FULL BAYESIAN POISSON-HIERARCHICAL MODELS FOR CRASH DATA ANALYSIS: INVESTIGATING THE IMPACT OF MODEL CHOICE ON SITE-SPECIFIC PREDICTIONS

A Dissertation

by

SEYED HADI KHAZRAEE KHOSHROOZI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Chair of Committee,	Dominique Lord
Committee Members,	Valen Johnson
	Yunlong Zhang
	Luca Quadrifoglio
Head of Department,	Robin Autenrieth

August 2016

Major Subject: Civil Engineering

ABSTRACT

The Poisson-gamma (PG) and Poisson-lognormal (PLN) regression models are among the most popular means for motor-vehicle crash data analysis. Both models belong to the Poisson-hierarchical family of models, which provides a straightforward framework for interpretation of parameters. Over the last two decades, highway safety researchers have increasingly favored a full Bayesian approach to estimation of Poisson-hierarchical models due to its theoretical and computational advantages. While numerous studies have compared the overall performance of alternative Bayesian Poisson-hierarchical models, little research has addressed the impact of model choice on the expected crash frequency prediction at individual sites. This dissertation takes a microscopic approach to comparing the models' predictions and strives to identify possible trends e.g., that an alternative model's prediction for sites with certain conditions tends to be higher (or lower) than that from another model. The practical importance of such trends is reflected most clearly when alternative models are utilized to identify hazardous highway sites (e.g., roadway segments, intersection, etc.) by ranking the sites with respect to their expected crash frequency.

In addition to the PG and PLN models, this research formulates a new member of the Poisson-hierarchical family of models: the Poisson-inverse gamma (PIGam). The PIGam model was of special interest because of the heavy tail of the inverse gamma distribution and the conjectured potential of the PIGam model in dealing with highly over-dispersed data. Four field datasets (from Toronto, Texas, Michigan and Indiana) covering a wide range of over-dispersion characteristics were selected for analysis.

This study discovered that the disparities between the alternative models predictions are mainly associated with the sites where the observed crash frequency is significantly larger or smaller than expected for a site with similar traffic and physical characteristics.

For both scenarios, it was demonstrated that the PIGam model tends to predict a higher expectation for crash frequency than would the PLN and PG models, in order. In consequence, sites with unusually high number of observed crashes are likely to be ranked higher (in terms of expected crash frequency) when the PIGam model is used instead of the PLN model, and similarly when the PLN model is used instead of the PG model.

Furthermore, the disparities between alternative model predictions were found to be even more important when the calibrated models were applied to predict crash frequency at sites with no observed crash count. For all four datasets, the PIGam model tended to predict higher expected crash frequencies than did the PLN and PG models, in order.

Finally, a comparison between the models goodness-of-fit using the deviance information criterion (DIC) refuted the conjecture that models with heavy-tailed distributions will certainly perform better as the data become more over-dispersed. The author believes that the relative goodness-of-fit of alternative models to a given dataset is too complicated to be reliably predicted before actually fitting the models. However, the study demonstrated that models with similar measures of goodness-of-fit may predict considerably different crash frequencies at individual sites. This dissertation identified the relationships between alternative models' predictions at individual sites and described the resulting practical implications of choosing one model over another.

DEDICATION

Dedicated to Zahra, the love of my life

ACKNOWLEDGEMENTS

I shall first thank my dedicated advisor, Professor Dominique Lord, for his tremendous assistance and support throughout this dissertation and my PhD degree. He has been my academic role model and I could not possibly have attained this level of knowledge and experience without his guidance and inspiration.

I should extend my appreciation and gratitude to my mentor from the Department of Statistics, Professor Valen Johnson. His Bayesian statistics course encouraged me to focus my PhD research on Bayesian hierarchical models. In every step of this dissertation, I had the privilege of having numerous meetings with him and benefiting from his precious advice.

I would also like to thank Dr. Yunlong Zhang and Dr. Luca Quadrifoglio for serving in my dissertation committee and providing insightful advice during my preliminary and final defense sessions.

Finally, I acknowledge my wife's support throughout my PhD degree. Without Zahra's encouragement and patience, this dissertation would have never been completed.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

CHAPTER I
INTRODUCTION

This chapter introduces the research problem, lists the objectives, and outlines the structure of this dissertation.

1.1 PROBLEM STATEMENT

Over the past three decades, highway safety researchers have been concerned with developing statistical models to analyze motor vehicle crashes. The purpose of such models is to relate the frequency of crashes to the traffic, geometrical, and/or environmental characteristics of highway entities. Lord and Mannering (2010), and subsequently Mannering and Bhat (2014), provided excellent reviews on the alternative methods for statistical modeling of crash frequency data and described the inherent issues involved with crash data and different modeling approaches. Safety prediction models can be used to identify hazardous locations (a.k.a. hotspots) and to estimate the benefit of countermeasures in improving safety.

A mass of published work in the field of crash data modeling has focused on evaluating the application of countless statistical models by comparing their goodness-of-fit (GOF) to field datasets. These studies seek to answer the question of "which model performs better?" Measuring the GOF of statistical models is not a straightforward task; researchers use a plenty of different methods for assessing the GOF. These methods can yield contradictive results in determining the model that performs the best.

Even if there was a consensus among researchers in employing a unique method for GOF assessment, the performance of different models would depend on the characteristics of the dataset to which they are fitted. Based on the model structure and application to a limited number of datasets, some studies have suggested that certain models are expected to perform better than other specified models for data with certain characteristics. For example, Geedipally et al. (2012) predicted that the heavy-tailed negative binomial-Lindley (NB-L) distribution is expected to outperform the negative binomial model for datasets with abundant zero crash observations and a long/heavy tail. While rules of thumb like the aforementioned may be valid, it is difficult to predict with certainty which model performs better before the models are actually fitted to the data. Therefore, one shall be very cautious when making conclusions of the type "one model is better than the other" based on a comparison of the models' GOF to a few datasets.

GOF analyses in the literature of crash data modeling have been focused on the overall fit of a model to an entire dataset. The important question that has been overlooked is "how are the site-specific predictions for expected crash frequency affected by model choice?" Is there a trend in the difference between the predicted crash frequencies of different models, e.g. that a model's prediction for sites with certain conditions is higher (or lower) than that from a different model? This dissertation documents such a microscopic analysis for three hierarchical-Poisson regression models including the two most popular models in crash data analysis i.e., the Poisson-Gamma (PG) and Poisson-lognormal (PLN), and a new model developed by the author: the Poisson-inverse gamma (PIGam*). The Poisson-Inverse Gamma model is formulated and added to this study to investigate the potential benefit of having a long/heavy-tailed mixing distribution (i.e., inverse Gamma) in handling datasets with unusually high crash count observations.

---

* The Poisson-Inverse Gamma model is abbreviated as "PIGam" to avoid confusion with the Poisson-Inverse Gaussian (PIG) model, whose application to crash data has been investigated by Zha et al. (2016).

The impact of model choice on the predicted crash frequencies is most importantly revealed when crash prediction models are used to identify hazardous sites and rank them based on their crash proneness and hence priority for treatment. Rather than considering several candidate models and employing the best one (based on GOF comparisons) to rank highway entities for safety treatment, this study aims to add depth to our understanding of how the ranking of sites is influenced if each of the three considered models is selected.

1.2 RESEARCH OBJECTIVES

The objectives of this research are as follows:

1) Identify individual sites in each dataset for which the three alternative models predict significantly different expected crash frequencies and detect their common characteristics. Exploring the data and expected crash frequencies predicted by alternative models may reveal important trends regarding the relative performance of the model, e.g., that a model will likely predict higher (or lower) crash frequencies than another model will for sites with certain characteristics.
2) Investigate any possible relationship between the relative predictions of the models and the dispersion characteristics of datasets. For this purpose, the models will be fitted to four crash datasets ranging from mildly over-dispersed to severely over-dispersed.
3) Bridge the gap between the relative predictions of the models and the fundamental traits of the models' mixing distributions (namely, gamma, lognormal, and inverse-gamma).

1.3 DISSERTATION OUTLINE

The outline of this dissertation is as follows:

Chapter II provides background information regarding the challenges in crash data modeling and the characteristics of Bayesian hierarchical models. Models with widespread application in highway safety research (i.e., Poisson-gamma and Poisson-lognormal) are introduced and Bayesian methods for estimating the models parameters are briefly discussed.

Chapter III formulates the three alternative models considered in this research and describes the common structure of the models as well as their distinctive properties.

Chapter IV describes the four crash datasets and the functional forms selected for modeling each dataset.

Chapter V covers the assumptions made for fitting the alternative models to the data and presents the results of parameter estimation for every dataset.

Chapter VI compares the predicted expectation for crash frequency at individual sites across the alternative models and investigates the data conditions that contribute to significantly different predictions between the considered models. In addition to the analysis for the existing sites in every dataset, application of the calibrated models for new sites without observed crash frequency is scrutinized and compared across the models.

Chapter VII compares the fitted models from the GOF standpoint and explains the site-specific conditions that result in a model providing a better fit than another.

Chapter VIII summarizes the analysis in this dissertation and draws conclusions regarding the differences between the predictions of alternative models based on the research findings.

CHAPTER II
BACKGROUND

This chapter provides the background information and literature review needed to understand the importance of the models selected for investigation in this dissertation, the mathematical structure of the models, and the methods for model estimation and evaluation. Section 2.1 provides background information regarding cross-sectional regression models for crash frequency prediction and reviews the important applications of these models in highway safety research. Section 2.2 presents the mathematical description of the structure of the Poisson-hierarchical class of regression models and introduces the models with widespread application in the body of highway safety literature. Section 2.3 covers the common methods for estimating the parameters of crash regression models and discusses the advantages and limitations of the full Bayesian approach which is taken in this research. Finally, Section 2.4 introduces the common techniques to assess the performance of Bayesian hierarchical models.

2.1 CROSS-SECTIONAL MODELS FOR CRASH PREDICTION

The regression models that will be studied in this research are all of the cross-sectional type, meaning that the crash count observations are made from different units but all during a fixed period of time. In cross-sectional crash-frequency models, each observation comes from a unique geographical site (such as a highway segment, an intersection, etc.), the response variable is the number of observed crashes, and covariates (regressors) can include a range of traffic, geometrical, and environmental variables or factors.

The response variable of crash-frequency models is a non-negative integer. As such, the traditional linear regression model, for instance, is not suitable. Numerous statistical

techniques are available for analysis of count data (Cameron and Trivedi, 1998; Hilber, 2014). Among the most recognized methods is the generalized linear model (GLM) (McCullagh and Nelder, 1989). Most crash-frequency models are of this type where the mean of the response variable (number of crashes) is related to the linear predictor via a link function (such as a log link) and the response variable itself follows a discrete distribution from the exponential family (such as Poisson). The GLM framework is adopted in this research.

Cross-sectional safety prediction models mainly serve two purposes: explaining the system and prediction. These models can explain the relationship between crash-proneness (unsafety) and different characteristics of highway entities. In GLMs, for example, the calibrated model parameters directly indicate the magnitude of the influence of each variable on the expected number of crashes. Several researchers have used cross-sectional models to quantify the effect of changes in highway traits (by engineering interventions for instance) on the expected number of crashes (Tarko et al., 1998; Lord and Bonneson, 2007; Wu et al., 2015; Wu and Lord, 2016). However, such inferences have been questioned by other researchers, such as Hauer (2010), who argue that regression models can only explain correlation but not causation.

The predictive capability of safety performance models is crucial from the engineering aspect. Part C of the Highway Safety Manual (HSM, 2010), for example, is entirely devoted to predictive models and their applications in highway safety. Safety prediction models are used for the following important purposes:

1) Assist highway designers to quantify the influence of each deign feature (lane width, lighting, etc.) on the (un)safety and compare alternative designs from the safety standpoint.
2) Rank highway sites based on crash-proneness and identify the most hazardous locations (e.g., for safety treatment).

3) Provide an estimate for crash-proneness of sites with similar characteristics when using the EB method to estimate the impact of safety countermeasures on the expected number of crashes (Hauer, 1997).

2.2 POISSON-HIERARCHICAL MODELS

Because crashes are random events and typically independent of one another, the Poisson distribution is the intuitive means to describe the randomness in crash counts. In practice, however, the Poisson is rarely an appropriate distribution because the crash data are often over-dispersed (Poch and Mannering, 1996; Hauer, 2001; Mitra and Washington, 2006) i.e., the conditional variance of observed crash counts is greater than the mean, whereas the Poisson distribution is equi-dispersed (i.e., the mean equals the variance).

The over-dispersion in crash count data is mainly attributed to the heterogeneity among the different sites (highway segments, intersections, etc.) where crash data are collected (Hauer, 2001; Washington et al., 2003). Although the effect of some important factors (such as traffic volume) on the expected number (mean) of crashes is usually accounted for by a regression model, some degree of heterogeneity is always believed to remain unobserved due to factors either unknown or known but hard to collect and include in the model. In addition, Lord et al. (2005) showed that the theory behind the fundamental crash process can per se give rise to over-dispersion. They argued that a crash count is the sum of a series of Bernoulli random variables, the trial being a vehicle/driver going through a site, with unequal probabilities of success (i.e., experiencing a crash), which gives rise to an over-dispersed distribution of crash counts.

To accommodate the possible over-dispersion in crash data, researchers have extensively used mixed-Poisson models, where crash counts are assumed to have a Poisson distribution with a variable mean that follows an underlying distribution often referred to

as the mixing distribution. Mixed-Poisson models are indeed hierarchical, where at the first level of hierarchy, conditional on the mean, the observed crash counts are mutually independent and Poisson-distributed and at the second level, the unobservable mean of crash counts varies across sites with an assumed probability distribution.
Importantly, the hierarchical Poisson models particularly suit the theoretical nature of crash frequency data. Such conceptual suitability, often referred to as "goodness-of-logic", has been sacrificed for a better statistical goodness-of-fit (GOF) in a number of regression models proposed for crash data analysis. For instance, Lord et al. (2005, 2007) criticized the application of popular zero-inflated count models (as used by Shankar et al., 1997; Lee and Mannering, 2002; Kumara and Chin, 2003; Qin et al., 2004, etc.), arguing that the inherent dual-state assumption in these models is inconsistent with the nature of crash data because no highway entity is completely safe and thus a long-term mean equal to zero is impossible. In contrast, the hierarchical Poisson models offer an interpretable structure where the error term captures the unmodeled differences between sites.

Two types of mixed-Poisson models have gained extensive attention among highway safety researchers, Poisson-gamma and Poisson-lognormal. These models are introduced below but their mathematical formulation is presented in the next chapter.

1) *Poisson-Gamma* (*PG*): When the Poisson parameter is assumed to have a gamma probability distribution, the mixed-Poisson distribution will have a closed form probability density function (pdf) which turns out to be that of the Negative Binomial (NB). The over-dispersion parameter (α) of the NB distribution captures the unmodeled heterogeneity (Miaou and Lord, 2003). The value of this parameter defines the relationship between the distribution mean and variance as $\mathrm{Var}(y_i) = \mathrm{E}[y_i] + \alpha \mathrm{E}[y_i]^2$. The simplicity of model fitting and parameter interpretation has made the Poisson-Gamma/NB the most popular model in crash data analysis (Lord and Mannering, 2010). In its classical applications, modelers estimated the parameters

using the Maximum Likelihood Estimate (MLE) method (e.g. Maycock and Hall, 1984; Hauer et al., 1989; Bonneson and McCoy, 1993; Vogt and Bared, 1998). Recently, the Poisson-Gamma models have also been estimated using Bayesian methods (Schluter et al., 1997; Miaou and Lord, 2003; Miaou and Song, 2005; Lord and Miranda-Moreno, 2008).

2) *Poisson-Lognormal* (*PLN*): The Poisson-lognormal model results when the Poisson parameter is assumed to follow a lognormal distribution. Unlike the Poisson-Gamma model, the marginal distribution of the Poisson-lognormal model does not have a closed form and the model parameters cannot be estimated directly using the Maximum Likelihood Estimates (MLE) method. Despite the availability of Hinde's (1982) numerical integration method to approximate the MLE parameters, the Poisson-lognormal safety performance models in the published studies have all been estimated using the Bayesian approach. The Poisson-lognormal model is potentially more flexible than the Poisson-Gamma (Lord and Mannering, 2010) and the model has become increasingly popular in crash data analysis over the past few years (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2008; Aguero-Valverde, 2013). Highway safety researchers have also used multivariate Poisson-lognormal models to jointly model crash frequency by severity while accounting for the correlation among different severity levels (Park and Lord, 2007; Ma et al., 2008; El-Basyouny and Sayed, 2009).

2.3 MODEL ESTIMATION METHODS

The MLE method has traditionally been used to estimate safety performance model parameters. Nonetheless, the MLE method cannot be used straightforwardly where the marginal likelihood function is difficult to characterize such as in the Poisson-lognormal model. However, the development of such models and others with complex functional forms was greatly facilitated after the rediscovery of Markov Chain Monte Carlo

(MCMC) simulation methods (Besag et al., 1995; Gilks et al., 1996; Robert and Casella, 1999) for model estimation from the Bayesian perspective. Unlike the traditional frequentist approach, the Bayesian approach to statistics aims to estimate the probability distribution of model parameters using the information in the observed data as well as the prior knowledge about model parameters. The drastic growth in the processing speed of personal computers and availability of Bayesian software programs, such as WinBUGS (Spiegelhalter et al., 2003) and MLwiN (Yang et al., 1999) has helped significantly in the increasing popularity of Bayesian models.

The full (hierarchical) Bayesian estimation of safety performance models has been explored only in the last two decades (Schluter et al., 1997; Davis and Yang, 2001; Miaou and Lord, 2003; Carriquiry and Pawlovich, 2004; Miaou and Song, 2005; Park and Lord, 2007). Prior to FB models, however, the empirical Bayes method was introduced into the highway safety literature (Hauer and Persaud, 1983; Hauer, 1986, 1992). The EB method basically uses the Bayes rule to combine the information from some reference population (or results of a regression model) with the observed crash counts at a certain site to estimate the expected (long-term) mean of crashes. There is extensive documentation and application of the EB method in highway safety especially in before-after studies to estimate the effect of safety countermeasure (Hauer, 1997; Persaud, 1998; Harwood et al., 2002; Persaud and Lyon, 2007; Fitzpatrick and Park, 2009) and also in identification of hotspot locations (Persaud et al., 1999; Heydecker and Wu, 2001; Miranda-Moreno et al., 2005; Lord and Park, 2008).

The full Bayesian approach has a fundamental advantage over the MLE and EB methods; it takes into account the uncertainty associated with model parameters and provides exact measures of uncertainty (Miaou and Lord, 2003). In the MCMC method, this is carried out by sampling from the posterior distribution of model parameters. The MLE and EB methods, on the other hand, ignore this uncertainty and thus overestimate the model precision (Carriquiry and Pawlovich, 2004; Goldstein, 2010; Park et al.,

2010). This advantage of the full Bayesian approach is especially important when the sample size is relatively small (Miaou and Lord, 2003).

The main challenge in taking the full Bayesian approach is the specification of prior distributions for model parameters. Many modern statisticians have investigated and documented this matter (e.g., Gelman et al, 2003; Rao, 2003; Carlin and Louis, 2008; Lee, 2012). In a hierarchical Poisson model, the mixing distribution is indeed the prior distribution for the Poisson parameter. The Poisson parameter prior distribution (gamma, lognormal, etc.) itself has one or more parameters for which so-called hyper-prior distributions needs to be presumed. In the absence of prior knowledge, it is recommended to use non-informative (a.k.a. vague or diffuse) hyper-priors with the idea to let the data "speak for itself." The utilization of such hyper-priors would minimize the influence of the prior knowledge on the posterior distribution of model parameters. In crash data analysis, hierarchical Bayesian models with non-informative hyper-priors have been used by numerous studies (e.g., Miaou et al. 2003; Davis and Yang, 2001; Song et al., 2006; Miranda-Moreno et al, 2005; Park et al., 2010; El-Bayouny and Sayed, 2012).

Before closing this section, it is necessary to remind that all Poisson mixture distributions (and the respective Poisson-hierarchical modes) are obviously over-dispersed and hence incapable of handling data characterized by under-dispersion i.e., variance less than the mean. Nevertheless, crash data exhibiting under-dispersion (conditional on the mean) are fairly rare and may be a sign of over-fitting i.e., including too many variables in the model. However, modelers should always be wary about the possibility of under-dispersion especially when working with crash data with low sample means, as under-dispersion is more likely to prevail in such conditions (Oh et al., 2006; Khazraee et al., 2015).

2.4 BAYESIAN MODEL ASSESSMENT TECHNIQUES

The performance of safety prediction models is often evaluated and compared using statistical GOF measures. In the Bayesian paradigm, measures such as the Bayes factor, Bayesian information criterion (BIC), and Watanabe-Akaike information criterion (WAIC), among others are used to assess and compare the performance of fitted models (Gelman et al. 2013). However, the GOF of Bayesian hierarchical regression models is most commonly compared using the deviance information criterion (DIC). This study adopts DIC as the main tool for GOF comparison between alternative models.

Proposed by Spiegelhalter et al. (2002), the DIC is a Bayesian generalization of the Akaike Information Criterion (AIC), and is defined as:

$$DIC = \bar{D} + P_D \tag{2.1}$$

where $\bar{D} = \mathrm{E}(-2\log(\Pr(y|\theta))$ is the expectation of the model deviance under the posterior distribution of the model parameters (collectively denoted as θ), and P_D is the effective number of parameters, defined as:

$$P_D = \bar{D} - D(\bar{\theta}) \tag{2.2}$$

where $D(\bar{\theta})$ is the deviance under the posterior expectation of parameters.

$\bar{D}$ is a classical estimate of fit; a smaller $\bar{D}$ indicates a better fit as it corresponds to a greater log-likelihood. P_D is indeed a penalty for model complexity and ensures a fair comparison between competing models with different degrees of complexity. DIC is particularly useful when the posterior distributions of model parameters are obtained via MCMC simulation, which is the case for the models in this study.

2.5 CHAPTER SUMMARY

Cross-sectional regression models are widely used in highway safety research to predict the expected number of crashes at individual road sites (roadway segments, intersections, etc.). These models assume the crash count observations at different sites to be independent and model the observed crash frequency as a function of site-specific characteristics including traffic, geometrical, and environmental variables.

Although crashes are random events, the Poisson distribution usually fails to suit crash data analysis because crash data are often plagued by over-dispersion. The over-dispersion in crash data is thought to originate from the unobserved heterogeneity among the sites, and can be addressed by assuming that crash counts at individual sites are Poisson-distributed with a mean that itself follows a continues distribution. Such assumption gives rise to mixed Poisson models which are hierarchical in nature. The two most commonly used models for crash data analysis i.e., Poisson-gamma and Poisson-lognormal, are of this type.

A Bayesian approach for model estimation is taken in this study because of its theoretical appeal as well as the capability of the MCMC algorithm to estimate models whose marginal likelihood function may not be algebraically characterized (including the Poisson-lognormal and Poisson-inverse gamma models). The DIC is adopted as a common tool for GOF comparison between Bayesian hierarchical models. The next chapter describes the characteristics of the alternative models for analyzing crash data.

CHAPTER III
ALTERNATIVE MODELS

This chapter specifies the alternative Poisson-hierarchical regression models that are selected for analysis in this dissertation. Section 3.1 presents the general modeling assumptions and mathematical formulation of the models. Section 3.2 explains how the common structure of the models provides for a fair and facilitated comparison between models predictions. Finally, Section 3.3 describes the fundamental properties of the mixing distributions as the main source of variation between the models predictions.

3.1 MODELS SPECIFICATION

This research thoroughly evaluates and compares the crash frequency predictions of the following hierarchical Poisson regression models using a full Bayesian approach:

1) Poisson-Gamma (PG)
2) Poisson-Lognormal (PLN)
3) Poisson-Inverse Gamma (PIGam)

The first two models are commonly used in highway safety analyses, whereas the latter is new and its appropriateness for crash data modeling is to be examined.

Let y_i denote the number of crashes observed at the i'th site (road segment, intersection, etc.) during the study period. In Poisson hierarchical models, y_i's, when conditional on their mean m_i, are assumed to be Poisson distributed:

$$y_i \mid m_i \sim \text{Poisson}(m_i) \qquad i = 1, 2, \ldots, n \qquad (3.1)$$

On the second level of hierarchy, the mean of the Poisson distribution is variable with an underlying mixing distribution: gamma, lognormal, and Inverse Gamma, respectively. In all models under study, the mean of the mixing distribution, μ_i, is modeled as a log-linear function of the prevailing traffic, geometric, and/or environmental variables:

$$\mathrm{E}(m_i) = \mu_i = \exp(X_i\beta) \tag{3.2}$$

where X_i is the vector of covariates for Site i and β is the vector of unknown coefficients. Every mixing distribution selected for this study has two parameters. Below, these distributions are reparametrized in terms of their mean (μ_i) and a remaining hyper-parameter (shape or scale) to structure the regression models:

1) Poisson-Gamma (PG): The Poisson parameter (mean) follows a gamma distribution with shape parameter φ and scale parameter λ_i:

$$\Pr(m_i|\varphi,\lambda_i) = \frac{1}{\Gamma(\varphi)\lambda_i^{\varphi}} m_i^{\varphi-1}\exp(\frac{-m_i}{\lambda_i}) \qquad \varphi,\ \lambda_i > 0 \tag{3.3}$$

$$\mathrm{E}(m_i) = \mu_i = \varphi\lambda_i = \exp(X_i\beta) \tag{3.4}$$

$$\Pr(m_i|\varphi,\lambda_i) = \frac{\varphi^{\varphi}}{\Gamma(\varphi)\exp(\varphi.(X_i\beta))} m_i^{\varphi-1}\exp(\frac{-m_i\varphi}{\exp(X_i\beta)}) \tag{3.5}$$

2) Poisson-Lognormal (PLN): The Poisson parameter follows a lognormal distribution with location parameter υ_i and shape parameter σ^2:

$$\Pr(m_i|\upsilon_i,\sigma^2) = \frac{1}{m_i\sigma\sqrt{2\pi}}\exp(-\frac{(\ln(m_i)-\upsilon_i)^2}{2\sigma^2}) \qquad \sigma^2 > 0 \tag{3.6}$$

$$\mathrm{E}(m_i) = \mu_i = \exp(\upsilon_i + \sigma^2/2) = \exp(X_i\beta) \tag{3.7}$$

$$\Pr(m_i \mid \mu_i, \sigma^2) = \frac{1}{m_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(m_i) - X_i\beta + \sigma^2/2)^2}{2\sigma^2}\right) \quad (3.8)$$

3) Poisson-Inverse Gamma (PIGam): The Poisson parameter follows an inverse gamma distribution with shape parameter φ and scale parameter λ_i:

$$\Pr(m_i \mid \varphi, \lambda_i) = \frac{\lambda_i^{\varphi}}{\Gamma(\varphi)} m_i^{-\varphi-1} \exp\left(\frac{-\lambda_i}{m_i}\right) \qquad \varphi,\ \lambda_i > 0 \quad (3.9)$$

$$\mathrm{E}(m_i) = \mu_i = \frac{\lambda_i}{\varphi - 1} = \exp(X_i\beta) \qquad \text{for } \varphi > 1 \quad (3.10)$$

$$\Pr(m_i \mid \varphi, \mu_i) = \frac{[(\varphi-1)\exp(X_i\beta)]^{\varphi}}{\Gamma(\varphi)} m_i^{-\varphi-1} \exp\left(\frac{-(\varphi-1)\exp(X_i\beta)}{m_i}\right) \quad \text{for } \varphi > 1 \quad (3.11)$$

Please note that the PIGam model was developed by the author and no prior application of this model was found in the literature. To distinguish between the parameters of the PG and PIGam models, the shape parameters are accompanied by subscripts denoting the model name i.e., φ_{PG} for the Poisson-gamma and φ_{PIGam} for the Poisson-inverse gamma.

3.2 COMMON MODEL STRUCTURE

At this point, it is essential to fully understand the common structure of the selected hierarchical models and appreciate the important features shared by all three models. First, it is critical to distinguish between the two "mean" parameters used in the parameterization of the models: m_i and μ_i. m_i is the site-specific expected crash frequency which, a priori, follows a model-specific continuous distribution with mean μ_i and a model-specific hyper-parameter (φ_{PG}, φ_{PIGam}, or σ^2). μ_i captures the effect of the covariates on the expected crash frequency through the regression model and may be interpreted as the prior expected crash frequency for sites with the same covariate values

(i.e., traffic, geometric, and/or environmental conditions) as those of Site *i*. The posterior distribution of m_i (i.e., $m_i|X,y$) combines the prior data from posterior μ_i with the observed crash frequency, and thus every $E(m_i|X,y)$ deviates from the respective $E(\mu_i|X,y)$.

In the hierarchical models in this study, each m_i is indeed a model parameter, whereas μ_i's are determined based on a given fixed vector of covariates (X_i) and a variable vector of regression parameters (β). Therefore, every one of the alternative models have a total of $n+p+1$ parameters: n m_i's each from a unique site in the dataset, p regression coefficients, and an additional hyper-parameter which captures the dispersion characteristic of the data (φ_{PG}, φ_{PIGam}, or σ^2). Compared to a non-hierarchical Poisson-Gamma/Negative Binomial model with $p+1$ parameters, for example, it is evident that Bayesian hierarchical models are far more complicated.

The similar number of parameters and structure provides for a reasonable comparison between the models' performance and facilitates the comparison between their predictions for the expected crash frequency. No model can be presumed to take advantage of the higher number of parameters for its flexibility to fit the data. Also, site-specific m_i's and μ_i's as well as their relationship under each model can be directly and meaningfully compared, as carried out in Chapter 6.

3.3 MIXING DISTRIBUTION PROPERTIES

Given the common structure of the Poisson hierarchical models, the distinctive performance of the models is attributable to the fundamental characteristics of their mixing distributions (i.e., prior distribution for m_i's).

Unlike the gamma distribution, the lognormal and inverse-gamma distributions are classified as heavy-tailed because their right tails are not exponentially bounded.

Moreover, the inverse-gamma distribution has a thicker tail than the lognormal distribution and, therefore, the lognormal distribution lies between the gamma and inverse-gamma distribution in terms of the tail thickness. The thicker tail, which is representative of greater probabilities for higher expected crash frequencies, can theoretically be beneficial when modeling data with frequent unusually high crash observations.

The inverse gamma distribution is the distribution of the reciprocal of a variable that follows a gamma distribution. Since the gamma distribution is light-tailed, the inverse of a gamma-distributed variable has a very low probability in the vicinity of zero. This notion is evident in Figure 1, which compares the shape of the gamma, lognormal, and inverse gamma probability distribution functions (pdf's) for four different combinations of mean and variance. It is interesting to note that the lognormal distribution lies between the gamma and inverse-gamma distributions almost in every important aspect including the probability density in vicinity of zero, probability density for very large values (i.e., tail thickness), mode, probability at mode, skewness, etc. As we move from the gamma distribution to lognormal and then to inverse-gamma, near-zero values for expected crash frequencies (m_i's) become less likely, and extremely high values for the expected crash frequency become more likely.

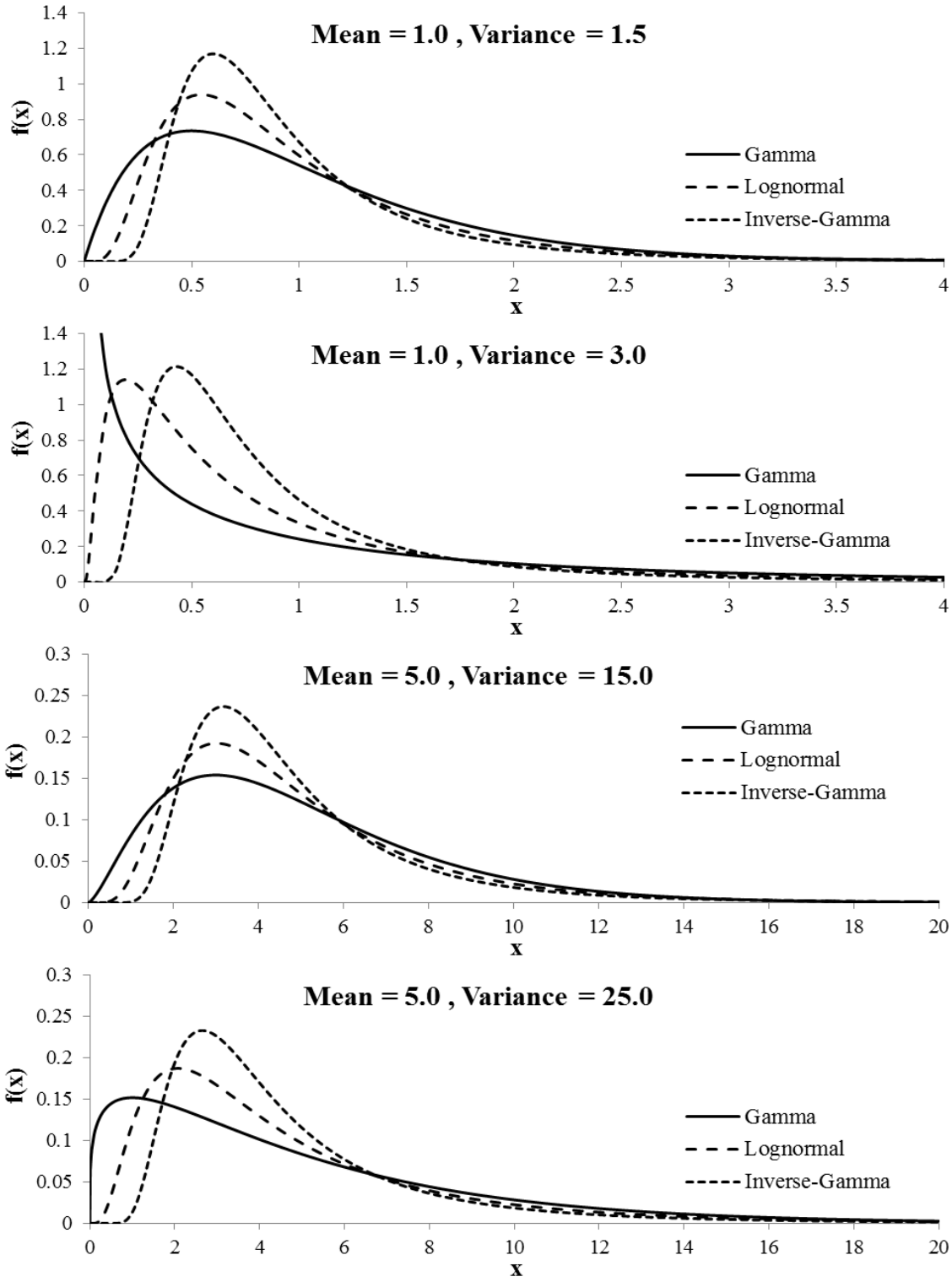


Figure 1. Probability distribution function of the gamma, lognormal, and inverse-gamma distributions for different mean-variance combinations

3.4 CHAPTER SUMMARY

This chapter focuses on three specific models belonging to the Bayesian Poisson-hierarchical class of count regression models, namely Poisson-gamma, Poisson-lognormal, and Poisson-inverse gamma. The three models share a common structure in the sense that crash counts at individual sites are assumed to be Poisson-distributed with site-specific means that follow the model's mixing distribution i.e., gamma, lognormal, and inverse gamma distribution, respectively. Furthermore, all selected models include only one hyper-parameter, in addition to the common number of regression parameters, which describes the shape of the mixing distribution. Such common structure and number of parameters facilitates the comparisons between the models predictions and provide a level ground for assessing the models GOF.

CHAPTER IV
DATA DESCRIPTION AND MODELS FUNCTIONAL FORM

This chapter describes the four datasets that are used for analysis in this dissertation and presents the functional forms assumed for modeling each of the crash datasets. All datasets are relatively large and are named after the city/state they are collected from: Toronto (Canada), Texas, Michigan, and Indiana. Section 4.1 presents the summary statistics for the four datasets. Section 4.2 to 4.5 presents further information about the Toronto, Texas, Michigan, and Indiana datasets, respectively.

4.1 SUMMARY STATISTICS FOR THE FOUR DATASETS

Table 1 presents the summary statistics for the variables in each dataset. The datasets were specifically selected to cover a commonplace range of dispersion characteristics. The over-dispersion parameter ($\alpha = 1/\varphi$) of the traditional negative binomial (NB) GLM (estimated using the MLE method) is considered as a rough measure of conditional dispersion in the data (i.e., dispersion of the crash counts conditional on the modeled mean). The NB model over-dispersion parameter ($\alpha_{NB(MLE)}$) is 0.140, 0.391, 0.654, and 0.888 for the Toronto, Texas, Michigan, and Indiana data, respectively. The source of data and the adopted functional form for each dataset are described in separate sections in the following.

