors with very large variances. The choice of a diffuse prior is not unique, and some diffuse priors can lead to improper posterior estimates (i.e., posteriors do not integrate to a finite number). It is therefore essential that the chosen diffuse prior $\pi(\eta)$ leads to a proper posterior $\pi(\theta_i \mid \eta)$.

SIMULATION FRAMEWORK FOR THE POISSON-GAMMA MODEL

This section describes characteristics of the simulation framework for the Poisson-gamma models. The simulation was performed using a mixed distribution where the sample mean and the count data were simulated in a step-wise fashion. The Poisson-gamma data were generated using the following steps:

(i) Generate a mean μ_i for observation i given the sample mean λ;

$$\mu_i = \lambda$$

(ii) Generate a discrete count Y_i given the fact that the mean for observation i is gamma distributed with the inverse dispersion parameter ϕ:

$$\begin{aligned} &\mathrm{Y_i} \sim \mathrm{Poisson}(\tilde{\mu}_i) \\ &\tilde{\mu}_i = \mu_i \exp(\varepsilon_i) \\ &\exp(\varepsilon_i) \sim \mathrm{Gamma}(\phi, \phi) \end{aligned}$$

(iii) Repeat steps i and ii "n" times, where n is equal to the sample size.

Note that the steps described above are the same as if one were to simulate data using $\tilde{\mu}_i \sim \text{Gamma}(\phi, \phi/\mu_i)$ with $E(\mu_i) = \mu_i$. The parameterization of the gamma distribution above, $\theta \sim gamma(a,b)$, is used when its mean and variance are defined as $E(\theta) = a/b$ and $Var(\theta) = a/b^2$, respectively. It can be shown that when $E(\theta) = 1$ and $Var(\theta) = 1/\phi$ (where $a = \phi$ and $b = \phi$), the Poisson-gamma function gives rise to a NB distribution with $Var(Y) = \mu + \mu^2/\phi$ (Cameron and Trivedi, 1998).

For this study, the same scenarios as the ones employed by Lord (2006) for evaluating the effects of low sample mean values and small sample size for the MLE method were used. The data were generated using the software R (Venebles et al., 2005), a statistical software very popular among statisticians. The scenarios include the following:

- n = Sample size = **50**, **100** and 1000 observations
- ϕ = Inverse dispersion parameter = 0.5, 1.0 and 2.0
- λ = **0.5**, **1.0** or 10

The values in bold character characterize data subjected to extreme conditions: low sample mean and/or small sample sizes. The other values are used to assess the asymptotic properties of the data for $n \times \lambda \to \infty$. These properties are more significant for the MLE method than for the Bayesian method.

For each combination of sample size, dispersion parameter, and sample mean, the simulation was replicated 200 times. At the end of the replications, the standard statistics, such as the mean, standard deviation, and maximum and minimum values were computed.

Once the data were simulated, the posterior mean of the inverse dispersion parameter ϕ was estimated using the following hierarchical Poisson-gamma model (which can be also seen as a HNB model):

(i) $Y_i \sim \text{Poisson}(\tilde{\mu}_i)$
(ii) $\tilde{\mu}_i = \mu_i \exp(\varepsilon_i)$
(iii) $\exp(\varepsilon_i) \sim \text{Gamma}(\phi, \phi)$
(iv) $\phi \sim \text{Gamma}(a, b)$

In this model, another level of random variability is added to the hierarchical model. That is, ϕ is assumed to be random, following a gamma PDF with parameters "a" and "b". These parameters have to be fixed and they can have a relevant impact under some specific conditions. Here, note also that one could assume different variations of this model. For instance, ϕ could follow an exponential PDF.

All values were estimated using WinBUGS (Spiegelhalter et al., 2003). Usual convergence statistics, such as the MC Error and Rhat, were evaluated as part of the analysis. For the different simulation scenarios note that:

- From each simulation, $\hat{\phi}$ was estimated as the posterior mean of ϕ. That is, $\hat{\phi} = M^{-1}\sum_{j=1}^{M}\phi$, where M is the number of interactions after burn-in interactions.

- For the MCMC simulations, the number of Markov chains = 3, number of total iterations per chain = 10,000, length of burn in = 5,000, number of iterations between saving of results 2,500 for each chain (number of thin = 2). Thus, 7,500 iterations were used for parameter estimations.

SIMULATION RESULTS FOR THE POISSON-GAMMA MODEL

This section presents the results of the simulation analyses for the Poisson-gamma models. The first section presents the results for the non-informative prior distribution. The second section shows the results for the informative prior distribution.

Non-informative prior distribution on ϕ (large variance)

For this part of the analysis, the following parameterization, $\phi \sim gamma(0.01, 0.01)$, was selected as the hyper-prior. This parameterization means that a gamma distribution has a mean equal to 1 and a (large) variance =100 (variability range of ϕ is pretty large and concentrated to 1.0, see e.g., Figure 1). This non-informative prior has been suggested in the highway safety literature (Miaou et al., 2003).

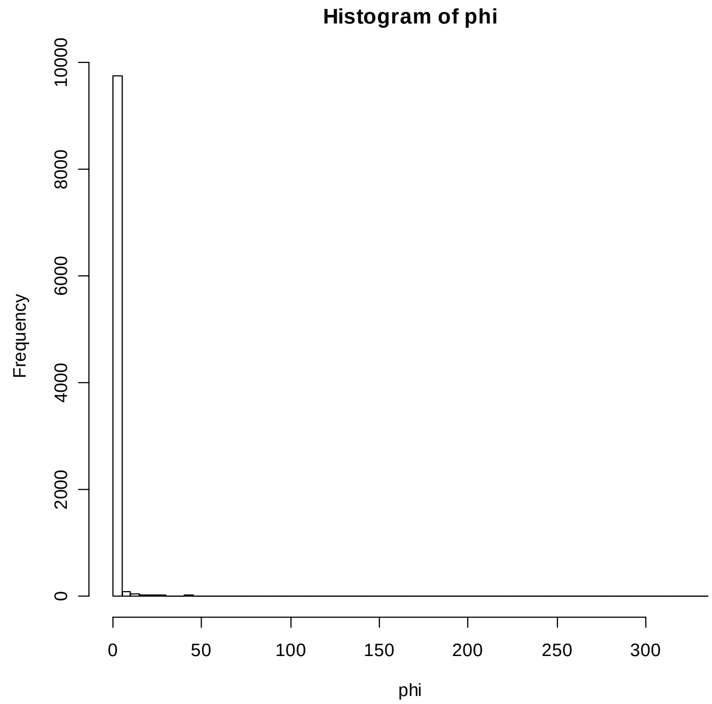

Figure 1. Distribution of $\phi \sim gamma(0.01, 0.01)$, drawing 10000 samples.

