

2018

BAYESIAN SPATIO-TEMPORAL ANALYSIS OF ROAD TRAFFIC CRASHES

Amin Azimian

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Azimian, Amin, "BAYESIAN SPATIO-TEMPORAL ANALYSIS OF ROAD TRAFFIC CRASHES" (2018). *Graduate Theses, Dissertations, and Problem Reports*. 7158.
<https://researchrepository.wvu.edu/etd/7158>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

BAYESIAN SPATIO-TEMPORAL ANALYSIS OF ROAD TRAFFIC CRASHES

Amin Azimian

**Dissertation submitted to the College of
Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Doctor of Philosophy
in
Civil Engineering**

**V. Dimitra Pyrialakou Ph.D., Committee Chairperson
Kakan Dey Ph.D.
David Martinelli Ph.D.
Steven Lavrenz Ph.D.
Ashish Nimbarte Ph.D.
Sijin Wen Ph.D.**

Department of Civil and Environmental Engineering

**Morgantown, West Virginia
2018**

**Keywords: transportation safety, spatial, temporal, Bayesian, traffic crash,
microscopic, macroscopic**

Copyright 2018 Amin Azimian

Abstract

BAYESIAN SPATIO-TEMPORAL ANALYSIS OF ROAD TRAFFIC CRASHES

Amin Azimian

Road traffic crashes are one of the major causes of death and serious injury in the US, leading to economic losses and human suffering. In recent years, several research efforts have been made to screen areas and locate hotspots/zones and to identify the factors contributing to traffic crashes of different severity levels. Such research typically aggregates crash locations into spatial units at the macro level, such as counties, or at the micro level, such as road segments or intersections.

Many crash estimation methods have been proposed in the literature. These methods range from classical approaches, such as the linear, Poisson, negative binomial, and logistic regression methods, to more state-of-the-art approaches, such as the empirical Bayesian (EB), spatial autoregressive, and full Bayesian methods. A considerable drawback of classical methods is that they cannot account for the regression-to-the-mean bias and potential unobserved heterogeneity, which can result in unstable and biased parameter estimates. Additionally, EB and spatial autoregressive methods are unable to address multilevel data and group-level random effects. By contrast, the full Bayesian framework is more flexible and can be easily extended to include random effect terms that can act as proxies for unobserved or missing covariates that have a spatial or temporal structure.

Although the full Bayesian framework appears to be a promising methodology for dealing with crash data, the few previous studies in the area of road traffic safety that used this framework had some major limitations. Some studies performed their analyses in

univariate settings that did not account for correlations among crash severities, whereas other studies did not address spatial and/or temporal effects properly. Additionally, the key factors that might affect the number and/or severity of crashes, such as urbanity and hospital accessibility, were not explored. Finally, studies used either micro- or macro-level data to perform safety analyses for detecting hotspots. No comprehensive research combining the two has been conducted. All of these limitations might result in the loss of available information, as well as in biased and inconsistent findings.

This dissertation proposes an integrated safety screening approach that combines macro- and a micro-level analyses and accounts for the above-mentioned issues. This approach is illustrated using the case of West Virginia and is expected to provide a comprehensive and effective framework for use in transport safety planning.

In the macro-level data analysis, various multivariate Bayesian models (Poisson lognormal models) were estimated to relate various county-level socio-economic and transport-related factors with crash occurrence/frequency while accounting for unobserved heterogeneity in a multivariate setting. In the micro-level data analysis, multi-level data were used to examine various Bayesian logit models, including nominal and ordered response models, in order to identify the micro-level factors contributing to fatality risk while accounting for random effects over space and time and between crashes. The results from the two levels were combined using a GIS-based approach to identify hotspots.

DEDICATION

This dissertation is lovingly dedicated to my family. Your support and encouragement mean so much to me, and I am forever grateful. Thank you!

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. V. Dimitra Pyrialakou. I am deeply honored to be her first Ph.D. student. I appreciate all her contributions in terms of her time, ideas, and efforts to make my Ph.D. experience truly productive and stimulating. The joy and enthusiasm she has for research were contagious and motivational for me, especially during the tough times in my Ph.D. pursuit.

I also thank my committee members, namely, Drs. David Martinelli, Kakan Dey, Sijin Wen, Ashish Nimbarte, and Steven Lavrenz, for all their advice and support.

Finally, I thank my family for their love, care, and support. Specifically, I express my immense gratitude to my lovely wife, Hanieh, for her patience and unparalleled support. My family is my biggest support system, materially and morally.

TABLE OF CONTENTS

Abstract	i
Dedication.....	iv
Acknowledgments.....	v
Table of Contents	vi
LIST OF FIGURES	vii
LIST OF TABLES	viii
LIST OF SYMBOLS/NOMENCLATURE.....	ix
CHAPTER 1: INTRODUCTION	1
1.1. Research Motivation and Problem Statement.....	1
1.2. Research Goal and Framework	3
1.3. Research Contributions.....	4
1.4. Dissertation Organization	5
CHAPTER 2: LITERATURE REVIEW	6
2.1. Statistical Methods Overview	6
2.2. Crash Factors.....	10
2.3. Macroscopic Traffic Safety Research	12
2.4. Microscopic Traffic Safety Research	28
2.5. Summary	27
CHAPTER 3: DATA COLLECTION AND EXPLORATORY DATA ANALYSIS	29
3.1. Macro-level Data.....	29
3.2. Micro-level Data	36
3.3. Exploratory Spatial Data Analysis	38
3.4. Summary	48
CHAPTER 4: METHODOLOGY	49
4.1. Bayesian Framework	49
4.2. Macro-level Data Analysis	50
4.3. Micro-level Data Analysis	56
4.4. Marginal Effects.....	59
4.5. Model Comparison.....	59
4.6. Integration of Macro- and Micro-level Data Analyses	60
4.7. Summary	63
CHAPTER 5: RESEARCH RESULTS	65
5.1. Macro-level Data Analysis	65
5.2. Micro-level Data Analysis	91
5.3. Integration of Micro- and Macro-level Data Analyses	103
5.4. Summary	106
CHAPTER 6: CONCLUSION	110
6.1. Summary	110
6.2. Research Contributions.....	111
6.3. Generalizability and Transferability	117
6.4. Summary	119
References	121
Appendix	130

LIST OF FIGURES

Figure 1.1 Research framework for the development of an integrated screening method	4
Figure 2.1 Illustration of regression to the mean and the empirical Bayes estimate.	7
Figure 2.2 Multilevel structure of crash data.....	11
Figure 3.1 Framework for estimating the least cost/travel time to the nearest hospital.....	30
Figure 3.2 Map of (a) the least travel time to the nearest hospital and (b) the average access time to the nearest hospital per county	31
Figure 3.3 Map of (a) the ground elevation from DEM and (b) the average ground elevation per county	34
Figure 3.4 Area-level data collection and preparation process.....	36
Figure 3.5 Procedure for relating crash points to census tracts	37
Figure 3.6 Histograms of the average over the years 2010–2015: (a) fatal crash rate, (b) injury crash rate, and (c) PDO crash rate.....	39
Figure 3.7 Boxmap of the (a) average fatal crash rate, (b) average injury crash rate, and (c) average PDO crash rate	41
Figure 3.8 Queen’s case contiguity	42
Figure 3.9 Average (a) fatal crash rate, (b) injury crash rate, and (c) PDO crash rate in West Virginia	43
Figure 3.10 Scatter plot of the fatal crash rate (per 10,000 population) against covariates.	45
Figure 3.11 Scatter plot of the injury crash rate (per 10,000 population) against covariates	46
Figure 3.12 Scatter plot of the PDO crash rate (per 10,000 population) against covariates	47
Figure 4.1 Schematic framework for calculating the combined metrics	61
Figure 4.2 Example of assigning the county’s weighted excess risk value to its sub-census tracts.....	62
Figure 4.3 Example of estimating the combined metrics for the census tracts in Braxton County	63
Figure 5.1 Distribution of traffic crashes by severity.....	80
Figure 5.2 Temporal effects on the exponential scale by crash severity under Model 9.....	88
Figure 5.3 Spatial distribution of the weighted excess risk for counties in 2010–2015.	91
Figure 5.4 Exponential of the temporal effects derived from the Bayesian logit model (Model 4).....	101
Figure 5.5 Spatial distribution of the weighted excess risk of fatality for the census tracts in 2010-2015.....	103
Figure 5.6 Conversion of county and census tract shapefiles to raster.....	105
Figure 5.7 Estimating the combined metrics through the raster calculator tool.....	106
Figure 5.8 Final map from the combination of macro- and micro-level results	106
Figure 1 MCMC output for parameters under Model 9.....	130
Figure 2 Posterior density curves of parameters under Model 9	135
Figure 3 MCMC output for parameters in the full ordered logit model.....	140
Figure 4 Posterior density curves of parameters under Model 4, micro-level analysis	141