Table 1. Summary statistics of the datasets in this study

Variable	Min.	Max.	Average (std. dev)	Total
Toronto Data				
Crash count	0	54	11.6 (10.0)	10030
Major approach AADT (in veh/day) ($AADT_{maj}$)	5469	72,178	28044.8 (10660.4)	--
Minor approach AADT(in veh/day) ($AADT_{min}$)	53	42644	11010.2 (8599.4)	--
Texas Data				
Crash count	0	97	5.2 (9.0)	5219
Segment length (in miles)	0.10	8.55	0.845 (1.051)	844.97
AADT(in veh/day)	264	56890	8611.7 (6660.4)	--
Michigan Data				
Crash count	0	29	0.8 (2.1)	796
Segment length (in miles)	0.001	4.04	0.19 (0.38)	193.02
AADT(in veh/day)	250	19990	4331.8 (3139.9)	--
Indiana Data				
Crash count	0	329	16.97 (36.30)	5737
Segment length (in miles)	0.009	11.53	0.89 (1.48)	300.09
AADT(in veh/day)	9,442	143422	30237.6 (28776.4)	--
Minimum friction reading in the segment (FRICTION)	15.9	48.2	30.51 (6.67)	--
Median width (in feet) (MW)	16	194.7	66.98 (34.17)	--
Pavement surface (1: asphalt, 0: concrete) (PAVEMENT)	0	1	0.77 (0.42)	--
Median barrier (1 if present, 0 if absent) (BARRIER)	0	1	0.16 (0.37)	--
Interior rumble strips (1 if present, 0 if absent) (RUMBLE)	0	1	0.72 (0.45)	--

4.1 TORONTO DATA

The Toronto dataset contains crash count data collected in 1995 at 868 four-legged signalized intersections in Toronto. The following common and simple functional form was adopted for this dataset:

$$\mu_i = \beta_0 F_{Maj\_i}^{\beta_1} F_{Min\_i}^{\beta_2} \tag{4.1}$$

where F_{Maj_i} and F_{Min_i} denote the average annual daily traffic (AADT) flow on the major and minor approaches to the intersection, respectively.

4.2 TEXAS DATA

The Texas dataset is a randomly selected sample of 1,000 rural multilane highway segments, where the crash counts indicate the total number of crashes observed during a five year period in the early 2000s. The data were originally obtained from Department of Public Safety (DPS) and the Texas Department of Transportation (TxDOT) and previously used in the *NCHRP 17-29* project to develop safety prediction models for inclusion in the first edition of the Highway Safety Manual (HSM) (AASHTO, 2011). The original data included a total of 3,220 segments.

Despite availability of other relevant covariates, simple traffic-only models were concentrated upon for this dataset. As specified earlier, the focus of this research is not to develop practice-ready crash prediction models but rather to evaluate and compare certain types of Poisson hierarchical models. The researcher is aware that flow-only models can be influenced by the omitted-variable bias (Lord and Mannering, 2010; Mitra and Washington, 2012), but this bias will likely not affect the results of the analyses presented in this work because the alternative models will have the same number of input variables. The adopted functional form for this dataset follows:

$$\mu_i = \beta_0 L F_i^{\beta_1} \tag{4.2}$$

where F_i is the AADT through the highway segment and L is the segment length. Here, the segment length is treated as an offset (i.e., its power is kept fixed at 1) because theoretically, the crash risk is expected to have a linear relationship with the segment length.

4.3 MICHIGAN DATA

The Michigan dataset contains the single-vehicle crashes that occurred on rural two-lane highways in Michigan in 2006. Similar to the Texas dataset, a randomly selected sample of 1000 highway segments was taken from the original data and used for analysis. The database was collected for the Federal Highway Administrations's (FHWA) Highway Safety Informations System (HSIS) and originally included a total of 33,970 roadway segments. The exactly same functional form as that for the Texas data was adopted for modeling the Michigan dataset:

$$\mu_i = \beta_0 LF_i^{\beta_1} \tag{4.3}$$

In this dataset, about 67% of segments experienced no crashes in 2006. The selection of this dataset provides an opportunity to examine the performance of different models with data characterized by a large number of zeros, which is specifically common in road safety analyses.

4.4 INDIANA DATA

The Indiana dataset contained data collected for a five-year period (1995-1999) from 338 rural interstate freeway segments in Indiana. The Indiana data is the most highly over-dispersed dataset considered in this study which includes a relatively large proportion of zero crash counts (about 36%) along with very high crash counts observed at some other sites. Unlike the previously mentioned simple model forms, as Table 1 indicates, several roadway characteristic variables are included in the dataset to be used as additional covariates. The following functional form is used for Indiana data:

$$\mu_i = \beta_0 LF_i^{\beta_1} \exp(\sum_{j=2}^{6} \beta_j x_j) \tag{4.4}$$

where x_j's are the additional covariates listed in Table 1.

4.5 CHAPTER SUMMARY

This chapter presented the features of the four dataset used in this dissertation and outlined the functional forms adopted to model each dataset. The datasets and their covariates were specifically chosen with the intention that they cover a wide range of dispersion characteristic, from mildly over-dispersed to highly over-dispersed. This choice of data will provide the opportunity to compare the alternative models under various data scenarios. The next chapter describes the model fitting procedure and presents the model estimation results.

CHAPTER V
MODEL ESTIMATION

This chapter describes the assumptions used to fit the previously specified models to the datasets and presents the results of parameter estimation. Section 5.1 delineates the hyper-priors assumed for the hyper-parameters in each model, Section 5.2 describes the MCMC algorithms used to estimate models parameters, and Section 5.3 summarizes the posterior estimates for models parameters when fitted to each dataset.

5.1 CHOICE OF HYPER-PRIORS

A critical step in Bayesian modeling is the selection of hyper-prior distributions (Gelman et al. 2013; Lee 2012). In order to control for the potential influence of prior information on the performance of the competing models, non-informative (vague) prior distributions were considered for the models' hyper-parameters in this step of the research. However, for large datasets such as the ones studied here, the choice of hyper-priors will not make a significant change in the parameter estimates. The following prior distributions were used:

- Regression coefficients (all models):

  $\beta_j \sim N(0\ ,\ \sigma = 100)$ for $j = 0,1, 2, \ldots, p$

- Poisson-gamma model:

  $\varphi_{PG} \sim$ Uniform $(0,+\infty)$

- Poisson-lognormal model:

  $\sigma^{-2} \sim \Gamma(0.01,0.01)$

- Poisson-inverse gamma:

  $\varphi_{PIGam} \sim$ Uniform $(1,+\infty)$

5.2 MCMC SIMULATION FOR POSTERIOR INFERENCE

Posterior distributions of model parameters were estimated using the basic Markov Chain Monte Carlo (MCMC) simulation methods. The MCMC algorithms for all models were coded in the software R (R Core Team, 2014). An example of R codes used for estimating the models are presented in Appendix B. Except for m_i's in the PG model which were updated using a Gibbs sampler, sampling from other parameters was carried out using the Metropolis algorithm. The posterior distribution of each parameter was estimated using three chains with different initial values. A total of 20,000 MCMC iterations were obtained for each chain, the first 10,000 of which were discarded to diminish the influence of the starting values. Thus, a total of 30,000 (10,000 draws from three chains) were used to construct the posterior distributions. Gelman's $\widehat{R}$ factor (Gelman et al., 2013) and visual inspection were utilized to ensure sufficient convergence.

5.3 PARAMETER ESTIMATION RESULTS

Appendix A presents the MCMC traceplots and the posterior distributions for the parameters in each model and dataset. Tables 2-5 summarize the posterior inference with the mean and standard deviation of the parameters.

Table 2. Posterior Estimates of Model Parameters for Toronto Data

Parameter	Respective Variable	**PG**	**PLN**	**PIGam**
		Mean (Std Dev.)	Mean (Std Dev.)	Mean (Std Dev.)
$\ln(\beta_0)$	Intercept	-10.3066 (0.4702)	-10.2786 (0.4541)	-10.2502 (0.4602)
β_1	$AADT_{maj}$	0.6267 (0.0464)	0.6167 (0.0451)	0.6069 (0.0452)
β_2	$AADT_{min}$	0.6852 (0.0227)	0.6935 (0.0221)	0.7016 (0.0216)
φ_{PG}		7.207 (0.6765)		
σ^{-2}			7.266 (0.6525)	
φ_{PIGam}				8.439 (0.7217)

Table 3. Posterior Estimates of Model Parameters for Texas Data

Parameter	Respective Variable	PG	PLN	PIGam
		Mean (Std Dev.)	Mean (Std Dev.)	Mean (Std Dev.)
$\ln(\beta_0)$	Intercept	-7.0321 (0.3486)	-6.9996 (0.3547)	-6.9520 (0.3573)
β_1	AADT	0.9790 (0.0383)	0.9759 (0.0392)	0.9713 (0.0393)
φ_{PG}		3.364 (0.3304)		
σ^{-2}			3.426 (0.3283)	
φ_{PIGam}				8.439 (0.7217)

Table 4. Posterior Estimates of Model Parameters for Michigan Data

Parameter	Respective Variable	PG	PLN	PIGam
		Mean (Std Dev.)	Mean (Std Dev.)	Mean (Std Dev.)
$\ln(\beta_0)$	Intercept	-2.6153 (0.6075)	-2.6632 (0.6237)	-2.6200 (0.5611)
β_1	AADT	0.4959 (0.0738)	0.5029 (0.7579)	0.5001 (0.0679)
φ_{PG}		1.6123 (0.296)		
σ^{-2}			1.6023 (0.269)	
φ_{PIGam}				2.6823 (0.3222)

Table 5. Posterior Estimates of Model Parameters for Indiana Data

Parameter	Respective Variable	PG	PLN	PIGam
		Mean (Std Dev.)	Mean (Std Dev.)	Mean (Std Dev.)
$\ln(\beta_0)$	Intercept	-4.4795 (1.3987)	-3.5237 (1.4019)	-3.5516 (1.4174)
β_1	AADT	0.6909 (0.1300)	0.6224 (0.1335)	0.6287 (0.1297)
β_2	FRICTION	-0.0264 (0.0105)	-0.0300 (0.0109)	-0.0297 (0.0102)
β_3	PAVEMENT	0.4201 (0.1913)	0.4947 (0.2106)	0.4781 (0.1988)
β_4	MW	-0.0052 (0.0020)	-0.0075 (0.0022)	-0.0077 (0.0021)
β_5	BARRIER	-3.0291 (0.2968)	-4.1862 (0.4952)	-5.2219 (0.3980)
β_6	RUMBLE	-0.4005 (0.1812)	-0.3185 (0.2038)	-0.1779 (0.1941)
φ_{PG}		1.1054 (0.1417)		
σ^{-2}			1.0767 (0.1488)	
φ_{PIGam}				1.9052 (0.2053)

As Tables 2-5 suggest, only for the Indiana data the model choice resulted in significantly different estimates for the regression coefficients; in the other three datasets, the regression coefficients estimates varied only slightly across the alternative models. This finding was quite expected due to the relatively great number of input variables and the substantial over-dispersion remaining in the data even after including so many variables. From the practical point of view, the more input variables included in the modeling, the more important the choice of the model becomes especially if the regression coefficients are going to be used to determine crash modification factors (CMFs).

5.4 CHAPTER SUMMARY

The Bayesian models in this research were fitted to the study datasets using MCMC simulation algorithms for sampling from the posterior distribution of models' hyper-parameters. Diffuse hyper-prior were adopted to eliminate the potential influence of prior information on models performance and provide for a sensible comparison between models predictions. The model choice caused only small variation in estimates of the regression coefficients for the Toronto, Texas, and Michigan datasets, and significantly different estimates for the Indiana data, which included a greater number of covariates compared to the other datasets, and exhibited the highest degree of over-dispersion. The next chapter compares the alternative models in terms of their site-specific predictions. Chapter 7 will compare the models in terms of their goodness-of-fit to each dataset.

CHAPTER VI
COMPARISON OF MODEL PREDICTIONS

This chapter presents the most important analyses in this dissertation as it concentrates on individual sites and the variations in predicted crash frequencies resulting from application of the alternative models. In other words, the analysis focuses on the posterior distribution of individual m_i's (i.e., $E(m_i|X,y)$) and how they are affected by the model choice. Section 6.1 provides the background information and introduces the two types of comparison between the models predictions that are carried out in this research: for sites with observed crash data in each dataset, and for new sites without observed crash data. These comparisons are presented in Section 6.2 and 6.3, respectively.

6.1 INTRODUCTION

Each dataset consists of n sites with an observed crash frequency during the specified study period. Given the Bayesian structure of the models, the posterior distribution of individual m_i's depend on the model parameters as well as the site-specific observed crash frequency, Y_i. Section 6.2 investigates the relationship between the distribution of $(m_i|X,y)$, μ_i (the loglinear predictor determined by the covariates and model parameters), and Y_i under each of the models. Such relationship affects the ranking of sites based on their crash proneness.

Furthermore, it is common for safety analysts to use calibrated crash prediction models to predict the expected crash frequency at planned facilities that have not been constructed yet. Here, the main objective is to estimate the safety impact of each design feature (e.g., lane width, type of median barrier, signal phasing, etc.) and help decision making in the design/planning stage. Also, the observed crash frequency may not be available at existing sites for reasons such as limited resources for data collection, etc.

For such applications, the models' prediction is not affected by the site-specific observed crash frequency and, as shown in Section 6.3, the pattern of the difference between the models' predictions is quite different from that for the sites with observed crash count.

In the comparisons in this chapter, m and μ variables are superscripted with the abbreviated name of the model with which they are associated; for example, μ_i^{PLN} is the μ of the i'th site when the Poisson-lognormal model is used.

6.2 SITES WITH CRASH DATA

This section includes three parts. First, the distinctions between the models predictions for the expected crash frequency at the sites in each dataset are analyzed in details. Second, the impact of these disparities on the ranking of sites based on crash proneness is manifested. Finally, the mean-variance relationship is plotted for the sites in each dataset.

6.2.1 Distinctions between Expected Crash Frequency Predictions

To investigate the distinctions between the predictions of the alternative models, the sites with the greatest absolute difference between the predicted $E(m_i|X,y)$ (expected crash frequency $E(Y_i) = E(m_i|X,y)$) by any two of the models were first identified. Table 6 lists these sites for each dataset. Also shown in Table 6 are the $E(\mu_i)$'s associated with these sites. $E(\mu_i)$ is indeed the model's prediction for the expected crash frequency based on the site characteristics before it is adjusted by the observed crash count (using the Bayesian structure of the model), which is determined from the posterior samples of the model regression parameters using the following equation:

$$E(\mu_i) = \frac{1}{T}\sum_{t=1}^{T}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_t) \qquad (6.1)$$

where $\boldsymbol{\beta_t}$ is a sample from the vector of regression coefficients drawn at the *t*'th iteration of the MCMC algorithm, and *T* is the total number of iterations.

All sites with considerably different predictions of expected crash frequency (shown in Table 6) share a common attribute: an observed crash frequency greatly different from the estimated E(μ_i). In other words, at all of these sites, the observed crash frequency (Y_i) is significantly smaller or larger than the expected crash frequency at a site with similar traffic and geometric characteristics (i.e., E(μ_i)).

As Figure 2 demonstrates, except for the Indiana dataset where the estimated vector of coefficients varies considerably across the models (see Table 5), the difference between the E(μ_i)'s estimated from different models were quite small for all sites. However, the main difference between the models arose from the different amount by which the posterior distribution of μ_i was shifted toward the observed crash count (Y_i) to form the posterior distribution of the site-specific expected crash frequency ($m_i|X,y$). This phenomenon is illustrated in Figure 3 for two sites in the Toronto dataset: Site #619 where the observed crash count (Y_{619} = 5) is much smaller than E(μ_{619}) estimated from every model, and Site #494 where the observed crash count (Y_{494} = 29) is much greater than E(μ_{494}) estimated from every model (see the top two rows in Table 6).

Table 6. Sites with greatest difference between any two models' predicted crash frequency

Site Number	Y_i	Poisson-Gamma		Poisson-Lognormal		Poisson-Inverse Gamma		Max. difference btw predicted E(m_i|X,y)
		E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	
Toronto Data								
619	5	28.48	9.73	28.61	11.41	28.77	13.18	3.45
494	29	10.07	21.10	9.98	22.58	9.91	24.01	2.91
424	42	16.61	34.31	16.78	36.11	16.97	37.22	2.90
757	48	19.80	40.48	19.74	41.80	19.70	43.35	2.87
533	53	25.30	46.89	25.46	48.46	25.64	49.63	2.74
42	8	29.65	12.23	29.83	13.37	30.04	14.86	2.64
15	1	16.08	5.65	16.22	6.97	16.36	8.11	2.46
612	4	20.05	8.25	20.04	9.28	20.06	10.37	2.12
Texas Data								
Site Number	Y_i	Poisson-Gamma		Poisson-Lognormal		Poisson-Inverse Gamma		Max. difference btw predicted E(m_i|X,y)
		E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	
905	97	19.73	85.77	19.79	91.54	19.85	93.54	7.78
900	22	3.91	13.64	3.92	16.53	3.92	18.21	4.58
177	14	2.14	6.76	2.15	9.01	2.15	10.58	3.82
855	24	7.00	18.48	7.05	19.94	7.12	20.83	2.34
75	15	3.80	9.76	3.81	11.16	3.83	11.97	2.20
170	20	5.51	14.51	5.55	15.88	5.60	16.63	2.12
642	16	4.31	10.88	4.32	12.19	4.33	12.93	2.05
537	9	1.39	3.62	1.39	4.68	1.41	5.66	2.04
Michigan Data								
Site Number	Y_i	Poisson-Gamma		Poisson-Lognormal		Poisson-Inverse Gamma		Max. difference btw predicted E(m_i|X,y)
		E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	
109	7	0.68	2.59	0.69	3.86	0.70	4.64	2.05
997	7	0.84	2.98	0.85	4.10	0.87	4.82	1.84
64	5	0.25	0.89	0.25	1.70	0.25	2.53	1.63
498	6	0.65	2.19	0.65	3.14	0.66	3.74	1.55
544	5	0.39	1.30	0.39	2.04	0.40	2.74	1.44
761	3	14.81	4.16	15.06	4.56	15.32	5.45	1.29
542	11	2.45	7.62	2.48	8.48	2.53	8.87	1.24
895	29	15.43	27.73	15.57	28.36	15.89	27.24	1.12
Indiana Data								
Site Number	Y_i	Poisson-Gamma		Poisson-Lognormal		Poisson-Inverse Gamma		Max. difference btw predicted E(m_i|X,y)
		E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	E(μ_i)	E(m_i|X,y)	
247	14	0.26	2.88	0.09	8.69	0.03	12.17	9.28
258	27	2.64	19.74	2.66	24.69	3.06	25.24	5.50
288	11	0.71	4.68	0.23	6.96	0.09	9.05	4.37
302	122	116.30	121.94	148.39	122.97	173.99	118.90	4.07
297	215	188.96	214.91	245.48	216.75	287.52	212.70	4.05
67	108	61.55	107.19	72.64	105.03	85.55	108.49	3.46
299	265	155.65	264.28	202.72	266.32	237.48	263.30	3.02
63	121	47.82	119.20	53.69	118.77	54.84	121.60	2.84

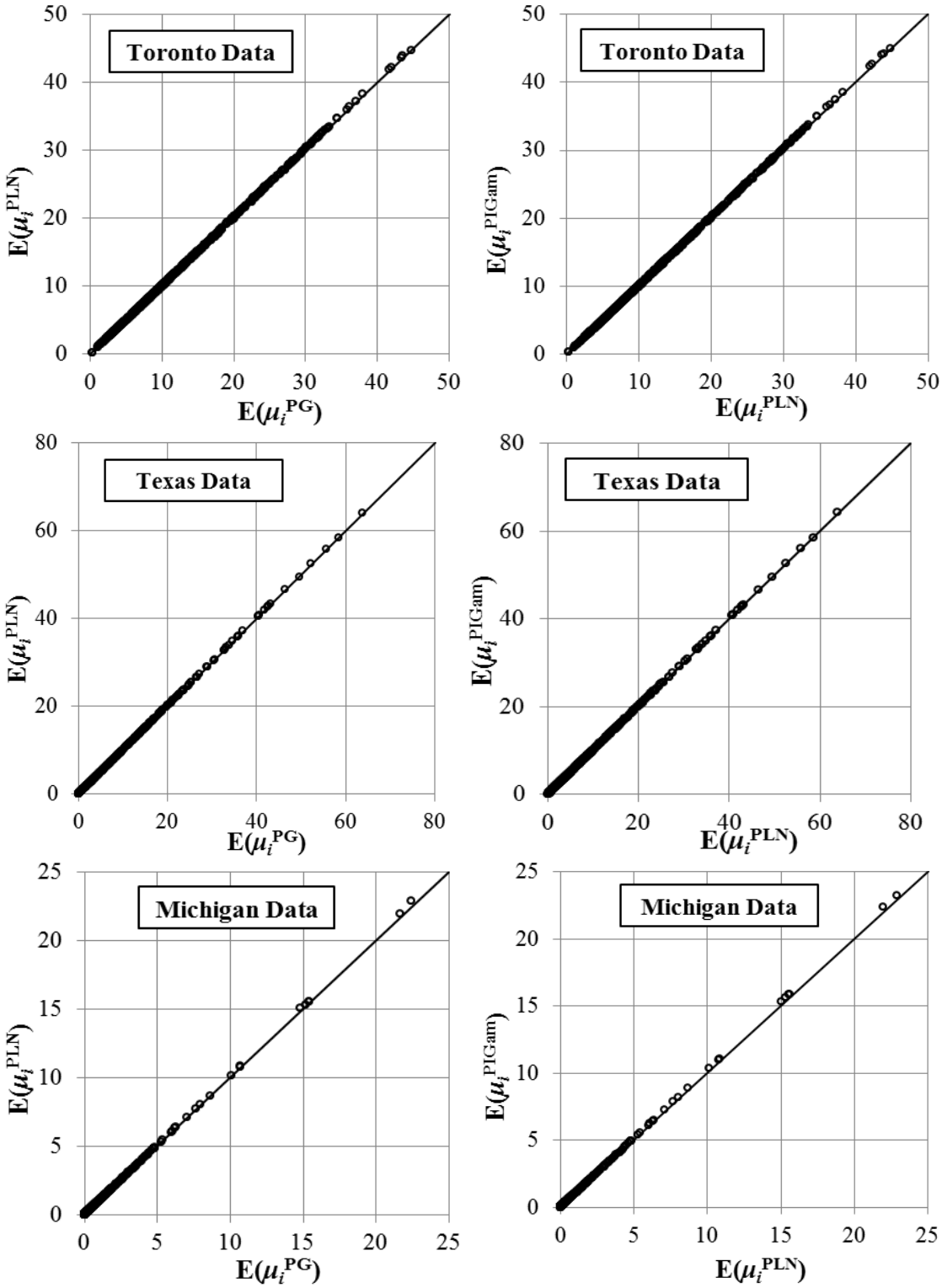


Figure 2. Comparison of posterior $E(\mu_i)$'s obtained by each model (except for Indiana data)

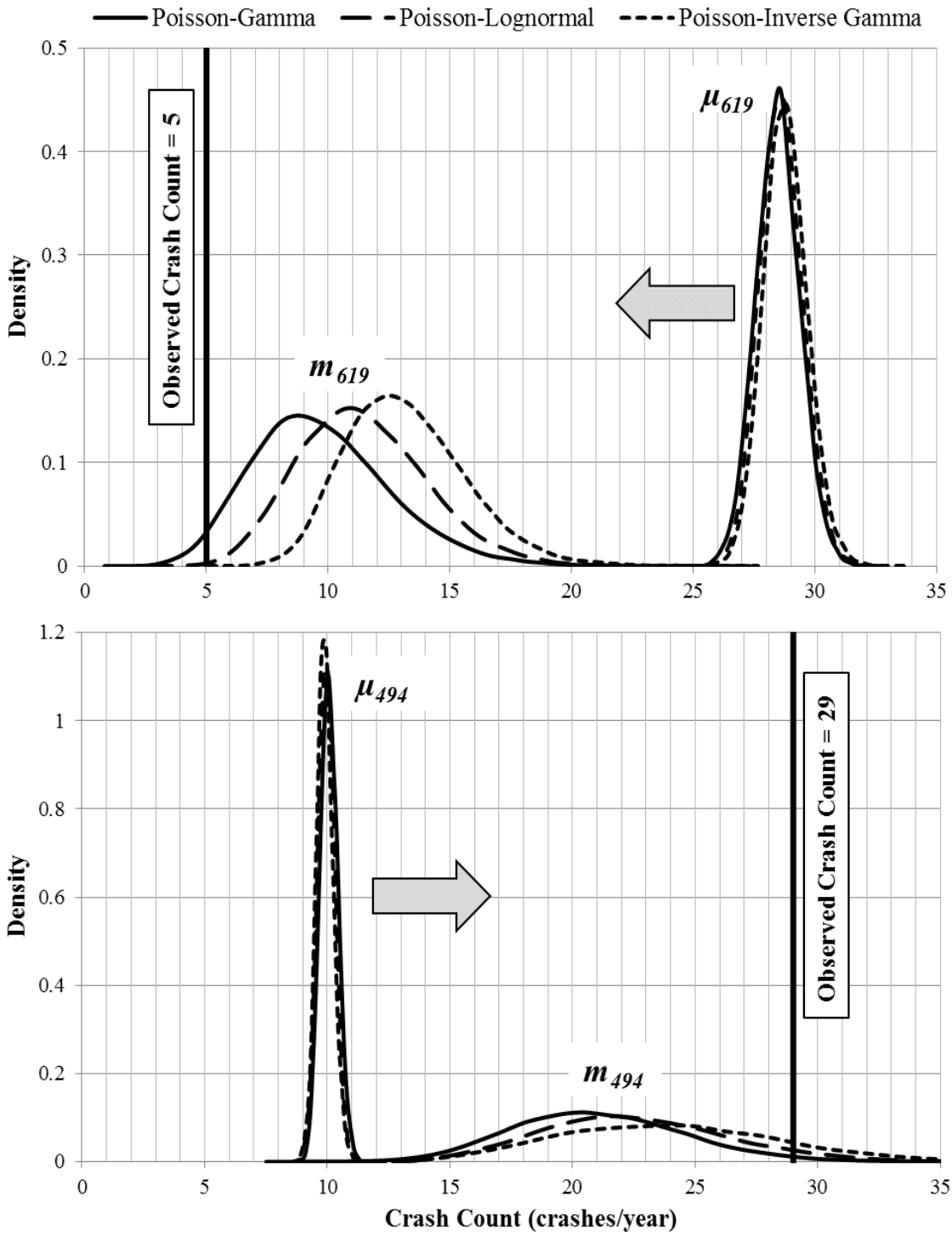


Figure 3. Posterior distribution of expected crash frequency ($m_i|X,y$) and μ_i for Site #619 & Site #494 in the Toronto dataset

For both sites in Figure 3, as well as all other sites except those in the Indiana dataset, the posterior distribution of μ_i was quite similar across the three models. However, for Site #619, which experienced substantially fewer crashes than the estimated $E(\mu_i)$, the distribution of expected crash frequency from the PG model (m_i^{PG}) shifted toward the observed crash count more than did the distribution of m_i^{PLN} and m_i^{PIGam}, in order. Conversely, for Site #494, which experienced substantially more crashes than the estimated $E(\mu_i)$, the distribution of m_i^{PG} shifted toward the observed crash count less than did the distribution of m_i^{PLN} and m_i^{PIGam}, in order. In both cases, the aforementioned phenomenon would result in $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$. As Table 6 indicates, this trend was consistent among all sites (in Toronto, Texas, and Michigan) where the models predictions for the expected crash frequency were significantly different.

It is of interest at this point to examine whether or not the aforementioned trend exists for all sites. Figure 4 to Figure 6 are constructed to illustrate how the relationship between $E(m_i|X,y)$ from different models is affected by the difference between the observed crash frequency and the posterior $E(\mu_i)$. For each dataset (except Indiana), $E(m_i|X,y)$'s from every pair of models are compared in a separate scatter plot. In all these plots, the horizontal axis represents the difference between the observed crash frequency (Y_i) and the average of the $E(\mu_i)$'s estimated by using each of the two models considered. It must be noted that reference to $Y_i - E(\mu_i)$ as "residual" is avoided to prevent confusion with $Y_i - E(m_i|X,y)$ (both terms may be interpreted as residual). As discussed earlier, the difference between the $E(\mu_i)$'s estimated using any two of the models were quite small; therefore, the horizontal axis is a measure of the difference between the observed crash frequency and $E(\mu_i)$ estimated from either of the models. The vertical axis indicates the difference between the $E(m_i|X,y)$'s estimated using the two models being compared.

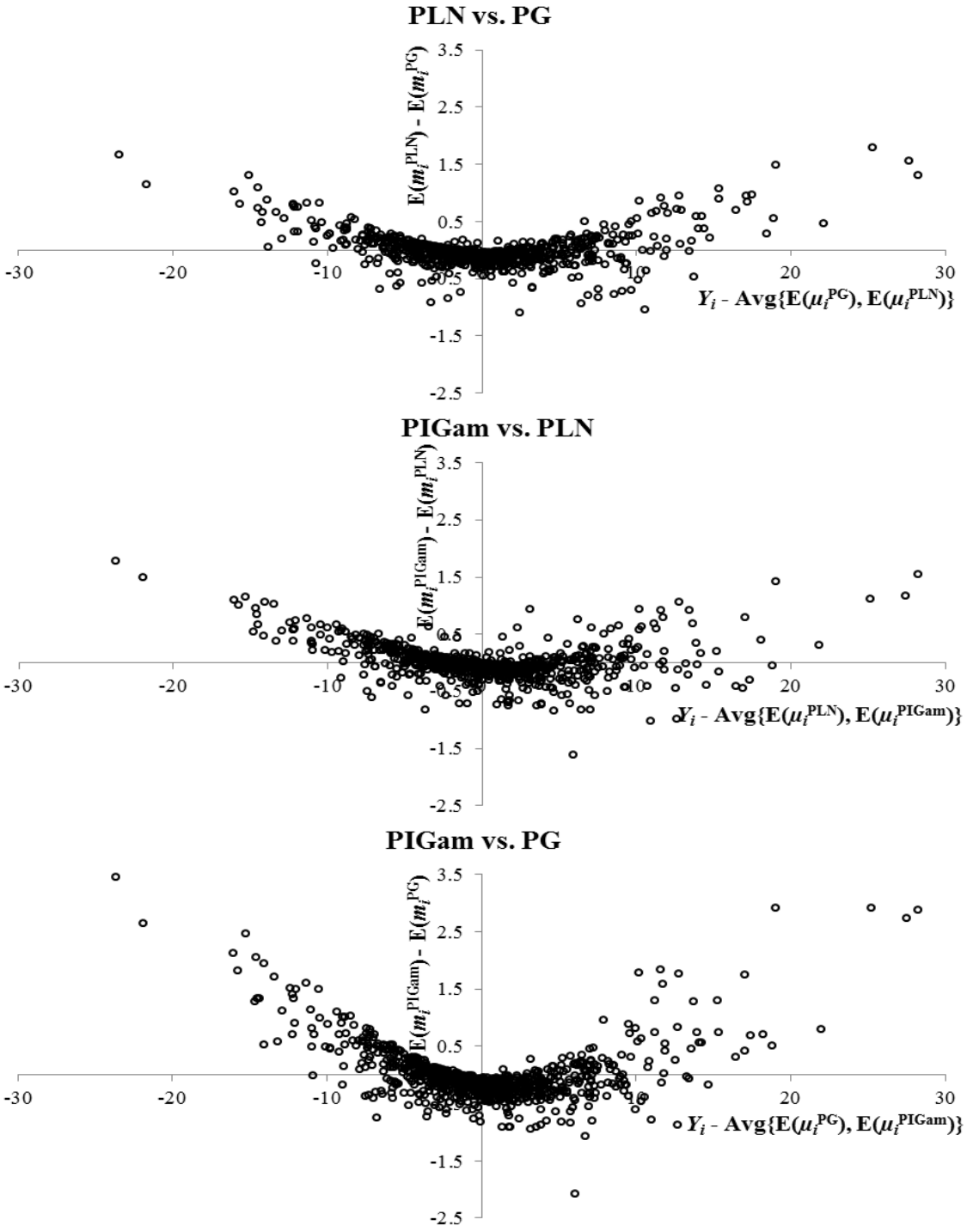


Figure 4. Difference between expected crash frequencies ($E(m_i|X,y)$'s) estimated using any two of the candidate models as a function of the difference between observed crash frequency and the posterior $E(\mu_i)$ estimated using either of the models for the Toronto dataset

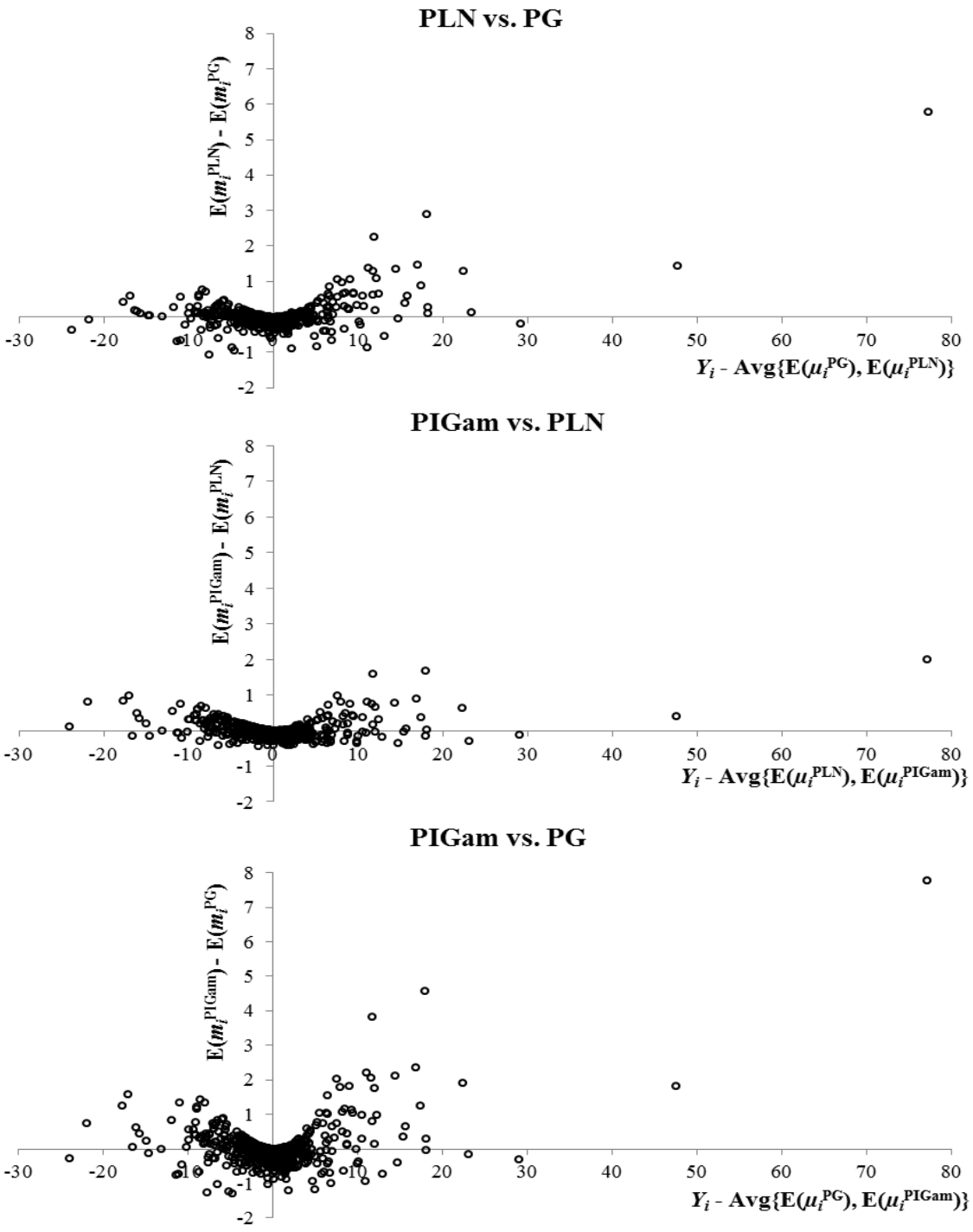


Figure 5. Difference between expected crash frequencies ($E(m_i|X,y)$'s) estimated using any two of the candidate models as a function of the difference between observed crash frequency and the posterior $E(\mu_i)$ estimated using either of the models for the Texas dataset

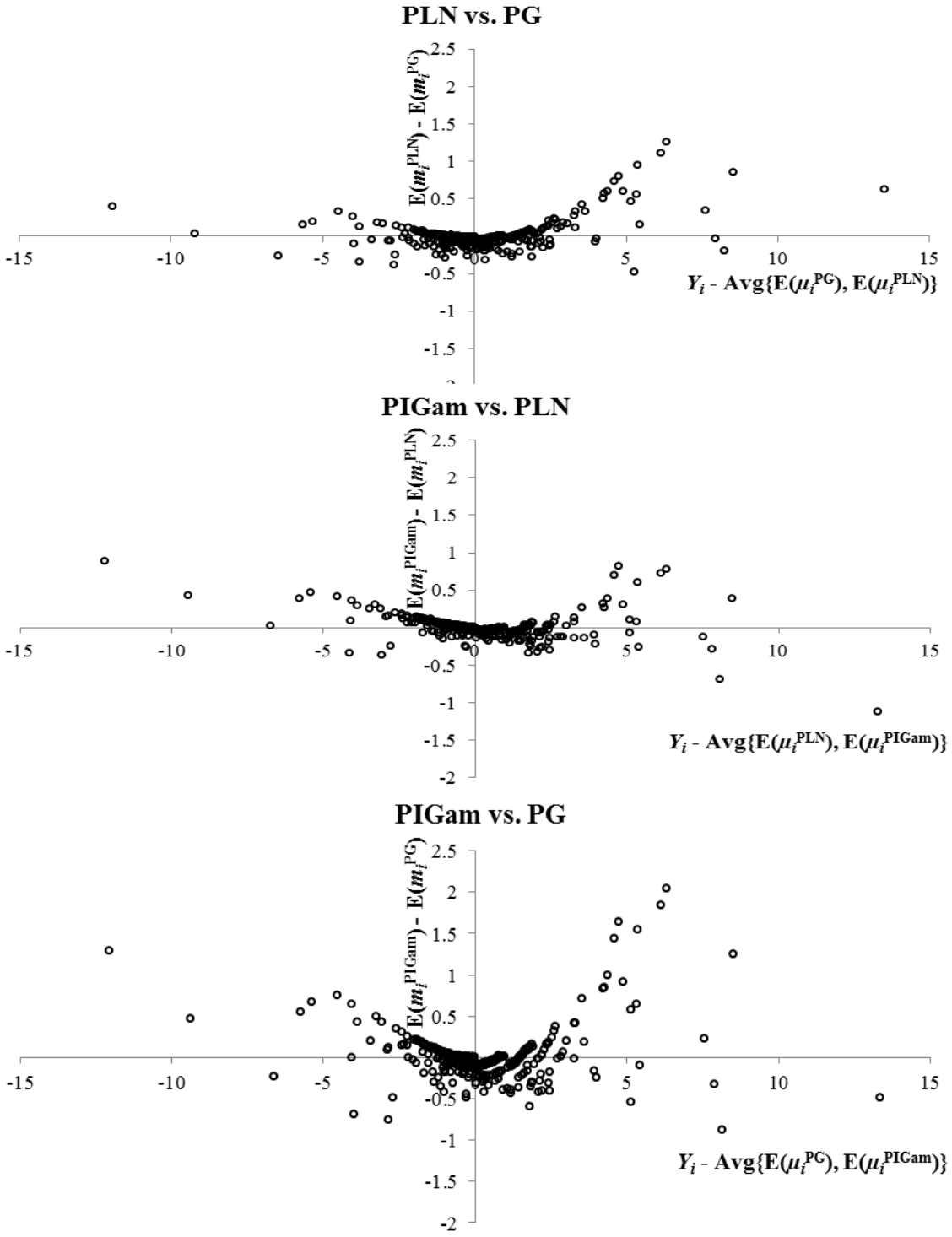


Figure 6. Difference between expected crash frequencies (E(m_i|X,y)'s) estimated using any two of the candidate models as a function of the difference between observed crash frequency and the posterior E(μ_i) estimated using either of the models for the Michigan dataset

Figure 4–6 reveal an important pattern regarding the site-specific predictions of the alternative models relative to one another: the previously described pattern (exemplified by Figure 3), which results in $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$, becomes more prevalent and pronounced as the difference between the observed crash count and posterior $E(\mu_i)$ increases. The trend dissipates as Y_i - $E(\mu_i)$ becomes smaller and even reverses as it approaches zero, resulting in a large number of sites where $E(m_i^{PG}|X,y)$ is slightly larger than $E(m_i^{PLN}|X,y)$, and $E(m_i^{PLN}|X,y)$ is slightly larger than $E(m_i^{PIGam}|X,y)$. Consequently, the positive and negative differences between the predictions of different models balance off and none the models predict an overall higher or lower crash frequency compared to another. As Table 7 indicates, the total expected crash frequency (over all sites combined) is virtually equal across the three models and very close to the total observed crash frequency, indicating that none of the models predicts significantly biased m_i's for any of the datasets (including the Indiana dataset).