The simulation results are presented in Tables 1 to 3 for $\lambda = 10.0$, $\lambda = 1.0$, and $\lambda = 0.5$, respectively. These tables show that both low sample mean values and a small sample size can have important effect on the posterior mean estimate of ϕ. Table 1 shows that when the sample mean is high, the posterior mean is properly estimated even for a sample size equal to 50. It is interesting to observe in Table 2 that even for a large sample size, low sample mean values can affect the posterior estimates of ϕ (the cell in the last row and last column). The simulation results concur with the outcome found for the MLE estimates, in which the probability that the inverse dispersion parameter becomes mis-estimated increases significantly as the sample mean values become lower and the sample size becomes smaller (Clark and Perry, 1989; Piegorsh, 1990; Dean, 1994; Lord, 2006).

TABLE 1. Results of $\hat{\phi}$ for $\lambda = 10.0$

n=50			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.52	1.06	2.22
Stand. Dev.	0.13	0.25	0.66
Max	1.27	2.02	5.09
Min	0.28	0.64	1.01
n=100			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.50	1.04	2.04
Stand. Dev.	0.07	0.16	0.36
Max	0.74	1.40	3.53
Min	0.34	0.69	1.26
n=1000			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.50	1.01	2.01
Stand. Dev.	0.02	0.05	0.11
Max	0.57	1.18	2.44
Min	0.43	0.88	1.71

TABLE 2. Results of $\hat{\phi}$ for $\lambda = 1.0$

n=50			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	1.26	5.50	14.45
Stand. Dev.	2.35	7.37	10.85
Max	17.23	30.55	42.69
Min	0.18	0.38	0.87
n=100			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.59	1.95	7.12
Stand. Dev.	0.23	2.68	8.08
Max	2.18	22.83	40.18
Min	0.24	0.46	0.64
n=1000			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.50	1.05	2.14
Stand. Dev.	0.05	0.13	0.37
Max	0.62	1.46	3.92
Min	0.40	0.77	1.51

TABLE 3. Results of $\hat{\phi}$ for $\lambda = 0.5$

n=50			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	5.04	11.37	18.62
Stand. Dev.	6.69	8.75	9.74
Max	30.67	33.95	39.81
Min	0.10	0.42	0.51
n=100			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	1.28	7.21	17.88
Stand. Dev.	2.41	8.23	12.09
Max	21.75	36.03	54.58
Min	0.21	0.35	0.71
n=1000			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.53	1.06	2.81
Stand. Dev.	0.09	0.15	2.73
Max	0.86	1.62	28.18
Min	0.41	0.69	1.16

In this part of the analysis, the stability of the MCMC simulation for the different scenarios was also examined. The stability was examined using the diagnostic tools, such as the MC error and Rhat, provided by WinBUGS (Spiegelhalter et al., 2003). As expected, the simulation runs for the extreme conditions ($n = 50$ and $\lambda \leq 1.0$) were unstable; on the other hand, the simulation runs for high sample means and/or large sample sizes (see Table 1) were stable. Figure 2 illustrates an example of an unstable MCMC simulation for $n = 50$, $\lambda = 1$, and $\phi = 2$. This figures shows that the MC error is very large indicating that the simulation is erratic. In fact, the graph in the middle confirms the instability of the simulation.

Node statistics

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	7.508	27.26	1.479	21.65%	1.024	7079.00%	3001	9000
deviance	69.11	10.69	0.5499	49.3	69.05	86.2	3001	9000

dic.stats()

DIC

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
y	69.11	60.556	8.554	77.664
total	69.11	60.556	8.554	77.664

history(*,C://history.odc)

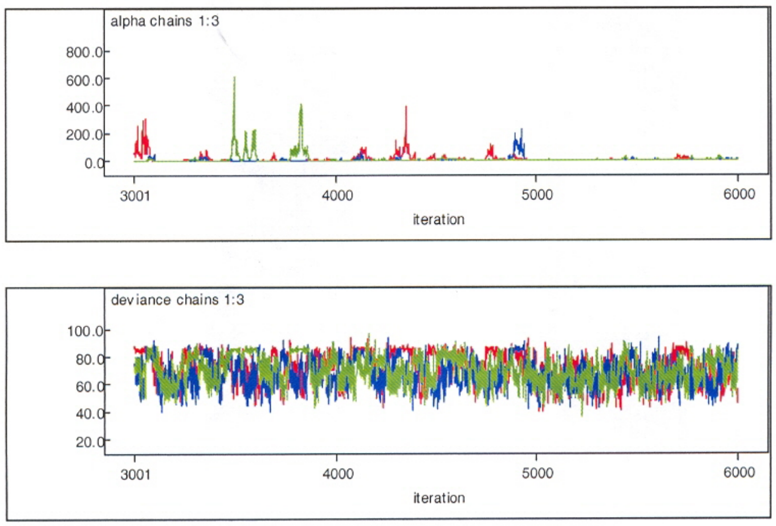

Figure 2. Example of an MCMC output for $n = 50$, $\lambda = 1$, and $\phi = 2$.

Informative prior distribution on ϕ (small variance)

For this part of the analysis, the parameterization of the gamma parameter was the following $\phi \sim gamma(0.1,1.0)$. This parameterization has been proposed by several researchers for hierarchical Poisson-gamma models (e.g., see Carlin and Louis, 2000). With the parameters of the gamma, that is $a = 0.1$ and $b = 1.0$, the variability range of ϕ is small and is concentrated to the lowest categories (see Figure 3).