LIST OF TABLES

TABLE 3.1 Descriptive statistics of the variables used in the macroscopic data analysis (330 observations).....	35
TABLE 3.2 Descriptive statistics of the variables used in the microscopic data analysis (3,724 observations).....	38
TABLE 3.3 Global Moran’s I statistics.....	42
TABLE 3.4 Pearson correlation among crash severities	42
TABLE 4.1 Summary of the frequency models proposed in this study.....	52
TABLE 5.1 DIC values associated with the models.....	67
TABLE 5.2 Parameter estimates for Model 1	69
TABLE 5.3 Parameter estimates for Model 2.....	70
TABLE 5.4 Parameter estimates for Model 3.....	71
TABLE 5.5 Parameter estimates for Model 4.....	72
TABLE 5.6 Parameter estimates for Model 5.....	73
TABLE 5.7 Parameter estimates for Model 6.....	74
TABLE 5.8 Parameter estimates for Model 7.....	75
TABLE 5.9 Parameter estimates for Model 8.....	76
TABLE 5.10 Parameter estimates for Model 9.....	77
TABLE 5.11 Negative binomial models’ parameter estimates	81
TABLE 5.12 Comparison of Bayesian and NB models	82
TABLE 5.13 Marginal effects of significant covariates in Model 9	86
TABLE 5.14 Estimated variance-covariance matrices for Model 9	89
TABLE 5.15 DIC values for the nominal response models	93
TABLE 5.16 Parameter estimates for the basic nominal response model (Model 1).....	93
TABLE 5.17 Parameter estimates for the full nominal response model (Model 2)	94
TABLE 5.18 DIC values for the ordered response models	95
TABLE 5.19 Parameter estimates for the basic ordered logit model (Model 3)	95
TABLE 5.20 Parameter estimates for the full ordered logit model (Model 4)	96
TABLE 5.21 Classical ordered logit model parameter estimates.....	97
TABLE 5.22 Comparison of Bayesian and classical ordered logit models	97
TABLE 5.23 Marginal effects of significant variables in the full ordered logit model.....	99

LIST OF SYMBOLS/NOMENCLATURE

I	Moran's index
N	Number of spatial units
w	Element of the spatial weight matrix
y	Crash rate
Obs	Observed number of crashes
C	Expected number of crashes
λ	Poisson parameter
E	Exposure variable (population)
β	Vector of coefficients for the model of crash severity s
X	Vector of covariates
θ	Unstructured spatial random effect term
φ	Structured spatial random effect term
ν	Unstructured temporal random effect term
δ	Structured temporal random effect term
γ	Spatio-temporal interaction
ρ	Crash-level random effect
r	Within-area correlation
σ	Variance
Y	Crash severity indicator
P	Fatality risk
τ	Precision parameter
c	Threshold (cutoff value)
ε	Weighted excess risk
DIC	Deviance information criterion

CHAPTER 1: INTRODUCTION

1.1. Research Motivation and Problem Statement

Over the past few decades, the use of transport systems has significantly grown because of the development of the highway system and the availability of inexpensive private vehicles to mid-income social classes. Although such developments have contributed to economic growth and urban mobility, they have imposed direct and indirect costs on our communities, such as noise and air pollution, traffic congestion and delays, and road traffic crashes.

Road traffic crashes are one of the major causes of death and serious injury in the US, causing economic losses and human suffering (CDC 2016). According to the National Safety Council, around six million traffic crashes occurred in the US in 2016, of which 34,439 were fatal and around two million were injury crashes that led to physical, social, and psychological consequences. The estimated cost of motor vehicle deaths, injuries, and property damage in 2016 was US\$432.5 billion (National Safety Council, 2016.); this figure includes medical and work loss costs. Reducing crashes and their related consequences is therefore a critical issue.

In recent years, many statistical models have been established to analyze traffic crashes at the macroscopic and microscopic levels. Macro-level crash studies have primarily focused on the quantification of the impacts of demographic, socio-economic, and area-level transport-related factors on aggregated traffic crashes (i.e., number of crashes per county or another geographical unit) by using count models. On the other hand, micro-level safety analyses have been used to analyze traffic crashes in particular road sections (e.g., segments, intersections), either focusing on crash count or on crash severity models; that is, count models aim to identify the relationship between crash frequency and road geometric features (e.g., lane width, presence of median barriers), whereas crash severity models are used to determine the effects of individual-level factors, such as age,

gender, and road condition, on severity outcomes. As macro- and micro-level analyses focus on two different dimensions, combining these two analyses could overcome some of the shortcomings of frameworks that use the two methods separately.

Despite the substantial research efforts in the traffic safety field, major limitations have not been addressed properly in the literature. From a methodological standpoint, for example, conventional crash count models, such as the Poisson and negative binomial (NB) regression models, have been widely used in previous studies to establish the relationship between crash frequency and other potential factors. Compared with Poisson models, NB models are more common in traffic safety studies, as they account for the overdispersion in crash data. Nevertheless, NB and other classical count models ignore correlated unmeasured risk factors across space and time. Recent developments in statistical computations have allowed researchers to model spatial correlation using the conditional autoregressive model (CAR) in the Bayesian framework. In addition to CAR models, other spatial models, such as spatial autoregressive and spatial error models, account for spatial correlations; however, they are less common in traffic safety research, as introducing random effect terms is difficult. Additionally, classical models assume that crash occurrences are independent across different severity levels, which may result in biased parameter estimates. By contrast, the univariate CAR model can be extended to the multivariate conditional autoregressive (MCAR) model, which accounts for the correlation across different severity outcomes.

Although crash data are of a hierarchical nature, most methods proposed by the literature do not account for this fact. A hierarchical structure here refers to either intrinsic spatial and temporal patterns and/or the fact that multiple individuals in multiple vehicles are typically involved in traffic crashes. Compared with conventional methods, the hierarchical Bayesian framework is more flexible and can be easily extended to include random effect terms that can act as proxies for unobserved or

missing factors; these factors can affect crashes and might have a spatial or temporal structure. Therefore, the hierarchical Bayesian framework appears to be a promising methodology for dealing with crash data.

Additionally, another limitation of recent studies that utilize a full Bayesian framework is the omission of key factors that might affect the number and/or severity of crashes, such as urbanization levels and hospital accessibility.

In light of the above, this dissertation proposes an integrated zonal screening approach that aims to identify sites with promise incorporating macro- and micro-level traffic safety analyses. The proposed framework will allow agencies to more accurately analyze the overall traffic safety of an area. Additionally, this dissertation seeks to investigate the relationship between traffic crashes and various macro- and micro-level factors by utilizing a hierarchical Bayesian approach. The case study of West Virginia (WV) is used to illustrate the framework and methods proposed in this dissertation.

1.2. Research Goal and Framework

The overarching goal of this dissertation is to present a screening approach by integrating both macro- and micro-level data in order to provide a comprehensive and effective method for use in transport safety planning. A comprehensive framework, shown in Figure 1.1, was developed to address this goal.

The following are the five main tasks involved in the framework:

1. Develop various Bayesian multivariate models (Poisson lognormal models) to relate various macro-level variables, such as socio-economic and transport-related factors, with crash occurrence/frequency while accounting for possible temporal effects and spatial autocorrelations

2. On the basis of the results of task 1, identify the hotspots in terms of number of crashes (i.e., counties that experienced a significant amount of crashes) and prioritize the high-risk counties based on the excess risk estimated from the best model in task 1
3. Establish different hierarchical Bayesian logit models from micro-level data and identify the individual-level and road-specific factors contributing to fatal injuries in each census tract
4. Identify high-risk census tracts in terms of risk of fatality of drivers involved in a fatal crash on the basis of the results from the best model in task 3
5. Develop an integrated and comprehensive screening method by combining the results from tasks 2 and 4 using a GIS-based approach