Table 7. Total expected crash frequency ($E(m_i)$) over all sites predicted by each model for each dataset

Dataset	Total Expected Crash Frequency, $\sum_{i=1}^{n} E(m_i)$			Total Observed Crash Frequency, $\sum_{i=1}^{n} Y_i$
	PG Model	PLN Model	PIGam Model	
Toronto	10031.23	10026.98	10030.67	10030
Texas	5217.45	5221.59	5221.27	5219
Michigan	795.77	795.54	797.22	796
Indiana	5736.61	5738.24	5736.41	5737

As Figure 4 to Figure 6 suggest, it is only for the sites with observed crash frequencies significantly different from their $E(\mu_i)$ that the models may deliver considerably different predictions; the $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship observed for most of such sites is, therefore, much more important than the relationship for the other sites. However, it should be noted that while the models may deliver significantly different predictions only for the sites with substantially different than expected

observed crash counts (as indicated in Table 6), such unusually higher or lower than expected observed crash counts do not necessarily guarantee a $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship.

Finally, an explanation is due for why the models did not exhibit the same prediction patterns for the Indiana dataset. Among the considered datasets, the Indiana dataset was unique in that it contained six covariates (compared to one or two in the other datasets) and was plagued by extreme over-dispersion (conditional on the mean). These traits gave rise to a vector of regression coefficients that is highly sensitive to the model choice. As a result, the posterior $E(\mu_i)$'s could vary drastically across the alternative models, invalidating the type of analysis in Figure 3 to Figure 6. In reality, a crash dataset with such very large over-dispersion given the number of covariates (six) is relatively rare. The Indiana dataset was included in the analysis to explore the comparative performance of the models under extreme over-dispersion scenarios, and the results warn that the patterns described in this section must not be extended to any dataset before ensuring that the models estimate fairly similar values for the vector of coefficients.

6.2.2 Impact on Ranking of Sites by Expected Crash Frequency

As described earlier, the distinctions between the alternative models predictions will directly influence an important application of crash prediction models: ranking of sites by their expected crash frequency and hence priority for safety treatment.

The major distinction between the rankings of sites by the alternative models is attributable to the sites at which the alternative models predict substantially different expectations for crash frequency. Previously, these sites were characterized as those for which Y_i is much smaller or greater than the posterior $E(\mu_i)$. In this section, since the focus is on hazardous sites, the analysis will be confined to sites with a Y_i much greater than $E(\mu_i)$. These sites can be interpreted as sites that have experienced an unusually

high number of crashes given the prevailing traffic/geometric/environmental conditions. Therefore, this section compares the hazardous site rankings across the alternative models for the sites with the highest $Y_i - E(\mu_i)$ (i.e., most unusually high number of observed crashes) in each dataset.

Tables 8–10 list the 30 sites with the greatest $Y_i - E(\mu_i)$ in the Toronto, Texas, and Michigan datasets, respectively, and indicate the 'unsafety' rank of every site by each of the alternative models. Similar to the previous analysis, $E(\mu_i)$ in $Y_i - E(\mu_i)$ was estimated as the average of $E(\mu_i)$'s predicted by each of the three alternative models, which are indeed very close to one another (see Figure 2). The Indiana dataset was again excluded because $E(\mu_i)$'s varied considerably across the models.

According to Tables 8–10, at all sites (with large $Y_i - E(\mu_i)$) with significantly different ranks by the alternative models, the PIGam model assigned a higher rank than did the PLN and PG models, in order. At sites where the aforementioned pattern was not observed, the site rank varied only slightly across the alternative models. Please note that from the practical standpoint, the variation in the ranking of sites is less important for lower-ranked (i.e., less hazardous) sites.

The results of the analysis clearly demonstrate that sites with unusually high number of observed crashes (given the site's traffic/geometric/environmental features) are likely to be ranked higher (in terms of expected crash frequency) when the PIGam model is used instead of the PLN model, and similarly when the PLN model is used instead of the PG model. The author believes that the described trend in ranking of sites by crash proneness is the most important practical implication of using different members of the Bayesian Poisson-hierarchical family of regression models in highway safety analysis. Tables 11–14 list the most hazardous sites and their rank by each of the alternative models for each dataset. As indicated by these tables, the model choice can have a significant impact on the ranking of hazardous sites.

Table 8. Alternative model rankings of sites with the greatest Y_i – Avg[E(μ_i)] in the Toronto dataset

Site Number	Y_i	E(μ_i)			Y_i – Avg[E(μ_i)]*	E(m_i|X,y)			Rank		
		PG	PLN	PIGam		PG	PLN	PIGam	PG	PLN	PIGam
757	48	19.80	19.74	19.70	28.26	40.48	41.80	43.35	15	12	9
533	53	25.30	25.46	25.64	27.53	46.89	48.46	49.63	7	5	3
424	42	16.61	16.78	16.97	25.21	34.31	36.11	37.22	28	24	20
402	54	31.76	32.03	32.33	21.96	49.89	50.36	50.68	1	1	1
494	29	10.07	9.98	9.91	19.01	21.10	22.58	24.01	125	104	91
480	42	23.09	23.16	23.24	18.84	37.48	38.04	37.99	19	18	19
101	51	32.47	32.75	33.06	18.24	47.66	47.96	48.36	5	6	5
217	38	20.48	20.60	20.73	17.40	33.41	34.39	34.10	34	28	29
456	29	11.84	11.93	12.03	17.07	22.53	23.49	24.28	105	93	86
644	40	22.76	22.97	23.19	17.02	35.83	36.69	36.25	23	20	23
32	36	19.56	19.55	19.56	16.44	31.59	32.30	31.90	40	37	38
170	33	17.65	17.64	17.65	15.35	28.57	29.46	29.31	54	51	50
446	26	10.64	10.72	10.81	15.27	19.83	20.92	21.12	144	128	119
10	34	19.14	19.32	19.51	14.67	30.00	30.23	29.83	47	46	48
543	35	20.58	20.71	20.86	14.28	31.29	31.68	31.85	44	39	39
395	38	23.70	23.90	24.12	14.09	34.67	35.27	35.23	26	26	26
669	39	24.93	25.05	25.18	13.95	35.82	36.20	36.55	24	22	21
175	47	33.07	33.24	33.44	13.75	44.49	45.08	45.76	8	8	8
88	45	31.13	31.43	31.75	13.57	42.38	41.91	42.83	11	11	11
85	37	23.31	23.54	23.79	13.45	33.78	33.94	33.72	30	29	31
11	30	16.55	16.68	16.83	13.31	25.92	25.91	25.90	71	69	69
317	22	9.09	9.18	9.28	12.82	16.34	17.05	18.11	190	183	169
452	24	11.24	11.29	11.35	12.71	19.00	19.95	19.83	156	140	143
661	46	33.14	33.29	33.47	12.70	43.78	43.89	42.91	9	9	10
548	29	16.39	16.41	16.44	12.59	25.16	25.88	25.42	78	70	74
719	36	23.93	24.11	24.31	11.88	33.18	33.83	33.71	35	30	32
792	29	17.02	17.13	17.26	11.86	25.47	25.67	25.88	76	73	70
537	29	17.01	17.15	17.30	11.85	25.44	25.44	25.46	77	77	73
847	17	5.26	5.24	5.22	11.76	10.23	11.01	11.81	365	331	307
732	44	32.06	32.31	32.59	11.68	41.87	41.78	41.74	12	13	13

* Avg[E(μ_i)] = Average[E(μ_i^{PG}), E(μ_i^{PLN}), E(μ_i^{PIGam})]

Table 9. Alternative model rankings of sites with the greatest Y_i – Avg[E(μ_i)] in the Texas dataset

Site Number	Y_i	E(μ_i)			Y_i – Avg[E(μ_i)]*	E(m_i|X,y)			Rank		
		PG	PLN	PIGam		PG	PLN	PIGam	PG	PLN	PIGam
905	97	19.73	19.79	19.85	77.21	85.77	91.54	93.54	1	1	1
12	81	33.29	33.37	33.44	47.63	76.63	78.08	78.46	2	2	2
769	72	42.78	42.87	42.93	29.14	69.92	69.74	69.61	3	3	3
171	64	40.58	40.77	41.00	23.22	62.20	62.33	62.03	4	4	4
446	34	11.48	11.56	11.65	22.44	28.88	30.18	30.80	27	23	22
159	38	19.72	19.80	19.90	18.19	35.33	35.60	35.61	16	16	16
709	53	34.61	34.80	35.04	18.18	51.31	51.42	51.27	8	8	8
900	22	3.91	3.92	3.92	18.08	13.64	16.53	18.21	100	67	59
511	29	11.50	11.54	11.58	17.46	25.04	25.91	26.29	34	33	32
855	24	7.00	7.05	7.12	16.94	18.48	19.94	20.83	58	51	46
839	31	15.20	15.25	15.30	15.75	28.13	28.74	28.77	28	28	28
19	38	22.43	22.51	22.60	15.49	35.98	36.37	36.33	14	14	14
375	58	43.23	43.23	43.16	14.80	56.87	56.82	56.47	6	6	6
170	20	5.51	5.55	5.60	14.45	14.51	15.88	16.63	91	75	67
179	32	18.88	18.99	19.11	13.01	29.98	29.43	29.25	24	26	25
143	20	7.59	7.59	7.57	12.42	16.23	16.89	17.20	71	64	64
480	17	4.85	4.89	4.94	12.11	12.05	13.15	13.81	110	103	96
262	25	12.91	12.94	12.97	12.06	22.51	22.70	22.66	42	40	40
177	14	2.14	2.15	2.15	11.85	6.76	9.01	10.58	218	162	135
585	20	8.20	8.22	8.24	11.78	16.60	17.23	17.39	68	62	63
642	16	4.31	4.32	4.33	11.68	10.88	12.19	12.93	123	111	105
75	15	3.80	3.81	3.83	11.19	9.76	11.16	11.97	145	121	114
818	28	16.85	16.95	17.09	11.03	26.12	25.28	25.20	33	34	34
108	22	11.35	11.37	11.37	10.64	19.58	19.89	19.86	53	52	51
595	16	5.40	5.44	5.49	10.56	11.95	12.56	12.94	111	108	104
76	31	20.71	20.82	20.96	10.17	29.60	29.38	29.22	26	27	27
591	35	24.84	24.96	25.12	10.03	33.79	33.66	33.36	17	17	17
817	43	33.07	33.10	33.09	9.91	42.12	42.46	42.11	10	10	10
708	14	4.49	4.51	4.55	9.48	9.94	10.59	10.97	141	131	125
764	14	4.60	4.61	4.63	9.39	10.03	10.73	11.15	140	128	120

* Avg[E(μ_i)] = Average[E(μ_i^{PG}), E(μ_i^{PLN}), E(μ_i^{PIGam})]

Table 10. Alternative model rankings of sites with the greatest Y_i – Avg[E(μ_i)] in the Michigan dataset

Site Number	Y_i	E(μ_i)			Y_i – Avg[E(μ_i)]*	E(m_i|X,y)			Rank		
		PG	PLN	PIGam		PG	PLN	PIGam	PG	PLN	PIGam
895	29	15.43	15.57	15.89	13.37	27.73	28.36	27.24	1	1	1
542	11	2.45	2.48	2.53	8.51	7.62	8.48	8.87	10	10	10
899	19	10.67	10.85	11.03	8.15	17.92	17.73	17.04	4	4	4
218	15	7.01	7.10	7.23	7.89	13.49	13.46	13.17	6	6	6
911	12	4.37	4.44	4.52	7.56	9.94	10.29	10.17	8	8	8
109	7	0.68	0.69	0.70	6.31	2.59	3.86	4.64	59	32	26
997	7	0.84	0.85	0.87	6.14	2.98	4.10	4.82	46	27	22
334	9	3.53	3.56	3.63	5.43	7.29	7.45	7.19	11	11	12
498	6	0.65	0.65	0.66	5.35	2.19	3.14	3.74	74	42	32
399	7	1.66	1.68	1.71	5.32	4.38	4.95	5.03	24	21	20
114	16	10.71	10.79	11.02	5.16	15.29	14.82	14.75	5	5	5
389	7	1.83	1.87	1.90	5.13	4.57	5.05	5.15	23	19	19
963	6	1.09	1.10	1.12	4.90	3.09	3.69	4.00	43	33	29
64	5	0.25	0.25	0.25	4.75	0.89	1.70	2.53	205	104	65
544	5	0.39	0.39	0.40	4.61	1.30	2.04	2.74	134	84	57
559	5	0.62	0.63	0.64	4.37	1.86	2.46	2.85	94	67	52
665	5	0.73	0.74	0.75	4.26	2.06	2.64	2.90	86	58	47
955	5	0.75	0.76	0.77	4.24	2.12	2.63	2.96	79	59	46
353	7	2.96	3.00	3.06	3.99	5.58	5.55	5.34	17	15	17
549	7	3.03	3.05	3.11	3.94	5.62	5.54	5.45	16	16	15
943	5	1.37	1.38	1.41	3.61	3.05	3.38	3.24	44	38	39
124	4	0.45	0.45	0.46	3.55	1.23	1.67	1.94	142	107	89
713	4	0.70	0.71	0.72	3.29	1.71	2.04	2.12	104	83	81
182	4	0.71	0.72	0.73	3.28	1.73	2.02	2.14	101	87	80
762	5	1.69	1.72	1.75	3.28	3.39	3.51	3.38	36	35	37
702	4	0.95	0.97	0.99	3.03	2.11	2.28	2.31	81	72	74
774	4	1.07	1.07	1.10	2.92	2.25	2.43	2.31	72	68	75
691	4	1.16	1.17	1.19	2.83	2.36	2.49	2.37	65	64	73
359	4	1.25	1.27	1.29	2.73	2.46	2.57	2.45	63	61	68
857	3	0.33	0.34	0.34	2.66	0.79	1.02	1.17	240	174	152

* Avg[E(μ_i)] = Average[E(μ_i^{PG}), E(μ_i^{PLN}), E(μ_i^{PIGam})]

Table 11. Alternative model rankings of the most hazardous sites in the Toronto dataset

Site Number	Y_i	$E(\mu_i)$			$E(m_i\|X,y)$			Rank		
		PG	PLN	PIGam	PG	PLN	PIGam	PG	PLN	PIGam
402	54	31.76	32.03	32.33	49.89	50.36	50.68	1	1	1
393	51	43.36	43.68	44.04	49.89	49.07	49.68	2	2	2
856	50	43.53	43.88	44.26	49.08	48.61	47.00	3	3	6
738	50	41.90	42.23	42.60	48.79	48.54	48.62	4	4	4
101	51	32.47	32.75	33.06	47.66	47.96	48.36	5	6	5
832	48	44.72	44.79	44.90	47.54	46.91	46.60	6	7	7
533	53	25.30	25.46	25.64	46.89	48.46	49.63	7	5	3
175	47	33.07	33.24	33.44	44.49	45.08	45.76	8	8	8
661	46	33.14	33.29	33.47	43.78	43.89	42.91	9	9	10
418	44	37.02	37.25	37.51	42.86	42.07	41.79	10	10	12
88	45	31.13	31.43	31.75	42.38	41.91	42.83	11	11	11
732	44	32.06	32.31	32.59	41.87	41.78	41.74	12	13	13
853	43	31.76	31.95	32.16	40.88	41.12	40.10	13	14	15
667	42	34.39	34.66	34.96	40.68	39.99	40.26	14	15	14
757	48	19.80	19.74	19.70	40.48	41.80	43.35	15	12	9
663	40	34.45	34.70	34.98	39.02	38.81	38.82	16	17	17
653	41	30.35	30.65	30.96	38.94	37.89	38.47	17	19	18
535	41	29.88	30.15	30.45	38.80	39.44	39.03	18	16	16
480	42	23.09	23.16	23.24	37.48	38.04	37.99	19	18	19
497	39	28.81	28.99	29.20	36.96	36.44	36.35	20	21	22
721	38	28.28	28.47	28.68	36.06	35.36	35.68	21	25	24
426	35	41.60	41.93	42.29	35.92	36.11	35.51	22	23	25
644	40	22.76	22.97	23.19	35.83	36.69	36.25	23	20	23
669	39	24.93	25.05	25.18	35.82	36.20	36.55	24	22	21
490	36	29.52	29.70	29.90	34.74	33.81	34.58	25	31	27
395	38	23.70	23.90	24.12	34.67	35.27	35.23	26	26	26
557	36	29.30	29.45	29.63	34.63	35.14	34.32	27	27	28
424	42	16.61	16.78	16.97	34.31	36.11	37.22	28	24	20
470	36	26.82	27.02	27.24	34.06	33.33	33.98	29	34	30
85	37	23.31	23.54	23.79	33.78	33.94	33.72	30	29	31

Table 12. Alternative model rankings of the most hazardous sites in the Texas dataset

Site Number	Y_i	$E(\mu_i)$			$E(m_i\|X,y)$			Rank		
		PG	PLN	PIGam	PG	PLN	PIGam	PG	PLN	PIGam
905	97	19.73	19.79	19.85	85.77	91.54	93.54	1	1	1
12	81	33.29	33.37	33.44	76.63	78.08	78.46	2	2	2
769	72	42.78	42.87	42.93	69.92	69.74	69.61	3	3	3
171	64	40.58	40.77	41.00	62.20	62.33	62.03	4	4	4
922	59	63.76	64.00	64.28	59.27	58.41	58.03	5	5	5
375	58	43.23	43.23	43.16	56.87	56.82	56.47	6	6	6
340	52	58.43	58.49	58.48	52.35	51.75	51.33	7	7	7
709	53	34.61	34.80	35.04	51.31	51.42	51.27	8	8	8
897	47	41.76	41.88	42.01	46.63	45.80	45.46	9	9	9
817	43	33.07	33.10	33.09	42.12	42.46	42.11	10	10	10
465	41	42.57	42.65	42.72	41.13	40.92	40.47	11	11	11
791	38	46.38	46.52	46.67	38.53	38.16	37.87	12	12	12
739	36	52.24	52.44	52.66	37.01	37.21	37.05	13	13	13
19	38	22.43	22.51	22.60	35.98	36.37	36.33	14	14	14
23	35	49.57	49.55	49.42	35.92	35.95	35.78	15	15	15
159	38	19.72	19.80	19.90	35.33	35.60	35.61	16	16	16
591	35	24.84	24.96	25.12	33.79	33.66	33.36	17	17	17
29	33	40.45	40.58	40.72	33.61	32.55	32.35	18	21	20
698	33	36.98	37.11	37.27	33.35	33.04	32.68	19	18	19
421	32	55.70	55.86	56.01	33.32	32.95	33.05	20	19	18
699	33	32.77	32.86	32.96	33.01	32.56	32.14	21	20	21
74	30	40.37	40.57	40.81	30.82	30.59	30.37	22	22	23
315	31	24.51	24.53	24.52	30.19	30.14	29.78	23	24	24
179	32	18.88	18.99	19.11	29.98	29.43	29.25	24	26	25
32	31	22.64	22.64	22.60	29.89	29.55	29.24	25	25	26
76	31	20.71	20.82	20.96	29.60	29.38	29.22	26	27	27
446	34	11.48	11.56	11.65	28.88	30.18	30.80	27	23	22
839	31	15.20	15.25	15.30	28.13	28.74	28.77	28	28	28
475	29	22.01	22.10	22.20	28.04	27.38	27.05	29	30	30
94	28	25.37	25.45	25.53	27.67	27.49	27.11	30	29	29

Table 13. Alternative model rankings of the most hazardous sites in the Michigan dataset

Site Number	Y_i	E(μ_i)			E(m_i|X,y)			Rank		
		PG	PLN	PIGam	PG	PLN	PIGam	PG	PLN	PIGam
895	29	15.43	15.57	15.89	27.73	28.36	27.24	1	1	1
906	20	22.44	22.91	23.24	20.14	19.76	19.39	2	2	2
120	18	21.63	21.94	22.34	18.24	17.90	17.56	3	3	3
899	19	10.67	10.85	11.03	17.92	17.73	17.04	4	4	4
114	16	10.71	10.79	11.02	15.29	14.82	14.75	5	5	5
218	15	7.01	7.10	7.23	13.49	13.46	13.17	6	6	6
289	12	10.08	10.13	10.35	11.71	11.44	11.11	7	7	7
911	12	4.37	4.44	4.52	9.94	10.29	10.17	8	8	8
232	9	15.37	15.58	15.87	9.60	9.35	9.38	9	9	9
542	11	2.45	2.48	2.53	7.62	8.48	8.87	10	10	10
334	9	3.53	3.56	3.63	7.29	7.45	7.19	11	11	12
467	6	15.15	15.29	15.60	6.88	6.91	7.34	12	12	11
401	6	8.59	8.68	8.85	6.42	6.18	5.94	13	13	13
555	6	6.20	6.28	6.40	6.01	5.75	5.52	14	14	14
516	6	4.74	4.81	4.90	5.68	5.40	5.28	15	17	18
549	7	3.03	3.05	3.11	5.62	5.54	5.45	16	16	15
353	7	2.96	3.00	3.06	5.58	5.55	5.34	17	15	17
398	5	6.03	6.12	6.23	5.21	4.96	4.86	18	20	21
603	5	5.96	5.98	6.11	5.20	4.92	4.77	19	22	24
793	6	3.47	3.50	3.57	5.20	5.09	4.80	20	18	23
254	4	7.94	8.01	8.18	4.67	4.56	4.66	21	23	25
669	5	3.78	3.81	3.89	4.62	4.38	4.18	22	25	27
389	7	1.83	1.87	1.90	4.57	5.05	5.15	23	19	19
399	7	1.66	1.68	1.71	4.38	4.95	5.03	24	21	20
705	4	5.25	5.31	5.41	4.31	4.07	4.01	25	28	28
24	5	2.87	2.90	2.96	4.24	4.15	3.82	26	26	30
666	5	2.75	2.77	2.83	4.18	4.03	3.77	27	29	31
761	3	14.81	15.06	15.32	4.16	4.56	5.45	28	24	16
403	4	4.21	4.25	4.34	4.06	3.87	3.62	29	31	34
10	5	2.52	2.55	2.60	4.04	3.92	3.73	30	30	33

Table 14. Alternative model rankings of the most hazardous sites in the Indiana dataset

Site Number	Y_i	$E(\mu_i)$			$E(m_i\|X,y)$			Rank		
		PG	PLN	PIGam	PG	PLN	PIGam	PG	PLN	PIGam
65	329	197.76	235.96	278.80	328.25	329.69	329.40	1	1	1
299	265	155.65	202.72	237.48	264.28	266.32	263.30	2	2	2
297	215	188.96	245.48	287.52	214.91	216.75	212.70	3	3	3
73	204	181.87	230.27	270.39	203.93	202.42	203.57	4	4	4
300	145	155.80	203.37	238.69	145.04	144.86	142.28	5	5	5
302	122	116.30	148.39	173.99	121.94	122.97	118.90	6	6	8
63	121	47.82	53.69	54.84	119.20	118.77	121.60	7	8	6
301	119	136.72	177.36	207.57	119.19	119.98	120.67	8	7	7
67	108	61.55	72.64	85.55	107.19	105.03	108.49	9	9	9
30	103	36.85	43.38	51.20	100.99	100.11	101.63	10	12	10
291	100	118.62	154.32	180.82	100.21	100.39	100.26	11	10	11
75	99	78.22	94.74	112.48	98.64	100.31	98.91	12	11	12
74	97	59.18	68.22	80.44	96.09	95.02	94.87	13	13	13
306	95	77.56	90.53	106.79	94.67	94.18	93.62	14	14	14
42	94	67.42	76.59	91.72	93.47	92.94	93.30	15	15	15
76	90	77.09	96.29	113.24	89.82	88.46	90.59	16	16	16
268	87	53.43	64.58	76.12	86.34	85.10	85.67	17	17	17
72	79	60.60	72.52	84.78	78.70	79.55	77.30	18	18	19
71	79	38.23	45.79	53.70	77.88	77.77	78.30	19	19	18
14	76	101.12	121.85	143.73	76.33	75.19	75.44	20	20	20
295	75	87.56	114.68	134.51	75.10	74.26	75.01	21	21	21
8	67	96.90	118.95	139.93	67.34	66.37	67.10	22	22	22
253	66	53.47	63.52	74.90	65.61	65.21	64.91	23	23	23
15	62	99.23	119.03	140.27	62.42	60.29	61.70	24	27	25
102	62	49.54	63.24	74.29	61.67	62.40	60.10	25	24	26
111	61	68.78	85.16	100.04	61.05	62.06	62.02	26	25	24
29	60	65.61	78.59	92.62	60.13	60.33	59.39	27	26	27
337	57	50.71	58.97	69.51	56.83	56.79	56.05	28	29	28
338	57	50.71	58.97	69.51	56.79	56.60	55.44	29	30	30
234	57	44.30	56.25	66.02	56.74	56.95	55.72	30	28	29

6.2.3 Mean-Variance Relationship

The mean-variance relationship for the observed crash frequency (Y_i) can be established by using the conditional variance identity:

$$\mathrm{Var}(Y_i) = \mathrm{E}(\mathrm{Var}(Y_i \mid m_i)) + \mathrm{Var}(\mathrm{E}(Y_i \mid m_i)) \tag{6.2}$$

$$\mathrm{Var}(Y_i) = \mathrm{E}(m_i) + \mathrm{Var}(m_i) \tag{6.3}$$

The posterior samples of m_i's were used to determine the posterior mean-variance relationship for sites in each dataset. As Figure 7 indicates, the mean-variance relationship for sites in the datasets was not noticeably affected by the model being used.

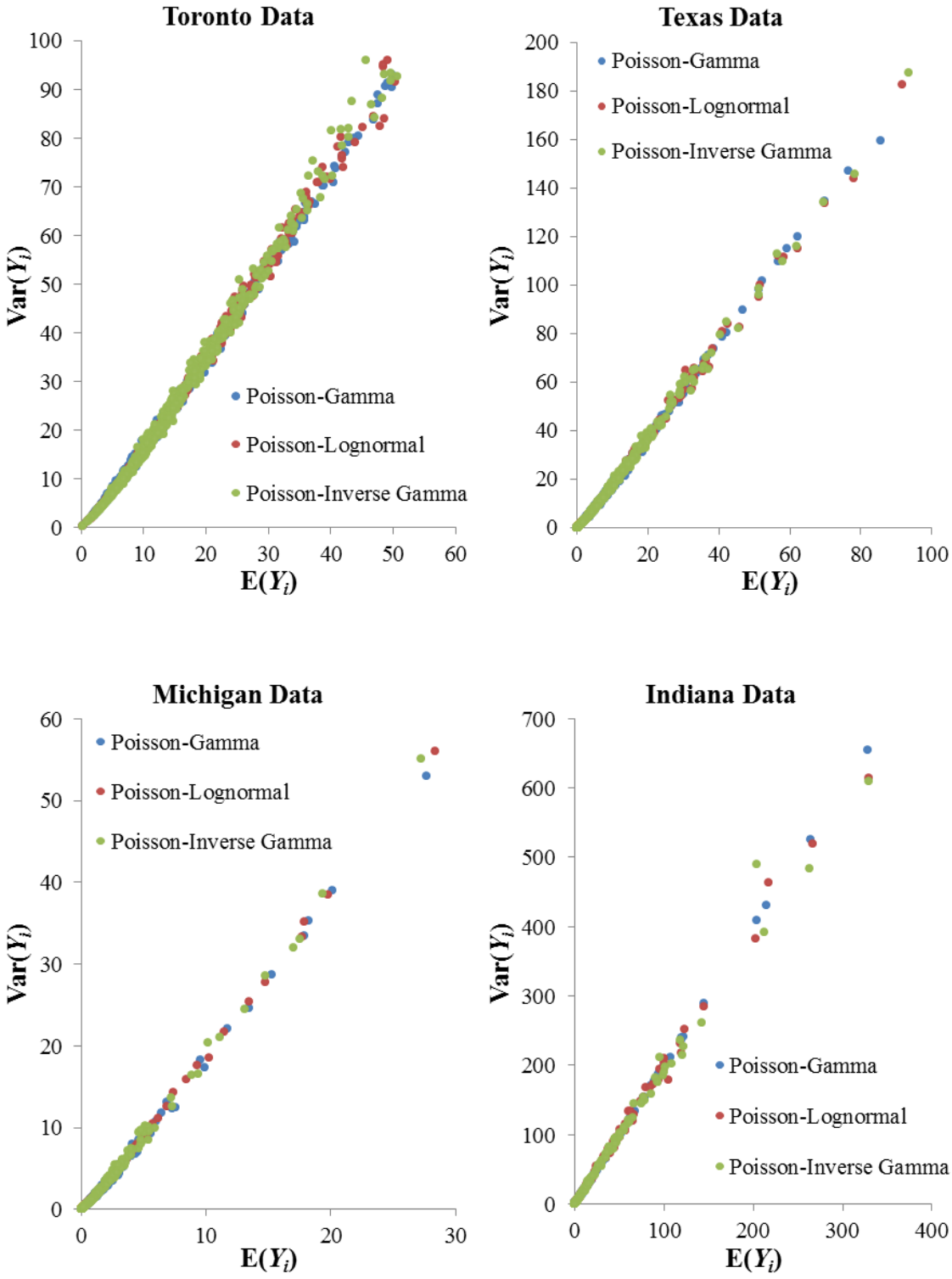


Figure 7. Posterior mean-variance relationship for each dataset and model

6.3 SITES WITHOUT CRASH DATA

This section includes two parts. First, the distinctions between the models predictions for expected crash frequency are analyzed for new sites without observed crash data. Second, the mean-variance relationship is plotted for the each model-dataset.

6.3.1 Distinctions between Expected Crash Frequency Predictions

The expected crash frequency at sites without crash history information (e.g. planned sites considered for construction) are typically predicted by regression models calibrated using a sample of sites of the same type and in the same jurisdiction. Without a site-specific observed crash count affecting the prediction, the expected crash frequency at a new site will follow a model-specific distribution (gamma, lognormal, or inverse-gamma) with mean $\mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$.

A convenient way to compare the models predictions for sites without observed crash count is to compare μ_i's, because μ_i is indeed equal to the expected crash frequency at a hypothetical new site with the same (modeled) characteristics as those of the *i*'th site in the dataset but with no available observed crash count. This comparison was carried out in Figure 2 and indicated that μ_i's did not differ drastically except for the Indiana dataset. However, an important trend will be revealed if the difference between the models predicted E(μ_i)'s are magnified. Figure 8 to Figure 11 are constructed to investigate the relationship between the models predicted E(μ_i)'s as a function of the magnitude of the predicted E(μ_i)'s. For each dataset, E(μ_i)'s from every pair of models are compared in a separate scatter plot. In all these plots, the horizontal axis represents the E(μ_i) predicted by one of the two model being compared, whereas the vertical axis indicates the difference between the two models predicted E(μ_i)'s.