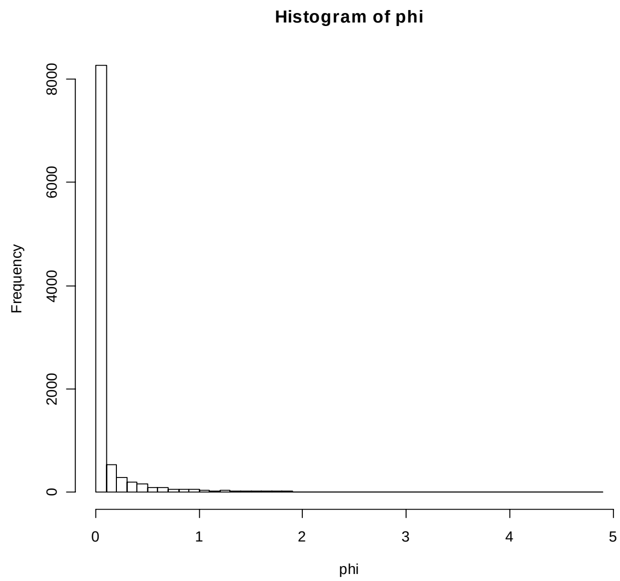

Figure 3. Distribution of $\phi \sim \text{Gamma}(0.1,\ 1.0)$**,**
drawing 10000 samples.

The results of the simulation runs are presented in Tables 4 to 6 for $\lambda = 10.0$, $\lambda = 1.0$, and $\lambda = 0.5$ respectively. The same set of simulation runs were generated for the informative prior distribution. [Note: For a few simulation runs, WinBUGS flagged warning messages. According to the WinBUGS tutorial, this error can be attributed to various reasons, such as: 1) initial values generated from a 'vague' prior distribution that may be numerically extreme; 2) numerically impossible values, such as the log of a non-positive number; and, 3) numerical difficulties in sampling. To circumvent these problems, it is possible to include: better initial values; more informative priors–uniform priors might still be used but with their range restricted to plausible values–; better parameterization to improve orthogonality; or, standardization of covariates to have mean 0 and standard deviation 1.]

TABLE 4. Results of $\hat{\phi}$ for $\lambda = 10.0$

n=50			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.51	1.04	1.91
Stand. Dev.	0.12	0.21	0.45
Max	0.95	1.84	3.64
Min	0.23	0.51	0.98
n=100			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.51	1.01	1.97
Stand. Dev.	0.08	0.15	0.32
Max	0.75	1.53	3.12
Min	0.32	0.70	1.15
n=1000			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.50	1.01	2.00
Stand. Dev.	0.02	0.05	0.11
Max	0.58	1.14	2.31
Min	0.45	0.85	1.76

TABLE 5. Results of $\hat{\phi}$ for $\lambda = 1.0$

n=50			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.57	1.07	1.57
Stand. Dev.	0.23	0.38	0.53
Max	1.38	2.72	3.16
Min	0.16	0.37	0.46
n=100			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.55	1.09	1.75
Stand. Dev.	0.17	0.32	0.48
Max	1.26	2.44	3.22
Min	0.28	0.49	0.84
n=1000			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.50	1.01	2.00
Stand. Dev.	0.05	0.12	0.31
Max	0.66	1.31	3.46
Min	0.40	0.72	1.44

TABLE 6. Results of $\hat{\phi}$ for $\lambda = 0.5$

n=50			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.65	0.92	1.28
Stand. Dev.	0.30	0.35	0.39
Max	1.83	2.06	2.33
Min	0.13	0.26	0.34
n=100			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.54	1.03	1.47
Stand. Dev.	0.22	0.35	0.42
Max	1.38	2.36	3.01
Min	0.20	0.28	0.47
n=1000			
	phi = 0.5	**phi = 1.0**	**phi =2.0**
Mean	0.51	1.01	2.04
Stand. Dev.	0.06	0.17	0.46
Max	0.66	1.61	3.39
Min	0.37	0.70	1.15

The results of the simulation study show that, in most cases, the posterior mean estimate $\hat{\phi}$ is more accurately estimated and is more stable, as determined by the diagnostic tools of WinBUGS (Spiegelhalter et al., 2003) (not shown here). However, there are two exceptions. First, for the scenarios with $\phi = 2.0$, and $\lambda = 0.5$ and $\lambda = 1.0$, the posterior value is significantly underestimated. This is explained by the fact that a is too low for these conditions. To circumvent this problem, it is suggested to use a above 1. Second, for the scenario with $\lambda = 0.5$ and $n = 50$, some numerical difficulties in the sampling occurred for a few simulation runs (the MCMC estimates could not be computed). In this situation, the gamma hyper-prior for the dispersion parameter has a very low expected values for E[inv. disp. Parameter]=0.1 and Var[inv. disp. Parameter]=0.1. Thus, it is possible that this prior specification may generate extremely small values for the dispersion parameter (approximately 0).

SIMULATION FRAMEWORK FOR THE POISSON-LOGNORMAL MODEL

This section describes characteristics of the simulation framework for the HPL models. One objective of this research was to evaluate whether a better mixed-Poisson distribution could be used for modeling crash data characterized by low sample mean values and small sample size. The Poisson-lognormal data were generated using the following steps:

i) Generate a mean μ_i for observation i given the sample mean λ (μ_i is fixed for all i's);

$$\mu_i = \lambda$$

ii) Generate a discrete count Y_i given the fact that the mean for observation i is distributed with a lognormal distribution and variance σ^2:

$$Y_i \sim \text{Poisson}(\tilde{\mu}_i)$$
$$\tilde{\mu}_i = \mu_i \exp(\varepsilon_i)$$
$$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$$

iii) Repeat steps i and ii "n" times, where n is equal to the sample size

It should be pointed out that simulation framework is the same as saying that $\tilde{\mu}_i \sim \text{LogNormal}(\log(\mu), \sigma^2)$. Similarly, note that in this case $E(\tilde{\mu}_i) = \exp(\mu_i + \frac{1}{2}\sigma^2)$, which is different than the $E(\tilde{\mu}_i) = \exp(\mu_i)$ for the Poisson-gamma model. Thus, the mean from the Poisson-gamma will not be the same as for the Poisson-lognormal. For instance, for the same values of μ_i, one can see in the tables below that $E(\tilde{\mu}_i)$ from the Poisson-lognormal is greater than the Poisson-gamma model. For this part of the analysis, only the scenarios describing extreme conditions were used:

- n = Sample size = 50 observations
- σ = Dispersion parameter = 0.56, 0.69 and 0.83, which means that $\text{Var}[\exp(\varepsilon_i)] = 0.5$, 1.0 and 2.0 (as for the Poisson-gamma).
- λ = 0.5

Similar to the Poisson-gamma simulation study, for each combination of sample size, dispersion parameter, and sample mean, the simulation was replicated 200 times. At the end of the replications, the standard statistics, such as the mean, standard deviation, and maximum and minimum values were computed.

Once the data were simulated, the variance of the lognormal distribution σ^2 was estimated using the following hierarchical Poisson-lognormal model:

(i) $Y_i \sim \text{Poisson}(\tilde{\mu}_i)$

(ii) $\tilde{\mu}_i = \mu_i \exp(\varepsilon_i)$

(iii) $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$

$\sigma^{-2} \sim gamma(a, b)$

Note that one can use $\exp(\varepsilon_i) \sim \log \text{Normal}(0, \sigma^2)$. Here, σ^2 is assumed to follow an Inverse-gamma distribution with hyper-parameters a and b. This model has been widely suggested in the literature for mapping of diseases in small area estimation (e.g., see Rao, 2003). One could easily incorporate spatial-time effects for modeling more complex data (see Miaou et al., 2003; Miaou and Song, 2005). This model can also be easily extended to the multivariate hierarchical Poisson model.

SIMULATION RESULTS FOR THE POISSON-LOGNORMAL MODEL

This section presents the results of the simulation analyses for the Poisson-lognormal models. In order to see the effect of using non-informative prior distribution, the following assumption was used: $\sigma^{-2} \sim \text{Gamma}(0.01, 0.01)$. This means that a gamma distribution with mean = 1 and large variance =100. This non-informative prior has been suggested in many applications (e.g. Spiegelhalter et al., 1994 & 2003). The results are summarized in Tables 7 and 8 for $\lambda = 1.0$ and $\lambda = 0.5$, respectively. These tables show that the Poisson-lognormal model is very stable even when the sample size is very small and the sample mean is very low.

TABLE 7. Results of $\hat{\sigma}$ for $\lambda = 1.0$

n=50			
	sigma = 0.56	**sigma = 0.69**	**sigma =83**
Mean	0.48	0.61	0.82
Stand. Dev.	0.17	0.20	0.22
Max	0.95	1.03	1.32
Min	0.20	0.24	0.27

TABLE 8. Results of $\hat{\sigma}$ for $\lambda = 0.5$

n=50			
	sigma = 0.56	**sigma = 0.69**	**sigma =83**
Mean	0.45	0.57	0.71
Stand. Dev.	0.17	0.22	0.27
Max	0.94	1.12	1.37
Min	0.22	0.24	0.24

In order to confirm the stability of the Poisson-lognormal model detailed in the tables above, this model was fitted using the output of the simulation run for the Poisson-gamma (that is the count distribution used in Tables 2 and 3) for the following conditions:

λ = 0.5, 1.0
n = Sample size = 50 observations
ϕ = Inverse dispersion parameter = 0.5, 1.0 and 2.0

It is important to keep in mind at this point that the Poisson-lognormal and Poisson-Gamma models cannot be directly compared, because the variance structure between the two models is very different. The goal of this exercise was to examine the stability of the Poisson-lognormal model when it is estimated using extreme values generated by a different distribution. The results of the Poisson-lognormal model is shown in Tables 9 and 10 for $\lambda = 1.0$ and $\lambda = 0.5$, respectively.

TABLE 9. Results of $\hat{\sigma}$ and $\lambda = 1.0$

n=50			
	phi = 0.5	**phi = 1.0**	**phi = 2.0**
Mean	0.88	0.71	0.45
Stand. Dev.	0.28	0.19	0.15
Max	1.53	1.12	0.89
Min	0.36	0.39	0.23

TABLE 10. Results of $\hat{\sigma}$ and $\lambda = 0.5$

n=50			
	phi = 0.5	**phi = 1.0**	**phi = 2.0**
Mean	0.70	0.45	0.37
Stand. Dev.	0.26	0.18	0.11
Max	1.11	0.99	0.72
Min	0.27	0.23	0.26

Tables 9 and 10 show that the posterior mean value of the dispersion parameter of the Poisson-lognormal model is still very stable despite the fact that, using the same count distribution, the posterior mean of the dispersion parameter of the Poisson-gamma model was not.

DISCUSSION

The results of the analysis presented above raise a few important issues that merit further discussion. First, the overall results of this study show that a Poisson-gamma model estimated using a hierarchical Bayes framework is as affected by a small sample size and low sample mean values as models estimated using the MLE approach. This problem is particularly significant when a non-informative prior is used for characterizing the gamma hyper-parameter of the model. However, the mis-estimation (or bias) of the inverse dispersion parameter starts occurring at a lower sample mean value and smaller sample size than that found for the MLE (Lord, 2006). This outcome suggests that a transportation safety analyst may still be better off, in general, to estimate the coefficients of crash prediction models using a hierarchical Bayes framework rather than the MLE when the data are characterized by low sample mean values and small sample size. Obviously, the posterior mean value is more likely to be properly estimated if the analyst uses an informative prior, as shown in Tables 4 to 6. It is important to point out that the Baysian modeling framework has not been used often in highway safety studies.