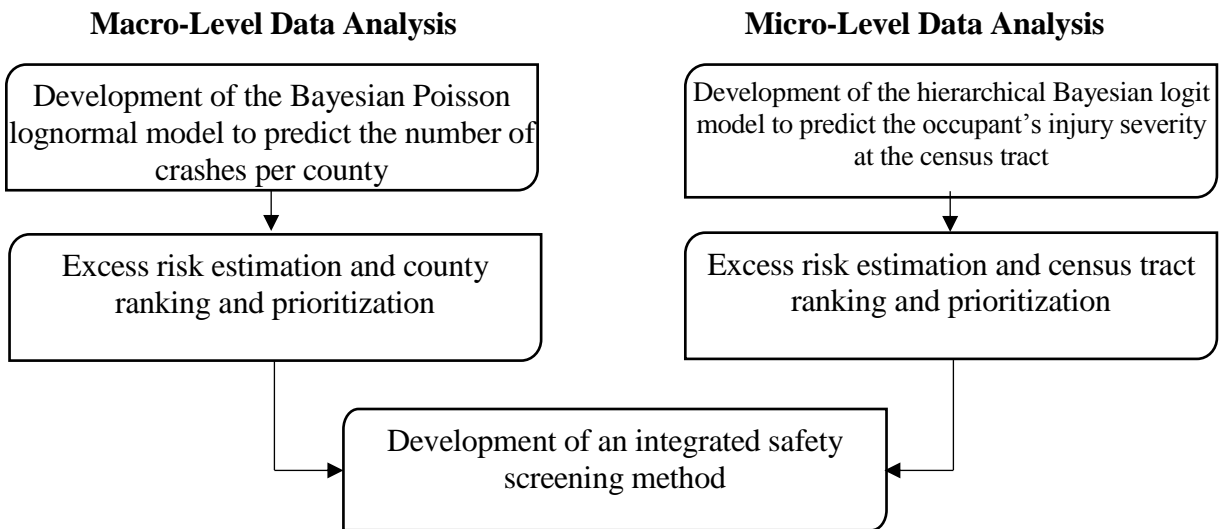


Figure 1.1 Research framework for the development of an integrated screening method

1.3. Research Contributions

The original contributions of this dissertation to the state of the art are as follows: (1) the development of comprehensive spatio-temporal models that account for the unique nature of crash data, which can provide a more accurate crash estimation, (2) the integration of macro- and micro-

level analyses to identify high-risk areas, and (3) the quantification of the impacts of various factors contributing to road traffic crashes. In a bid to avoid biased and inconsistent parameter estimates and to draw valid inferences about safety phenomena, the proposed models will account for unobserved heterogeneity. This approach can help state agencies to effectively handle their budget for improving road safety in high-risk areas and to inform crash prevention and control programs.

1.4. Dissertation Organization

The dissertation is divided into six chapters that follow this introductory chapter, namely, the Literature Review, Data Collection and Exploratory Data Analysis, Methodology, Research Results, and Conclusions. Chapter 2 (Literature Review) synthesizes the specifications and techniques used in crash prediction, as well as highlights high-risk locations and the factors contributing to traffic crashes. Chapter 3 (Data Collection and Exploratory Data Analysis) describes the data processing techniques used to prepare the data collected for analysis and provides a description of the final data sets used for this dissertation. Chapter 4 (Methodology) presents the proposed macro- and micro-level data analysis techniques used. It also describes the proposed integrated screening approach and explains the planned methodology to combine the results of the two analyses. Chapter 5 (Research Results) analyzes the estimation output from the proposed models and reports the high-risk locations for the case study of WV. Chapter 6 (Conclusions) provides the conclusions of this dissertation and discusses its contributions from both theoretical and empirical perspectives. It describes the planning and policy implications for traffic safety improvement. Finally, the chapter discusses the limitations of this dissertation and provides recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

The literature review presented in this chapter consists of three main sections. First, applied statistical methods in traffic safety are discussed. Second, past research efforts related to traffic safety at the macroscopic level are presented. Finally, previous traffic safety studies at the microscopic level are summarized.

2.1. Statistical Methods Overview

To date, many statistical methods have been developed to perform macroscopic and microscopic traffic safety analyses. Generally, such methods can be classified into classical methods (e.g., nominal safety measures, count, linear regression, and logit models), empirical Bayesian (EB) methods, spatial autoregressive models, and full Bayesian models.

Classical methods, such as nominal safety measures (crash rate/frequency), have been primarily used by transportation agencies to detect and rank hazardous road segments and intersections. However, these methods ignore locations with zero crashes in the time analyzed. Other classical methods, such as Poisson, NB, linear regression, and logit models, have mostly been used to identify and establish the relationship between potential factors (either at the macro or micro level) and traffic crashes/severities, but not to rank them. Generally, these models ignore the hierarchical nature of crash data and are designed in a univariate framework, that is, the occurrence of traffic crashes of different severity levels is considered independent; in other words, the methods do not account for the correlation between different crash severities, thus violating traditional Gauss-Markov assumptions (Darwiche, 2009). Nevertheless, the literature (see for example, Ma and Kockelman, 2006; Tunaru, 2002) has demonstrated the strong correlation among different types of crashes because of shared site-specific unobserved factors. Although significant studies examine

this topic in micro-level crash analysis, limited research has addressed these potential correlations between each type of crash in macroscopic safety analysis.

The **EB method** has been widely used in micro-level safety studies; it has been proposed as an alternative to address many of the limitations of classical methods because it estimates the expected crash frequency in road segments or intersections by combining observed and predicted crash frequencies from a safety performance function (Hauer, 1986).

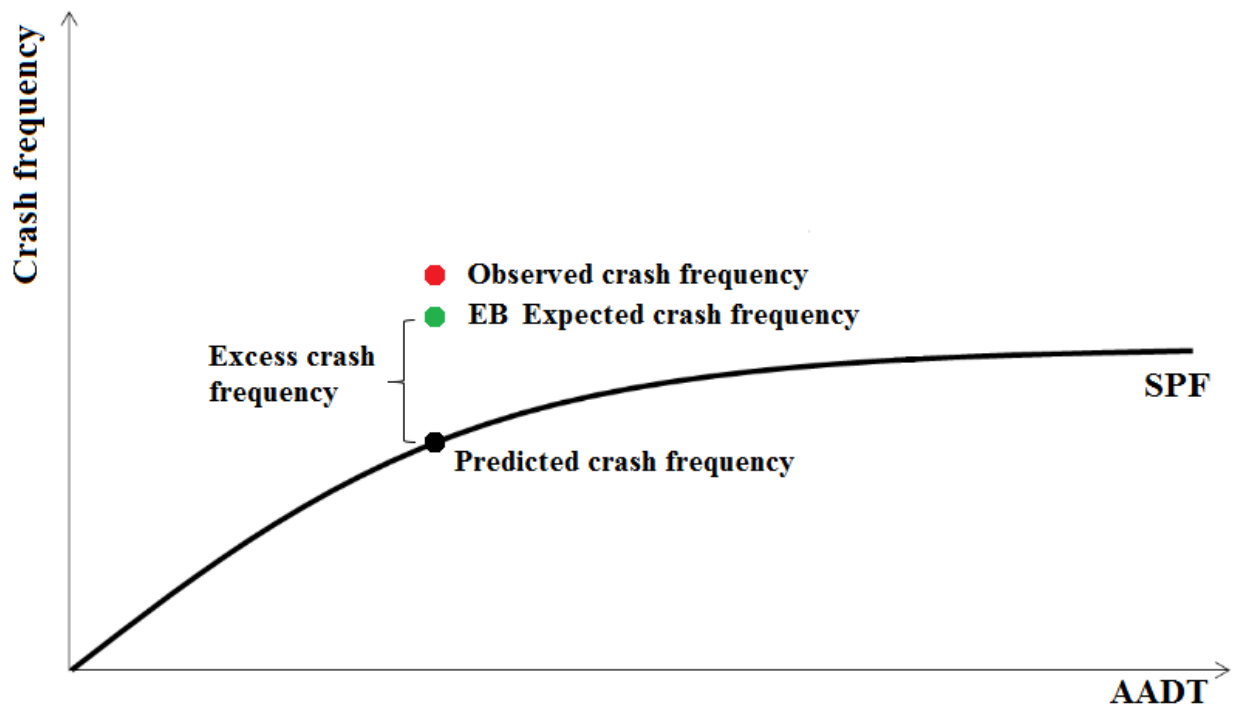


Figure 2.1 Illustration of regression to the mean and the empirical Bayes estimate (revised from Gross et al. (2010))

Figure 2.1 shows that the difference between expected and predicted values refers to excess crash or the potential for safety improvement. In the EB method, the safety performance function (SPF) is usually developed using NB, which estimates the predicted crash frequency as a function

of road characteristics, such as the annual average daily traffic (AADT) and segment length. The main advantage of the EB method is that it accounts for the regression-to-the-mean bias. For example, the method recognizes that a road segment or an intersection with zero crash during a given time period is expected to experience some crashes, close to the average number of crashes in similar sites, in the long term. However, as the Institute of Transportation Engineers explains, the EB method is not a fully Bayesian model, and it is “complex and not ready for widespread implementation” (Pawlovich, 2003). Moreover, the development of an SPF requires a large database storing the characteristics of reference sites (e.g., AADT per road segment, segment length, lane width, shoulder width, road grade, presence of a horizontal curve).