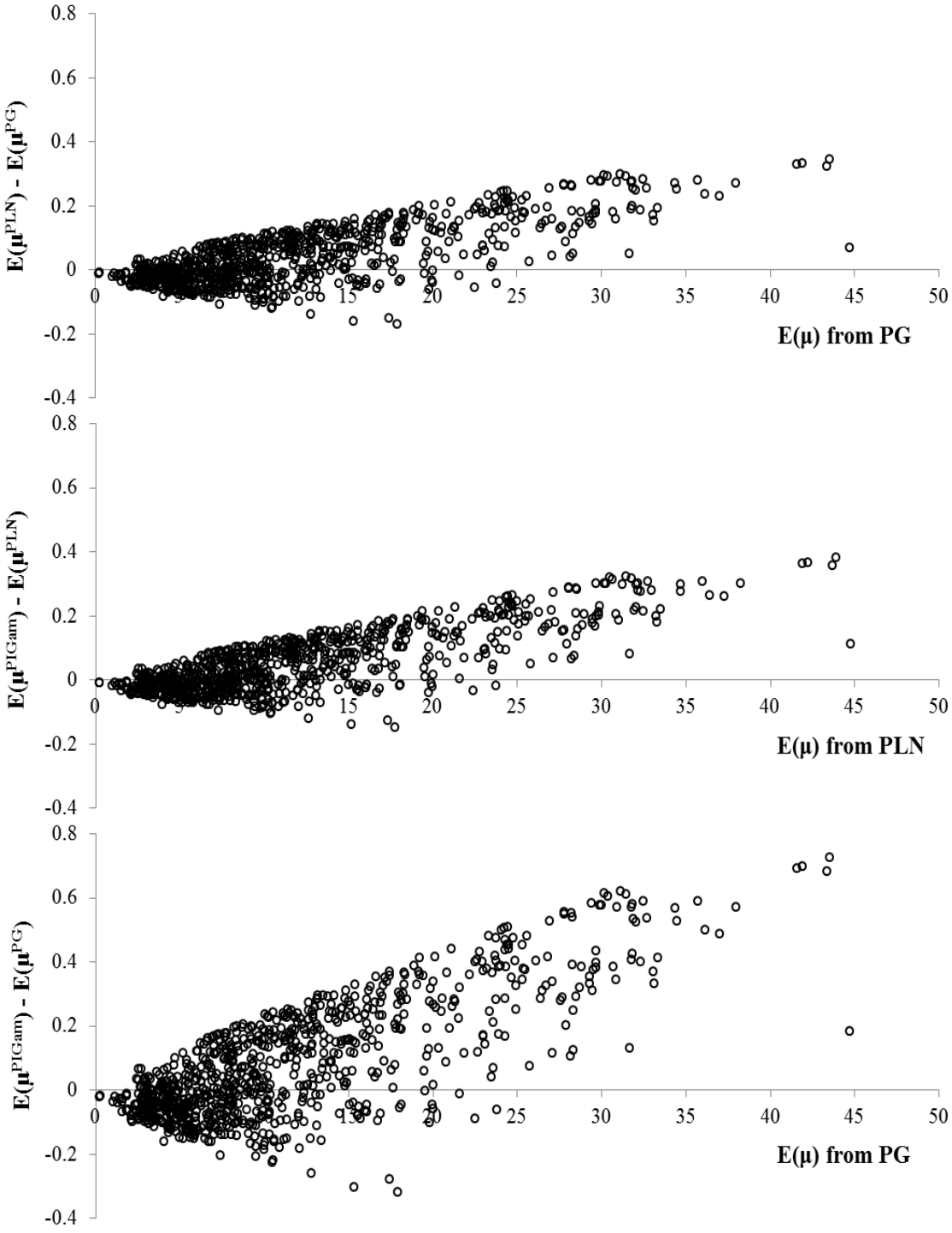


Figure 8. Difference between Toronto dataset E(μ_i)'s estimated using any two of the candidate models as a function of the E(μ_i)'s estimated from one of the two models considered

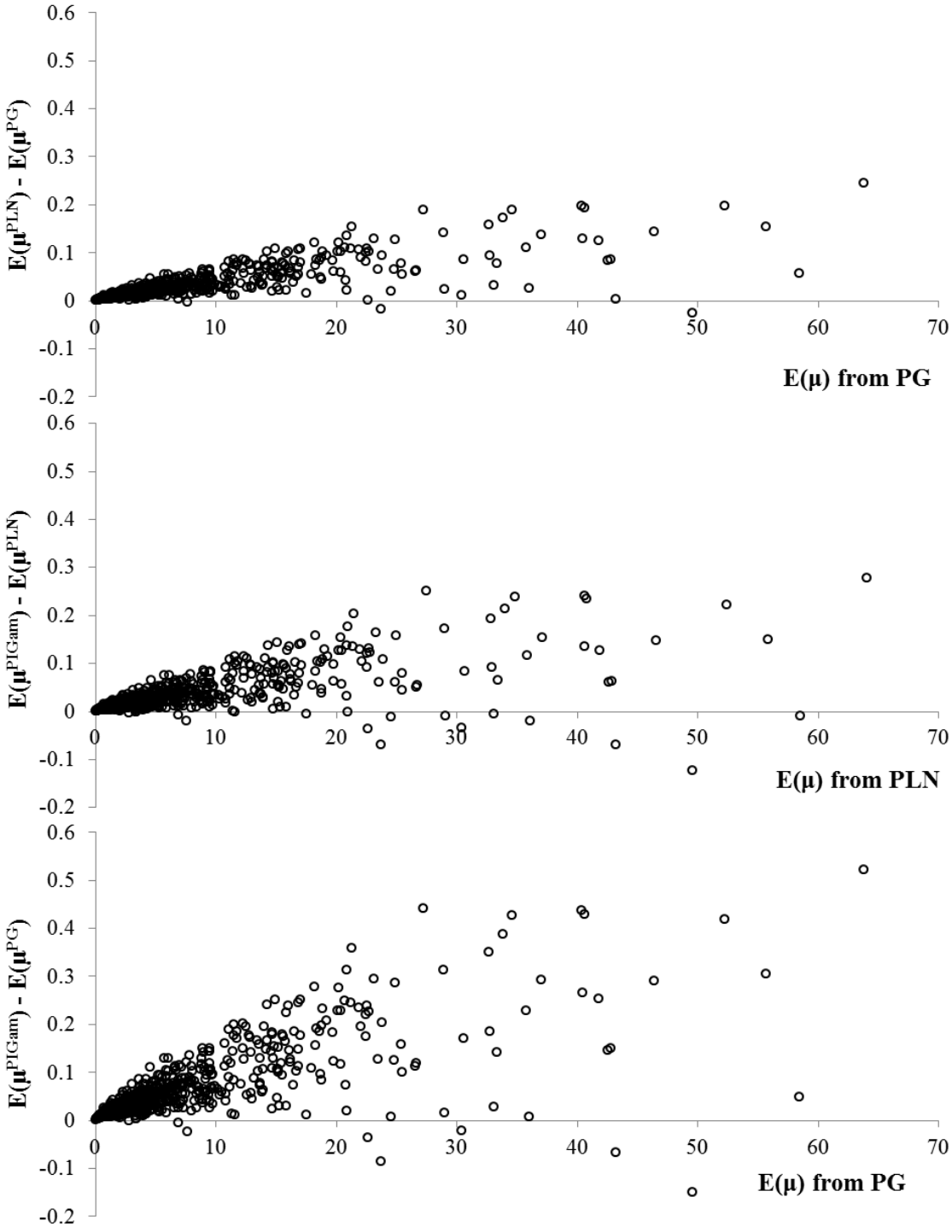


Figure 9. Difference between Texas dataset E(μ_i)'s estimated using any two of the candidate models as a function of the E(μ_i)'s estimated from one of the two models considered

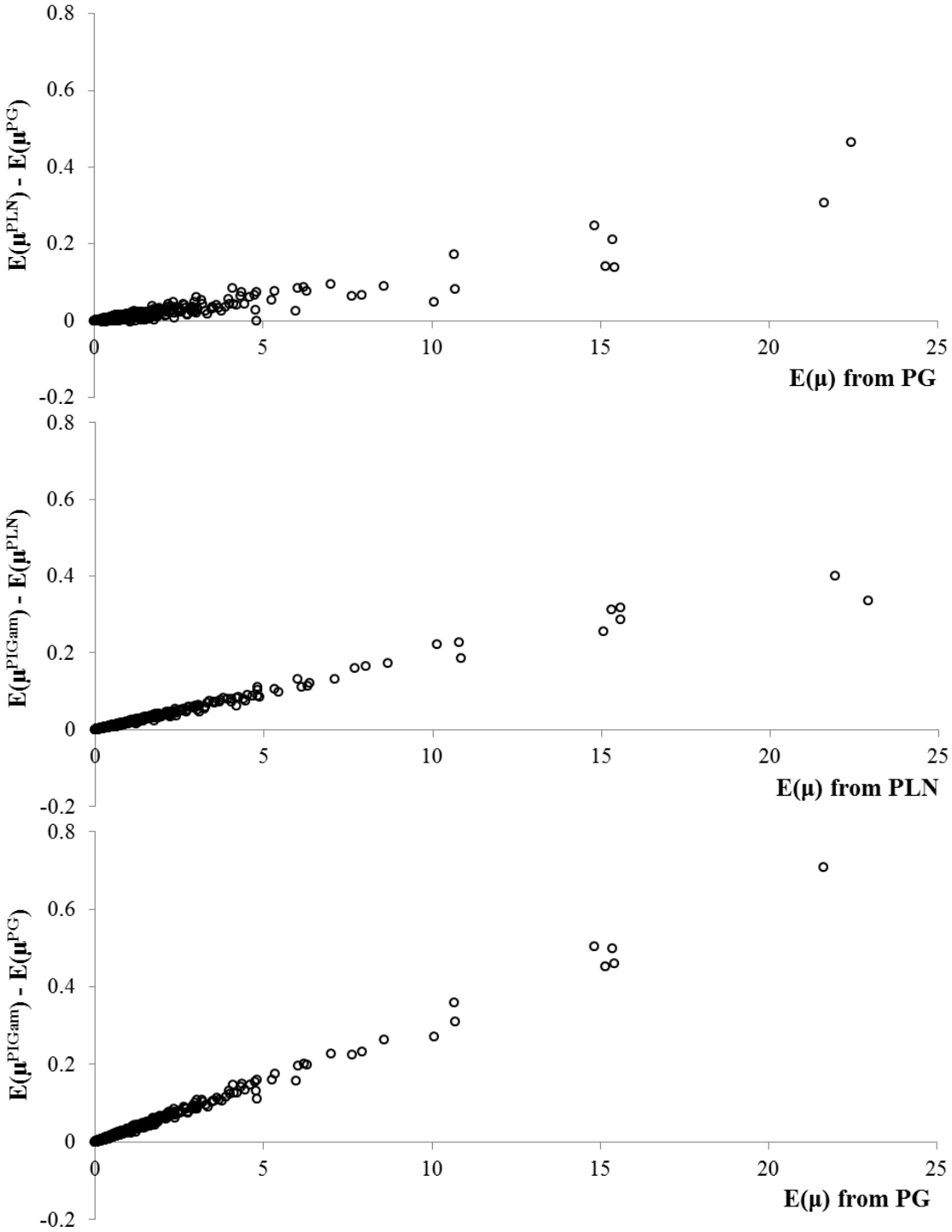


Figure 10. Difference between Michigan dataset E(μ_i)'s estimated using any two of the candidate models as a function of the E(μ_i)'s estimated from one of the two models considered

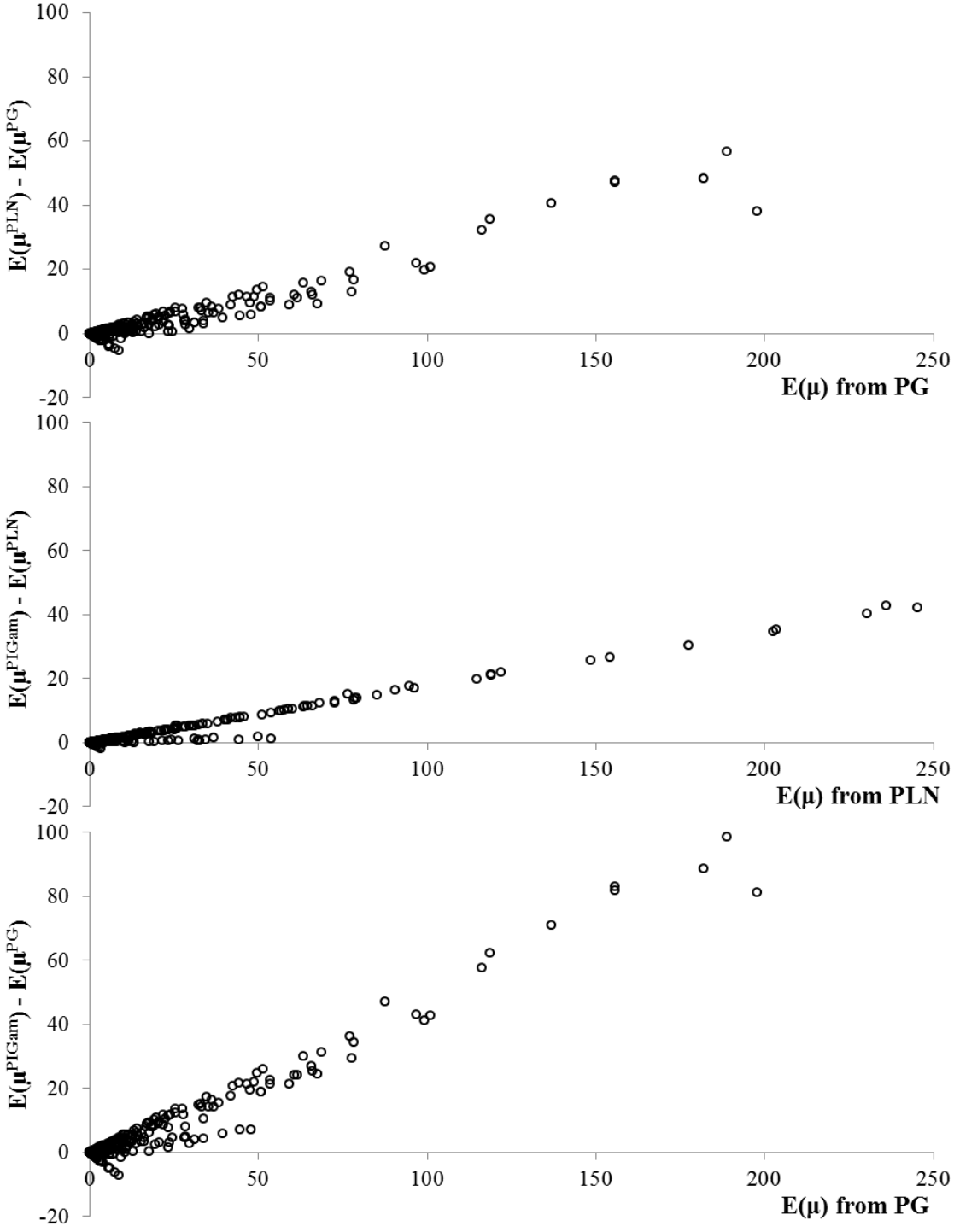


Figure 11. Difference between Indiana dataset E(μ_i)'s estimated using any two of the candidate models as a function of the E(μ_i)'s estimated from one of the two models considered

It was by no means unexpected that the difference between the models' $E(\mu_i)$ predictions increase with the increase in the magnitude of $E(\mu_i)$; the difference between the alternative models' coefficients for traffic volume causes a larger difference between the models' $E(\mu_i)$'s for sites with greater traffic volume and thus greater $E(\mu_i)$'s. However, Figure 8 to Figure 11 clearly indicate that the PLN model tends to predict larger μ_i's compared to the PG model and the PIGam model tends to predict larger μ_i's compared to PLN model. The figures also indicate that the $E(\mu_i{}^{PG}) < E(\mu_i{}^{PLN}) < E(\mu_i{}^{PIGam})$ relationship becomes more prevalent and pronounced for larger $E(\mu_i)$'s. Despite the magnitude of the differences between the models' $E(\mu_i)$'s being relatively small, the aforementioned relationship is an important finding because unlike the m_i predictions for sites with observed crash frequency (see Table 7), it results in each model over- or under-predicting μ_i's (collectively, over all sites combined) relative to the other two models.

Table 15. Total $E(\mu_i)$ over all sites predicted by each model for each dataset

Dataset	Total Expected Crash Frequency, $\sum_{i=1}^{n} E(\mu_i)$		
	PG Model	PLN Model	PIGam Model
Toronto	10003.38	10035.07	10075.94
Texas	5344.41	5366.73	5392.80
Michigan	824.07	833.63	849.61
Indiana	5306.06	6345.98	7332.74

It is evident from Table 15 that the total predicted $E(\mu_i)$ can vary significantly depending on the model being used. The PIGam model over-predicts the expected crash frequencies (for all sites combined) compared to the PLN model and the PG model, in order. Another important observation is that the degrees of variation between models' μ_i predictions increases as the datasets become more over-dispersed. As implied by Figure 8 to Figure 11 and Table 15, the proportional (rather than absolute) difference between

predicted E(μ_i)'s become more variant across the models in the following order of the datasets: Toronto, Texas, Michigan, and Indiana, the same ranking of the datasets with respect to their NB model over-dispersion parameter (see Chapter IV). As expressed earlier, the variation in the models predictions for the Indiana dataset is of a far greater extent compared to the other datasets.

6.3.2 Mean-Variance Relationship

For sites without observed crash frequency, the calibrated model provides prior information through the estimated parameters. Unlike the original sites in the dataset, the expected crash frequency at these sites is not affected by an observed crash frequency and thus m follows a model-specific mixing distribution with mean μ. The "i" subscript was dropped because here the general relationship between the variables is sought rather than the relationship for numbered sites in a given dataset. The general relationship between mean and variance of sites without crash data can be established using the conditional variance identity in Equation 6.3:

$$\mathrm{Var}(Y) = \mathrm{E}(m) + \mathrm{Var}(m) \tag{6.4}$$

While E(m) is equal to μ in all models, Var(m) depends on the type of model and is parameterized in terms of μ and the model-specific hyper-parameter as described below:

- PG: $m \sim Gamma(\text{shape} = \varphi_{PG}, \text{scale} = \lambda)$ (6.5)

$$\mu = \varphi_{PG}\lambda \tag{6.6}$$

$$\mathrm{Var}(m) = \varphi_{PG}\lambda^2 = \frac{1}{\varphi_{PG}}\mu^2 \tag{6.7}$$

- PLN: $m \sim Lognormal(\text{location} = \upsilon, \text{scale} = \sigma)$ (6.8)

$$\mu = \exp(\upsilon + \sigma^2/2) \tag{6.9}$$

$$\text{Var}(m) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} = (e^{\sigma^2} - 1)\mu^2 \quad (6.10)$$

- PIGam: $m \sim \textit{Inverse Gamma}(\text{shape} = \varphi_{PIGam}, \text{scale} = \lambda)$ (6.11)

$$\mu = \frac{\lambda}{\varphi_{PIGam} - 1} \quad \text{for } \varphi_{\text{PIGam}} > 1 \quad (6.12)$$

$$\text{Var}(m) = \frac{\lambda^2}{(\varphi_{\text{PIGam}} - 1)^2(\varphi_{\text{PIGam}} - 2)} = \frac{1}{\varphi_{\text{PIGam}} - 2}\mu^2 \quad \text{for } \varphi_{\text{PIGam}} > 2 \quad (6.13)$$

Please note that although the PIGam model is defined for $\varphi_{\text{PIGam}} > 1$, the variance of the inverse-gamma distribution (and hence that of the PIGam model) is indeterminate for $\varphi_{\text{PIGam}} > 2$, which means it cannot be determined using Equation 6.13.

For all three models, the variance of crash counts has a parabolic relationship with its mean. Hence, the mean-variance relationship can be summarized as:

$$\text{Var}(Y) = \mu + \alpha\mu^2 \quad (6.14)$$

where α is the model-specific dispersion parameter as defined below:

- $\alpha_{PG} = \frac{1}{\varphi_{PG}}$ (6.15)
- $\alpha_{PLN} = e^{\sigma^2} - 1$ (6.16)
- $\alpha_{PIGam} = \frac{1}{\varphi_{\text{PIGam}} - 2}$ for $\varphi_{\text{PIGam}} > 2$ (6.17)

Using the posterior samples of φ_{PG}, φ_{PIGam}, and σ^2, the expectation of the dispersion parameter, E(α), is determined for every model-dataset and presented in Table 16. Figure 12 plots the mean-variance relationships for each dataset. As a substitute for the indeterminate mean-variance relationship of the PIGam model for the Michigan and

Indiana data, the variance was estimated for hypothetical sites with the same (modeled) characteristics as those of the sites in the dataset (hence the same set of μ_i's) but with no available observed crash count. In order to do so, for each hypothetical new site *i*, posterior predictive $\tilde{m}_i$ samples were drawn from an inverse gamma distribution with mean μ_i and then, posterior predictive $\tilde{Y}_i$ samples were drawn from a Poisson distribution with mean $\tilde{m}_i$. Using the 30,000 posterior samples for each model parameters, 30000 samples of $\tilde{m}_i$ (and $\tilde{Y}_i$, in turn) were generated for each hypothetical new site. The $\tilde{Y}_i$ samples were used to estimate the expected variance of crash frequency at each site.

Table 16. Over-dispersion parameter (α) for each model-dataset

Dataset	α_{PG} Mean (Std Dev.)	α_{PLN} Mean (Std Dev.)	α_{PIGam} Mean (Std Dev.)
Toronto	0.1399 (0.0127)	0.1489 (0.0143)	0.1573 (0.0179)
Texas	0.3001 (0.0292)	0.3431 (0.0381)	0.4201 (0.0574)
Michigan	0.6398 (0.1109)	0.9088 (0.2010)	Indeterminate*
Indiana	0.9192 (0.1161)	1.5995 (0.3537)	Indeterminate*

* because some φ_{PIGam} samples are less than 2, for which the variance is indeterminate as indicated in Equation 6.13.

Table 16 and Figure 12 indicates that the difference between the mean-variance relationship of the three models increases as the data become more over-dispersed (ranging from the least over-dispersed data i.e., Toronto, to the most highly over-dispersed data i.e., Indiana). In other words, in terms of the crash frequency prediction for sites with no crash data, the choice of the model is more important and consequential for more over-dispersed datasets.

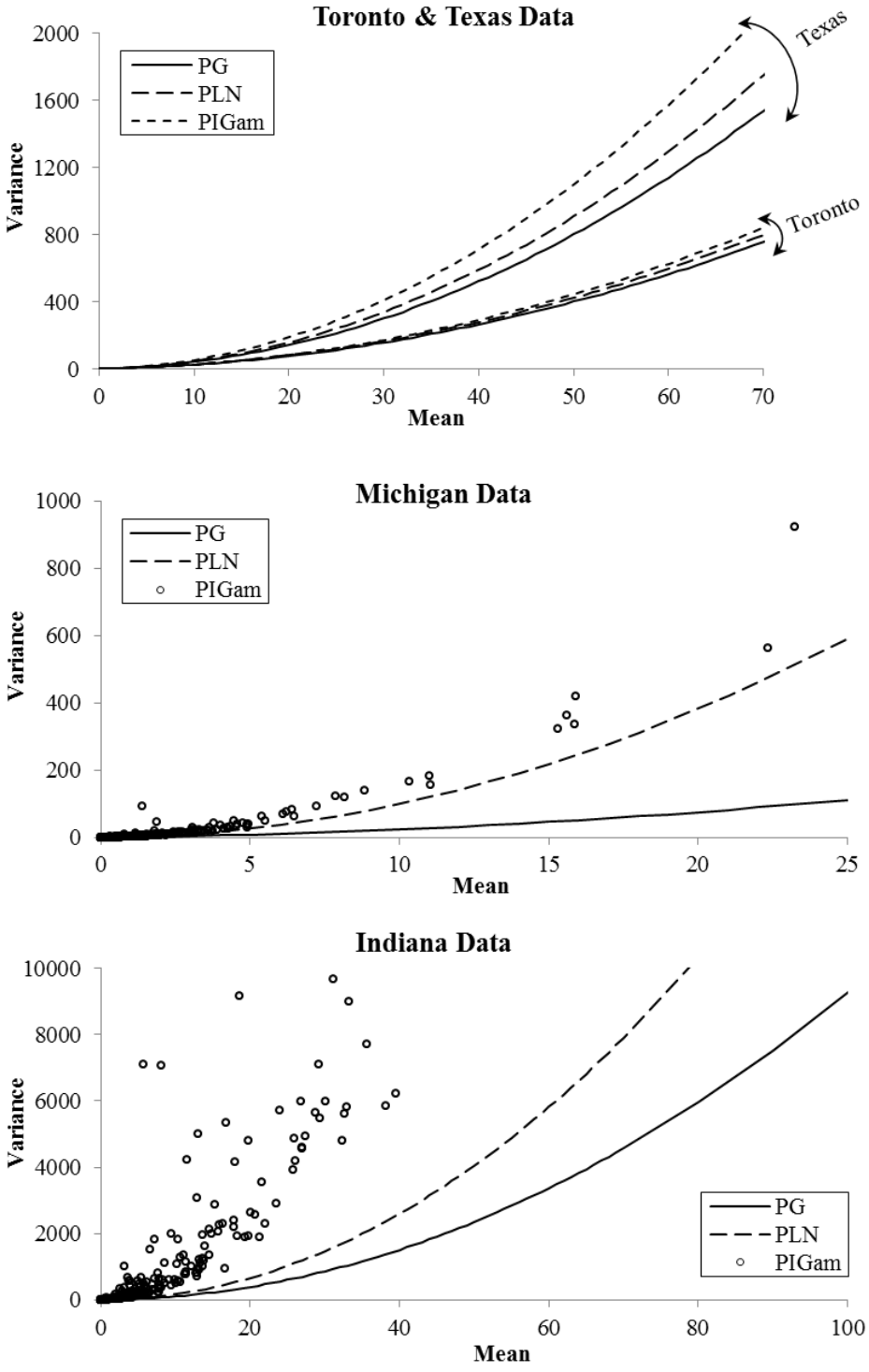


Figure 12. Mean-variance relationship for sites with no observed crash data

6.4 CHAPTER SUMMARY

In this chapter, the alternative models were compared based on their expected crash frequency predictions for individual sites. The analysis was carried out in two separate sections: Section 6.1 for the existing sites in each dataset, where models predictions are affected by the observed crash frequency, and Section 6.2 for new (or existing) sites where crash data are not available and the calibrated models are to be used for crash frequency prediction.

For the sites with crash data, it was discovered that the disparities between the alternative models predictions were mainly associated with the sites where the observed crash frequency was significantly larger or smaller than expected for a site with similar traffic and physical characteristics. For both scenarios, it was demonstrated that the Poisson-gamma model inclined to predict a lower expectation for crash frequency than would the Poisson-lognormal and Poisson-inverse gamma models, in order. For sites where the observed crash frequency was in the vicinity of the expectation for the given site characteristics, however, the reverse of the aforementioned pattern was observed. Nonetheless, from the practical standpoint, the latter trend is not nearly as important as the former because it results in relatively small variations between the alternative models predictions.

The disparities between alternative model predictions were found to be even more important when the calibrated models were applied to predict crash frequency at sites with no observed crash count. For all four datasets, the Poisson-inverse gamma model tended to predict higher expected crash frequencies than did the Poisson-lognormal and Poisson-gamma models, in order. The difference between the models predictions grew larger for sites with higher crash frequency expectations. Furthermore, the difference between the models predictions for the expectation and variance of crash frequency increased in the same order that the over-dispersion in the datasets did, implying that the

model choice becomes more consequential for more over-dispersed data. Next chapter will compare the alternative models in terms of their goodness-of-fit.

CHAPTER VII
GOODNESS OF FIT ANALYSIS

This chapter compares the goodness-of-fit (GOF) of the Bayesian hierarchical regression models in this dissertation. Section 7.1 presents the conventional GOF assessment where the models overall fit to datasets are compared using the deviance information criterion (DIC), as introduced in Chapter 3. The main goal is to test the conjecture that models with heavy/long-tailed distributions would outperform a light-tailed distributions such as the tradition Poisson-gamma in terms of statistical fit to highly over-dispersed data. Section 7.2 proposes a site-specific approach to GOF assessment where the models fit to data will be explored at individual sites level. The objective of the latter analysis is to detect latent trends in relative GOF of alternative models to observed crash count at individual sites with common features. Simply stated, the research strives to answer the following question: "Are there sites with certain common characteristics where a certain model can be expected to provide relatively better (or worse) fit to the site-specific observed crash count than would the other alternative models?"

7.1 OVERALL GOODNESS-OF-FIT

As explained in Chapter 2, the DIC is selected for GOF analysis in this research because of its popularity in Bayesian hierarchical model selection and its straightforward calculation when posterior distribution of model parameters are estimated using MCMC simulations. The mathematical expression of the DIC was described in Section 2.4. Denoting the vector of model parameters with $\boldsymbol{\theta}$, $\bar{\boldsymbol{\theta}}$ is conveniently estimated by averaging the posterior $\boldsymbol{\theta}$ samples, and $\bar{D}$ is estimated as the average of the model deviance for each $\boldsymbol{\theta}$ sample. Table 17 presents the DIC values obtained for each model-dataset, as well as the DIC breakdown into its two components, $\bar{D}$ and $P_{D.}$

Table 17. Deviance information criterion and its components for each model-dataset

Dataset	Poisson-Gamma			Poisson-Lognormal			Poisson- Inverse Gamma		
	$\bar{D}$	P_D	DIC	$\bar{D}$	P_D	DIC	$\bar{D}$	P_D	DIC
Toronto	4321.0	458.2	4779.2	4340.8	456.3	4797.0	4376.7	447.2	4823.9
Texas	3496.4	397.7	3894.1	3488.5	403.7	3892.2	3503.2	397.5	3900.8
Michigan	1567.4	178.3	1745.7	1556.1	203.3	1759.3	1576.9	202.7	1779.6
Indiana	1298.2	190.4	1488.5	1273.4	189.6	1463.0	1287.9	180.2	1468.1

The three models in this study have the same number of parameters and very similar structures. However, as Table 17 confirms, the same number of parameters for these hierarchical models does not necessarily translate to similar *effective* number of parameters (P_D). Since every site's expected crash frequency (m_i) is a parameter, posterior m_i's that are too close to the observed Y_i's may be a sign that a model is over-fitting the data.

A rule of thumb in comparing the DICs of alternative models is that a difference of more than 10 constitutes a disparity between the models' fit to the data (MRC Biostatistics Unit, 2016). As such, for both the Toronto and Michigan datasets, the PG model performs significantly better than the PLN model and the PLN model significantly better than the PIGam model. For the Texas dataset, the PG and PLN models perform equally well and only slightly better than the PIGam model. Finally, for the Indiana dataset, the PLN model performs slightly better than the PIGam model while both models fit the data substantially better than the PG model. It is also noteworthy that unlike most other modeling aspects discussed so far, the DIC of the PLN model does not necessarily fall between that of the PG and PIGam models.

As argued in the introduction to this study, it is virtually impossible to forecast the relative GOF of alternative models for a given dataset before the models are actually fit to the data. The results in Table 17 refute the conjecture that a heavy-tailed distribution

will certainly perform better as the over-dispersion in data increases. For example, the Michigan data is more over-dispersed than the Texas data (as measured by $\alpha_{NB(MLE)}$, see Chapter 4), but the light-tailed PG model performs better (relative to the other two models) for the Michigan dataset.

7.2 SITE-SPECIFIC GOODNESS-OF-FIT

While the DIC values can provide little insight into the fundamental differences between the alternative models, it is useful to take a microscopic look at the constituent components of the DIC and seek a relationship between the overall GOF of the models and their site-specific predictions for the expected crash frequency. This analysis is in alignment with the general focus of this study on the distinctions between the site-specific predictions of the considered models.

Replacing P_D from Equation 2.2 into Equation 2.1, the DIC can be reformulated as:

$$\text{DIC} = 2\bar{D} - D(\bar{\boldsymbol{\theta}}) \tag{7.3}$$

The following equations describe how the DIC of the considered models can be broken down into site-specific components:

$$\text{DIC} = 2 \times \text{E}[\,-2\log(\Pr(\boldsymbol{Y}|\boldsymbol{\theta})] - [-2\log(\Pr(\boldsymbol{Y}|\bar{\boldsymbol{\theta}})] \tag{7.4}$$

$$\text{DIC} = 2 \times {}^{1}\!/_{T} \sum_{t=1}^{T} [-2\log(\prod_{i=1}^{n} \Pr(Y_i|m_{i,t})] - [-2\log(\prod_{i=1}^{n} \Pr(Y_i|\bar{m}_i)] \tag{7.5}$$

$$\text{DIC} = 2 \times {}^{1}\!/_{T} \sum_{t=1}^{T} [-2\sum_{i=1}^{n} \log(\Pr(Y_i|m_{i,t}))] - [-2\sum_{i=1}^{n} \log(\Pr(Y_i|\bar{m}_i))] \tag{7.6}$$

$$\text{DIC} = 2 \times \sum_{i=1}^{n} [-2\,{}^{1}\!/_{T} \sum_{t=1}^{T} \log(\Pr(Y_i|m_{i,t}))] - [-2\sum_{i=1}^{n} \log(\Pr(Y_i|\bar{m}_i))] \tag{7.7}$$

where $m_{i,t}$ is the t'th sample from the posterior distribution of m_i, and T is the total number of MCMC samples. In deriving Equation 7.5 from Equation 7.4, please note that m_i's are assumed to be mutually independent (see Chapter 3). Finally, let $D_i = -2\log(\Pr(Y_i \mid m_i))$ define the component of the deviation for site #i, as opposed to the total model deviation, $D = -2\log(\Pr(\boldsymbol{Y} \mid \boldsymbol{m}))$. Then, the total DIC can as well be expressed as a sum of site-specific components (DIC_i's):

$$\text{DIC} = \sum_{i=1}^{n} DIC_i = \sum_{i=1}^{n} [2\bar{D}_i - D_i(\bar{m}_i)] \tag{7.8}$$

The breakdown of the DIC into site-specific components is important because it provides the opportunity to investigate the contribution of each site (observation) to the overall GOF of each model. The lower a DIC_i value, the better the model has fit the crash count observation at site #i. The study of the DIC_i's across the models may reveal sites with certain common characteristics where a certain model provide relatively better (or worse) fit than the other alternative models.

Since $Y_i \mid m_i$ is Poisson-distributed under all three models, a posterior E(m_i) closer to the observed Y_i will generally result in a larger log-likelihood and thus a smaller DIC_i. Simply stated and quite intuitive, the model fit improves as the predicted expectation for crash frequency approaches the observed crash count. The pattern of DIC_i's across the models is therefore closely related to the differences between the models predictions for m_i's, which were discussed in detail in Chapter 6. It was demonstrated in Chapter 6 that the distinctions between the models' m_i prediction for a specific site is a function of the difference between the respective μ_i (i.e., expected crash frequency at a site with similar modeled characteristics but with no crash count observation) and the observed crash frequency (Y_i). Thus, in the GOF analysis herein, the variations of DIC_i's across the models are also investigated as a function of Y_i – E(μ_i).

Due to the large number of data points and relatively small differences between DIC_i's from the three models, a cumulative DIC plot will be more useful in illustrating the differences between the models and revealing the trends. These plots, presented in Figure 13 for the Toronto, Texas, and Michigan dataset, are constructed as follows.

First, the sites in each dataset are ranked based on their $Y_i - E(\mu_i)$. As shown in Figure 2, $E(\mu_i)$'s from the three models are very close to one another (except for the Indiana dataset which is not included here) and thus, it is reasonable to determine $[Y_i - E(\mu_i)]$'s based on the average of the $E(\mu_i)$'s obtained from each of the three models. Once the sites are ranked in the increasing order of $Y_i - E(\mu_i)$, the cumulative DIC (i.e., the sum of DIC_i's for all sites with a smaller or equal $Y_i - E(\mu_i)$) is calculated for each of the models. In this analysis, the interest is in the difference between the DIC's of the models and thus, instead of the cumulative DIC function for each model, the difference between the DIC functions is plotted in Figure 13. The PG model is assumed as the base model and the vertical axis indicates how much larger or smaller the cumulative DIC of the PLN and PIGam models are compared to that of the PG model. The cumulative DIC plots exhibit a similar trend for all three datasets; four intervals of $Y_i - E(\mu_i)$ are identified and the respective regions are numbered in every plot and explained below.