However, over the last few years, more researchers are starting to use this framework for modeling motor vehicle crashes (e.g., see Qin et al, 2003; Miaou and Lord, 2003; Miranda-Morenon et al., 2005; Miaou and Song, 2006 among others).

Second, this simulation study demonstrates that the use of the MCMC sampling methods under diffuse priors (or non-informative) can provide inaccurate posterior estimates. Such inaccuracies are not necessarily limited to low sample mean values, but may in fact be present under the small sample sizes. Considering alternative prior specifications for the inverse dispersion parameter of the Poisson-gamma models, it was shown that the MCMC estimates either failed to reproduce direct Bayes on average, or suffer from unacceptably large simulation variances, even for reasonably a large burn-in period. This occurred in all the cases investigated: $\phi \sim gamma(a,b)$ where $a = 0.01$, $b = 0.01$ and $a = 0.001$, $b = 0.001$. Better results were obtained when assuming a small variance on ϕ (what we call it "semi-informative prior"). This is the case for the hyper-parameter where $a = 0.1$ and $b = 1.0$. Obviously, the largest variance assumed on ϕ, the more inaccurate posterior estimates will be under the mentioned critical conditions.

The third issue is related to the source for the instability of Poisson-gamma models, particularly for the estimation of the inverse dispersion parameter. This instability can be attributed to the fact that the gamma prior distribution is sharply peaked near zero when $a = b \to 0$, distorting posterior inferences. Based on this study, it is not recommended to use the $gamma(a,b)$ family of non-informative prior distributions (e.g., $a = b = 0.01, 0.001$, etc.) because, as discussed in previously where $\lambda \to 0$ and $n \to 0$, the resulting inferences can be very sensitive to the choice of the hyper-parameters a and b.

For the fourth issue, given the instability of Poisson-gamma models, alternative prior specifications should be explored for modeling motor vehicle collisions subjected to low sample mean values and a small sample size. One potential solution to the noted problem can be the use of informative priors or other prior distributors for the inverse dispersion parameter. Some researchers have, in fact, looked into alternative model prior distributions. For instance, Christiansen and Morris (1997) have suggested a prior distribution on ϕ that is given by $p(\phi)= z_0/(z_0+\phi)^2$. The case considered was the one proposed by Aul and Davis (2006), in which the hyper-parameter $z_0 = 1$. Although not shown in this study, the hierarchical Poisson-gamma model suggested by these authors was also investigated. From the analysis, the MCMC estimates for this choice had the same problems under the critical conditions described in this research. In other words, the use of such prior did not provide stable results. Nonetheless, further work is needed to evaluate alternative prior specifications (e.g., see Gelman, 2005).

Fifth, when working with a flat distribution, the Poisson-lognormal model always performs better than the Poisson-gamma model. The stability of this model can be attributed to the fact that the Inverse-gamma prior assumed for the parameter σ^2 is a conjugate distribution of the Normal (Gaussian) distribution. Note that this convenient conjugate prior distribution produces a full conditional posterior of the same form (Roa, 2003). As mention by Winkelmann (2003), the Poisson-lognormal model has a natural interpretation since ε_i can capture the effect of several additive omitted variables. This can perform better than its competitors when there are many omitted (independent) site-attributes. In this situation, establishing normality of ε_i can be more robust.

Another related issue is that the output of the Poisson-lognormal (the dispersion parameter) cannot be easily used by practitioners. In this situation, the measure of dispersion is the parameter "σ" and cannot be used directly with the EB method proposed by Hauer (1997). This form of the EB uses the inverse dispersion parameter of the Poisson-gamma model. The equation is defined the following way:

$$\hat{\hat{\mu}}_i = \gamma_i \hat{\mu}_i + (1-\gamma_i) y_i \tag{8}$$

Where,

$\hat{\hat{\mu}}_i$ = the EB estimate of the expected number of crashes per year for site i;

$\hat{\mu}_i$ = the ML estimate produced from a Poisson-gamma model fitted using the reference population for site i (crashes per year);

$\gamma_i = \dfrac{1}{1+\hat{\mu}_i/\phi}$, the weight factor estimated as a function of the ML estimate and the inverse dispersion parameter; and

y_i = the observed number of crashes per year at site i.

As it can be seen above, the inverse dispersion parameter plays an important role for estimating the weight factor.

There has been a proposed EB estimator for the Poisson-lognormal model. Unfortunately, this estimator is an approximation of the posterior mean and does not have a closed form, which makes its estimate difficult to compute. In highway safety, Miranda-Moreno et al. (2005) used this estimator for identifying hazardous railway crossings. Given the limited research on this topic, further work is needed on the application of the EB approach using a Poisson-lognormal model, especially for cases when this model is applied for before-after studies. The reader is referred to Clayton and Kaldor (1987), Sohn (1994), Rao (2003), and Meza (2003) for more details.

It should be pointed out that when working with hierarchical Bayes models combined with the MCMC approach, one does not have to worry about two-step EB process: 1) estimating model parameters by MLE and 2) making posterior inference. Remember that in the EB approach, the unknown vector of model parameters (e.g. $\boldsymbol{\beta}, \phi$) are replaced by suitable MLE estimates ($\hat{\boldsymbol{\beta}}, \hat{\phi}$). The problem with the EB is that it does not make allowance for uncertainty in estimating $\hat{\boldsymbol{\beta}}, \hat{\phi}$ (model parameters are simply replaced by their estimates assuming these are error free). When working with hierarchical Bayes models under a MCMC approach, uncertainty in the parameters is introduced by introducing a full hyper-prior framework. Posterior distributions are then directly simulated using MCMC algorithms. This allows drawing directly the whole posterior densities of parameter of interest. From the posterior analysis, practitioners can obtain direct results for estimating accident modification factors (AMFs) from before-after studies (see Aul and Davis, 2006) and hazardous locations in the hotspot identification activity (see Miaou and Song, 2005).