The **spatial autoregressive model** is a method to analyze datasets that contain observations on geographical units. The spatial weight matrix is its core component; this component allows researchers to explore how an event at one location is directly and indirectly affected by its neighbors. This model can be written in two main forms. The first form is the spatial lag model in which the spatial weight matrix is applied to the dependent variable along with a spatial autoregressive parameter, which can be estimated from the data. This approach can be written when an observation at a given location is affected by the observations measured at nearby locations. The second approach is used when only the error terms across different spatial units are spatially correlated. Spatial autoregressive models are well suited to the maximum likelihood method and the generalized method of moments estimation, which make introducing random effect terms difficult (Anselin, 2013; Banerjee, 2009). Note that in the conditional autoregressive framework in a full Bayesian approach, the introduction of random effect terms is easy because of the hierarchical conditional representation.

The **full Bayesian method** estimates the posterior value of parameters by using sampling techniques, such as Markov chain Monte Carlo (MCMC) methods. Among others, the method can provide credible intervals for posterior values and account for the regression-to-the-mean bias. It can also address unobserved heterogeneity over space, time, and crash severities (Lord and Mannering, 2010; Lord et al., 2005; Mannering and Bhat, 2014; Mannering et al., 2016). Ignoring this unobserved heterogeneity can decrease the efficiency of the estimator (Ma and Kockelman, 2006; Mannering et al., 2016).

From a spatial viewpoint, unobserved spatial heterogeneity might be caused by unmeasured variation or differences across zones; for example, some factors that may affect traffic crashes but have not been included in the model (because of data availability), such as the percentage of roadways with sidewalk or the percentage of roadways that have adequate lighting, can systematically vary across zones. On the other hand, spatial correlation might be caused by correlated unobserved factors in neighboring zones; for example, traffic crashes in one zone might be induced by unobserved factors, such as the lack of attractiveness in neighboring zones, which prompt their residents to travel in that zone. Many authors demonstrated the presence of spatial autocorrelation and heterogeneity in crash data across different spatial units (see for example, Agüero-Valverde and Jovanis, 2010, 2006a; Wang and Abdel-Aty, 2010; Liu and Sharma, 2018; Wang and Abdel-Aty, 2006). Moreover, temporal correlation may exist in crash data because of the temporal nature of traffic laws, the economy, the weather, and travel demand and behavior et al. 2016; Liu and Sharma, 2018).

Crash occurrences across severity outcomes are not independent, so the use of a multivariate model when different crash severities need to be analyzed in the model is more suitable (Mannering and Bhat, 2014; Mannering et al., 2016). Many authors have utilized and compared Bayesian models

in univariate and multivariate settings, proving that multivariate models are superior to univariate models because they capture the within-area correlation among different crash severities (see for example, Agüero-Valverde, 2013; Barua et al., 2014; Boulieri et al., 2016; Wu and Donnell, 2016; Liu and Sharma, 2018; Y. Wang and Kockelman, 2013).

Bayesian models can be specified through multiple/hierarchical levels. The basic idea of hierarchical Bayes (known else as multilevel modeling) is to think of the lowest-level units as organized into a hierarchy of successively higher-level units. For example, in terms of a crash, the hierarchy can be described as follows: individuals are involved in crashes, crashes occur in road sites, and road sites are located in greater regions (e.g., census tracts, counties). In view of this, the outcomes of an individual can be described as the sum of the effects for the individual, for the crash, for the road site, and for the region. Because of their capability to account for this hierarchy, hierarchical Bayes models have the potential to provide accurate crash estimates and rank locations with promise.

2.2. Crash Factors

Numerous factors can contribute to road traffic crashes. The Government Accountability Office (US Government Accountability Office, 2003) classifies these factors as human, roadway environment, and vehicle factors. Human factors pertain to the behavior of road users (e.g., speeding, alcohol involvement), as well as driver limitations (e.g., decision errors, poor vision). Roadway environment factors involve design and road features, such as lane width, sight distance on vertical or horizontal curves, the presence of raised median barriers, types of intersections, and pavement conditions (e.g., rain, ice, snow, fog). Vehicle factors stand for vehicle characteristics (e.g., model type, model year) and any vehicle-related failures (e.g., undeployed airbag).

Many important factors affecting crashes, such as accessibility to hospitals, demographic characteristics, and highway infrastructure functional class, might systematically vary at the spatial scale and should be considered in crash modeling (Aguero-Valverde and Jovanis, 2006b). These potential factors might also be related with different hierarchical levels. (Huang and Abdel-Aty, 2010) reported that crash data can be structured in multiple levels, and they proposed a five-level hierarchy, as shown in Figure 2.2. The proposed hierarchy implies that the factors affecting traffic crashes can be explored at both the macroscopic and microscopic levels. Figure 2.2 shows that in this study, the macroscopic level refers to geographic/spatial units (e.g., countries, regions), whereas the microscopic level concerns the traffic crash, vehicle, and individual levels. Hence, the use of a hierarchical technique would be critical to account for the multilevel effects of traffic crashes.

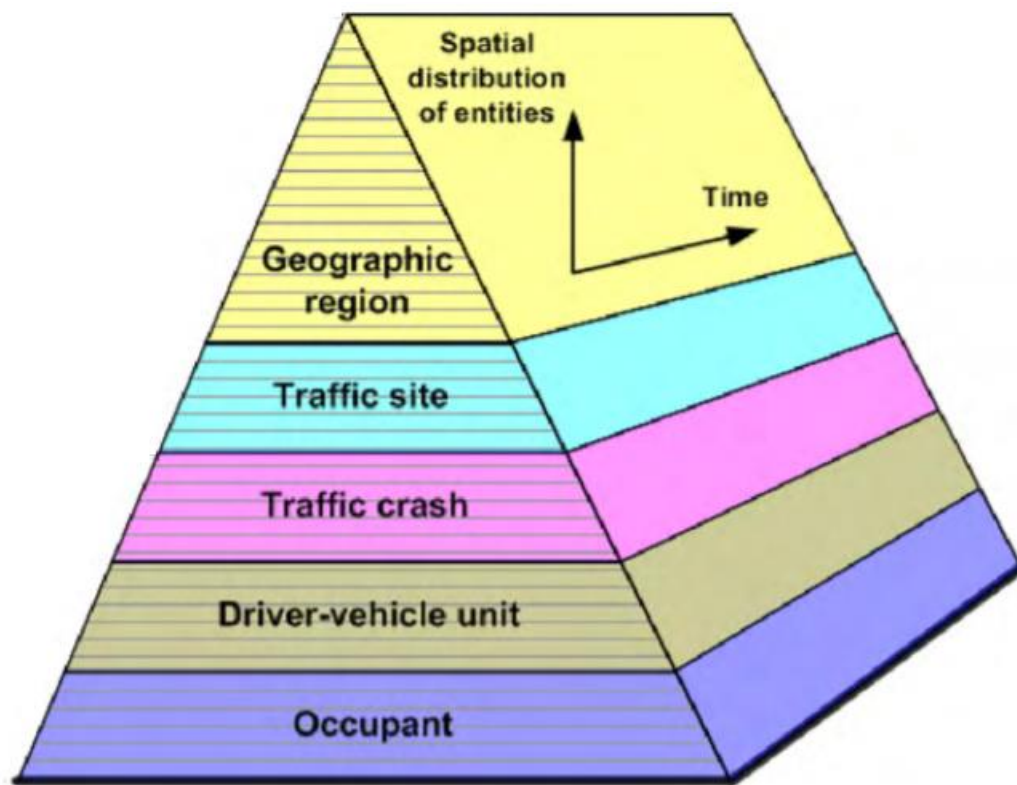


Figure 2.2 Multilevel structure of crash data (Adapted from (H. Huang and Abdel-Aty, 2010))

2.3. Macroscopic Traffic Safety Research

Both macroscopic and spatial data analysis techniques have been widely used in environmental sciences, enabling researchers to investigate important issues related to risk estimation, unmeasured confounding variables, and spatial dependence (Richardson, 1992). Later, such techniques were considered by traffic engineers to establish the relationship of traffic-related factors and socio-economic and demographic factors with crash frequencies and crash rates at various spatial units. In macro-level traffic safety studies, traffic crashes are generally grouped into spatial units, which range from counties to zip codes (Darwiche, 2009). As a pioneer work on traffic safety, (Smeed, 1949) developed an equation to predict the number of traffic fatalities (D) across different countries as a function of population (P) and the number of licensed motor vehicles (N). He assumed that the equation has a functional form of $D=AN^{\alpha}.P^{\beta}$, where A, α , and β were estimated through least square methods. The resulting equation based on the data from 20 countries in 1938 was as follows: $D= 0.0003(N.P^2)^{1/3}$.