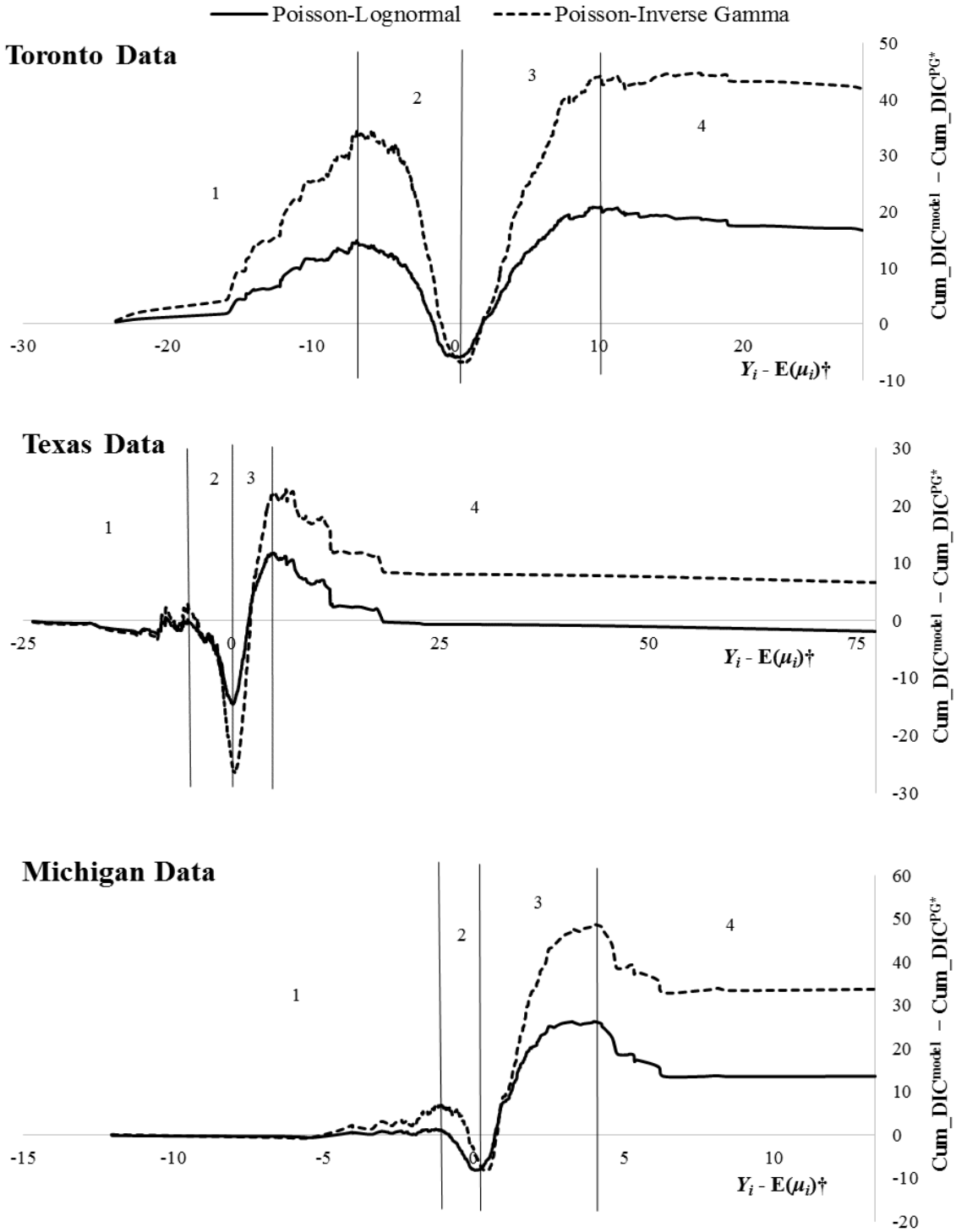


Figure 13. Difference between the cumulative DIC (defined as the sum of DIC_i's for all sites with a smaller or equal $Y_i - E(\mu_i)$) of PLN and PIGam models and that of the PG model as a function of $Y_i - E(\mu_i)$ for Toronto, Texas, and Michigan datasets

* Cumulative DIC of considered model (PLN or PIGam) – Cumulative DIC of PG model

† where $E(\mu_i)$ is calculated as average $E(\mu_i)$ obtained from PG, PLN, and PIGam models

Interval 1 corresponds with large negative values of $Y_i - E(\mu_i)$ and thus includes the sites where the observed crash frequency is much smaller than expected for a site with similar (modelled) characteristics. For such sites, as discussed in the previous chapter, a $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship is likely to be observed, which results in a closer fit (hence a smaller DIC_i) for the PG model compared to the PLN and PIGam models (in order). In consequence, the cumulative DIC function of the PLN and PIGam models (in order) will grow faster than that of the PG model. In Figure 13, this trend is very evident in the Toronto dataset, but not as noticeable in the Texas and Michigan datasets. Nonetheless, please note that the size of the difference between the cumulative DIC functions is quite small in both Texas and Michigan datasets. As mentioned in Chapter 6, a large difference between $E(\mu_i)$ and Y_i does not always result in a $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship. However, when this relationship is not observed, the difference between the $E(m_i|X,y)$'s from the three models (and hence the respective DIC_i's) are quite small.

Interval 2 starts at the approximate point where the trend in the differential cumulative DIC functions reverses. The boundary between Interval 1 and 2 is variant across the datasets and is loosely defined based on shape of the cumulative DIC functions. In contrast, the endpoint of Interval 2 (and the beginning of Interval 3) is strictly defined as zero ($Y_i - E(\mu_i) = 0$). Therefore, Interval 2 includes the sites where the observed crash frequency is slightly or moderately (depending upon the dataset) smaller than expected for a site with similar characteristics. As described in Chapter 7, for sites with small negative or positive value of Y_i - $E(\mu_i)$, a $E(m_i^{PG}|X,y) > E(m_i^{PLN}|X,y) > E(m_i^{PIGam}|X,y)$ relationship is usually observed, which translates to $m_i^{PIGam}|X,y$ being closer to Y_i than is $m_i^{PLN}|X,y$ and $m_i^{PIG}|X,y$ (in order). The resulting relationship between the DIC_i's (i.e., $DIC_i^{PIGam} < DIC_i^{PLN} < DIC_i^{PG}$) causes the differences between the cumulative DIC's of the models to decrease.

In terms of the relationship between the models' E(m_i) predictions and the closeness of E(m_i)'s to Y_i's (and thus site-specific DIC$_i$'s), Intervals 3 and 4 are the mirror image of Interval 2 and 1, respectively. In Interval 3, the $E(m_i^{PIGam}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PG}|X,y)$ relationship is prevalent, while in Interval 4, a $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship is likely to occur. Because in both of these intervals Y_i is greater than E(μ_i), the aforementioned relationships results in $DIC_i^{PG} < DIC_i^{PLN} < DIC_i^{PIGam}$ in Interval 3 and $DIC_i^{PIGam} < DIC_i^{PLN} < DIC_i^{PG}$ in Interval 4. Consequently, the site-specific GOF trends described in this section can be integrated with the analysis results in Section 6.1 and summarized in Figure 14. In interpreting this figure, it is critical to note that the displayed trends are the relationships that are merely more likely to occur than the other possible relationships; one shall not expect all observations falling in the same interval to exhibit similar trend.

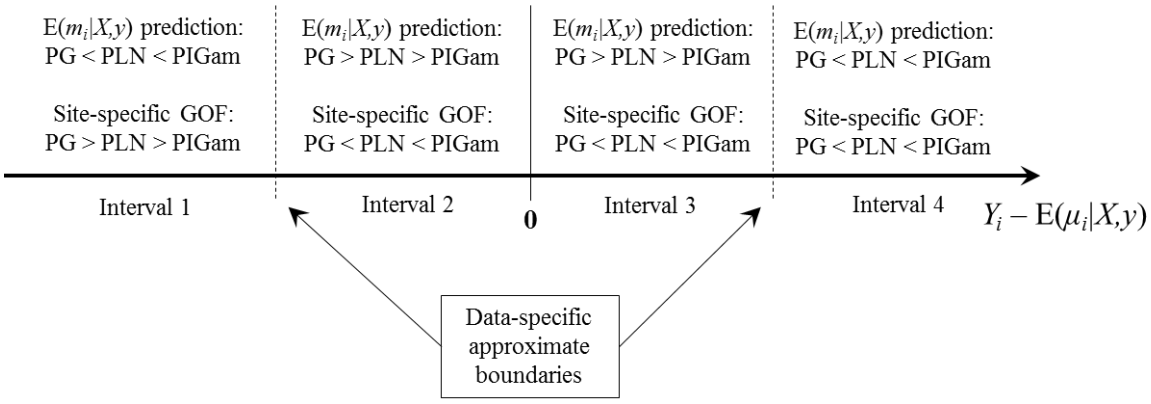


Figure 14. General trends for alternative models predictions and site-specific goodness of fit as a function of Y_i - E(μ_i)

Finally, it is noteworthy that in Figure 13, the amount of change in the differential cumulative DIC functions in Intervals 2 and 3 is greater than that in Intervals 1 and 4. This phenomenon is caused not because the differences between the DIC$_i$'s of the three models are greater for sites in Intervals 2 and 3 (it is indeed the opposite), but because these intervals include many more sites than Intervals 1 and 4. Table 18 indicates the approximate boundaries between the four intervals of Y_i – E(μ_i) and reports the

percentage of sites falling in each interval. It is interesting to note that among the three datasets considered here, the one for which the PLN and PIGam models fit the best (i.e., Texas) is the dataset which includes the greatest share of sites with very large [Y_i – E(μ_i)]'s. The PIGam and PLN models (in order) are expected to provide better fit for such observations.

Table 18. Four intervals of Y_i – E(μ_i) defined for each dataset (except Indiana) and the percentage of sites falling in each interval

Interval	Toronto Data		Texas Data		Michigan Data	
	Range	% of sites	Range	% of sites	Range	% of sites
1	(-∞,-6.7]	9%	(-∞,-5.7]	6%	(-∞,-1.1]	7%
2	(-6.7,0]	46%	(-5.7,0]	56%	(-1.1,0]	69%
3	(0,10.0]	41%	(0,5.3]	32%	(0,4.0]	22%
4	(10.0,+∞)	5%	(5.3,+∞)	6%	(4.0,+∞)	2%

It must be reminded that the Indiana data was not included in the analysis in this section for the same reason that it was excluded in the analysis in Section 6.1: the significantly different predictions of E(μ_i) across the alternative models prevents categorization of sites (observations) based on their Y_i – E(μ_i).

7.3 CHAPTER SUMMARY

In this chapter, first the overall GOF of the alternative models were compared for every dataset using the DIC. The results did not signal an apparent relationship between the models relative GOF and the conditional (on mean) over-dispersion of datasets (as measured by the dispersion parameter of the traditional negative binomial generalized linear model). Next, the overall DIC was broken down into site-specific constituents which were then compared across the models. It was demonstrated that, similar to the trends for the models predictions for expected crash frequency (see Section 6.1), the

relationship between site-specific GOFs of alternative models were a function of Y_i – E(μ_i) i.e. the difference between the observed crash frequency and the expected crash frequency at a site with similar characteristics. Accordingly, the range of [Y_i – E(μ_i)]'s can be divided into four intervals with a unique combination of site-specific prediction and GOF trend that is likely to be observed for most sites in the interval.

CHAPTER VIII
SUMMARY AND CONCLUSIONS

This chapter summarizes the methodology and findings of this dissertation and draws conclusions based on the research results. Section 8.1 provides a summary of the research methodology, Section 8.2 summarizes the important analysis results, and Section 8.3 concludes the research by listing the practical implications of the study findings in the practice of highway safety analysis.

8.1 SUMMARY OF METHODOLOGY

This dissertation focused on the application of Bayesian Poisson-hierarchical regression models for motor-vehicle crash data analysis with the objective to explore the distinctions between alternative models in terms of their crash frequency predictions at individual sites with given characteristics. The Bayesian Poisson-hierarchical family of models was chosen for this dissertation for two reasons: their popularity and theoretical appropriateness for crash count data analysis. In these models, crash counts at individual sites (Y_i's) are assumed to be mutually independent and Poisson-distributed with a Poisson parameter/mean (m_i) that itself follows a model-specific continuous distribution (called mixing distribution) with mean μ_i. The generalized linear modeling (GLM) framework was adopted to model μ_i as a loglinear function of traffic and physical characteristics of the highway site (roadway segments or intersections).

The research was limited to three alternatives models: the two most commonplace models for crash data analysis i.e., the Poisson-gamma (PG) and Poisson lognormal (PLN), and a new model formulated and proposed for the purpose of this study, the Poisson-inverse gamma (PIGam). The hierarchical Bayesian structure of the models adjusted every μ_i (i.e., the crash frequency expectation given the input variables) using

the respective crash count observation (Y_i) to deliver a posterior crash frequency expectation (m_i) that often lied between μ_i and Y_i. In addition to a common structure, the three models comprised of the same number of parameters (= number of observations + number of input variables + 2), which provided for a facilitated and sensible comparison between models predictions.

Given the common structure of the alternative models, all differences in models performance may be attributed to the fundamental characteristics of their mixing distributions: gamma, lognormal, and inverse-gamma. Unlike the gamma distribution, the inverse-gamma distribution has a long/thick tail, while the lognormal distribution lies in between the gamma and inverse-gamma distributions in terms of tail thickness (and other shape properties for a given mean and variance). Following the claims of a few other studies, it was hypothesized that the models corresponding with heavy-tailed distributions (PLN and PIGam, in order) will provide better statistical fit as the over-dispersion in data increases. This conjecture was tested by selecting four datasets from Toronto, Texas, Michigan, and Indiana that covered a wide range of conditional over-dispersion. The conditional over-dispersion (i.e., over-dispersion conditional on the modeled mean) was loosely estimated using the over-dispersion parameter of the traditional frequentist negative binomial GLM.

The dissertation deviated from the conventional model selection studies in the following sense: rather than ranking the alternative models based on their overall goodness-of-fit (GOF) and predictive performance, the research focused on the magnitude of models predictions for the expected crash frequency at individual sites. The goal was not to determine the model that outperforms the others (globally or for certain datasets), but to investigate how the model choice can affect the expected value for crash frequency at sites with certain characteristics. Lacking a preset agenda, the author enjoyed the freedom to explore the modeling results to identify trends that would help safety analysts to comprehend the practical implications of opting for one model over the other.

The Bayesian models were estimated using MCMC simulation from the posterior distribution of parameters. Non-informative hyper-priors were assumed for the models' hyper-parameters to eliminate the potential effect on the models estimation and let the data speak for themselves.

8.2 SUMMARY OF RESEARCH FINDINGS

For the Toronto, Texas, and Michigan data, the alternative models resulted in very close estimates for the regression coefficients and hence μ_i's. The functional form of the model for the Toronto dataset included three regression coefficients, while that for the Texas and Michigan datasets included only two. The Indiana dataset was very different from the other datasets in the sense that it included eight covariates that were all found to be statistically significant. The relatively great number of regression coefficients, coupled with the extreme over-dispersion remaining even after inclusion of so many covariates, resulted in considerably different estimates for the regression coefficients of the alternative models, and hence μ_i's that may vary drastically across the models. It was interesting to note that for all datasets, the posterior estimates of the PLN model coefficients were in between those of the PG and PIGam models, indicating an apparent link between the shape properties of the three mixing distributions and the performance of their respective Poisson-hierarchical models.

By identifying the sites with the greatest difference between the models m_i predictions, an important trend was discovered. In every dataset except Indiana, all sites at which the posterior E(m_i) varied significantly across the alternative models shared a common trait: an observed crash frequency (Y_i) that is substantially smaller or larger than the expectation for a site with similar modeled characteristics, as estimated by the posterior E(μ_i). This phenomenon was caused by the different amounts by which the posterior distribution of μ_i, which varied only slightly across the models, shifted toward the observed crash count under different alternative models. For either the case where E(μ_i)

was substantially larger or smaller than Y_i, the described phenomenon would result in the PG model predict a lower $E(m_i|X,y)$ than the PLN and PIGam models, in order.

It is important to note that while the aforementioned trend was observed for all sites (except those in Indiana) where the posterior $E(m_i)$ varied significantly across the models, not every site with substantially different Y_i and $E(\mu_i)$ indicated a $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship. Based on the analysis results, it can only be said that for sites where Y_i differs significantly from $E(\mu_i)$, the $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship was more prevalent than any other relationship between $E(m_i|X,y)$ of the alternative models. In addition, the likelihood of observing the $E(m_i^{PG}|X,y) < E(m_i^{PLN}|X,y) < E(m_i^{PIGam}|X,y)$ relationship and the difference between $E(m_i|X,y)$ of the alternative models generally (not absolutely) increased as the difference between Y_i and $E(\mu_i)$ grew larger.

On the other hand, for most sites where the posterior $E(\mu_i)$ was relatively close to Y_i, the previously described relationship was reversed to $E(m_i^{PG}|X,y) > E(m_i^{PLN}|X,y) > E(m_i^{PIGam}|X,y)$, although the differences between $E(m_i|X,y)$'s were quite small. The opposite relationships between $E(m_i|X,y)$ predictions of alternative models for the two sets of sites (where Y_i is significantly different from $E(\mu_i)$ or relatively close to it) resulted in a balanced total expected crash frequency (for all sites combined) across the alternative models. Consequently, none of the models predicted significantly biased posterior m_i's compared to another model, although the $E(m_i|X,y)$ predictions at certain sites varied considerably. Similar to the coefficient estimates, the PLN model predictions for $E(m_i|X,y)$'s usually remained between those from the PG and PIGam models, signaling another link between the models performance and fundamental properties of their respective mixing distributions.

Furthermore, the deviance information criterion (DIC) was used to compare the overall fit of the alternative models to each dataset. The results did not indicate a clear

relationship between the conditional over-dispersion in the datasets and the GOF of the alternative models, refuting the conjecture that mixing distributions with thicker tails will necessarily result in a better fit as the conditional over-dispersion in data grows. The author believes that the relative GOF of the alternative models for a given dataset is too complicated to be predicted with certainty based on a simple statistic before the model are actually fitted to data.

However, in alignment with the microscopic approach of this research, the DIC was broken down into site-specific components. The purpose was to analyze the contribution of each site/observation to the overall "lack-of-fit" of the model as measured by DIC. The analysis in Chapter 7 indicated that similar to the trends for the models predictions for expected crash frequency, the relationship between site-specific GOFs of alternative models is a function of $Y_i - E(\mu_i)$ i.e. the difference between the observed crash frequency and the expected crash frequency at a site with similar characteristics. Accordingly, the range of $[Y_i - E(\mu_i)]$'s can be divided into four intervals with a unique combination of site-specific prediction and GOF trend that is likely to be observed for most sites in the interval. Figure 14 illustrated these intervals and summarized the most important findings of this research.

It is important to note that the aforementioned patterns could not be examined for the Indiana data because the alternative models resulted in very different regression coefficients and thus posterior $E(\mu_i)$'s, preventing the type of analyses carried out for other datasets as explained above. The disparity between the models regression coefficients for the Indiana data is attributable to the relatively great number of covariates (six) and the substantial conditional over-dispersion (even after the six covariates were included in the models).

Finally, this research found the disparities between alternative model predictions to be even more important when the calibrated models were applied to predict crash frequency

at sites with no observed crash count. For all four datasets, the Poisson-inverse gamma model tended to predict higher expected crash frequencies than did the Poisson-lognormal and Poisson-gamma models, in order. The difference between the models predictions grew larger for sites with higher crash frequency expectations. Furthermore, the difference between the models predictions for the expectation increased in the same order that the over-dispersion in the datasets did, implying that the model choice becomes more consequential for more over-dispersed data.

8.3 PRACTICAL IMPLICATIONS

Based on the analyses in this dissertation, the choice of the mixing distribution in the Poisson-hierarchical family of generalized linear models can have important practical implications especially for two specific applications.

The first important application is when regression models are used to identify hazardous sites and rank them based on their crash proneness and hence priority for treatment. For this application, the important disparity between the models performance is associated with sites where the observed crash frequency is unusually higher or lower than what is typically expected at a site with similar characteristics.

For example, consider a very low-traffic intersection that experiences 20 crashes in a given year and a major arterial intersection that experiences no crashes in the same year. The alternative models are more likely to predict considerably different crash frequency expectations for these sites than they are for other sites with normal observed crash frequency (e.g., major intersections with many crashes and minor intersections with few crashes). Based on the analysis results, for sites with either unusually high or low observed crash frequency, the PG model is likely to predict a lower crash frequency expectation than would the PLN and PIGam models, in order. Although the aforementioned relationship does not hold for all sites with unusually high or low

observed crash frequencies, the likeliness of the relationship to hold is expected to increase as the difference between expected and observed crash frequencies grow (i.e., observed crash frequency becomes more unusual). Consequently, the PIGam model is inclined to rank the sites with unusual observed crash frequencies higher (in terms of crash proneness) than would the PLN and PG models, in order.

However, highway safety analysts should be warned that the patterns explained above cannot be expected to hold true if the alternative models estimate considerably different regression coefficients for a certain dataset. As was the case for the Indiana dataset in this study, disparate estimates for regression parameters are prevalent when a relatively great number of covariates are included in modeling. For such cases, one may not straightforwardly determine whether the observed crash frequency for a site is too different from expected because the crash frequency expectation can differ drastically across the models.

The second important application is when the alternative models are used to predict the expected crash frequency at new sites with no crash data (such as planned facilities that have not been constructed yet). For such conditions, the PIGam model tends to predict higher expected crash frequencies than does the PLN and PG models, in order. The difference between the models predictions is likely to grow larger for more over-dispersed datasets. It is critical for practitioners to be aware of such relative bias when applying different models in the Poisson-hierarchical family for prediction purposes.

8.4 FUTURE RESEARCH

Although this research revealed important trends regarding the predictions of the alternative models at individual sites, it did not find a specific relationship between the overall GOF of the alternative models and simple characteristics of the datasets. Clearly, appropriateness of the alternative models for given datasets is too complicated to be

indicated by a simple statistic such as the conditional over-dispersion (as examined in this study). However, important relationships between the relative GOF of the alternative models may be discovered if the analysis is carried out using simulated data. Data simulation will provide control over different data characteristics. Therefore, future work will use extensive simulated data to further investigate the relative GOF of the alternative models as a function of different data conditions.

REFERENCES

American Association of State Highway and Transportation Officials (AASHTO), Highway Safety Manual, 2010. 1st Edition.

Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transportation Research Record 2061, 55–63.

Aguero-Valverde, J., 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. Accident Analysis & Prevention 50, 289-297.

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems (with discussion). Statistical Science 10, 3–66.

Bonneson, J.A., McCoy, P.T., 1993. Estimation of safety at two-way stop-controlled intersections on rural highways. Transportation Research Record 1401, 83-89.

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

Carlin, B.P., Louis, T.A., 2008. Bayesian Methods for Data Analysis. 3rd Edition. Chapman and Hall.

Carriquiry, A., Pawlovich, M.D., 2004. From empirical Bayes to full Bayes: methods for analyzing traffic safety data. http://www.iowadot.gov/crashanalysis/pdfs/eb_fb_ comparison_whitepaper_october2004.pdf (accessed March 7th, 2014).

Davis, G.A., Yang, S., 2001. Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. Transportation Research Record 1746, 84-89.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate poisson-lognormal regression. Accident Analysis & Prevention 41, 820–828.

El-Basyouny, K., Sayed, T., 2012. Measuring safety treatment effects using full Bayes non-linear safety performance intervention functions. Accident Analysis and Prevention 45, 152–163.

Fitzpatrick, K., Park, E.S., 2009. Safety effectiveness of HAWK pedestrian treatment. Transportation Research Record 2140, 214-223.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. 3rd Edition. Chapman and Hall/CRC, New York.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall.

Goldstein, H, 2010. Multilevel Statistical Models. 4th Edition. John Wiley & Sons, West Sussex, England.

Harwood, D.W., Bauer, K.M., Potts, I.B., Torbic, D.J., Richard, K.R., Kohlman Rabbani, E.R., Hauer, E., Elefteriadou, L., 2002. Safety effectiveness of intersection left- and right-turn lanes, Report No. FHWA-RD-02-089. Federal Highway Administration (FHWA), Washington, D.C.

Hauer, E., Persaud, B.N., 1983. A common bias in before and after accident comparisons and its elimination. Transportation Research Record 905, 164-174.

Hauer, E., 1986. On the estimation of the expected number of accidents. Accident Analysis & Prevention 18, 1-12.

Hauer, E., Ng, J.C.N., Lovell, J., 1989. Estimation of safety at signalized intersections. Transportation Research Record 1185, 48-61.

Hauer, E., 1992. Empirical Bayes approach to the estimation of unsafety: The multivariate regression approach. Accident Analysis & Prevention 24 (5), 456-478.

Hauer, E., 1997. Observational before–after studies in road safety: Estimating the effect of highway and traffic engineering measures on road safety. Pergamon Press, Elsevier Science, Ltd., Oxford, United Kingdom.

Hauer, E., 2001. Overdispersion in modeling accidents on road sections and in empirical Bayes estimation", Accident Analysis and Prevention 33(6), 799-808.

Hauer, E., 2010. Cause, effect and regression in road safety: A case study. Accident Analysis and Prevention 42, 1128-1135.

Heydecker B.G., Wu, J., 2001. Identification of sites for accident remedial work by Bayesian statistical methods: An example of uncertain inference. Advances in Engineering Software 32, 859-869.

Hilbe, 2014. Modeling count data. Cambridge University Press, Cambridge, United Kingdom.

Hinde, J., 1982. Compound Poisson regression models, in R. Gilchrist, ed., GLIM 82: Proceedings of the International Conference on Generalized Linear Models, New York, Springer-Verlag.

Khazraee, S.H., Saez-Castillo, A.J., Geedipally S.R., Lord, D., 2014. Application of the hyper-Poisson generalized linear model for analyzing motor vehicle crashes. Risk Analysis 35 (5), 919-930.

Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Injury Prevention 3 (4), 53–57.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34 (2), 149–161.

Lee, P., 2012. Bayesian statistics: An introduction. 4th Edition. John Wiley & Sons, West Sussex, England.

Lord, D., Bonneson, J.A., 2007. Development of accident modification factors for rural frontage road segments in Texas. Transportation Research Record 2023, 20-27.

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 37(1), 35-46.

Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero inflated models in highway safety. Accident Analysis and Prevention 39(1), 53–57.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46 (5), 751–770.

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 40 (4), 1441-1457.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44, 291–305.

Ma, J., Kockelman, K., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis & Prevention 40, 964–975.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods in Accident Research 1, 1-22.

Maycock, G., Hall, R.D., 1984. Accidents at four-arm roundabouts. Laboratory Report LR 1120, Transport Research Laboratory, Crowthorne, Berkshire, U.K.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. 2nd Edition. Chapman and Hall.

Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transportation Research Record 1840, 31–40.

Miaou, S.-P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space time modeling approach. Journal of Transportation and Statistics 6 (1), 33–57.

Miaou, S.-P., Song J.J., 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, 699-720.

Miranda-Moreno, L.F., Fu, L., Saccomanno, F.F., Labbe, A, 2005. Alternative risk models for ranking locations for safety improvement. Transportation Research Record 1908, 1-8.

Mitra, S., Washington, S., 2006. On the nature of over-dispersion in motor vehicle crash prediction models. Accident Analysis and Prevention 39, 459–468.

Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accident Analysis and Prevention 49, 439–448.

MRC Biostatistics Unit, Cambridge Biomedical Campus. DIC: Deviance Information Criterion. http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic. Accessed April 3rd, 2016.

Oh, J., Washington S.P., Nam, D., 2006. Accident prediction model for railway–highway interfaces. Accident Analysis and Prevention 38(2), 346–356.

Park, E. S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record: Journal of the Transportation Research Board 2019, 1–6.

Park, E.S., Park, J., Lomax, T.J., 2010. A fully Bayesian multivariate approach to before-after safety evaluation. Accident Analysis and Prevention 42, 1118-1127.

Persaud, B.N., 1988. Do traffic signals affect safety? Some methodological issues. Transportation Research Record 1185, 37–47.

Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. Transportation Research Record 1665, 7-12.

Persaud, B., Lyon, C., 2007. Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions. Accident Analysis & Prevention 39, 546–555.

Poch, M., Mannering, F.L., 1996. Negative binomial analysis of intersection accident frequency. Journal of Transportation Engineering 122(2), 105-113.

Qin, X., Ivan, J.N., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevention 36 (2), 183–191.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org.

Rao, J., 2003. Small Area Estimation. John Wiley and Sons, West Sussex, England.

Robert, C.P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer-Verlag, New York, NY.

Schluter, P.J., Deely, J.J., Nicholson, A.J., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. The Statistician 46, 293–316.

Shankar, V., Milton, J., Mannering, F.L., 1997. Modeling accident frequency as zero altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29 (6), 829–837.

Song, J. J., Ghosh, M., Miaou, S., & Mallick, B., 2006. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. Journal of Multivariate Analysis 97, 246–273.

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 64, 583–640.

Spiegelhalter, D.J., Thomas, A., Best, N.G., Lun, D., 2003. WinBUGS Version 1.4.3 User Manual. MRC Biostatistics Unit, Cambridge. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.

Tarko, A., Eranky, S., Sinha, K., 1998. Methodological considerations in the development and use of crash reduction factors. Paper presented at the 77th annual meeting of the Transportation Research Board, Washington D.C.

Vogt, A., Bared, J., 1998. Accident models for two-lane rural segments and intersections. Transportation Research Record 1635, 18-29.

Washington, S.P., Karlaftis, M., Mannering, F.L., 2003. Statistical and econometric methods for transportation data analysis. Chapman and Hall, Boca Raton.

Wu, L., Lord, D., 2016. Investigating the influence of dependence between variables on crash modification factors developed using regression models. Paper presented at the 95th annual meeting of the Transportation Research Board, Washington D.C.

Wu, L., Lord, D., Zou, Y., 2015. Validation of CMFs derived from cross sectional studies using regression models. Transportation Research Record 2514, 88–96.

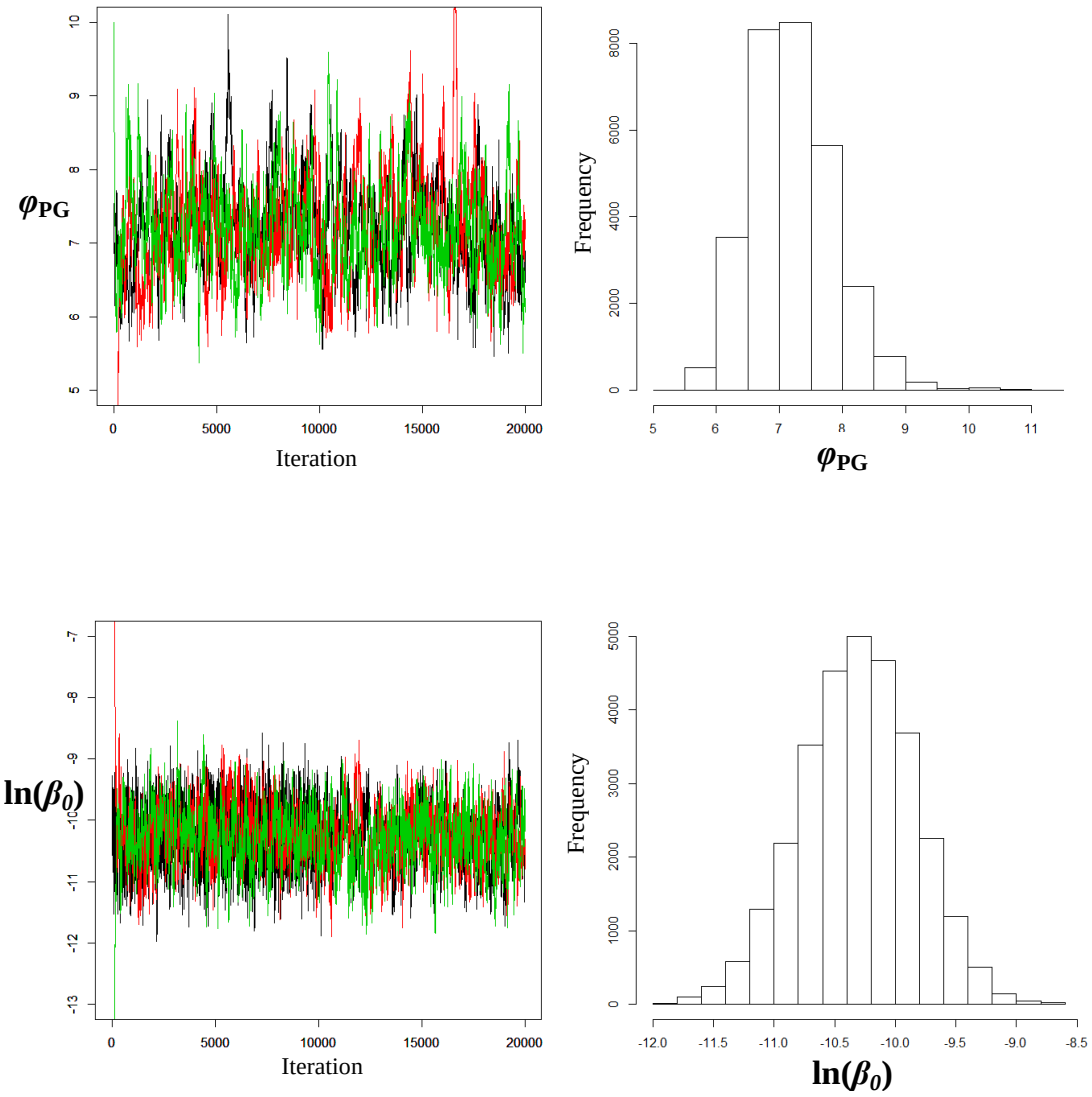
Yang, M., Rasbash, J., Goldstein, H., Barbosa, M., 1999. MLwiN Macros for Advanced Multilevel Modelling. Version 2.0a. Multilevel Models Project, Institute of Education, University of London, UK. <http://www.bris.ac.uk/cmm/software/mlwin/download/d-1-10/advmacma.pdf>

Zha, L., Lord, D., Zhou, Y., 2016. The Poisson Inverse Gaussian (PIG) Generalized Linear Regression Model for Analyzing Motor Vehicle Crash Data. Journal of Transportation Safety and Security 8 (1), 18-35.

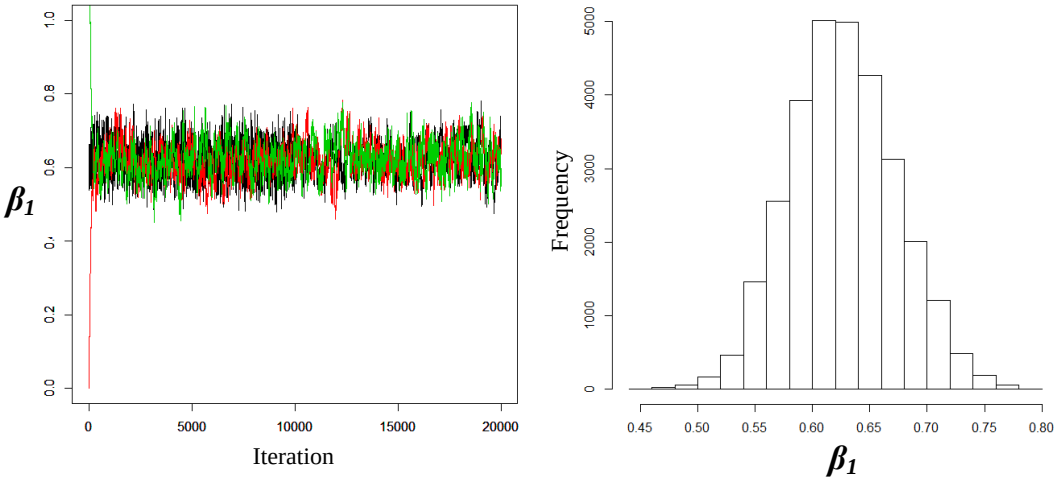
APPENDIX A
MODEL ESTIMATION RESULTS

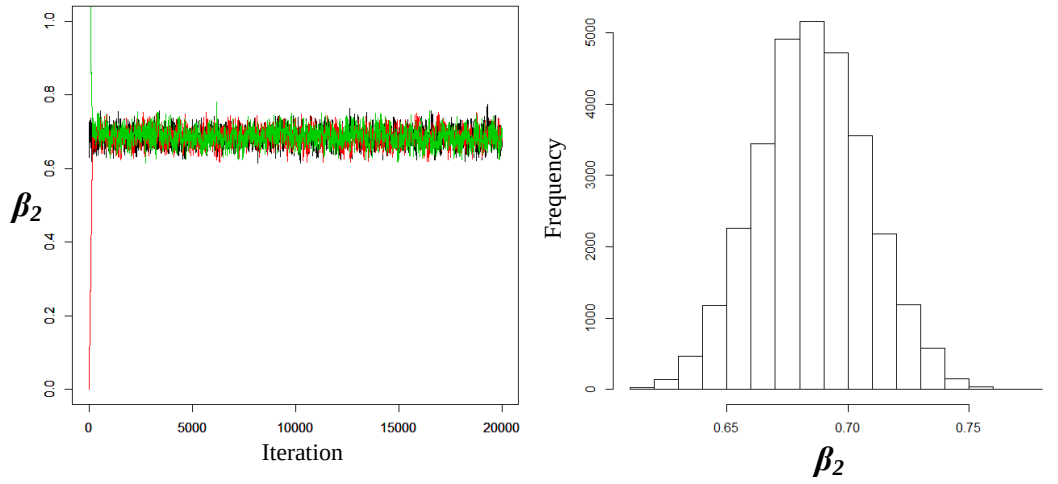
This appendix presents the MCMC traceplots and posterior distributions for the parameters in each model and dataset, where each chain is drawn with a distinct color (black, red, or green). The posterior distributions of the models' parameters are constructed using the samples from the three chains after the first 10000 sample from each chain are discarded.