The last issue is related to the recommended minimum sample size for developing models subjected to low sample mean values and small sample size using hierarchical Poisson-gamma models. This recommendation is for properly estimating the posterior mean of the inverse dispersion parameter; note that this also applies to the posterior mean of the coefficients as well. For non-information priors, it is recommended to use the values shown in Table 11. These values were confirmed using additional simulation work not shown here. For informative prior and Poisson-lognormal model, the recommended minimum sample should be 50 observations for sample mean values below 1.00.

Table 11. Recommended Minimum Sample Size† to Minimize an Unreliably Estimated Dispersion Parameter

Population Sample Mean (λ)	Minimum Sample Size
≥2.00	20
1.00	100
0.75	500
0.50	1,000
0.25	3,000

†The sample size refers to the number of observations, e.g. intersections or segments, in the data. It does not reflect the number of collisions collected at the sites that are part of the sample.

As discussed above, cases with limited-data are not rare, but rather quite common in highway safety, especially in the evaluation of countermeasures (before-after studies). In this context, there is a lack of knowledge on how to fix the hyper-parameters for prior distributions. As a practical solution, statisticians have adopted the use of diffuse proper priors in the unknown parameters, e.g. probability distributions with a very large spread. Despite the fact that this approach can work in many situations, introducing diffuse proper priors on all unknown parameters can be risky or dangerous under critical conditions. Transportation safety analysts should be warm about this issues and further research on how to integrate prior knowledge and past experiences in the modeling process deserve more investigation.

SUMMARY AND CONCLUSIONS

The objectives of this study sought to verify whether a small sample size and low sample mean values affect the estimation of the posterior mean of the dispersion parameter when the HNB model is used for developing crash prediction models, and determine whether better prior specifications or alternative mixed-Poisson models could be employed for minimizing biases caused by these extreme conditions. This paper was motivated by the fact that crash databases are often characterized by low sample mean values and/or a small sample size. Previous research has shown that low sample mean values can significantly affect the development of Poisson-gamma models, particularly the inverse dispersion parameter, estimated using the MLE method. Given the importance the dispersion parameter plays in various types of highway safety studies, there is a need to determine the circumstances in which these extreme conditions affect the posterior mean of the dispersion parameter of HNB models.

To accomplish the objectives of this study, a series of Poisson-gamma distributions were simulated using different values describing the mean, the dispersion parameter, the sample size, and the prior distribution. Non-informative and informative prior specifications were tested to determine the biases introduced by low sample mean values and small sample size. With their increase popularity, a series of Poisson-lognormal distributions were also simulated.

Two main conclusions are drawn from this research. First, crash data characterized by a low sample mean combined with a small sample size can seriously affect the estimation of the posterior mean of the inverse dispersion parameter of HNB models. The problem is very significant when a non-informative or diffuse prior is used to characterize the gamma hyper-parameter. On the other hand, using an informative prior for the same parameter reduces the likelihood of a mis-estimated posterior value.

Second, when crash data are characterized by low sample mean values and a small sample size, Poisson-lognormal models offer a better alternative than Poisson-gamma models. The simulation results have shown that these models are more stable for the extreme conditions described in this research. This type of model is however not useful for practitioners who are interested in applying the EB method commonly used in highway safety. In conclusion, it is hoped that this paper offers the necessary information for transportation safety modelers, who are frequently bound to use crash data characterized by low sample mean values and small sample sizes, about selecting the appropriate prior specifications or alternative mixed-Poisson models.

ACKNOWLEDGEMENTS

REFERENCES

Abdel-Aty, M., and M.F. Addella (2004) Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes: Generalized Estimating Equations for Correlated Data. Transportation Research Record 1897, pp. 106-115.

Abbess, C., D. Jarett, and C.C. Wright (1981) Accidents at Blackspots: estimating the Effectiveness of Remedial Treatment, With Special Reference to the "Regression-to-Mean" Effect. Traffic Engineering and Control, Vol. 22, No. 10, pp. 535-542.

Agrawal, R., and D. Lord (2006) Effects of Sample Size on the Goodness-of-fit Statistic and Confidence Intervals of Crash Prediction Models Subjected to low Sample Mean Values. Transportation Research Record, in press.

Anscombe, F.J. (1950) Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. Biometrika, Vol. 36, pp. 358-382.

Aul, N. and G.A. Davis (2006) Using Propensity Score Matching Method and Hybrid Bayesian Method to Estimate Crash Modification Factors of Signal Installation. Paper presented at the 85th Annual Meeting of the Transportation Research Board. Washington, D.C.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995) Bayesian Computation and Stochastic Systems (with Discussion). Statistical Science, Vol. 10, pp. 3-66.

Cameron, A.C., and P.K. Trivedi (1986) Econometric Models based on Count Data: Comparisons and Applications of Some Estimators and Tests. Journal of Applied Econometrics, Vol. 1, pp. 29-53.

Cameron, A.C., and P.K. Trivedi (1990) Regression-Based Tests for Overdispersion in the Poisson Model. Journal of Econometrics, Vol. 46, pp. 347-364.

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

Carlin, B. and T. Louis. (2001) *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed. Chapman and Hall.

Christiansen, C., and C. Morris. (1997). Hierarchical Poisson regression models. Journal of the American Statistical Association, Vol. 92, pp. 618–632.

Clark, S.J., and J.N. Perry (1989) Estimation of the Negative Binomial Parameter κ by Maximum Quasi-Likelihood. Biometrics, Vol. 45, pp. 309-316.

Clayton and Kaldor (1987). Empirical Bayes Estimates of Age-Standardized Relative Risk for use in Disease Mapping. Biometrics, Vol. 43, No. 3, pp. 671-681.

Congdon, C. (2001) *Bayesian Statistical Modelling*. John Wiley and Sons.

Davidian, M., and R.J. Carroll (1987) A note on Extended Quasi-Likelihood. Journal of the Royal Statistical Society: Series B, Vol. 82, pp. 1079-1091.

Davis, G.A. (2000) Accident Reduction Factors and Causal Inference in Traffic safety Studies. Accident Analysis & Prevention, Vol. 32, No. 1, pp. 95-109.