From a methodological standpoint, a wide range of macro-level methods has been utilized in the literature. These methods can be classified into (a) classical models, including NB models (Abdel-Aty and Radwan, 1998; Chin and Quddus, 2003; Hadayeghi et al., 2003; Noland and Oh, 2004; Noland and Quddus, 2004; Poch and Mannering, 1996), loglinear models (Abdel-Atyet al., 1998; Lee et al., 2006; Washington, 2006), (b) spatial autoregressive models (LaScala, et al., 2000; Levine et al., 1995; Quddus, 2008; Rhee et al., 2016), and (c) full Bayesian or hierarchical models (Aguero-Valverde, 2013; Aguero-Valverde and Jovanis, 2010, 2006a; Boulieri et al., 2016; Cheng et al., 2018; Huang et al., 2010; Lee et al. 2015; Liu and Sharma, 2018; Miaou, 1994; Siddiqui, et al., 2012; Truong et al., 2016). In the following sections, literature that has used each class of methods described above is summarized. This literature review is not intended to be comprehensive; rather, it aims to present key representative studies that influenced this dissertation.

2.3.1. Classical Methods

Abdel-Aty and Radwan (1998) conducted a county-level GIS-based crash analysis. The study found that counties with high population tend to have higher crash frequencies. The analysis concluded that rural counties tend to have more severe crashes than urban counties do.

Hadayeghi et al. (2003) developed different NB models for fatal and nonfatal crashes. They used demographic, traffic demand, and network data variables as explanatory variables in their models. The variables that had significant effects on crash occurrence are the number of households, road length, vehicle kilometers traveled, intersection density, speed limit, and volume–capacity ratio.

Noland and Quddus (2004) utilized NB models to analyze the association between demographic data and traffic fatalities. They reported that traffic analysis zones (TAZs) with higher employment density have more traffic crashes, whereas congested urbanized areas have fewer crashes. In a similar work, Noland and Oh (2004) used the NB model to predict county-level crashes based on demographic and infrastructure characteristics. The authors found that increases in the number of lanes appear to be associated with both increased traffic-related crashes and fatalities.

Kim et al. (2006) estimated an NB model to explore the relationship between land use, population, and economic development and crashes in a uniform grid structure. The results showed that areas with a high level of commercial activities have higher vehicle-to-vehicle crashes. In addition, areas around schools are associated with higher frequencies of overall crashes.

Loukaitou-Sideris et al. (2007) proposed an ordinary least squares (OLS) regression to model the frequency of pedestrian crashes and sociodemographic and land use characteristics at the census tract level. The study found that neighborhoods with high population and employment density, high traffic volumes, and a large concentration of commercial/retail and multifamily residential land uses have a higher probability for pedestrian crashes.

Kant (2008) analyzed the relationship between crash types and land use in Florida by using GIS. The study found that rear-end crashes and right turn crashes are more common on urban roads than on rural roads; however, run-off road crashes were more common on rural roads than on urban roads.

Khan et al. (2008) utilized county-level aggregate crash data for the state of Wisconsin to analyze the spatial clusters of crashes and the correlation between snow and clusters of weather-related crashes by using spatial statistics. The results identified the clustering of ice-related crashes around bridges in four counties with similar ice-related crash rates in southeast Wisconsin.

Wier et al. (2009) developed an aggregate-level crash prediction model by using an OLS model. They modeled the natural log of the number of vehicle–pedestrian injury collisions over a five-year period at 176 census tracts in San Francisco. The independent variables examined in their study included street, land use, and population characteristics, and their final model was able to explain approximately 72% of the systematic variation in vehicle–pedestrian injury collisions at the census tract level. The contributing factors were traffic volume, employee and resident populations, arterial streets without public transit, proportion of people living in poverty, and proportion of people aged 65 and over.

2.3.2. Spatial Autoregressive Models

In an early study, Levine et al. (1995) developed a spatial lag model to predict the crash counts in Honolulu TAZs based on socio-economic factors and roadway characteristics. The study found a positive correlation between road mileage and crashes. They also reported that the predictors of accident vary depending on trip activities and changes over time.

LaScala et al. (2000) used the spatial error model to explore the correlation between pedestrian crashes across census tracts in San Francisco. Their study suggested that pedestrian crash rates are associated with traffic flow, population density, and education.

Quddus (2008) developed three different models (the non-spatial NB regression model, the spatial error model, and the Bayesian model) to compare the relationship between dependent and independent variables. He found that the results from the NB and Bayesian models are similar. The results from the spatial error and lag models are generally consistent with those of the Bayesian models, except for the case of serious injuries. Finally, the authors reported that the results from the Bayesian hierarchical models are more consistent with the literature and more coherent in all cases.

Rhee et al. (2016) fitted various spatially autoregressive models to fatal and crash data. They concluded that the spatial error model outperforms the spatial lag model and an OLS baseline regression. The results also indicated that an increased length of roads with a speed limit below 30 km/h and a higher ratio of residents below the age of 15 are correlated with a lower traffic crash frequency, whereas a higher ratio of residents who moved to the TAZ, more vehicle kilometers traveled, and a greater number of access points with speed limit difference between side roads and the mainline above 30 km/h all increase the number of traffic crashes.

Soro et al. (2017) developed four models from panel data, which were the non-spatial model, the spatial autoregressive model, the spatial error model, and the fixed effect spatial autoregressive model with autoregressive disturbances (SARAR). Accident rate and injury rate were used as the dependent variables. Additionally, socio-economic factors, such as population percentage, unemployment rate, gross regional product, road length, and passenger and truck traffic, were considered as the regressors. A comparison of the models indicated that the SARAR model has better performance than the other models, as it generates a smaller akaike information criterion (AIC). The

gross regional product, population, unemployment rate, and passenger traffic were also found to be significant and had negative coefficients, whereas road length and truck traffic had positive coefficients.

2.3.3. Full Bayesian Models

Several different hierarchical Bayesian models can be used for the estimation of spatial risk patterns based on spatially aggregated count data. The Bayesian hierarchical model proposed by (Besag, 1974) has been widely used by various researchers in the case of areal count data (Aguero-Valverde and Jovanis, 2006b; Huang et al., 2002; Miaou, 1994).

Miaou et al. (2003) developed county-level spatial models of rural road crash frequency in Texas while controlling for roadway characteristics. Later, Aguero-Valverde and Jovanis (2006b) utilized full Bayes hierarchical models for county-level fatal crash counts in Pennsylvania. The authors explored the effects of county-level demographic (e.g., population, age, wealth) and weather condition variables (e.g., precipitation, total number of rainy days in a year). The conclusive results of the study indicate that population age cohort and road mileage have a positive impact on crash frequency, whereas vehicle miles traveled (VMT) decrease crash risk.

Song et al. (2006) explored several full Bayesian multivariate spatial models for estimating the crash rates from intersection, intersection-related, driveway-access, and non-intersection crashes in 254 counties in Texas. They only used three covariates in their models: a surrogate variable to represent weather variations, the presence of a curve as a surrogate variable to capture spatial variations in the percentage of sharp horizontal curves, and a surrogate variable indicating spatial variations in roadside conditions. The study found that these variables have significant impacts on intersection and intersection-related crashes.

Huang et al. (2010) developed various Bayesian models and compared their performance by utilizing different exposure variables (population versus daily vehicle miles traveled). The results showed that no significant differences exist between the parameter estimates in those models with different exposure variables. However, the findings suggested that utilizing the population as exposure can improve the goodness of fit of the model. One limitation of this study is that the effect of ADT or a combination of population and VMT as the exposure variable was not investigated.

Recently, various researchers have focused on the spatial analysis of non-motorist/pedestrian crashes by using Bayesian models (see for example Siddiqui et al., 2012 ; Wang and Kockelman, 2013). Siddiqui et al. (2012) investigated the effect of spatial correlation by using a Bayesian spatial framework to model pedestrian and bicycle crashes in TAZs. Poisson lognormal models were used to relate pedestrian and bicycle crashes to roadway characteristics and to various demographic and socio-economic factors. Wang and Kockelman (2013) examined the relationship between pedestrian crash counts across census tracts in Austin, Texas and various land use, network, and demographic attributes by using the Poisson lognormal conditional auto regressive model. The authors found that a more diverse area that includes residential and commercial land uses is associated with a higher pedestrian crash risk across different severity levels. The main limitation of these two studies is that they ignored the spatial and temporal correlation between observations. In addition, the correlations among the crash severities were not investigated.