Toronto Data, Poisson-Gamma Model

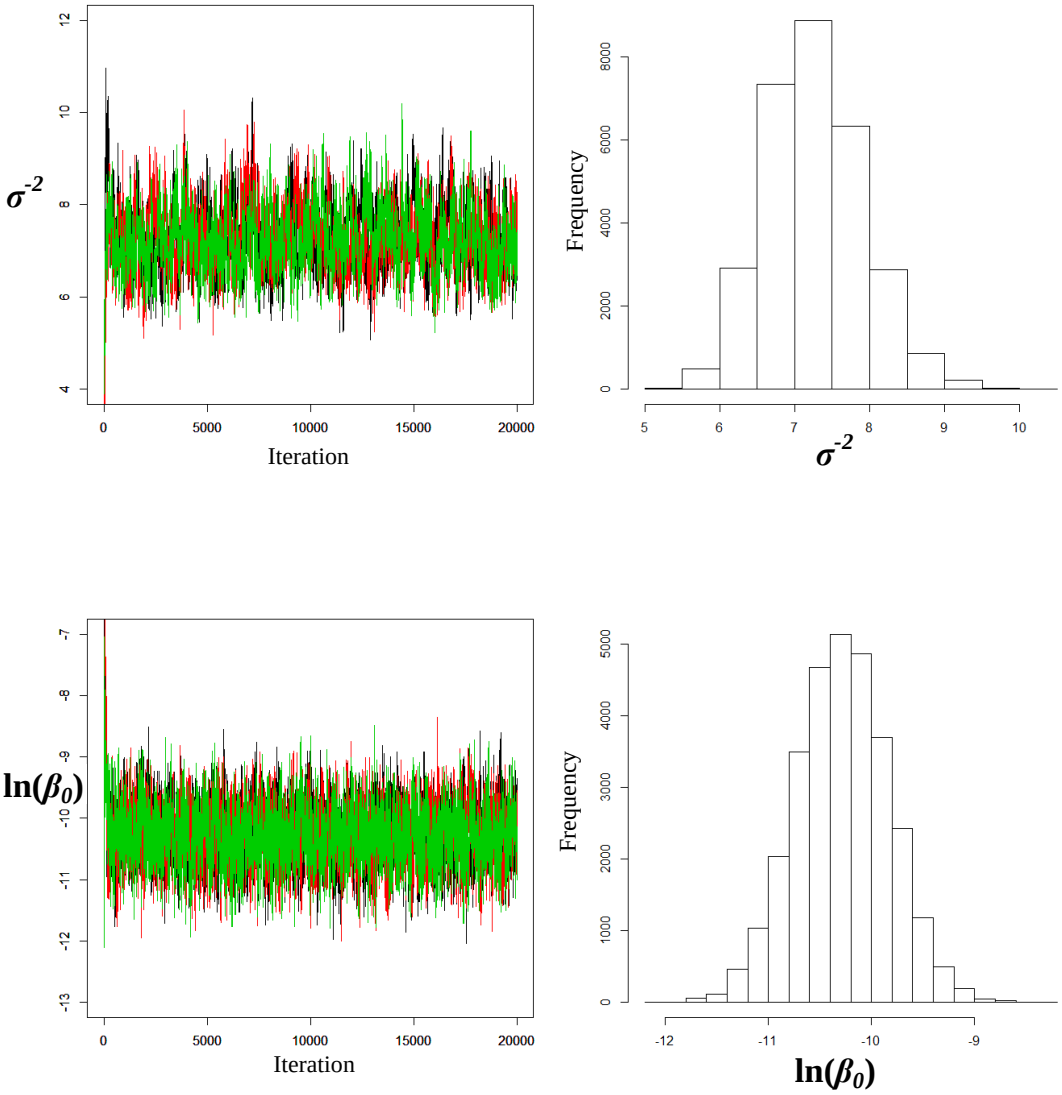


Toronto Data, Poisson-Gamma Model

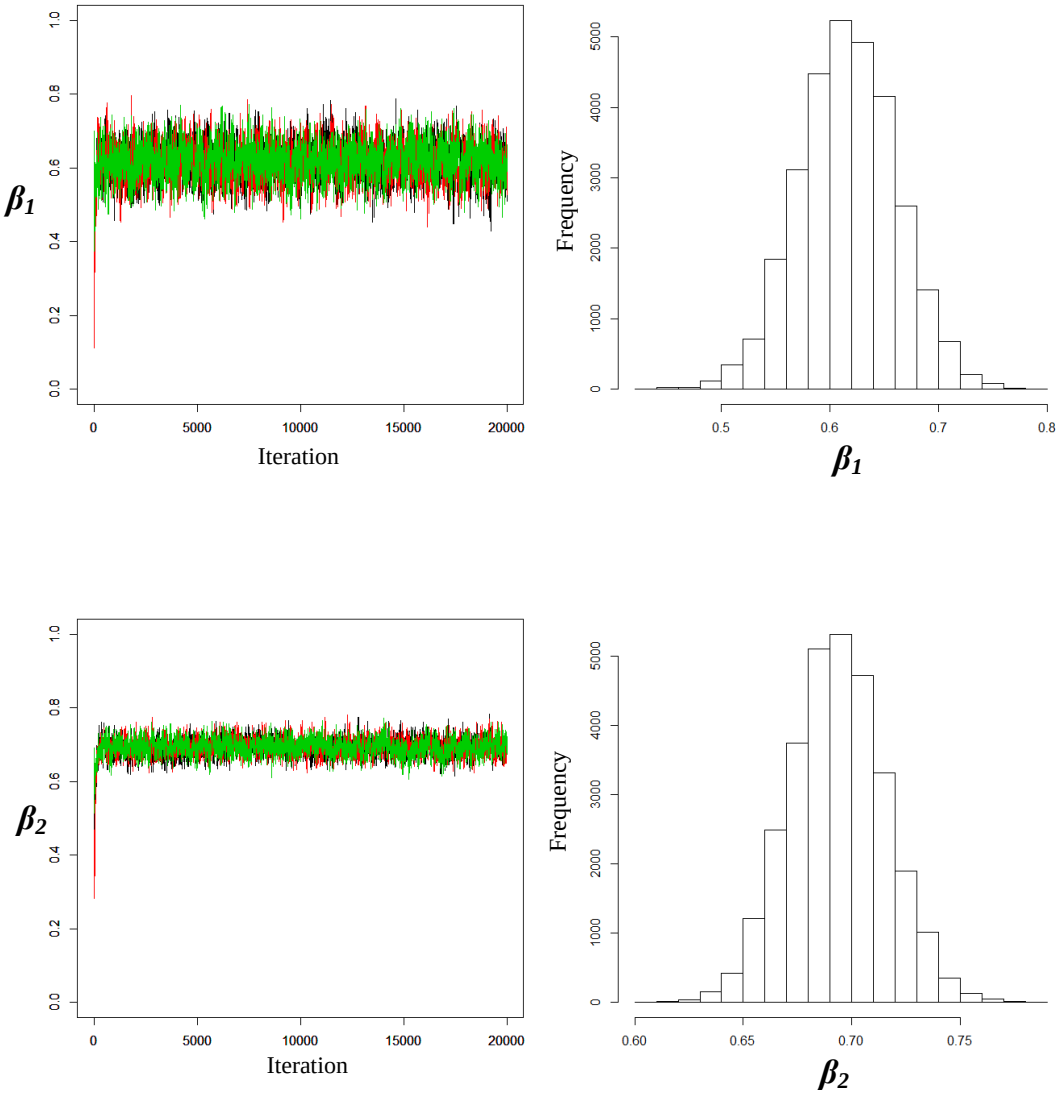




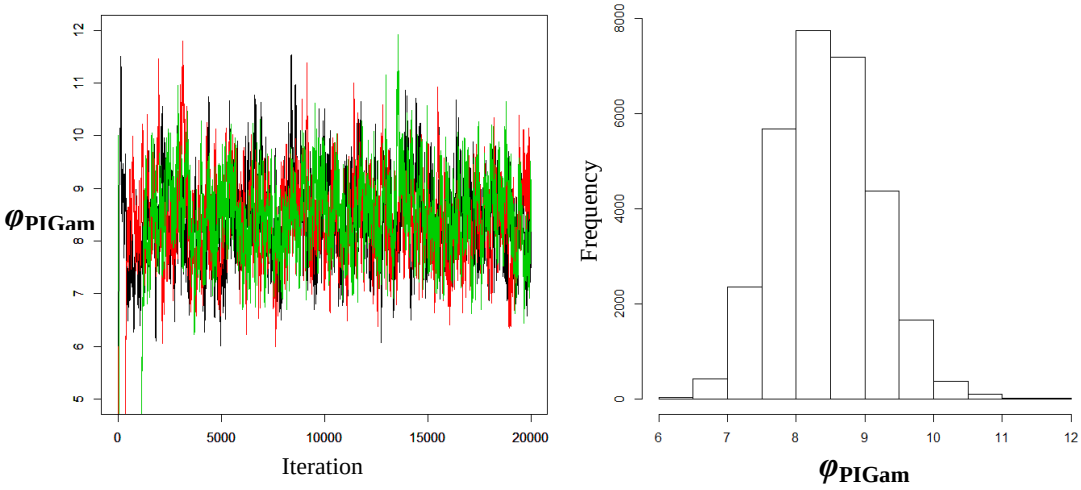
Toronto Data, Poisson-Lognormal Model

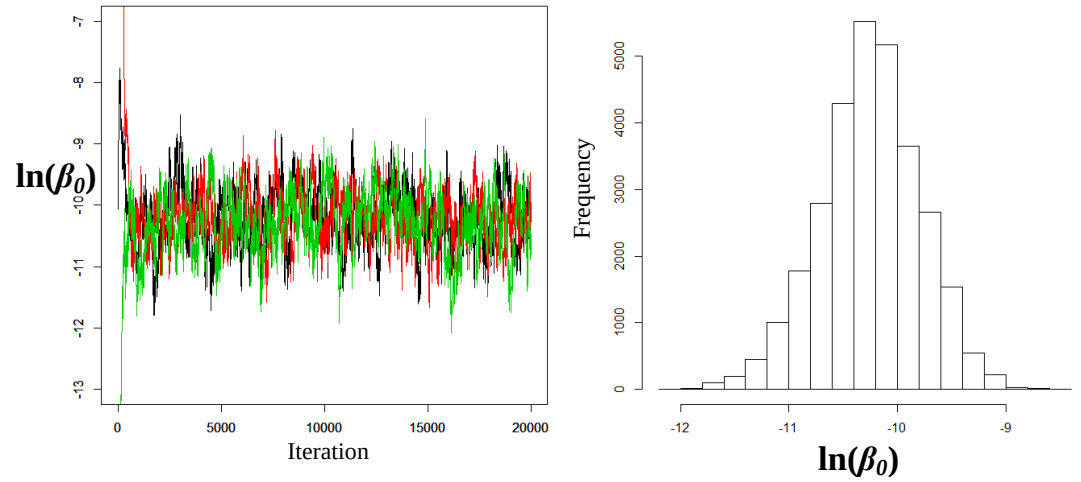


Toronto Data, Poisson-Lognormal Model

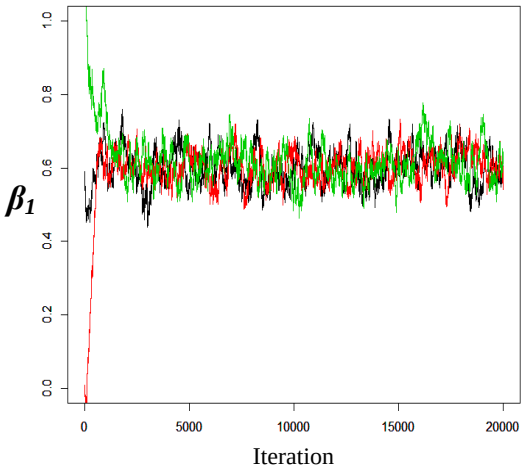


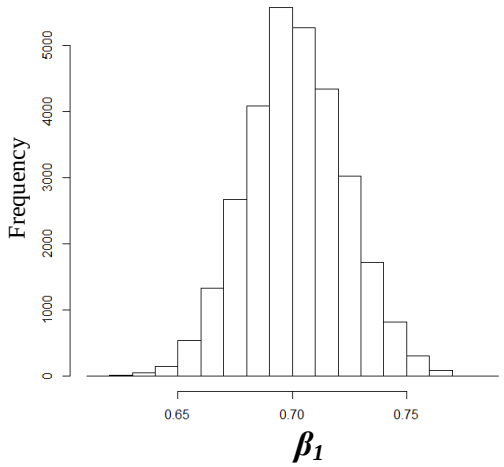
Toronto Data, Poisson-Inverse Gamma Model

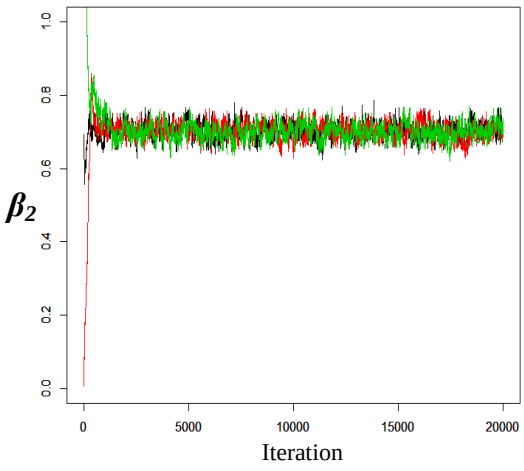


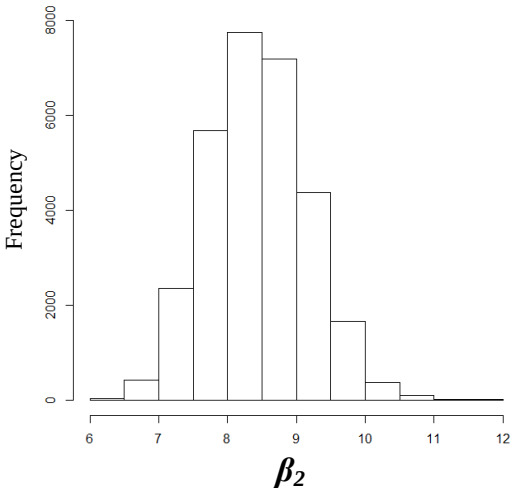


Toronto Data, Poisson-Inverse Gamma Model

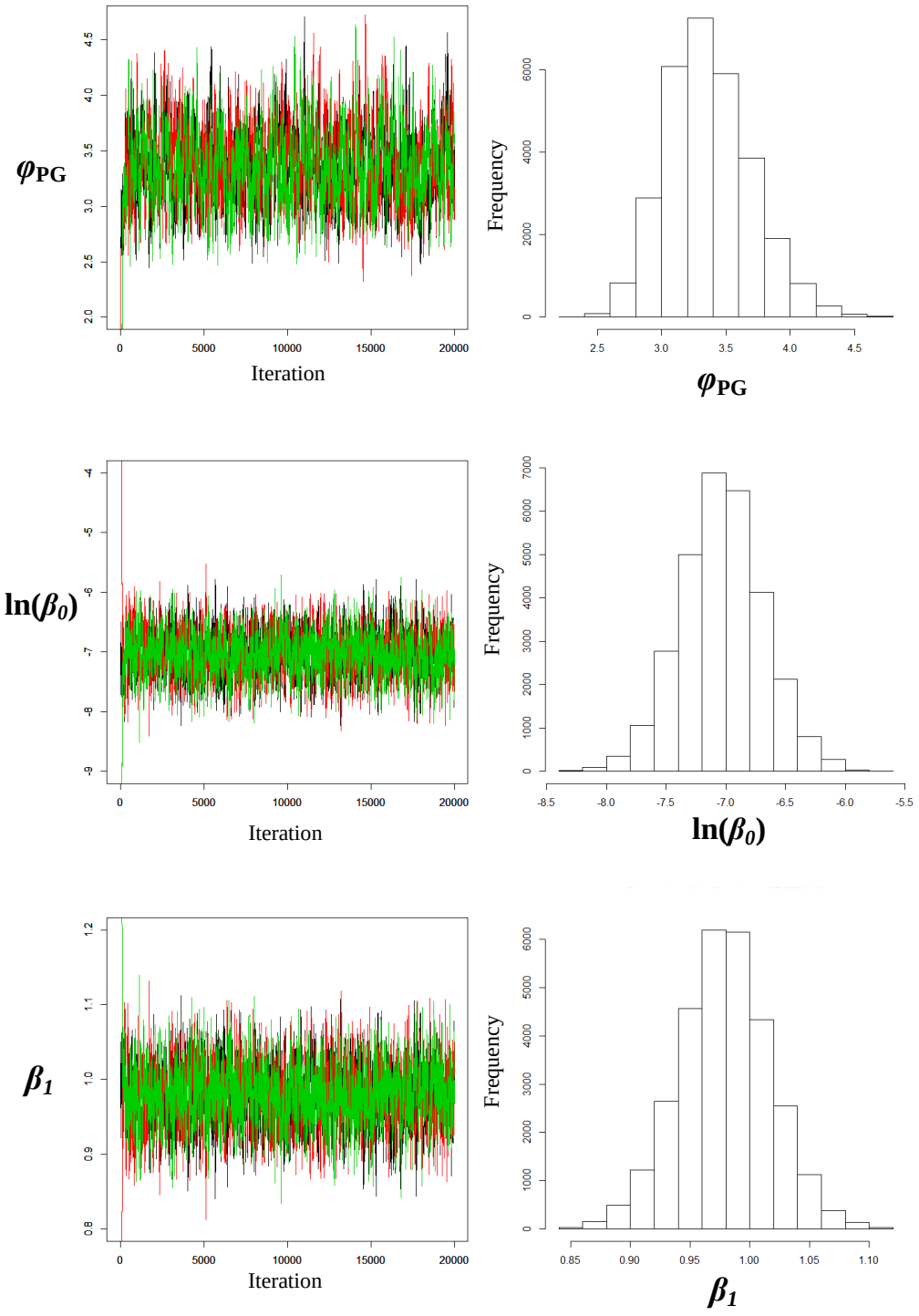




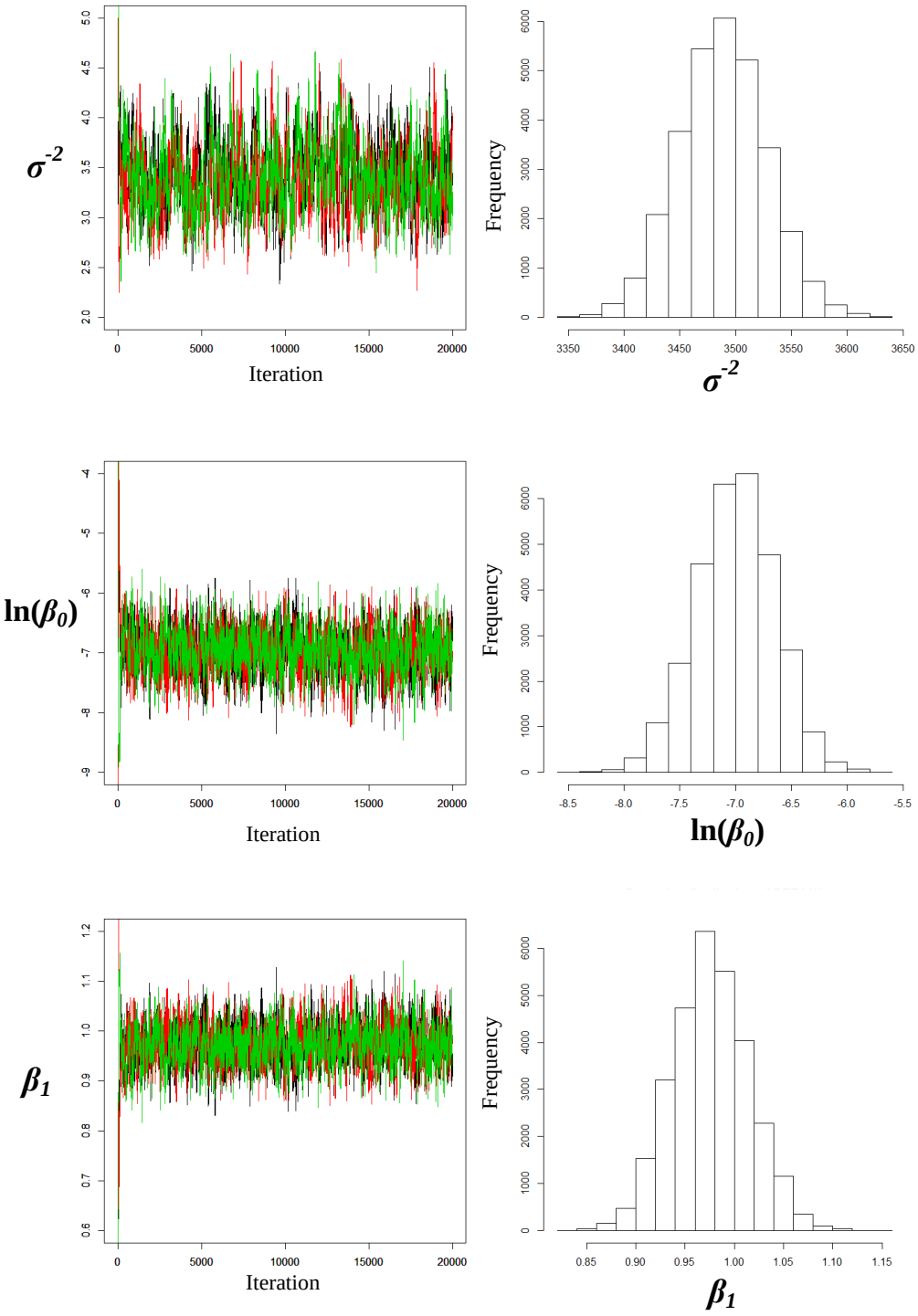




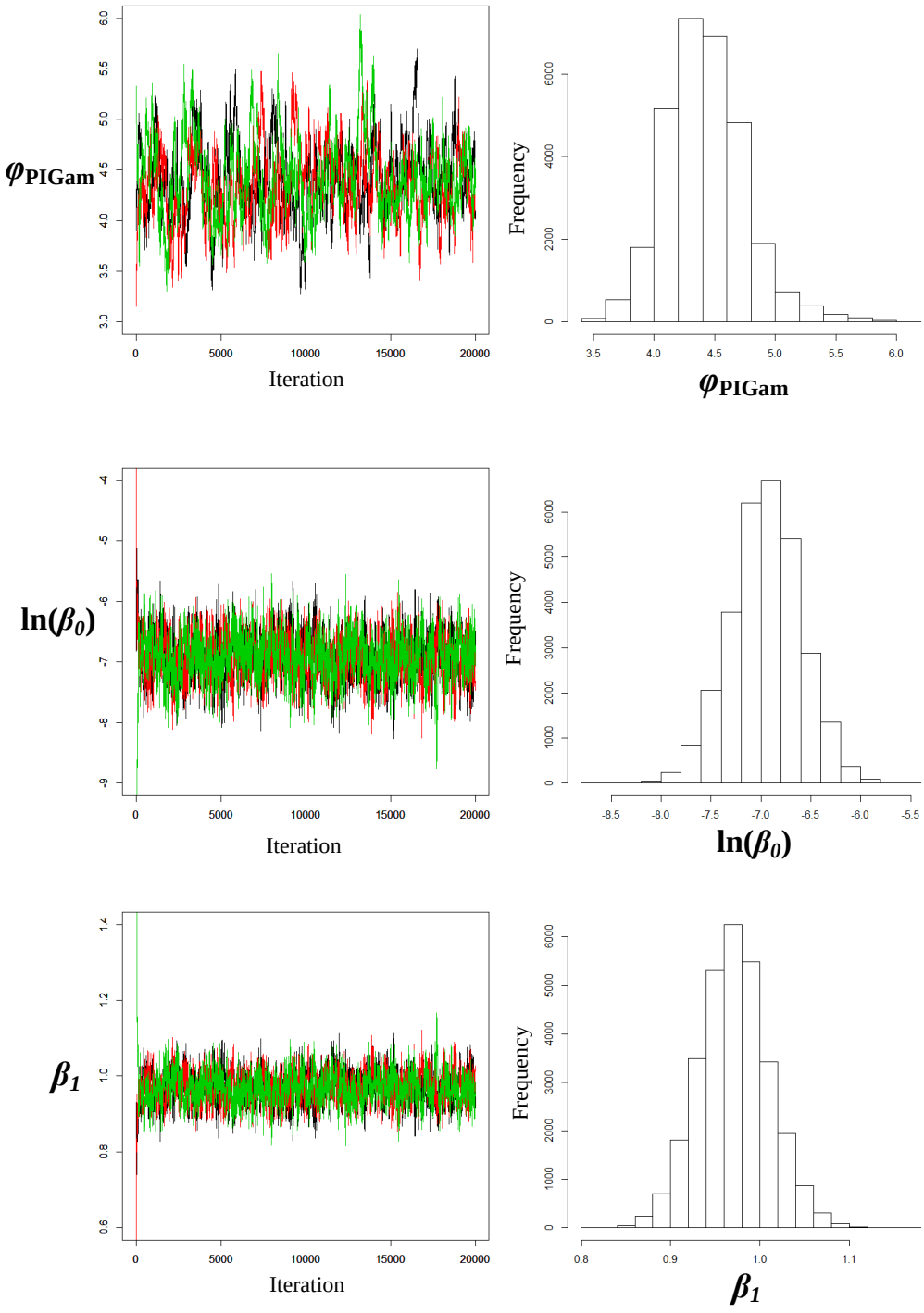
Texas Data, Poisson-Gamma Model



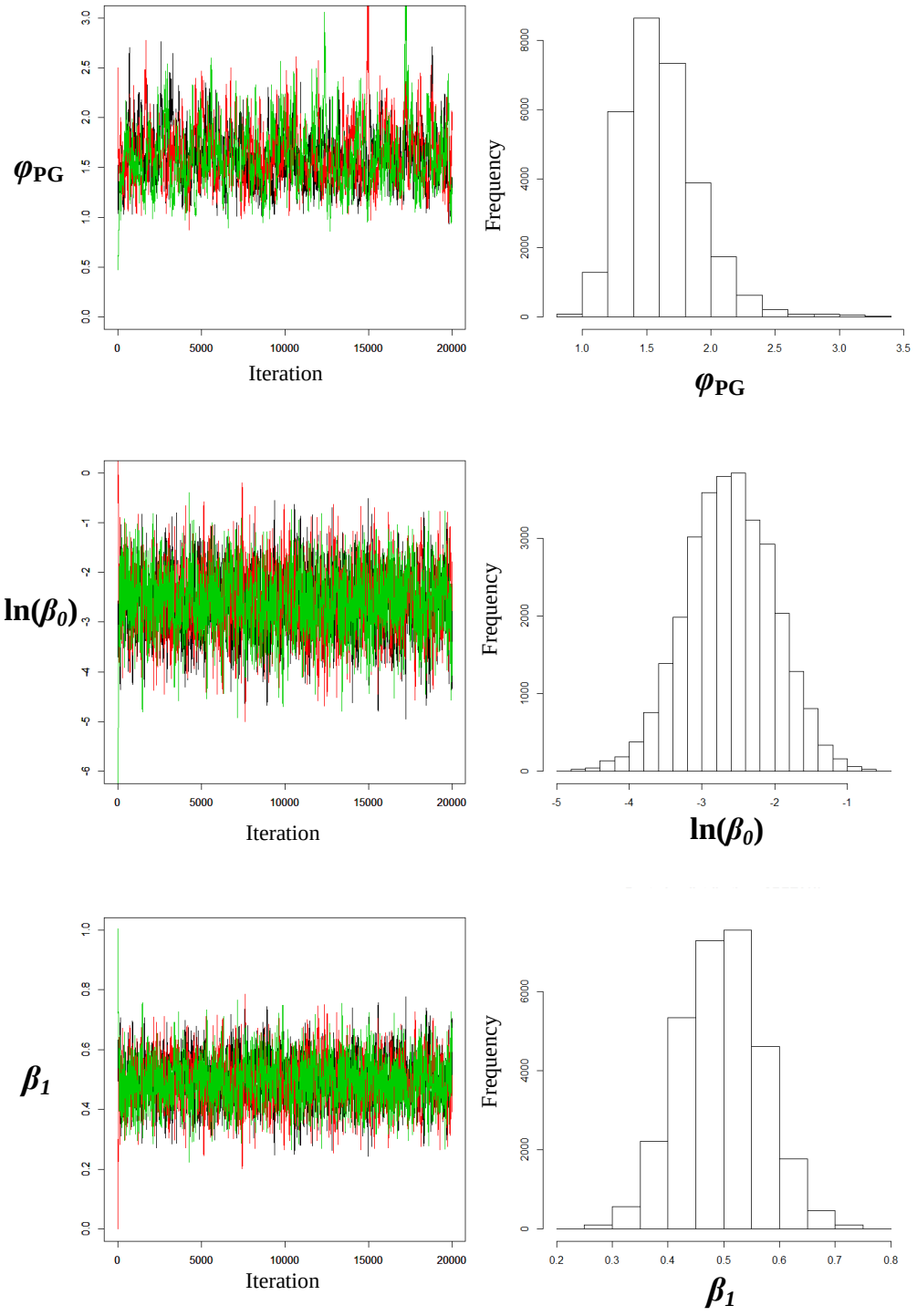
Texas Data, Poisson-Lognormal Model



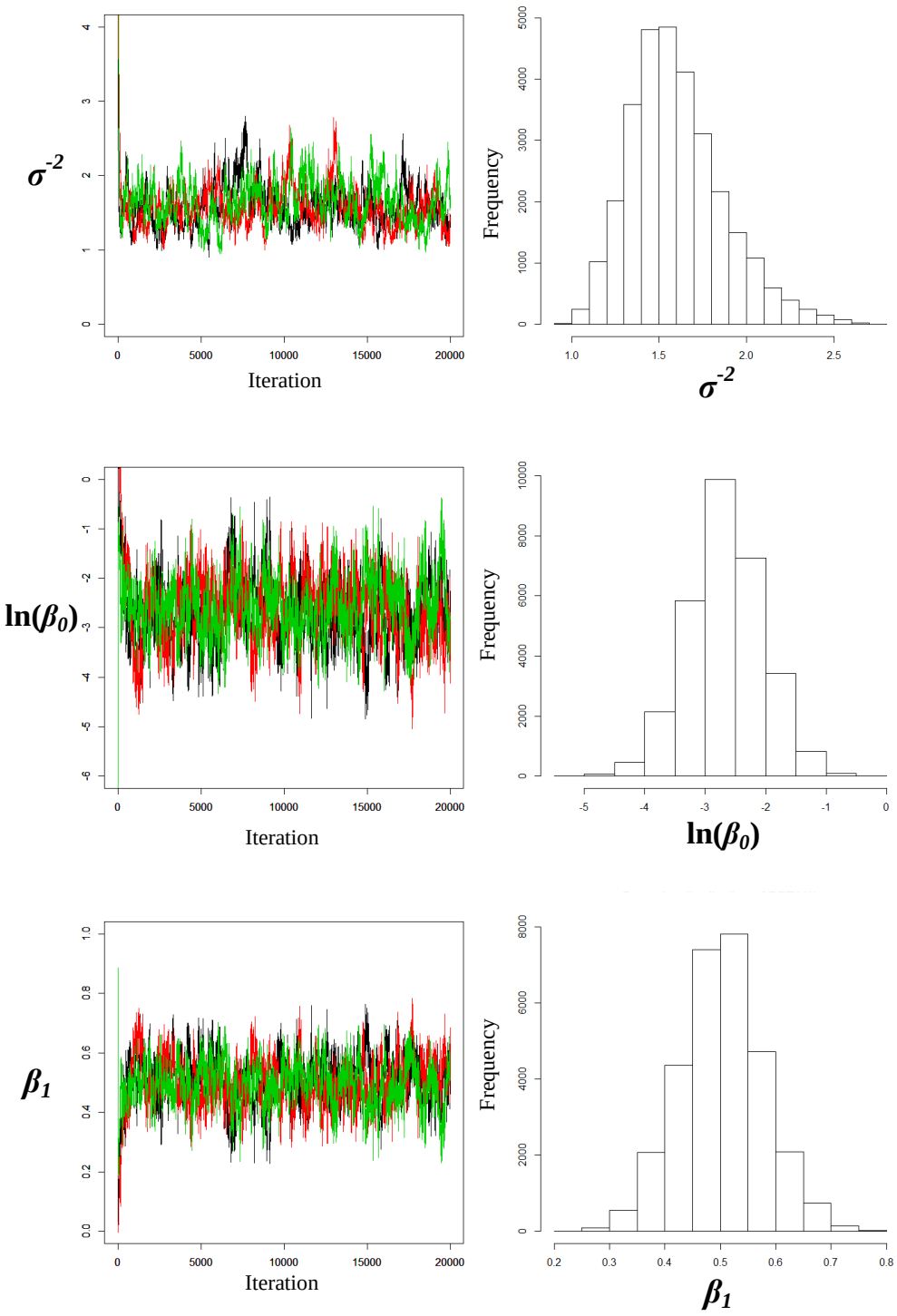
Texas Data, Poisson-Inverse Gamma Model



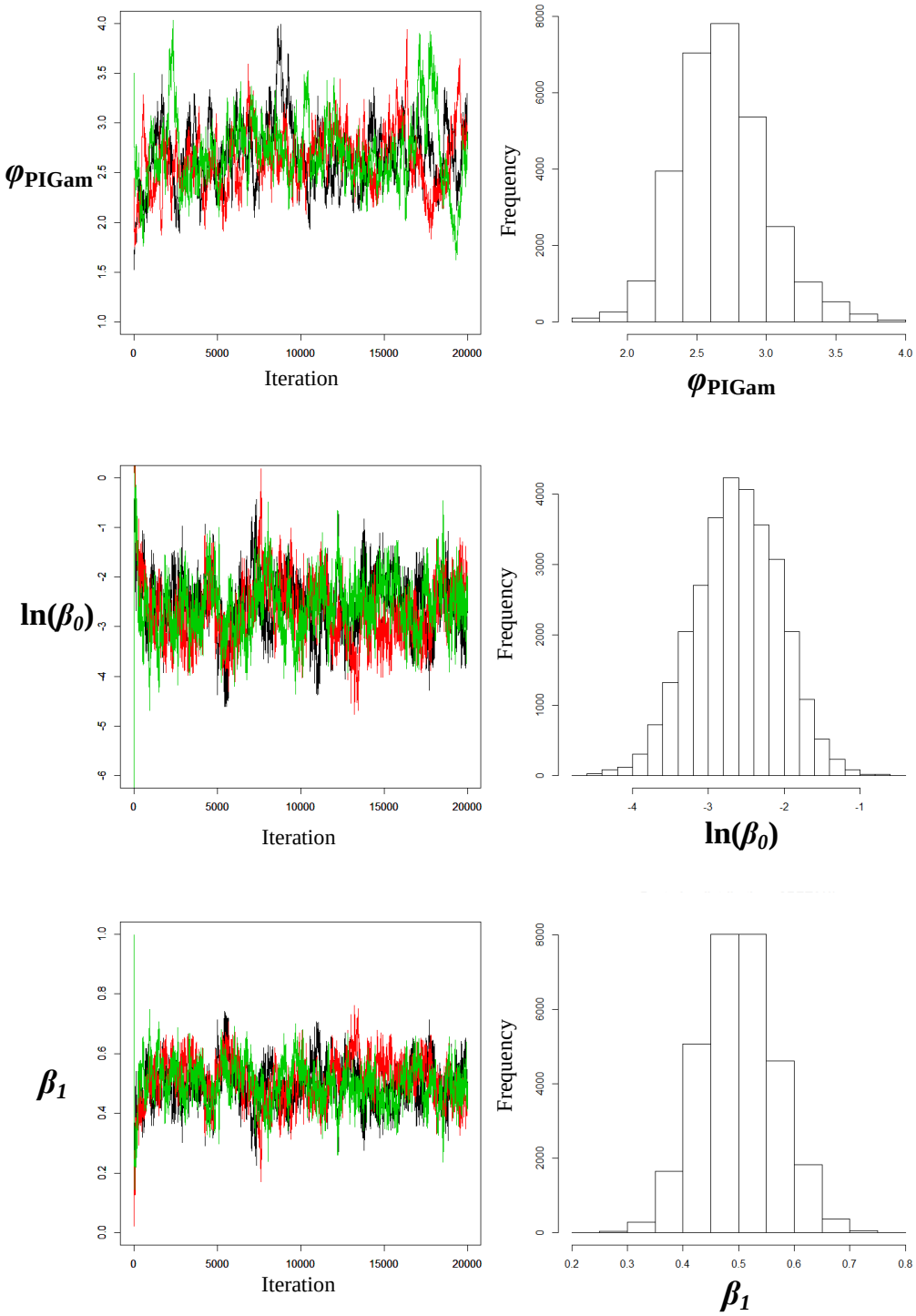
Michigan Data, Poisson-Gamma Model



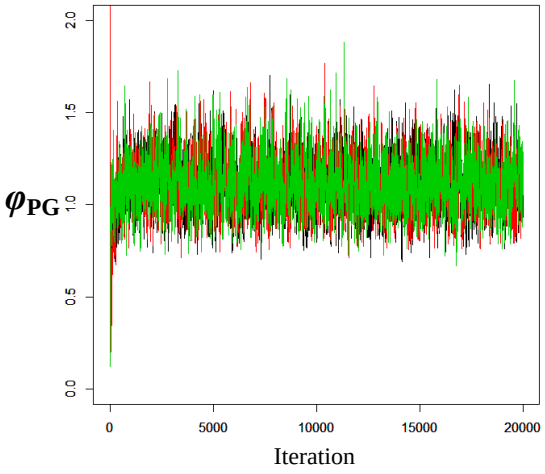
Michigan Data, Poisson-Lognormal Model

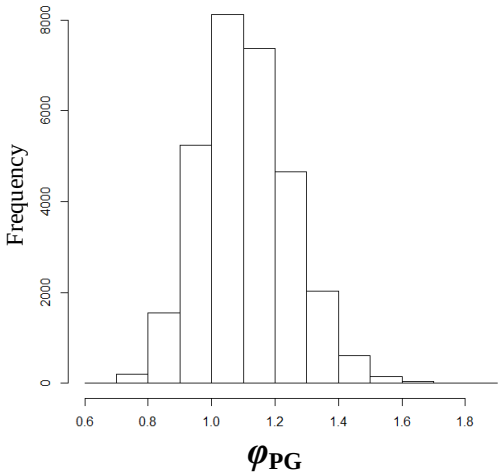


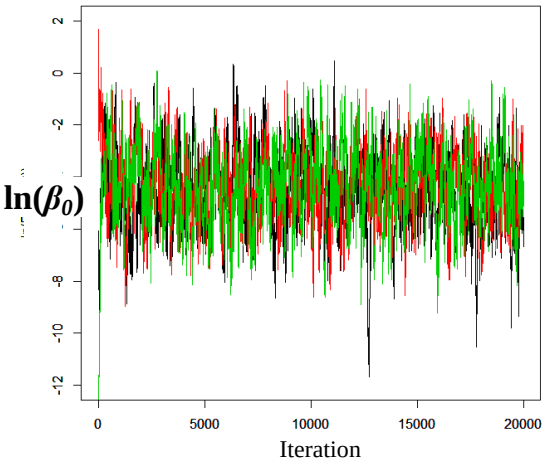
Michigan Data, Poisson-Inverse Gamma Model