Dean, C.B. (1994) Modified Pseudo-Likelihood Estimator of the Overdispersion Parameter in Poisson Mixture Models. Journal of Applied Statistics, Vol. 21, No. 6, pp. 523-532.

El-Basyouny K., and T. Sayed (2006) Comparison of two Negative Binomial Regression Techniques in Developing Accident Prediction Models. Presented at the 85th Annual Meeting of the Transportation Research Board, Washington, D.C.

Federal Highway Administration (2003) Module 1 – Network Screening. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites. U.S. Department of Transportation, Washington D.C., (web page: http://www.safetyanalyst.org/whitepapers/module1.pdf accessed on June 10th, 2005)

Fisher R.A. (1941) The Negative Binomial Distribution. Annals of Eugenics. Vol. 11, pp. 182-187.

Fridstrøm, L., J. Ifver, S. Ingebrigtsen, S., R. Kulmala, and L.K. Thomsen (1995) Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Analysis and Prevention, Vol. 27, No. 1, pp. 1–20.

Gelman A. (2005). Prior distributions for variance parameters in hierarchical Models. Bayesian Analysis. Vol. 1, No. 2, pp. 1–19.

Gelman A., A., J.B. Carlin, H.S. Stern, and D.B. Rubin. (2003) *Bayesian Data Analysis*. 2nd Edition. Chapman & Hall/CRC Press, New York, NY.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter. (1996) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, U.K.

Gourieroux, C., A. Monfort, and A. Trognon (1984a) Pseudo Maximum Likelihood Methods: Theory. Econometrika, Vol. 52, pp. 681-700.

Gourieroux, C., A. Monfort, and A. Trognon (1984b) Pseudo Maximum Likelihood Methods: Application. Econometrika, Vol. 52, pp. 701-720.

Gourieroux, C., and M. Visser (1986) A Count Data Model with Unobserved Heterogeneity. Journal of Econometrics, Vol. 79, pp. 247-268.

Hauer, E. (1996) Identification of Sites with Promise. Transportation Research Record 1542, pp. 54-60.

Hauer, E. (1997) *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Elsevier Science Ltd, Oxford.

Hauer, E., F.M. Council, and Y. Mohammedshah (2004) Safety Models for Urban Four-Lane Undivided Road Segments. Transportation Research Record 1897, pp. 96-105.

Hauer, E., J.C.N. Ng, and J. Lovell (1988) Estimation of Safety at Signalized Intersections. Transportation Research Record 1185, pp. 48-61.

Hauer, E., and B.P. Persaud (1984) Problem of Identifying Hazardous Locations using Accident Data. Transportation Research Record 975, pp. 36-43.

Heydecker, B.G., J. Wu (2001) Identification of Sites for Road Accident Remedial Work by Bayesien Statistical Methods: An Example of Uncertain Inference. Advances in Engineering Software, Vol. 32, pp. 859-869.

Ivan, J.N., C. Wang, and N.R. Bernardo (2000) Explaining Two-Lane Highway Crash Rates Using Land Use and Hourly Exposure. Accident Analysis & Prevention, Vol. 32, No. 6, pp. 787-795.

Kass, R.E., and Wassermann, L. (1996), Selecting Prior Distributions by Formal Rules. Journal of the American Statistical Association, 91, 1343–1370.

Kononov, J., and B.K. Allery (2004) Explicit Consideration of Safety in Transportation Planning and Project Scoping. Transportation Research Record 1897, pp. 116-125.

Kulmala, R. (1995) Safety at Rural Three- and Four-Arm Junctions: Development and Applications of Accident Prediction Models. VTT Publications 233, Technical Research Centre of Finland, Espoo.

Kumara, S.S.P., and H.C. Chin (2004) Study of Fatal Traffic Accidents in Asia Pacific Countries. Transportation Research Record 1897, pp. 43-47.

Kumara, S.S.P., H.C. Chin, and W.M.S.B. Weerakoon (2003) Identification of Accident Causal Factors and Prediction of Hazardousness of Intersection Approaches. Transportation Research Record 1840, pp. 116-122.

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

Lee, F. (1986) Specification Test for Poisson Regression Models. International Economic Review, Vol. 27, pp. 689-706.

Lord, D. (2000) The Prediction of Accidents on Digital Networks: Characteristics and Issues Related to the Application of Accident Prediction Models. Ph.D. Dissertation. Department of Civil Engineering, University of Toronto, Toronto, Ontario.

Lord, D. (2006) Modeling Motor Vehicle Crashes using Poisson-gamma Models: Examining the Effects of Low Sample MeanValues and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. Accident Analysis & Prevention. In press.

Lord, D., and J.A. Bonneson (2005) Calibration of Predictive Models for Estimating the Safety of Ramp Design Configurations. Transportation Research Record 1908, pp. 88-95.

Lord, D., A. Manar, and A. Vizioli (2005a) Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio for Rural and Urban Freeway Segments. Accident Analysis & Prevention. Vol. 37, No. 1, pp. 185-199.

Lord, D., and B.N. Persaud (2000) Accident Prediction Models with and without Trend: Application of the Generalized Estimating Equations Procedure. Transportation Research Record 1717, 102-108.

Lord, D., and B.N. Persaud (2004) Estimating the Safety Performance of Urban Transportation Networks. Accident Analysis & Prevention. Vol. 36, No. 2, pp. 609-620.

Lord, D., S.P. Washington, and J.N. Ivan (2005b) 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.

Lyon, C., J. Oh, B.N. Persaud, S.P. Washington, and J. Bared. (2003) Empirical Investigation of the IHSDM Accident Prediction Algorithm for Rural Intersections. Transportation Research Record 1840. pp. 78-86.

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

Maycock G., and R.D. Hall (1984) Accidents at 4-arm roundabouts. TRRL Laboratory Report 1120. Transportation and Road Research Laboratory, Crowthorne, Bershire.

McCullagh P., and J.A. Nelder (1989) *Generalized Linear Models*, 2nd Ed. Chapman and Hall, London, U.K.