Boulieri et al. (2016) extended a multivariate and univariate Bayesian framework to include temporal effects, and they compared different space-time models. The study concluded that the models with multivariate priors (i.e., correlated space effect models) outperform the univariate models. In addition, the findings of the study suggested that crash severities are spatially correlated. The major shortcomings of this study were as follows. First, intercept models were developed

without accounting for any potential risk factors; hence, this approach might not be sufficient to identify hotspot locations. Second, the temporal effects used in the proposed models do not allow the flexibility to fit the nonlinear effects of time. Third, similar to previous research, individual-level data were not considered.

Bhat et al. (2017) established a multivariate model to jointly analyze pedestrian crashes at four different severity levels in census tracts in Manhattan, New York City. They proposed spatial random coefficients in the model but did not account for temporal effects. Population cohort, ethnicity, educational attainment, land use, activity intensities, and commute mode shares were used as independent variables.

Finally, in a very recent study, Liu and Sharma (2018) developed a space-time model to analyze the annual county-level crash counts in Iowa from 2006 to 2015. They proposed spatial and temporal effects simultaneously. The results of the study showed that both spatial and temporal components contribute to the model's overall goodness of fit. The study explored the impact of a few variables, such as income, unemployment rate, rainfall, snowfall, temperature, and VMT, on fatal, major injury, and minor injury crash counts. The findings suggested that VMT has a significant and positive association with all crash severities, whereas the unemployment rate is significant and negative in major and minor injury models. The remaining variable is found to be insignificant.

2.4. Microscopic Traffic Safety Research

Microscopic analysis is one of the most widely used approaches to investigate the effects of drivers' and occupants' characteristics (e.g., age, gender) and road factors (e.g., geometric features, surface condition, lighting) on crash frequency/severity. Another important application of micro-level data analysis is in the selection of candidate road segments and intersections for improvement. The Highway Safety Manual (HSM) is one of the popular sources providing a

comprehensive framework for network screening and the identification of hazardous locations. HSM presents various methods and performance measures, such as nominal safety measures (crash rate and crash frequencies), and the EB method for measuring and evaluating roadways and intersections in terms of crash frequency. This section presents various methods discussed in HSM, as well as in other research papers. Again, the literature review is not intended to be comprehensive; rather, it aims to present key representative studies that influenced this dissertation.

2.4.1. Classical Methods

2.4.1.1. Nominal Safety Measures

A review of the literature and safety manuals developed by transportation agencies shows that safety practitioners have extensively used nominal safety measures, such as crash frequency and crash rates, to identify hazardous locations. Taylor and Thompson (1977) developed a hazardous rating formula to detect hazardous road segments by combining accident-related criteria (i.e., number of accidents per year, accident rate, accident severities) and non-accident criteria (i.e., number of complaints from citizens, traffic conflicts, erratic maneuvers).

Labadie and Barbaresso (1982) established a computerized priority program to rank roadside hazards. They considered different criteria (whether the roadway is curbed, the presence of horizontal curves, the presence of vertical curves, speed limit, distance from the pavement edge, roadway type) and related weight factors. They obtained the relative hazard rating by summing up the weighting factors associated with the criteria.

Spring and Hummer (1995) used a basic GIS approach to determine accident frequency on bridges, curves, and intersections. They defined a threshold for crash frequency to determine hazardous road elements by interviewing traffic safety experts.

2.4.1.2. Count Models

(Miaou et al. (1992) developed four Poisson regression models to predict truck-involved crashes on rural road segments based on AADT, shoulder width, surrogate measures for horizontal curvature (i.e., horizontal curvature change rate, mean absolute horizontal curvature, maximum absolute horizontal curvature), and surrogate measures for vertical grade (i.e., vertical grade change rate, mean absolute vertical grade, maximum absolute vertical grade). A comparison of their models suggested that the mean absolute curvature and the vertical grade perform better than the other surrogate measures and are positively correlated with the number of truck crashes. Additionally, increases in AADT and shoulder width for a given road segment, with other variables held constant, increase truck crashes for that road segment.

It should be noted that crash data tend to be overdispersed, that is, the variance exceeds the mean, whereas the Poisson model assumes equality of the mean and variance. Poisson models do not account for this and can underestimate the variance of parameter estimates. Hence, many safety practitioners use the NB model in lieu of Poisson models to account for overdispersion.

Hadi et al. (1995) investigated the impacts of roadway features, including lane width, shoulder width, median width, and median type on fatal and injury crashes. Depending on highway type, increasing the length of road elements can reduce traffic crashes. Moreover, a raised median barrier is more effective in reducing traffic crashes in four-lane highways than in two-lane highways.

Poch and Mannering (1996) used the NB model to explore the impacts of intersection characteristics on annual angle, rear-end, and approach-turn crashes in intersections. Traffic volume was found to increase all kinds of traffic crashes, whereas the presence of signal control, a permissive left turn, and stop control reduces the number of angle, rear-end, and approach turn crashes.

One of the major disadvantages of the NB model is that it is unable to distinguish between safe road segments that experienced zero crashes over time and unsafe road segments that had zero crashes at a given time, which can result in biased parameter estimates. To overcome this issue, many researchers utilized zero-inflated models, including the zero-inflated Poisson (ZIP) and zero-inflated negative binomial regression (ZINB) models. The logit component of the zero-inflated model separates safe and unsafe road segments (roads segments with zero and non-zero crash counts), whereas the Poisson or NB components predict the number of crashes. Shankar et al. (1997) investigated the applicability of ZIP and ZINB models to road segment crash frequencies. They found that the NB, ZINB, and ZIP models are a better fit to principal arterial, minor arterial, and collector arterial data.

2.4.1.3. Discrete Choice Models

Discrete choice models or logit models have been widely used by safety practitioners to understand the individual-level factors (e.g., occupants' characteristics and road features at crash locations) contributing to traffic severities, as well as to estimate the crash risk is discrete choice/logit model. For example, O'donnell and Connor (1996) used ordered logit models to estimate the linkages between road users' characteristics and traffic crashes at four different severity levels. Their results revealed that a left-rear seating position, being female, a blood alcohol content of more than 0.08, failure to use a seatbelt, a light truck, and head-on collisions are likely to increase the risk for severe crashes.

Bedard et al. (2002) used the multinomial logit model to identify the independent factors affecting drivers' fatality risk in single-vehicle crashes with fixed objects. The results showed that seatbelt use and vehicle speed are negatively and positively associated with the risk of fatal crashes, respectively.

Kockelman and Kweon (2002) applied ordered probit models to explore the risk of different injury levels in single- and two-vehicle crashes. The results suggested that pickups and sport utility vehicles are more likely to increase the risk of severe crashes under single-vehicle crash conditions. However, in two-vehicle crashes, these vehicle types are associated with less severe injuries for their drivers and more severe injuries for the occupants.

Srinivasan (2002) proposed the use of the ordered mixed logit model to model injury severity in a crash. The primary purpose of such a model is to accommodate random and correlated injury severity thresholds associated with various severity levels. The model results indicated that the injury severity thresholds for an individual are dependent on traffic, crash-related, and vehicle characteristics. Additionally, a significant unobserved variation in thresholds was observed.

Krull et al. (2000) examined the risk of a driver's injury in single-vehicle crashes in Michigan. They concluded that rollover involvement, passenger cars, failure to use a seatbelt, alcohol involvement, lighting, rural roads, higher speed limits, and dry pavements increase the risk of a driver's injury severity.

Khattak et al. (2003) used ordered probit models to investigate the risk factors that affect crash severities (from fatal, severe, moderate, and minor to no injury) in large truck rollovers in single-vehicle crashes. Traffic control violation, vehicle stopping, post-crash fire, driver drowsiness, and alcohol involvement were found to be associated with crash severity.

Milton et al. (2008) developed a modeling approach to better analyze the injury severity distributions of accidents on highway segments. They used the mixed logit model to determine the proportion of severity of each crash on a roadway. Variables, such as average daily traffic per lane, average daily truck traffic, truck percentage, interchanges per mile, and weather condition, were modeled as random parameters, whereas roadway characteristics, such as the number of horizontal

curves, number of grade breaks per mile, and pavement friction, were modeled as fixed parameters. The results showed that the effects of average daily traffic per lane, average annual snow, percentage of trucks, average daily truck traffic, and number of interchange per mile on property damage only (PDO) crashes, possible injury crashes, and injury crashes vary across road segments. However, pavement friction is likely to reduce the risk of injury, whereas the number of horizontal curves and grade friction are likely to reduce possible injury crashes.