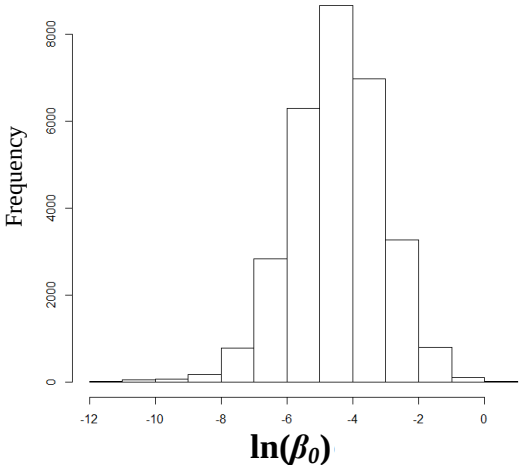


Indiana Data, Poisson-Gamma Model

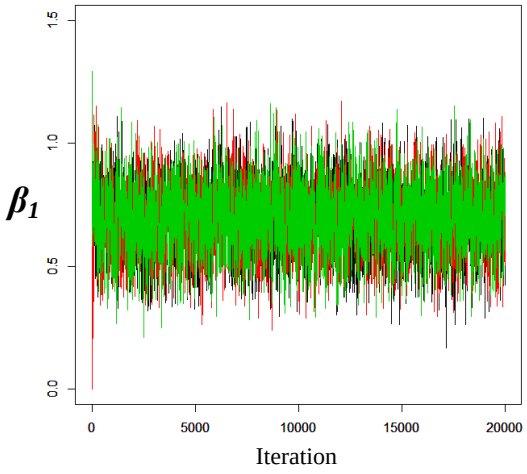


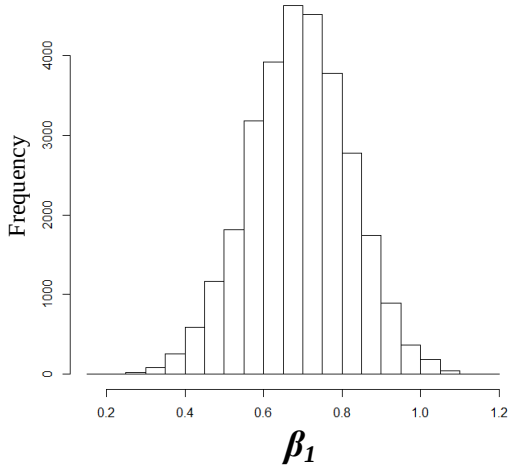


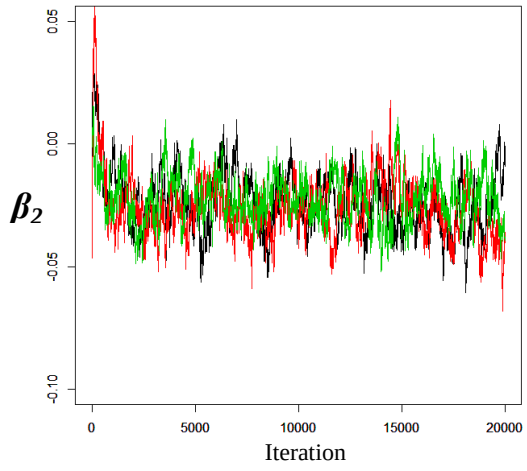


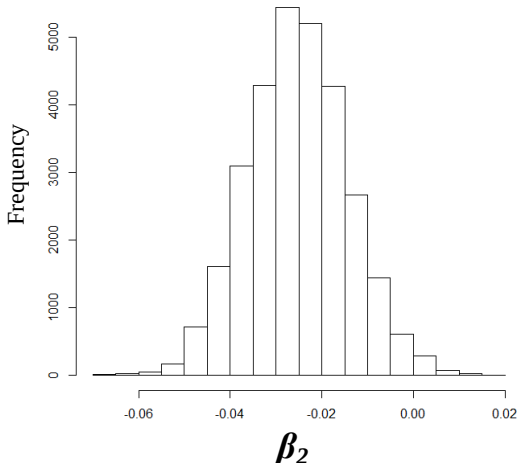


Indiana Data, Poisson-Gamma Model

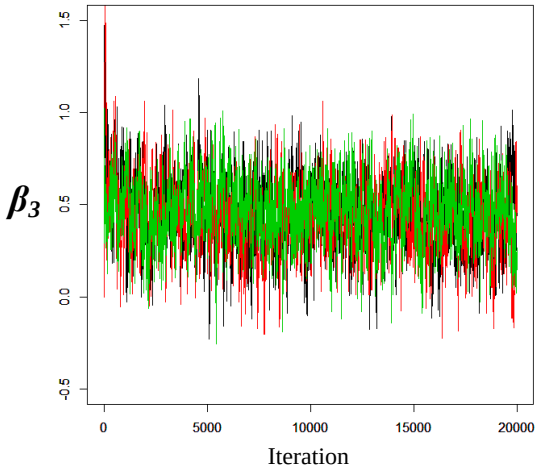


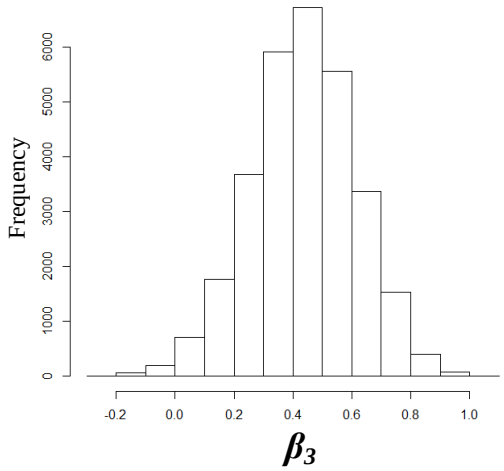


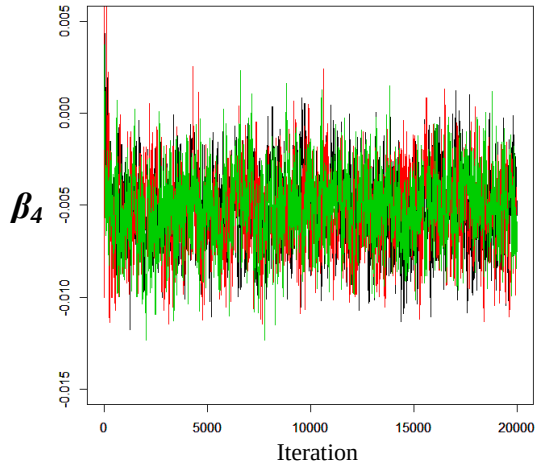


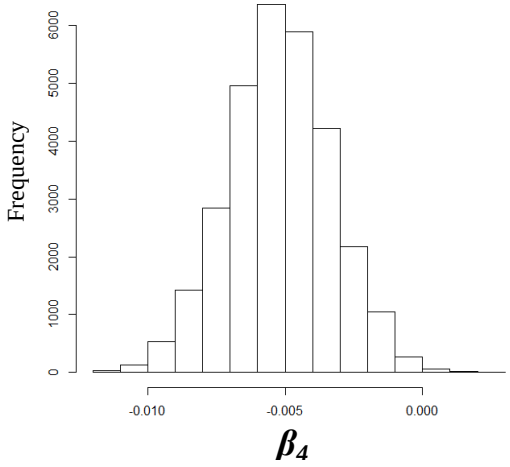


Indiana Data, Poisson-Gamma Model

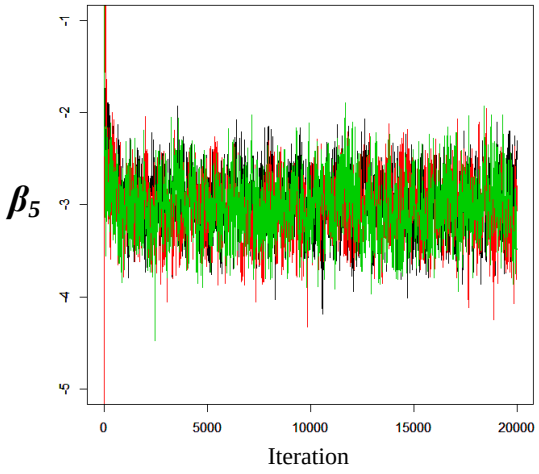


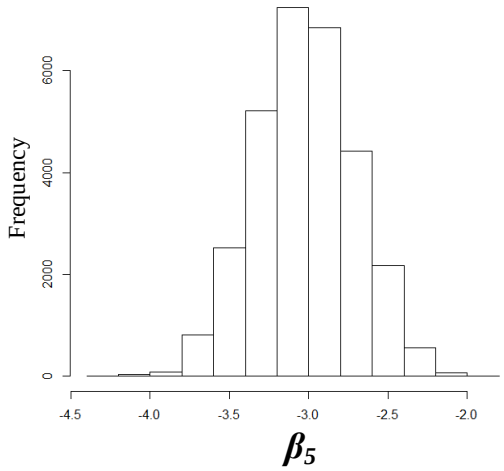


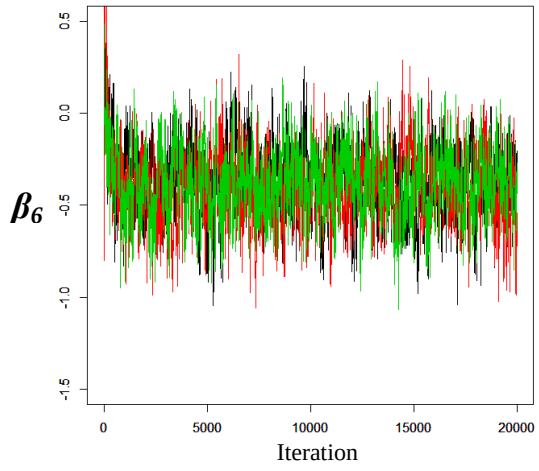


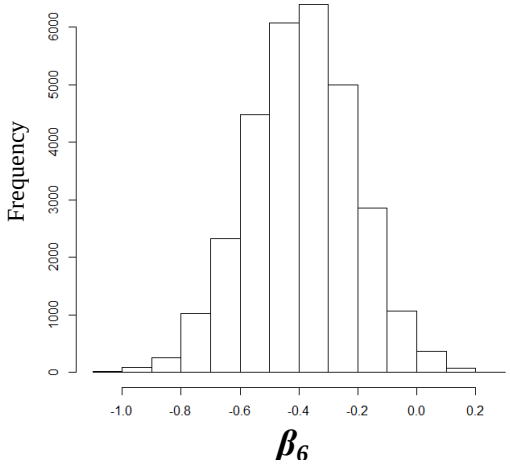


Indiana Data, Poisson-Gamma Model

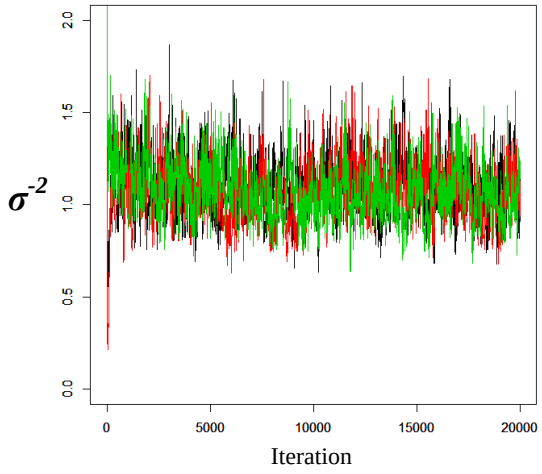


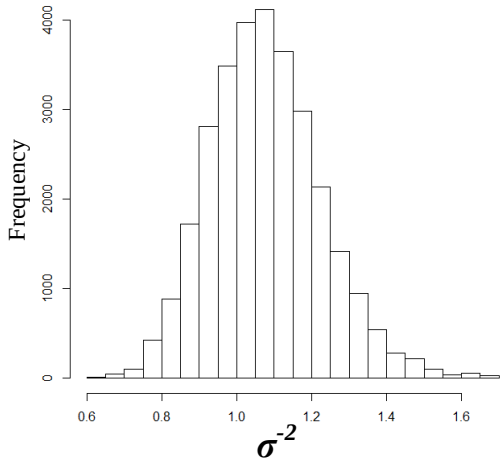


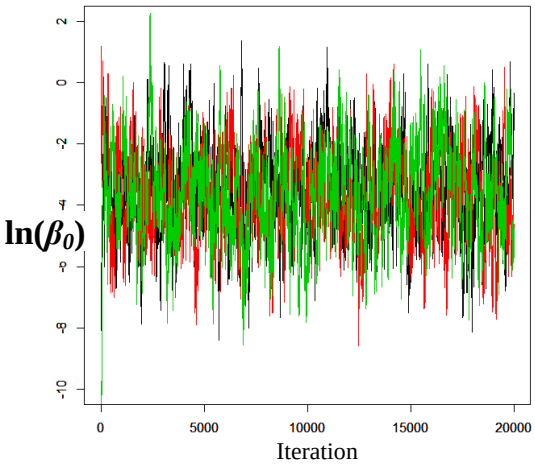


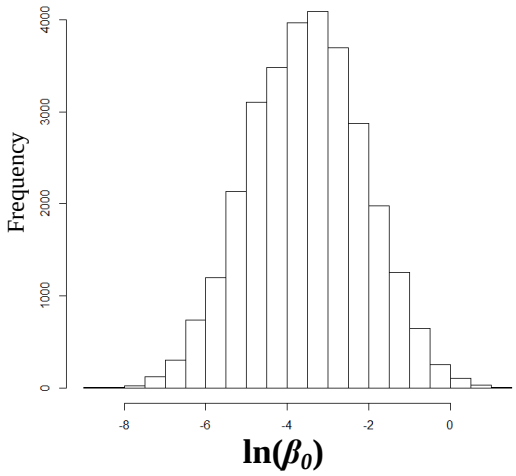


Indiana Data, Poisson-Lognormal Model

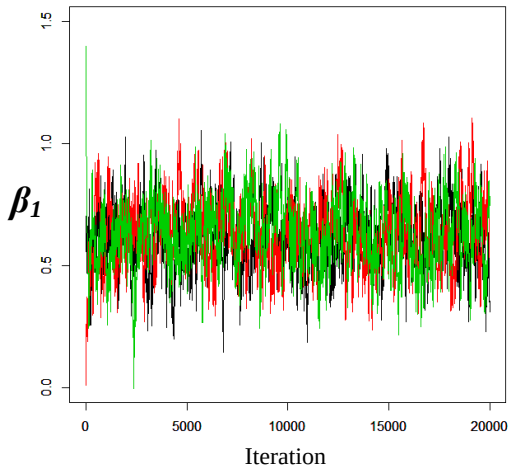


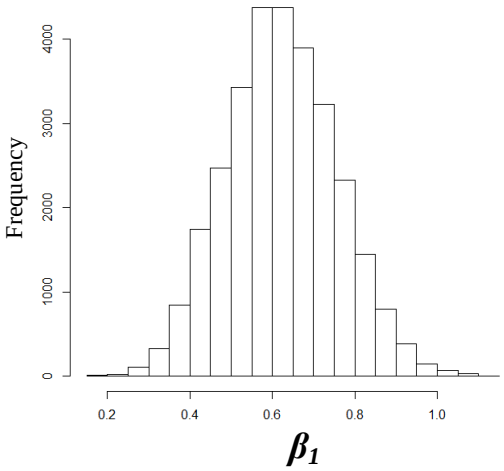


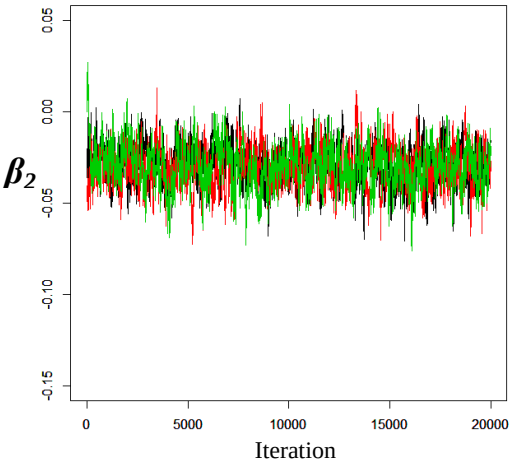


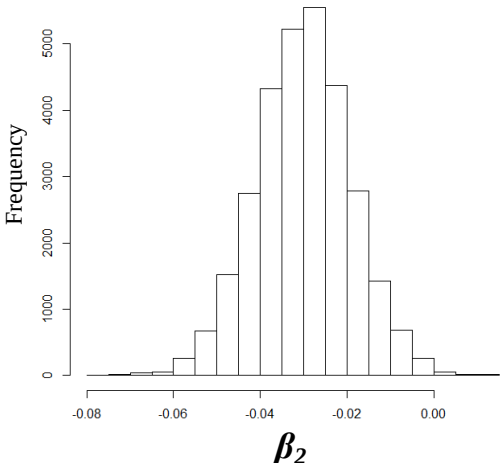


Indiana Data, Poisson-Lognormal Model

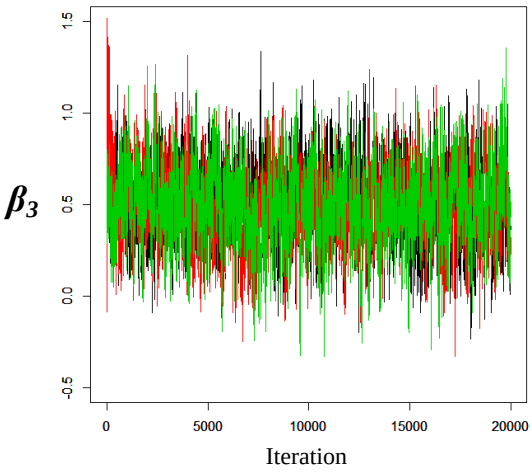


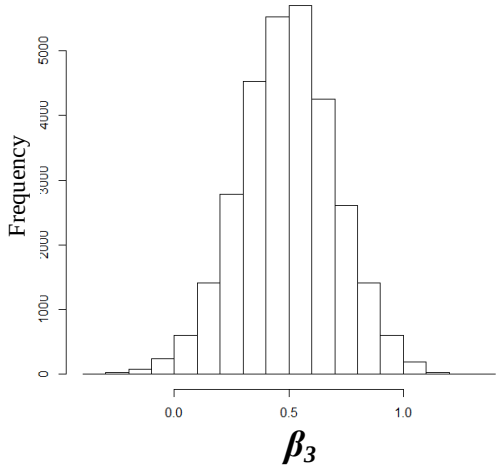


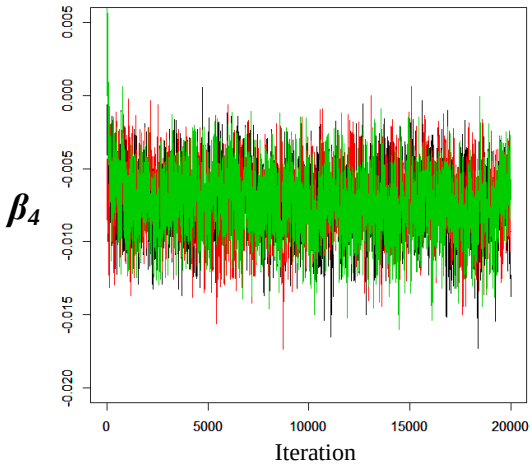


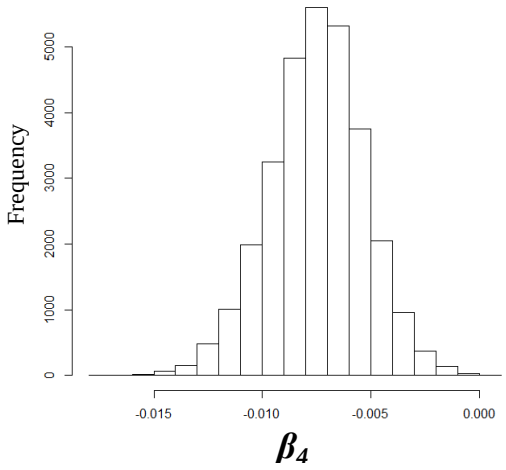


Indiana Data, Poisson- Lognormal Model

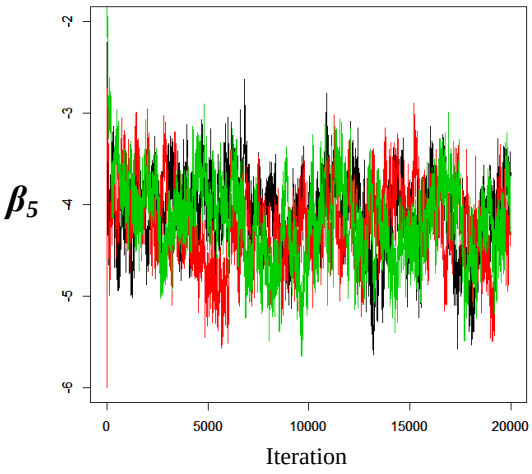


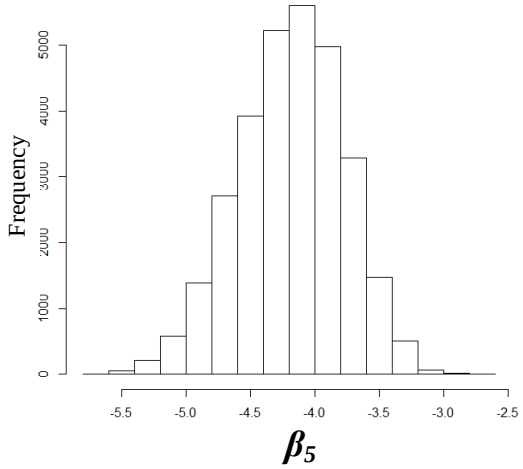


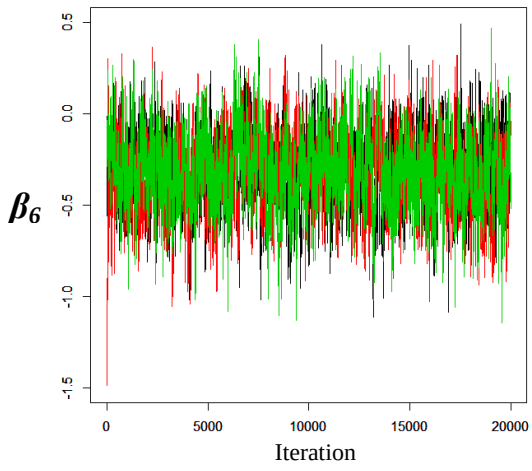


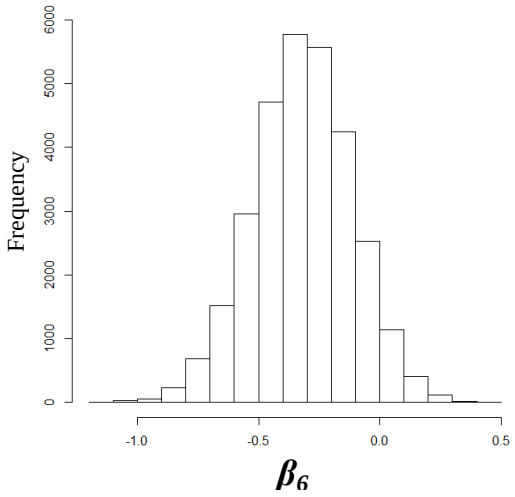


Indiana Data, Poisson-Lognormal Model

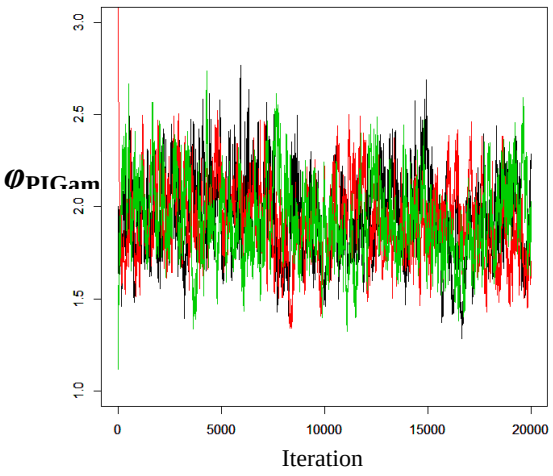


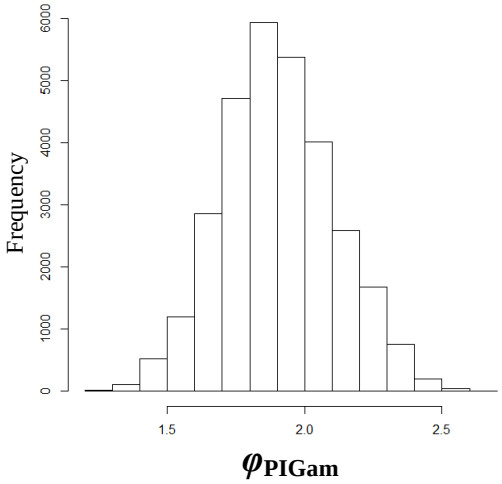


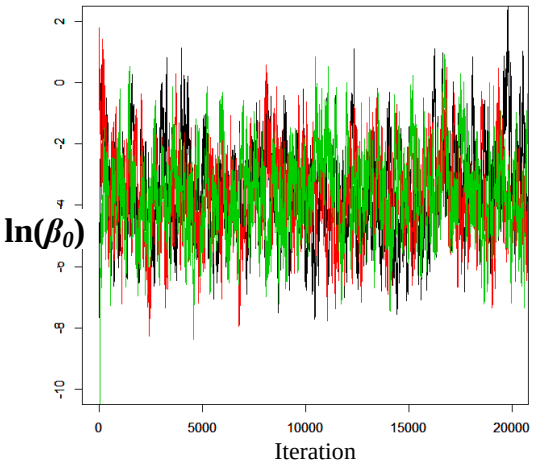


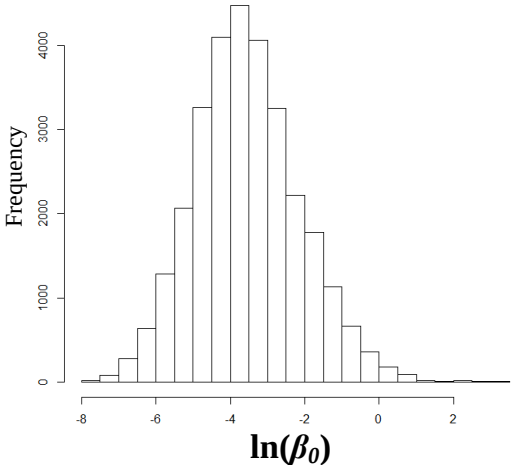


Indiana Data, Poisson-Inverse Gamma Model

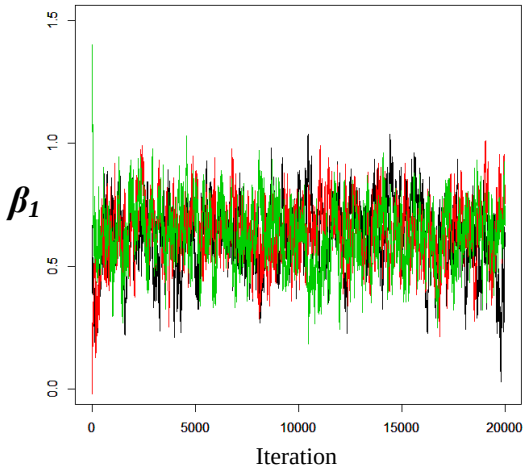


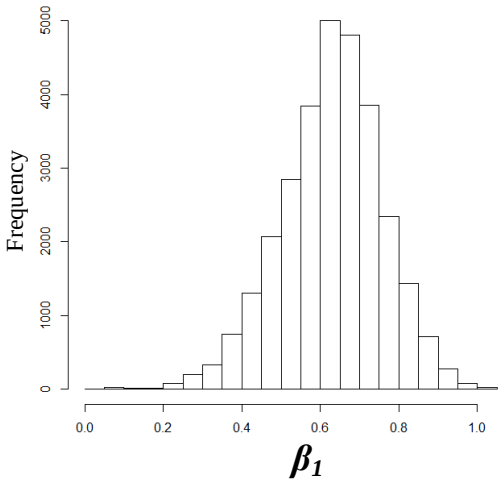


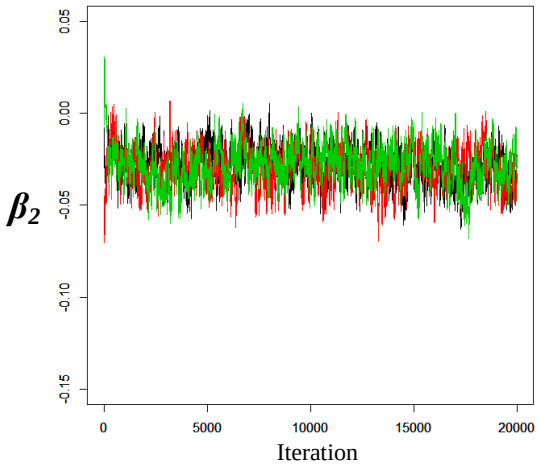


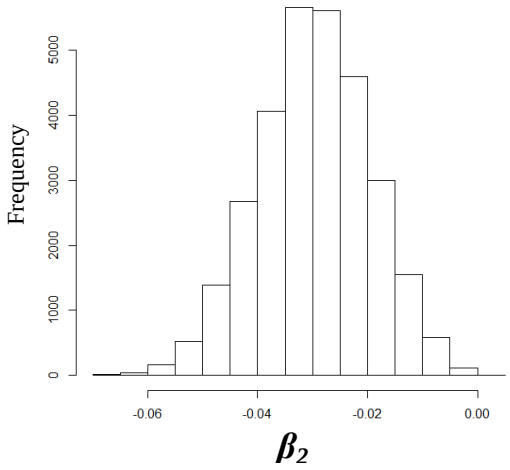


Indiana Data, Poisson-Inverse Gamma Model

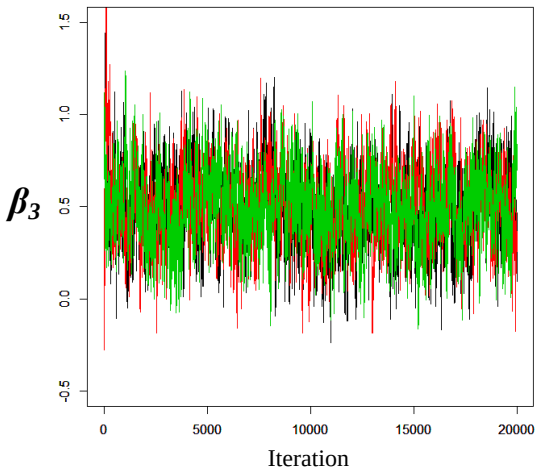


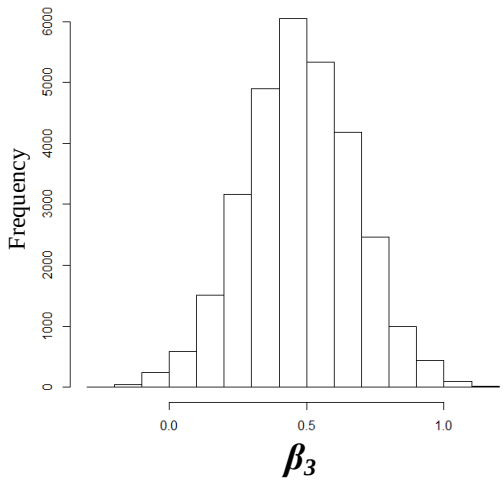


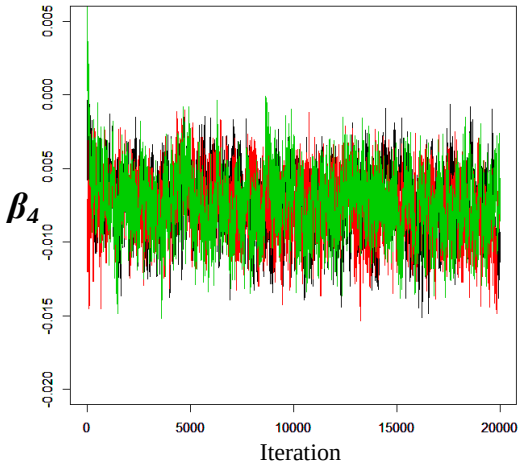


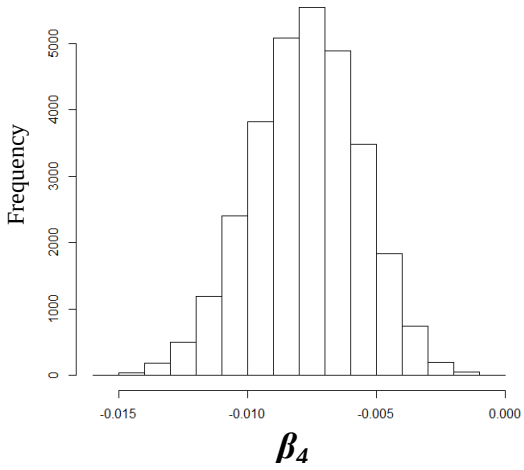


Indiana Data, Poisson-Inverse Gamma Model

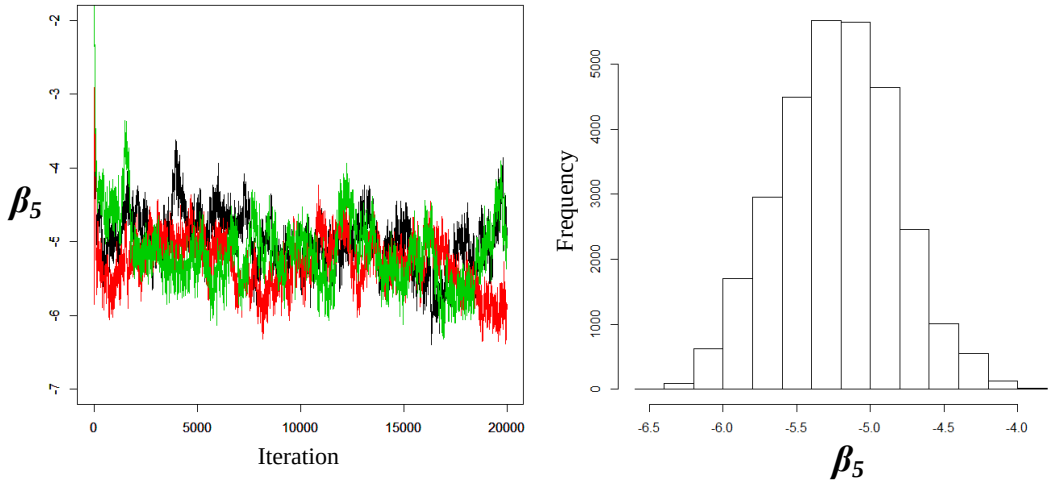


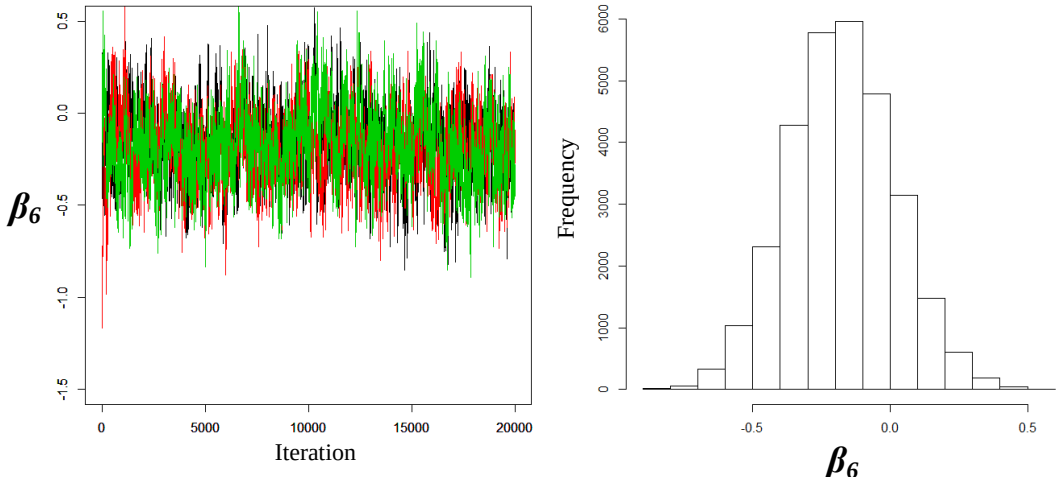






Indiana Data, Poisson-Inverse Gamma Model





APPENDIX B
R CODES FOR MODEL ESTIMATION

This appendix presents the R codes used to estimate the alternative models through the MCMC algorithm. As the purpose is illustrating the method, the codes used for the Indiana dataset are provided here. Separate codes were written for the Poisson-gamma, Poisson-lognormal, and Poisson-inverse gamma models.