Meza, J. (2003) Empirical Bayes Estimation Smoothing of Relative Risk in Disease Mapping. Journal of Statistical Planning and Inference, Vol. 112, pp. 43-62.

Miaou, S.-P., and D. Lord (2003) Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes. Transportation Research Record 1840, pp. 31-40.

Miaou, S.-P., and J.J. Song (2005) Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion and spatial dependence. Accident Analysis and Prevention, Vol. 37, No. 4, pp. 699-720.

Miaou, S.P., J.J. Song and B.K. Mallick (2003) Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. Journal of Transportation and Statistics. Vol. 6, No. 1, pp. 33-57.

Miranda-Moreno, L.F., L. Fu, F.F. Saccomano, and A. Labbe (2005) Alternative Risk Model for Ranking Locations for Safety Improvement. Transportation Research Record 1908, pp. 1-8.

Morris, C.N. (1997) Fitting Hierarchical Models. Workshop on Statistics and Epidemiology: Environment and Health, Minneapolis, MN. (web page: http://www.ima.umn.edu/summerstat/week6.html#wk6tue accessed on January 3rd, 2006)

Myers, R.H., D.C. Montgomery, and G.G. Vining (2002) Generalized Linear Models: with Applications in Engineering and the Sciences. Wiley Publishing Co., New York, N.Y.

Natarajan, R., and McCulloch, C. (1998). Gibbs Sampling with Diffuse Proper Priors: A Valid Approach to Data Driven Inference? Journal of Computational and Graphical Statistics 7, 3, pp 267- 277.

Noland, R.B., and M.A. Quddus (2004) Analysis of Pedestrian and Bycicle Casualties with Regional Panel Data. Transportation Research Record 1897, pp. 43-47.

Oh, J., C. Lyon, S.P. Washington, B.N. Persaud, and J. Bared (2003) Validation of the FHWA Crash Models for Rural Intersections: Lessons Learned. Transportation Research Record 1840. pp 41-49.

Payne, R.W. (ed.) (2000) The Guide to Genstat. Lawes Agricultural Trust, Rothamsted Experimental Station, Oxford, U.K.

Persaud, B.N., and L. Dzbik. (1993) Accident Prediction Models for Freeways. Transportation Research Record 1401, pp. 55-60.

Persaud, B.N., G. Bahar, C.J. Mollett, and C. Lyon (2004) Safety Evaluation of Permanent Raised Snow-Plowable Pavement Markers. Transportation Research Record 1897, pp. 148-163.

Persaud, B.N., D. Lord, and J. Palminaso (2002) Issues of Calibration and Transferability in Developing Accident Prediction Models for Urban Intersections. Transportation Research Record 1784, pp. 57-64.

Piegorsch, W.W. (1990) Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. Biometrics, Vol. 46, pp. 863-867.

Pieters, E.P., C.E. Gates, J.H. Martin, and W.L. Sterling (1977) Small-Sample Comparison of Different Estimators of Negative Binomial Parameters. Biometrics, Vol. 33, pp. 718-723.

Poch, M., and F.L. Mannering (1996) Negative Binomial Analysis Of Intersection-Accident Frequency. Journal of Transportation Engineering, Vol. 122, No. 2, 105-113.

Qin, X., J.N. Ivan, and N. Ravishankar (2004) Selecting Exposure Measures in Crash Rate Prediction for Two-Lane Highway Segments. Accident Analysis & Prevention, Vol. 36, No. 2, pp. 183-191.

Rao, J. (2003) *Small Area Estimation*. John Wiley and Sons.

Robert, C.P. and G. Casella. (1999) Monte Carlo Statistical Methods. Springer-Verlag, New York, N.Y.

Roberts, G.O. and J.S. Rosenthal (1998). Markov chain Monte Carlo: Some Practical Implications of Theoretical Results. Canadian Journal of Statistics. Vol. 26, pp. 5-31.

Ross, G.J.S., and D.A. Preece (1985) The Negative Binomial Distribution. The Statistician, Vol. 34, pp. 323-661.

SAS Institute Inc. (2002) Version 9 of the SAS System for Windows. Cary, NC.

Shenton, L.R., and P.A. Wallington (1962) The Bias of Moment Estimators with an Application to the Negative Binomial. Biometrika, Vol. 193-204.

Sohn, S.Y. (1994) A Comparative Study of Four Estimators for Analyzing the Random Event Rate of the Poisson Process. Journal of Statistical Computation and Simulation. Vol. 49, pp. 1-10.

Spiegelhalter, D.J., A. Thomas, N.G. Best, and D. Lun. (2003) WinBUGS Version 1.4.1 User Manual. MRC Biostatistics Unit, Cambridge. Available from http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.

Toft, N., G.T. Innocent, D.J. Mellor, and S.W.J. Reid (2006) The Gamma-Poisson Model as a Statistical Method to determine if Micro-Organisms are Randomly Distributed in a Food Matrix. Food Microbiology, Vol. 23, No. 1, pp. 90-94.

Venables, W. N., D. M. Smith, and the R Development Team. (2005) An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. Insightful Corporation, Seattle, WA.

Willson, L.J., J.L. Folks, and J.H. Young (1984) Multistage Estimation Compared with Sample-Size Estimation of the Negative Binomial κ. Biometrics, Vol. 40, pp. 109-117.

Willson, L.J., J.L. Folks, and J.H. Young (1984) Complete Suffeciency and Maximum Likelihood Estimation for the Two-Parameter Negative Binomial Distribution. Metrika, Vol. 33, pp. 349-362.

Winkelmann, R. (2003) Econometric *Analysis of Count Data*, 4th Edition, Springer, Heidelberg, New York.

Wood, G.R. (2002) Generalised Linear Accident Models and Goodness of Fit Testing. Accident Analysis & Prevention, Vol. 34, No. 4, pp. 417-427.

Yang, M., J. Rasbash, H. Goldstein, and M. Barbosa (1999) MLwiN Macros for Advanced Multilevel Modelling. Version 2.0. Multilevel Models Project, Institute of Education, University of London, U.K. (http://www.ioe.ac.uk/multilevel/)