Tay et al. (2011) proposed the use of the ordered multinomial logit model to examine the factors affecting the severity of pedestrian vehicle crashes in South Korea. Their results indicated that the factors increasing the probability of fatal injury to pedestrians include drivers' sex (reference: female), drivers' age and alcohol involvement, pedestrians' age and sex, pedestrians' location on the crosswalk, intersections, shoulder, outer (faster) lanes, freeways, provincial and national highways that are wider than 9 m, vehicle type and size, adverse weather conditions, dark hours, peak hours, and relatively rural areas.

2.4.2. Empirical Bayes

The EB method is based on Bayes' theorem, and "it was developed based on the recognition that accident counts are not the only clue to the safety of an entity" Hauer et al. (2002). The EB method aims to increase the precision of crash estimation and to account for the regression-to-the-mean bias. Hence, in the EB method, the expected crash counts on a road segment would be the weighted average of the observed crashes on a proposed road segment and the expected number of crashes in similar road segments, where the expected number of crashes for similar road segments can be obtained from classical regression models (e.g., NB regression), with traffic and geometric factors as the independent variables.

As an early work, Hauer and Persaud (1987) proposed an EB method to estimate the proportion of correctly and falsely identified deviant road segments. Later, Hagle et al. (1988) focused on the identification of hazardous road segments by using EB estimates of accident rates. They assumed that the crash rates in road segments follow Gamma distribution, which was the basis for developing the probability distribution function for the corresponding road segment. Hence, a road segment is classified as a hazardous location if the probability that its accident rate for a road segment exceeds a certain value is relatively large.

Hauer (1996) applied the EB method for the identification of sites with promise (SWiPs), which refer to hazardous locations. In the study, the author demonstrated that SWiPs are sites/locations whose number of crashes are higher than that of sites/location with similar characteristics.

Persaud et al. (1999) used the EB estimates of the expected number of crashes in signalized intersections and highway segments. They specified the NB method to model crash frequency based on AADT, segment length, and lane width, and the difference between the expected and observed number of crashes (which refers to excess crash frequency) was the basis to rank the sites.

2.4.3. Full Bayesian Models

Although the EB method has improved the process of crash estimation, it is not purely Bayesian (or it is not fully Bayesian), as it considers data for defining the prior distribution, and it does not indicate how the hyperparameter estimation error can be incorporated in the analysis. On the other hand, the hierarchical Bayes method incorporates such errors or randomness at different levels.

Davis (2000) used hierarchical Bayes and Gibbs sampling to estimate the accident rate by incorporating traffic volume errors. This model has improved the accuracy of accident rate estimation, and it is more reliable than classical models in identifying potential hazardous locations. Later, Davis and Yang (2001) combined the full Bayesian approach and induced exposure models to identify the dangerous intersections for older drivers.

Tunaru (2002) proposed a full Bayesian multivariate Poisson lognormal model to model the crash severities for road segments at two crash severity levels: fatality and minor injuries from single and multi-vehicle crashes. The author ranked road segments by using the expected crash frequency for each severity level. A null model was proposed, as no covariates were included in the analysis. The entries of the covariance matrix were found to be significant, suggesting that a significant correlation exists between the severity levels.

Kockelman and Damien (2008) developed a multivariate Poisson regression within a full Bayesian framework to estimate the number of victims per road segment in five different severity classes: fatal, disabling injury, non-disabling injury, possible injury, and non-injury. However, the authors assumed that the crash counts are not spatially correlated.

Aguero-Valverde and Jovanis (2009) used full Bayes multivariate Poisson lognormal models to estimate the expected crash frequency as a function of road characteristics, such as lane and shoulder width, for different crash severity levels in both univariate and multivariate settings. The findings suggested that the multivariate Poisson lognormal model outperforms the univariate model and improves the crash frequency estimates. The covariances and correlations between contiguous crash severity levels are also found to be high.

Jang et al. (2010) established four different models (Poisson, NB, ZIP, and ZINB) using a full Bayesian approach by including factors, such as degree of curvature and presence of a guiderail

and a median barrier, in all models. They compared model performance, and the results indicated that the ZIP model outperforms other models, as it has the highest posterior probability. The study also found that the Bayesian models utilizing power prior perform slightly better than the frequentist models do.

Ahmed et al. (2011) used Bayesian hierarchical models with spatial and random effects to predict the crash frequencies on mountainous road segments based on driver behavior, traffic and geometric characteristics, and adverse weather conditions. The study found that roadway geometry, segment downgrades, and the snow season are positively associated with crash risk.

Barua et al. (2014) investigated the effect of spatial and heterogeneous effects by using multivariate models of crash severities for two severity levels for 72 urban road segments. The study found that the model with both structured and unstructured spatial random effects has a better fit, suggesting that overdispersion and spatial correlation are present in the crash datasets. Similar to many other researchers, the authors did not investigate temporal effects in the crash dataset.

Usman et al. (2016) developed three hierarchical logit models, namely, a multinomial logit model, a sequential binary logit model, and an ordered logit model. These models were estimated for collision data at three levels: the occupant level, the vehicle level, and the collision level. A comparison of these models indicated that the multinomial logit model provides a better fit than the other models do, and the occupant-based data results are more reliable than the vehicle- and collision-based data results. The models did not explore spatial and temporal effects.

Wang and Huang (2016) used a univariate hierarchical model to estimate the crash frequency on intersection and road segments. Their model accounts for the spatial correlations between road entities, as well as the shared unobserved effects for the intersection and road segments in the same

TAZ. The study results showed that crashes are positively associated with AADT and with more lanes and more access points.

2.5. Summary

Various research efforts have been made to analyze traffic crashes at the macro and micro levels. The implemented methods can be classified into classical methods and more advanced methods, such as the EB, spatial autoregressive, and Bayesian approaches. As discussed earlier, some issues are associated with classical methods, as they cannot account for the regression-to-the-mean bias and spatial effects, which can cause poor crash estimation. In addition, both EB and spatial autoregressive methods are unable to address multilevel data and group-level random effects. By contrast, the full Bayesian framework is more flexible and can be easily extended to include random effect terms and accommodate data with multilevel structure. Nevertheless, even if the full Bayesian framework appears to be an appropriate approach for dealing with crash data, as discussed earlier in this chapter, previous studies using this method have had some major limitations. First, analyses have frequently been performed in univariate settings that do not account for correlations among crash severities, which can decrease the efficiency of the estimator Ma and Kockelman (2006). Second, temporal effects have not been addressed properly, which might also lead to poor crash frequency estimation. Third, a possible relationship between urban and rural crashes and their severity outcomes have not been investigated. Fourth, the nature of the relationship between fatal crashes in one area and the average time to the nearest hospital has not been explored in terms of accessibility. Finally, individual-level data have not been combined with area-based data to identify high-risk locations, which can result in the loss of available information or, in some cases, even in the incorrect identification of high-/low-risk location. In light of the above, this study builds upon the literature and proposes an integrated safety screening approach that can be used to identify

hotspots. This approach can provide insights and guide the development and/or improvement of crash prevention and control programs.

CHAPTER 3: DATA COLLECTION AND EXPLORATORY DATA ANALYSIS

In this study, two types of data are collected to perform the WV case study: macro or area-based data and micro or individual-level data. A detailed description of the data sources is given below.

3.1. Macro-level Data

Area-based data for six years (2010–2015) were obtained from different resources, such as the West Virginia Department of Transportation (WVDfarsOT), the American Community Survey (ACS), the West Virginia GIS Technical Center, and the West Virginia Department of Health and Human Resources (DHHR). The data collected were manipulated and classified into crash data, transport-related factors, socio-economic and demographic factors, and environmental factors, which are described as follows:

- **Crash data**

County-level crash counts were obtained from the traffic engineering division of the WVDOT. The database includes crash frequency at different severity levels (fatal, injury, and PDO) for each WV county.

- **Transport-related factors**

Average travel time to the nearest hospital: As the availability of hospitals can contribute to the survival chance of an individual with severe injury, the average travel time to the nearest hospital in each county was considered an explanatory variable for use in the fatal crash model. The hospital location and road network were obtained from the West Virginia GIS Technical Center and were processed using various GIS tools in ArcMap environment. First, the

road network was converted to a raster file, which was used as the basis for converting each cell to a cost value. Traffic crash victims were assumed to be transported to the nearest hospital with the speed of 70 mph from any location. Hence, the cost of traversing a unit distance would be $(1/70)$ h/mi or 5.33×10^{-4} min/meter. Second, as traffic crashes may occur anywhere on the road, the least accumulative travel time from any location to the nearest hospital/trauma center was calculated. Later, the average travel time to the nearest hospital was estimated for use as an independent variable in the fatal models. Figure 3.1 represents the framework for estimating the mean zonal travel time to the nearest hospital. The resulting map is shown in Figure 3.2.