Indiana Data, Poisson-Gamma Model

```
data <- read.csv("IN-data-norm-all.csv")

# n = number of units
n = dim(data)[1]
# y = the number of observed crashes, the first column of data
y = data[,1]
# off = the vector of offset values
off = data[,8]
# X = matrix of coveriate values (n rows, p columns)
X = data
X[,1] = c(rep(1,n))
X = X[c(-8)]
# p = number of model covariates (including intercept)
p = dim(X)[2]

# beta = vector of coefficients (p rows, 1 column)
# y~Poisson(m)
# m~Gamma (shape=phi,rate=phi/mu)
# mu(i) = exp(X(i)*beta+offset)
# beta(i) ~ Normal (nu(i),sigma(i))
# phi ~ uniform[0,1000]

#Prior parameters
nu = c(rep(0,p))
sigma = c(rep(100,p))

MCMC = 20000

#chain 1 initial values
m1 = c(rep(1,n))
beta1 = c(-2.448,0.687,-0.027,0.4296,-0.0052,-3.026,-0.398)
phi1 = 1.12
PHI1 = NULL
Dev1 = NULL
MCMC1 = MCMC
BURNIN1 = 10000
BETA1=matrix(rep(0,MCMC1*p),ncol=p)
M1=matrix(rep(0,MCMC1*n),ncol=n)

#chain 2 initial values
m2 = c(rep(1,n))
beta2 = c(-5,0,-0.05,0,-0.01,-6,-0.8)
phi2 = 2.12
PHI2 = NULL
Dev2 = NULL
MCMC2 = MCMC
BURNIN2 = 10000
```

```
BETA2=matrix(rep(0,MCMC*p),ncol=p)
M2=matrix(rep(0,MCMC2*n),ncol=n)

#chain 3 initial values
m3 = c(rep(1,n))
beta3 = c(0,1.4,0,0.8,0,0,0)
phi3 = 0.1
PHI3 = NULL
Dev3 = NULL
MCMC3 = MCMC
BURNIN3 = 10000
BETA3=matrix(rep(0,MCMC3*p),ncol=p)
M3=matrix(rep(0,MCMC3*n),ncol=n)

# Function to calculate sum of m(i)/exp(X(i)*beta+offset) for i=1:n
expsum <- function (beta,m){
  result = 0
          for (u in 1:n) {
          result = result + m[u]/exp(sum(X[u,]*beta)+off[u])
  }
  return(result)
}

#Chain 1
for (i in 1:MCMC1){
          # update m
          for (j in 1:n){
                    m1[j]=rgamma(1,y[j]+phi1,scale=(1+phi1/exp(sum(X[j,]*beta1)+off[j]))^(-1))
          }
          # update phi
          cand = phi1 + 0.1*rnorm(1)
          if (cand <=0) cand =phi1
          logr = n*(cand*log(cand)-log(gamma(cand))-phi1*log(phi1)+log(gamma(phi1))) - (cand -
phi1)*(sum(beta1*colSums(X))+sum(off) + expsum(beta1,m1) - sum(log(m1)))
          if( logr > log(runif(1)) ) phi1 = cand
          # update beta
          for (j in 1:p){
                    if ( j == 1 ) cand = beta1[j] + 0.08*rnorm(1)
                    if ( j == 2 ) cand = beta1[j] + 0.03*rnorm(1)
                    if ( j == 3 ) cand = beta1[j] + 0.003*rnorm(1)
                    if ( j == 4 ) cand = beta1[j] + 0.12*rnorm(1)
                    if ( j == 5 ) cand = beta1[j] + 0.0015*rnorm(1)
                    if ( j == 6 ) cand = beta1[j] + 0.2*rnorm(1)
                    if ( j == 7 ) cand = beta1[j] + 0.11*rnorm(1)
                    betacand = beta1
                    betacand[j] = cand
                    logr = -phi1*(colSums(X)[j]*(cand-beta1[j])+ expsum(betacand,m1) -
                    expsum(beta1,m1)) + log(dnorm(cand,nu[j],sigma[j]))-
                    log(dnorm(beta1[j],nu[j],sigma[j]))
                    if( logr > log(runif(1)) ) beta1[j] = cand
          }
          if( i>BURNIN1 ){
                    PHI1 = c(PHI1,phi1)
```

```
                BETA1[i-BURNIN1,]=beta1
                M1[i-BURNIN1,]=m1
                #Calculate conditional deviance
                loglikelihood  = 0
                for (z in 1:n){
                        loglikelihood = loglikelihood + dpois(y[z],m1[z],log=TRUE)
                }
                Dev1[i-BURNIN1]=-2*loglikelihood
        }
}

rej = c(rep(0,p))
for (r in 1:p){
        for (u in 2:MCMC1){
                if (BETA1[u,r] == BETA1[u-1,r]) rej[r] = rej[r] + 1
        }
}
rejbeta1 = 1-rej/MCMC1

rej=0
for (u in 2:MCMC1){
        if (PHI1[u-1] == PHI1[u]) rej = rej + 1
}
rejphi1 = 1 - rej/MCMC1

#Chain 2
for (i in 1:MCMC2){
        # update m
        for (j in 1:n){
                m2[j]=rgamma(1,y[j]+phi2,scale=(1+phi2/exp(sum(X[j,]*beta2)+off[j]))^(-1))
        }
        # update phi
        cand = phi2 + 0.1*rnorm(1)
        if (cand <=0) cand =phi2
        logr = n*(cand*log(cand)-log(gamma(cand))-phi2*log(phi2)+log(gamma(phi2))) - (cand -
        phi2)*(sum(beta2*colSums(X))+sum(off) + expsum(beta2,m2) - sum(log(m2)))
        if( logr > log(runif(1)) ) phi2 = cand
        # update beta
        for (j in 1:p){
                if ( j == 1 ) cand = beta2[j] + 0.08*rnorm(1)
                if ( j == 2 ) cand = beta2[j] + 0.03*rnorm(1)
                if ( j == 3 ) cand = beta2[j] + 0.003*rnorm(1)
                if ( j == 4 ) cand = beta2[j] + 0.12*rnorm(1)
                if ( j == 5 ) cand = beta2[j] + 0.0015*rnorm(1)
                if ( j == 6 ) cand = beta2[j] + 0.2*rnorm(1)
                if ( j == 7 ) cand = beta2[j] + 0.11*rnorm(1)
                betacand = beta2
                betacand[j] = cand
                logr = -phi2*(colSums(X)[j]*(cand-beta2[j])+ expsum(betacand,m2) -
                expsum(beta2,m2)) + log(dnorm(cand,nu[j],sigma[j]))-
                log(dnorm(beta2[j],nu[j],sigma[j]))
                if( logr > log(runif(1)) ) beta2[j] = cand
        }
```

```
        if( i>BURNIN2 ){
                PHI2 = c(PHI2,phi2)
                BETA2[i-BURNIN2,]=beta2
                M2[i-BURNIN2,]=m2
                #Calculate conditional deviance
                loglikelihood  = 0
                for (z in 1:n){
                        loglikelihood = loglikelihood + dpois(y[z],m2[z],log=TRUE)
                }
                Dev2[i-BURNIN2]=-2*loglikelihood
        }
}

rej = c(rep(0,p))
for (r in 1:p){
        for (u in 2:MCMC2){
                if (BETA2[u,r] == BETA2[u-1,r]) rej[r] = rej[r] + 1
        }
}
rejbeta2 = 1-rej/MCMC2

rej=0
for (u in 2:MCMC2){
        if (PHI2[u-1] == PHI2[u]) rej = rej + 1
}
rejphi2 = 1 - rej/MCMC2

#Chain 3
for (i in 1:MCMC3){
        # update m
        for (j in 1:n){
                m3[j]=rgamma(1,y[j]+phi3,scale=(1+phi3/exp(sum(X[j,]*beta3)+off[j]))^(-1))
        }
        # update phi
        cand = phi3 + 0.1*rnorm(1)
        if (cand <=0) cand =phi3
        logr = n*(cand*log(cand)-log(gamma(cand))-phi3*log(phi3)+log(gamma(phi3))) - (cand -
        phi3)*(sum(beta3*colSums(X))+sum(off) + expsum(beta3,m3) - sum(log(m3)))
        if( logr > log(runif(1)) ) phi3 = cand
        # update beta
        for (j in 1:p){
                if ( j == 1 ) cand = beta3[j] + 0.08*rnorm(1)
                if ( j == 2 ) cand = beta3[j] + 0.03*rnorm(1)
                if ( j == 3 ) cand = beta3[j] + 0.003*rnorm(1)
                if ( j == 4 ) cand = beta3[j] + 0.12*rnorm(1)
                if ( j == 5 ) cand = beta3[j] + 0.0015*rnorm(1)
                if ( j == 6 ) cand = beta3[j] + 0.2*rnorm(1)
                if ( j == 7 ) cand = beta3[j] + 0.11*rnorm(1)
                betacand = beta3
                betacand[j] = cand
                logr = -phi3*(colSums(X)[j]*(cand-beta3[j])+ expsum(betacand,m3) -
                expsum(beta3,m3)) + log(dnorm(cand,nu[j],sigma[j]))-
                log(dnorm(beta3[j],nu[j],sigma[j]))
```

```
                if( logr > log(runif(1)) ) beta3[j] = cand
        }
        if( i>BURNIN3 ){
                PHI3 = c(PHI3,phi3)
                BETA3[i-BURNIN3,]=beta3
                M3[i-BURNIN3,]=m3
                #Calculate conditional deviance
                loglikelihood  = 0
                for (z in 1:n){
                        loglikelihood = loglikelihood + dpois(y[z],m3[z],log=TRUE)
                }
                Dev3[i-BURNIN3]=-2*loglikelihood
        }
}

rej = c(rep(0,p))
for (r in 1:p){
        for (u in 2:MCMC3){
                if (BETA3[u,r] == BETA3[u-1,r]) rej[r] = rej[r] + 1
        }
}
rejbeta3 = 1-rej/MCMC3

rej=0
for (u in 2:MCMC3){
        if (PHI3[u-1] == PHI3[u]) rej = rej + 1
}
rejphi3 = 1 - rej/MCMC3


time <- proc.time() - ptm

PHI = c(PHI1[10001:20000],PHI2[10001:20000],PHI3[10001:20000])
BETA = rbind(BETA1[10001:20000,],BETA2[10001:20000,],BETA3[10001:20000,])
M = rbind(M1[10001:20000,],M2[10001:20000,],M3[10001:20000,])
```

Indiana Data, Poisson-Lognormal Model

```
data <- read.csv("IN-data-norm-all.csv")

# n = number of units
n = dim(data)[1]
# y = the number of observed crashes, the first column of data
y = data[,1]
# off = the vector of offset values
off = data[,8]
# X = matrix of coveriate values (n rows, p columns)
X = data
X[,1] = c(rep(1,n))
X = X[c(-8)]
# p = number of model covariates (including intercept)
p = dim(X)[2]

# beta = vector of coefficients (p rows, 1 column)
# y~Poisson(m)
# m~Lognormal (mu,sigma2)
# mu(i) = X(i)*beta + offset(i) - sigma2/2
# beta(i) ~ Normal (nu(i),sigma(i))
# invsig2 ~ gamma (k,rho)

#Prior parameters
nu = c(rep(0,p))
sigma = c(rep(100,p))
k=0.01
rho=0.01

MCMC=20000

#chain 1 initial values
m1 = c(rep(1,n))
beta1 = c(-2.448,0.687,-0.027,0.4296,-0.0052,-3.026,-0.398)
invsig2_1 = 4
INVSIG2_1 = NULL
Dev1 = NULL
MCMC1 = MCMC
BURNIN1 = 10000
BETA1 = matrix(rep(0,MCMC1*p),ncol=p)
M1 = matrix(rep(0,MCMC1*n),ncol=n)

#chain 2 initial values
m2 = c(rep(1,n))
beta2 = c(-5,0,-0.05,0,-0.01,-6,-0.8)
invsig2_2 = 5
INVSIG2_2 = NULL
Dev2 = NULL
MCMC2 = MCMC
```

```
BURNIN2 = 10000
BETA2 = matrix(rep(0,MCMC2*p),ncol=p)
M2 = matrix(rep(0,MCMC2*n),ncol=n)

#chain 3 initial values
m3 = c(rep(1,n))
beta3 = c(0,1.4,0,0.8,0,0,0)
invsig2_3 = 6
INVSIG2_3 = NULL
Dev3 = NULL
MCMC3 = MCMC
BURNIN3 = 10000
BETA3 = matrix(rep(0,MCMC3*p),ncol=p)
M3 = matrix(rep(0,MCMC3*n),ncol=n)

# Function to calculate sum of (log(m(i))- X(i)*beta + 0.5/invsigma2)^2 for i=1:n
funcsum <- function (beta,m,invsigma2){
  result = 0
        for (u in 1:n) {
        result = result + (log(m[u])-sum(X[u,]*beta) - off[u] +0.5/invsigma2)^2
  }
  return(result)
}

#Chain 1
for (i in 1:MCMC1){
        # update m
        for (j in 1:n){
                cand = m1[j] + 1*rnorm(1)
                if (cand <=0) cand = m1[j]
                        logr = (y[j]-1)*(log(cand)-log(m1[j])) - (cand - m1[j]) -
                        0.5*invsig2_1*((log(cand)-sum(X[j,]*beta1)-off[j]+0.5/invsig2_1)^2-
                        (log(m1[j])-sum(X[j,]*beta1)-off[j]+0.5/invsig2_1)^2)
                if( logr > log(runif(1)) ) m1[j] = cand
        }
        # update invsig2
        cand = invsig2_1 + 0.4*rnorm(1)
        if (cand <=0) cand = invsig2_1
                logr = 0.5*n*(log(cand)-log(invsig2_1)) - 0.5*(funcsum(beta1,m1,cand)*cand -
                funcsum(beta1,m1,invsig2_1)*invsig2_1) + log(dgamma(cand,k,rho)) -
                log(dgamma(invsig2_1,k,rho))
        if( logr > log(runif(1)) ) invsig2_1 = cand
        # update beta
        for (j in 1:p){
                if ( j == 1 ) cand = beta1[j] + 0.08*rnorm(1)
                if ( j == 2 ) cand = beta1[j] + 0.03*rnorm(1)
                if ( j == 3 ) cand = beta1[j] + 0.003*rnorm(1)
                if ( j == 4 ) cand = beta1[j] + 0.12*rnorm(1)
                if ( j == 5 ) cand = beta1[j] + 0.0015*rnorm(1)
                if ( j == 6 ) cand = beta1[j] + 0.2*rnorm(1)
                if ( j == 7 ) cand = beta1[j] + 0.11*rnorm(1)
                betacand = beta1
                betacand[j] = cand
```

```
		logr = -0.5*invsig2_1*(funcsum(betacand,m1,invsig2_1)-funcsum(beta1,m1,invsig2_1))
		if( logr > log(runif(1)) ) beta1[j] = cand
	}
	if( i>BURNIN1 ){
		INVSIG2_1 = c(INVSIG2_1,invsig2_1)
		BETA1[i-BURNIN1,]=beta1
		M1[i-BURNIN1,]=m1
		#Calculate conditional deviance
		loglikelihood  = 0
		for (z in 1:n){
			loglikelihood = loglikelihood + dpois(y[z],m1[z],log=TRUE)
		}
		Dev1[i-BURNIN1]=-2*loglikelihood
	}
}

rej = c(rep(0,p))
for (r in 1:p){
	for (u in 2:MCMC1){
		if (BETA1[u,r] == BETA1[u-1,r]) rej[r] = rej[r] + 1
	}
}
betaaccratio1 = 1-rej/MCMC1

rejinvsig2=0
for (u in 2:MCMC1){
	if (INVSIG2_1[u-1] == INVSIG2_1[u]) rejinvsig2 = rejinvsig2 + 1
}
invsig2accratio1 = 1 - rejinvsig2/MCMC1

#Chain 2
for (i in 1:MCMC2){
	# update m
	for (j in 1:n){
		cand = m2[j] + 1*rnorm(1)
		if (cand <=0) cand = m2[j]
			logr = (y[j]-1)*(log(cand)-log(m2[j])) - (cand - m2[j]) -
			0.5*invsig2_2*((log(cand)-sum(X[j,]*beta2)-off[j]+0.5/invsig2_2)^2-
			(log(m2[j])-sum(X[j,]*beta2)-off[j]+0.5/invsig2_2)^2)
		if( logr > log(runif(1)) ) m2[j] = cand
	}
	# update invsig2
	cand = invsig2_2 + 0.4*rnorm(1)
	if (cand <=0) cand = invsig2_2
		logr = 0.5*n*(log(cand)-log(invsig2_2)) - 0.5*(funcsum(beta2,m2,cand)*cand -
		funcsum(beta2,m2,invsig2_2)*invsig2_2) + log(dgamma(cand,k,rho)) -
		log(dgamma(invsig2_2,k,rho))
	if( logr > log(runif(1)) ) invsig2_2 = cand
	# update beta
	for (j in 1:p){
		if ( j == 1 ) cand = beta2[j] + 0.08*rnorm(1)
		if ( j == 2 ) cand = beta2[j] + 0.03*rnorm(1)
		if ( j == 3 ) cand = beta2[j] + 0.003*rnorm(1)
```

```
		if ( j == 4 ) cand = beta2[j] + 0.12*rnorm(1)
		if ( j == 5 ) cand = beta2[j] + 0.0015*rnorm(1)
		if ( j == 6 ) cand = beta2[j] + 0.2*rnorm(1)
		if ( j == 7 ) cand = beta2[j] + 0.11*rnorm(1)
		betacand = beta2
		betacand[j] = cand
		logr = -0.5*invsig2_2*(funcsum(betacand,m2,invsig2_2)-funcsum(beta2,m2,invsig2_2))
		if( logr > log(runif(1)) ) beta2[j] = cand
	}
	if( i>BURNIN2 ){
		INVSIG2_2 = c(INVSIG2_2,invsig2_2)
		BETA2[i-BURNIN2,]=beta2
		M2[i-BURNIN2,]=m2
		#Calculate conditional deviance
		loglikelihood  = 0
		for (z in 1:n){
			loglikelihood = loglikelihood + dpois(y[z],m2[z],log=TRUE)
		}
		Dev2[i-BURNIN2]=-2*loglikelihood
	}
}

rej = c(rep(0,p))
for (r in 1:p){
	for (u in 2:MCMC2){
		if (BETA2[u,r] == BETA2[u-1,r]) rej[r] = rej[r] + 1
	}
}
betaaccratio2 = 1-rej/MCMC2

rejinvsig2=0
for (u in 2:MCMC2){
	if (INVSIG2_2[u-1] == INVSIG2_2[u]) rejinvsig2 = rejinvsig2 + 1
}
invsig2accratio2 = 1 - rejinvsig2/MCMC2

#Chain 3
for (i in 1:MCMC3){
	# update m
	for (j in 1:n){
		cand = m3[j] + 1*rnorm(1)
		if (cand <=0) cand = m3[j]
			logr = (y[j]-1)*(log(cand)-log(m3[j])) - (cand - m3[j]) -
			0.5*invsig2_3*((log(cand)-sum(X[j,]*beta3)-off[j]+0.5/invsig2_3)^2-
			(log(m3[j])-sum(X[j,]*beta3)-off[j]+0.5/invsig2_3)^2)
		if( logr > log(runif(1)) ) m3[j] = cand
	}
	# update invsig2
	cand = invsig2_3 + 0.4*rnorm(1)
	if (cand <=0) cand = invsig2_3
		logr = 0.5*n*(log(cand)-log(invsig2_3)) - 0.5*(funcsum(beta3,m3,cand)*cand -
		funcsum(beta3,m3,invsig2_3)*invsig2_3) + log(dgamma(cand,k,rho)) -
		log(dgamma(invsig2_3,k,rho))
```

```
		if( logr > log(runif(1)) ) invsig2_3 = cand
		# update beta
		for (j in 1:p){
			if ( j == 1 ) cand = beta3[j] + 0.08*rnorm(1)
			if ( j == 2 ) cand = beta3[j] + 0.03*rnorm(1)
			if ( j == 3 ) cand = beta3[j] + 0.003*rnorm(1)
			if ( j == 4 ) cand = beta3[j] + 0.12*rnorm(1)
			if ( j == 5 ) cand = beta3[j] + 0.0015*rnorm(1)
			if ( j == 6 ) cand = beta3[j] + 0.2*rnorm(1)
			if ( j == 7 ) cand = beta3[j] + 0.11*rnorm(1)
			betacand = beta3
			betacand[j] = cand
			logr = -0.5*invsig2_3*(funcsum(betacand,m3,invsig2_3)-funcsum(beta3,m3,invsig2_3))
			if( logr > log(runif(1)) ) beta3[j] = cand
		}
		if( i>BURNIN3 ){
			INVSIG2_3 = c(INVSIG2_3,invsig2_3)
			BETA3[i-BURNIN3,]=beta3
			M3[i-BURNIN3,]=m3
			#Calculate conditional deviance
			loglikelihood  = 0
			for (z in 1:n){
				loglikelihood = loglikelihood + dpois(y[z],m3[z],log=TRUE)
			}
			Dev3[i-BURNIN3]=-2*loglikelihood
		}
}

rej = c(rep(0,p))
for (r in 1:p){
		for (u in 2:MCMC3){
			if (BETA3[u,r] == BETA3[u-1,r]) rej[r] = rej[r] + 1
		}
}
betaaccratio3 = 1-rej/MCMC3

rejinvsig2=0
for (u in 2:MCMC3){
		if (INVSIG2_3[u-1] == INVSIG2_3[u]) rejinvsig2 = rejinvsig2 + 1
}
invsig2accratio3 = 1 - rejinvsig2/MCMC3

INVSIG2 = c(INVSIG2_1[10001:20000],INVSIG2_2[10001:20000],INVSIG2_3[10001:20000])
BETA = rbind(BETA1[10001:20000,],BETA2[10001:20000,],BETA3[10001:20000,])
M = rbind(M1[10001:20000,],M2[10001:20000,],M3[10001:20000,])
```

Indiana Data, Poisson-Lognormal Model

```
data <- read.csv("IN-data-norm-all.csv")

# n = number of units
n = dim(data)[1]
# y = the number of observed crashes, the first column of data
y = data[,1]
# off = the vector of offset values
off = data[,8]
# X = matrix of covariate values (n rows, p columns)
X = data
X[,1] = c(rep(1,n))
X = X[c(-8)]
# p = number of model covariates (including intercept)
p = dim(X)[2]

# Function to calculate sum of exp(X(i)*beta+offset)/m(i) for i=1:n
expsum <- function (X,beta,m,off){
  result = 0
          for (u in 1:length(m)) {
          result = result + exp(sum(X[u,]*beta)+off[u])/m[u]
  }
  return(result)
}

# beta = vector of coefficients (p rows, 1 column)
# y~Poisson(m)
# m~Inverse-Gamma (mu,phi)
# mu(i) = exp(X(i)*beta)
# beta(i) ~ Normal (nu(i),sigma(i))
# phi ~ uniform[0,1000]

#Prior parameters
nu = c(rep(0,p))
sigma = c(rep(100,p))

MCMC=20000

#Chain 1 initial values
m1 = c(rep(1,n))
beta1 = c(-2.448,0.687,-0.027,0.4296,-0.0052,-3.026,-0.398)
phi1 = 2.12
PHI1 = NULL
Dev1 = NULL
MCMC1 = MCMC
BURNIN1 = 10000
BETA1=matrix(rep(0,MCMC*p),ncol=p)
M1=matrix(rep(0,MCMC*n),ncol=n)
```

```
#Chain 2 initial values
m2 = c(rep(1,n))
beta2 = c(-5,0,-0.05,0,-0.01,-6,-0.8)
phi2 = 3.12
PHI2 = NULL
Dev2 = NULL
MCMC2 = MCMC
BURNIN2 = 10000
BETA2=matrix(rep(0,MCMC*p),ncol=p)
M2=matrix(rep(0,MCMC*n),ncol=n)

#Chain 3 initial values
m3 = c(rep(1,n))
beta3 = c(0,1.4,0,0.8,0,0,0)
phi3 = 1.12
PHI3 = NULL
Dev3 = NULL
MCMC3 = MCMC
BURNIN3 = 10000
BETA3=matrix(rep(0,MCMC*p),ncol=p)
M3=matrix(rep(0,MCMC*n),ncol=n)

#Chain 1
for (i in 1:MCMC1){
        # update m
        for (j in 1:n){
                cand = m1[j] + 1*rnorm(1)
                if (cand <= 0) cand = m1[j]
                logr = -cand + m1[j] - (phi1-1)*exp(sum(X[j,]*beta1)+off[j])*(1/cand-1/m1[j]) + (-
                phi1+y[j]-1)*(log(cand)-log(m1[j]))
                if( logr > log(runif(1)) ) m1[j] = cand
        }
        # update phi
        cand = phi1 + 0.05*rnorm(1)
        if (cand <=1) cand = phi1
        logr = (cand - phi1)*(sum(beta1*colSums(X))+sum(off)) - expsum(X,beta1,m1,off)*(cand-phi1)
        +  sum(log(m1))*(-cand+phi1) + n*(log((cand-1)^cand/gamma(cand))-log((phi1-
        1)^phi1/gamma(phi1)))
        if( logr > log(runif(1)) ) phi1 = cand
        # update beta
        for (j in 1:p){
                if ( j == 1 ) cand = beta1[j] + 0.08*rnorm(1)
                if ( j == 2 ) cand = beta1[j] + 0.03*rnorm(1)
                if ( j == 3 ) cand = beta1[j] + 0.003*rnorm(1)
                if ( j == 4 ) cand = beta1[j] + 0.12*rnorm(1)
                if ( j == 5 ) cand = beta1[j] + 0.0015*rnorm(1)
                if ( j == 6 ) cand = beta1[j] + 0.2*rnorm(1)
                if ( j == 7 ) cand = beta1[j] + 0.11*rnorm(1)
                betacand = beta1
                betacand[j] = cand
                logr = phi1*colSums(X)[j]*(cand-beta1[j])-(phi1-1)*(expsum(X,betacand,m1,off) -
                expsum(X,beta1,m1,off)) + log(dnorm(cand,nu[j],sigma[j]))-
                log(dnorm(beta1[j],nu[j],sigma[j]))
```

```
                if( logr > log(runif(1)) ) beta1[j] = cand
        }
        if( i>BURNIN1 ){
                PHI1 = c(PHI1,phi1)
                BETA1[i-BURNIN1,]=beta1
                M1[i-BURNIN1,]=m1
                #Calculate deviance
                loglikelihood = 0
                for (z in 1:n){
                        loglikelihood = loglikelihood + dpois(y[z],m1[z],log=TRUE)
                }
                Dev1[i-BURNIN] = -2*loglikelihood
        }
}

rej = c(rep(0,p))
for (r in 1:p){
        for (u in 2:MCMC1){
                if (BETA1[u,r] == BETA1[u-1,r]) rej[r] = rej[r] + 1
        }
}
rejbeta1 = 1 - rej/MCMC1

rej=0
for (u in 2:MCMC1){
        if (PHI1[u-1] == PHI1[u]) rej = rej + 1
}
rejphi1 = 1 - rej/MCMC1

#Chain 2
for (i in 1:MCMC2){
        # update m
        for (j in 1:n){
                cand = m2[j] + 1*rnorm(1)
                if (cand <= 0) cand = m2[j]
                logr = -cand + m2[j] - (phi2-1)*exp(sum(X[j,]*beta2)+off[j])*(1/cand-1/m2[j]) + (-
                phi2+y[j]-1)*(log(cand)-log(m2[j]))
                if( logr > log(runif(1)) ) m2[j] = cand
        }
        # update phi
        cand = phi2 + 0.05*rnorm(1)
        if (cand <=1) cand = phi2
        logr = (cand - phi2)*(sum(beta2*colSums(X))+sum(off)) - expsum(X,beta2,m2,off)*(cand-phi2)
        +  sum(log(m2))*(-cand+phi2) + n*(log((cand-1)^cand/gamma(cand))-log((phi2-
        1)^phi2/gamma(phi2)))
        if( logr > log(runif(1)) ) phi2 = cand
        # update beta
        for (j in 1:p){
                if ( j == 1 ) cand = beta2[j] + 0.08*rnorm(1)
                if ( j == 2 ) cand = beta2[j] + 0.03*rnorm(1)
                if ( j == 3 ) cand = beta2[j] + 0.003*rnorm(1)
                if ( j == 4 ) cand = beta2[j] + 0.12*rnorm(1)
                if ( j == 5 ) cand = beta2[j] + 0.0015*rnorm(1)
```

```
		if ( j == 6 ) cand = beta2[j] + 0.2*rnorm(1)
		if ( j == 7 ) cand = beta2[j] + 0.11*rnorm(1)
		betacand = beta2
		betacand[j] = cand
		logr = phi2*colSums(X)[j]*(cand-beta2[j])-(phi2-1)*(expsum(X,betacand,m2,off) -
		expsum(X,beta2,m2,off)) + log(dnorm(cand,nu[j],sigma[j]))-
		log(dnorm(beta2[j],nu[j],sigma[j]))
		if( logr > log(runif(1)) ) beta2[j] = cand
	}
	if( i>BURNIN2 ){
		PHI2 = c(PHI2,phi2)
		BETA2[i-BURNIN2,]=beta2
		M2[i-BURNIN2,]=m2
		#Calculate deviance
		loglikelihood = 0
		for (z in 1:n){
			loglikelihood = loglikelihood + dpois(y[z],m2[z],log=TRUE)
		}
		Dev2[i-BURNIN2] = -2*loglikelihood
	}
}

rej = c(rep(0,p))
for (r in 1:p){
	for (u in 2:MCMC2){
		if (BETA2[u,r] == BETA2[u-1,r]) rej[r] = rej[r] + 1
	}
}
rejbeta2 = 1 - rej/MCMC2

rej=0
for (u in 2:MCMC2){
	if (PHI2[u-1] == PHI2[u]) rej = rej + 1
}
rejphi2 = 1 - rej/MCMC2

#Chain 3
for (i in 1:MCMC3){
	# update m
	for (j in 1:n){
		cand = m3[j] + 1*rnorm(1)
		if (cand <= 0) cand = m3[j]
		logr = -cand + m3[j] - (phi3-1)*exp(sum(X[j,]*beta3)+off[j])*(1/cand-1/m3[j]) + (-
		phi3+y[j]-1)*(log(cand)-log(m3[j]))
		if( logr > log(runif(1)) ) m3[j] = cand
	}
	# update phi
	cand = phi3 + 0.05*rnorm(1)
	if (cand <=1) cand = phi3
	logr = (cand - phi3)*(sum(beta3*colSums(X))+sum(off)) - expsum(X,beta3,m3,off)*(cand-phi3)
	+  sum(log(m3))*(-cand+phi3) + n*(log((cand-1)^cand/gamma(cand))-log((phi3-
	1)^phi3/gamma(phi3)))
	if( logr > log(runif(1)) ) phi3 = cand
```

```
	# update beta
	for (j in 1:p){
		if ( j == 1 ) cand = beta3[j] + 0.08*rnorm(1)
		if ( j == 2 ) cand = beta3[j] + 0.03*rnorm(1)
		if ( j == 3 ) cand = beta3[j] + 0.003*rnorm(1)
		if ( j == 4 ) cand = beta3[j] + 0.12*rnorm(1)
		if ( j == 5 ) cand = beta3[j] + 0.0015*rnorm(1)
		if ( j == 6 ) cand = beta3[j] + 0.2*rnorm(1)
		if ( j == 7 ) cand = beta3[j] + 0.11*rnorm(1)
		betacand = beta3
		betacand[j] = cand
		logr = phi3*colSums(X)[j]*(cand-beta3[j])-(phi3-1)*(expsum(X,betacand,m3,off) -
		expsum(X,beta3,m3,off)) + log(dnorm(cand,nu[j],sigma[j]))-
		log(dnorm(beta3[j],nu[j],sigma[j]))
		if( logr > log(runif(1)) ) beta3[j] = cand
	}
	if( i>BURNIN3 ){
		PHI3 = c(PHI3,phi3)
		BETA3[i-BURNIN3,]=beta3
		M3[i-BURNIN3,]=m3
		#Calculate deviance
		loglikelihood = 0
		for (z in 1:n){
			loglikelihood = loglikelihood + dpois(y[z],m3[z],log=TRUE)
		}
		Dev3[i-BURNIN3] = -2*loglikelihood
	}
}

rej = c(rep(0,p))
for (r in 1:p){
	for (u in 2:MCMC3){
		if (BETA3[u,r] == BETA3[u-1,r]) rej[r] = rej[r] + 1
	}
}
rejbeta3 = 1 - rej/MCMC3

rej=0
for (u in 2:MCMC3){
	if (PHI3[u-1] == PHI3[u]) rej = rej + 1
}
rejphi3 = 1 - rej/MCMC3

time <- proc.time() - ptm

PHI = c(PHI1[10001:20000],PHI2[10001:20000],PHI3[10001:20000])
BETA = rbind(BETA1[10001:20000,],BETA2[10001:20000,],BETA3[10001:20000,])
M = rbind(M1[10001:20000,],M2[10001:20000,],M3[10001:20000,])
```