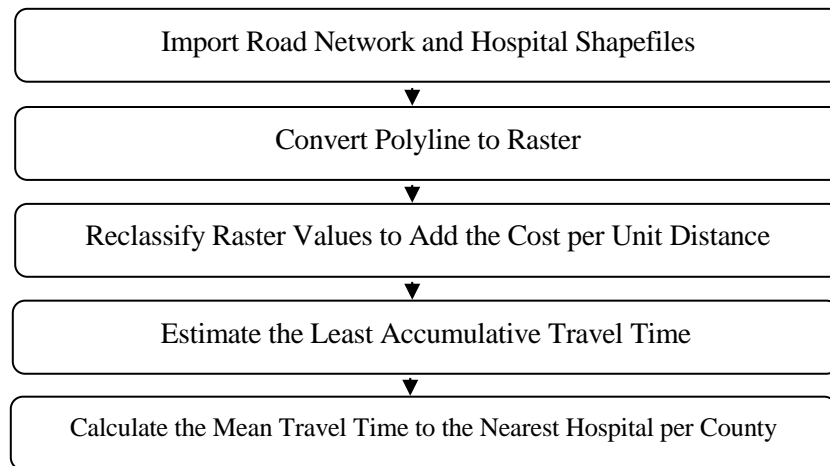
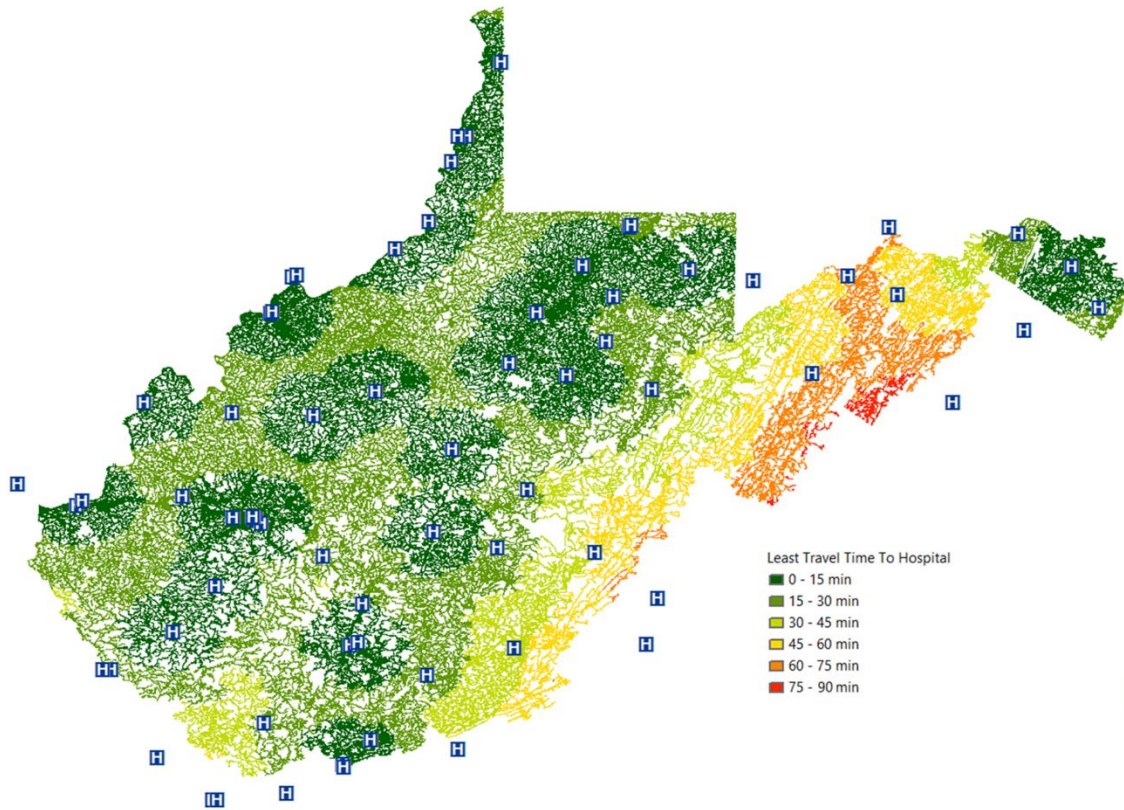
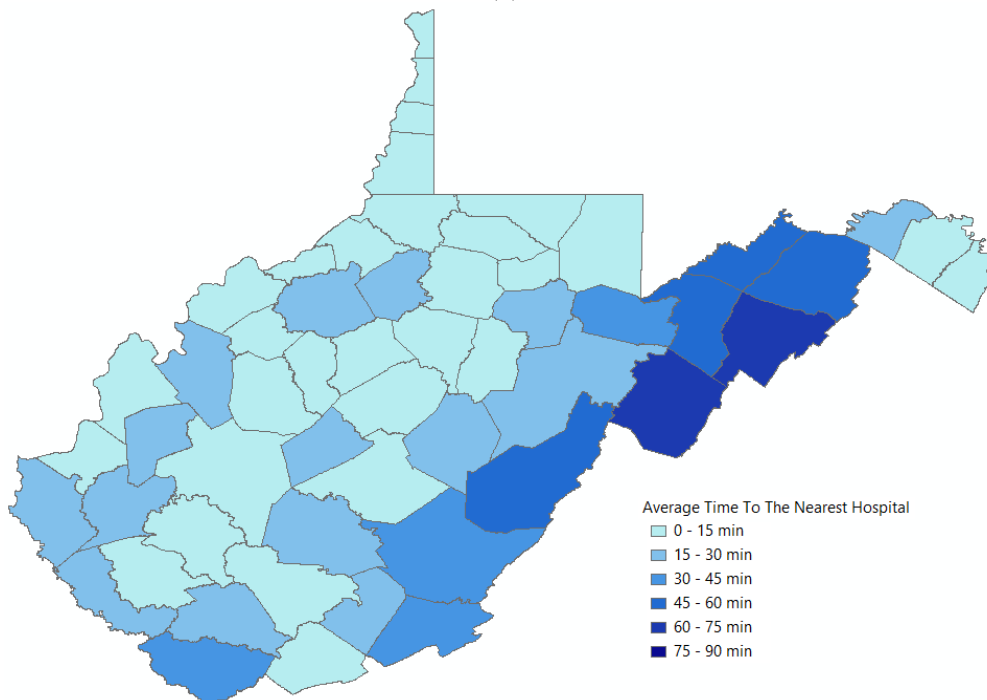


Figure 3.1 Framework for estimating the least cost/travel time to the nearest hospital



(a)



(b)

Figure 3.2 Map of (a) the least travel time to the nearest hospital and (b) the average access time to the nearest hospital per county

Median travel time to work: Median travel time may be associated with specific types of crashes. For instance, Lee et al. (2014) found that the portion of commuters with a commute time less than 15 minutes is negatively associated with at-fault crashes Lee et al. (2014). The median travel time to work was obtained from the ACS in the website of the US Census Bureau for the years 2010–2015.

Intersection and road density: About 40% of pedestrian crashes occur at intersections, which may result in fatal or severe injuries Lord et al. (1998). Moreover, according to the Fatality Analysis Reporting System (FARS) encyclopedia, about 70% of fatalities occur on primary roads (i.e., arterial and interstates), and the remaining ones occur on secondary roads. Primary and secondary roadways have different speed limits and traffic loads that can affect crash frequency and crash severity. Thus, intersection density (number of intersections divided by county area) and primary and secondary roadway density (length of roadways divided by county area) were created using GIS tools for use in the models. The road network was obtained from the WV GIS Technical Center.

- **Socio-economic and demographic factors**

Socio-economic and demographic factors are generally recognized as an important predictor of traffic crashes, and they are used by transportation engineers to apply mitigation strategies and plan for the future. The US census website is an excellent repository of state- and county-level socio-economic and demographic statistics, such as population estimates by age, percentage of the population under poverty, median household income, median age, and percentage of population in the labor force. These variables were obtained from the ACS in the website of the US Census Bureau as population estimates per county for the years 2010–2015.

Driving under influence (DUI) arrests (per 10,000 population): The DUI arrest rate in a county is used to relate alcohol consumption to crash frequency at different levels. Data were obtained from the website of the West Virginia DHHR for the years 2010–2015.

School density: According to the FARS, 131 people died in school-related crashes in 2016–2015. Of these, 8% were riding in buses, and 21% were pedestrians, bicyclists, and others. Hence, identifying the relationship between school density and different crash outcomes is important. School density was estimated by dividing the number of schools per county over the county area. The school shapefile was obtained from the West Virginia GIS Technical Center.

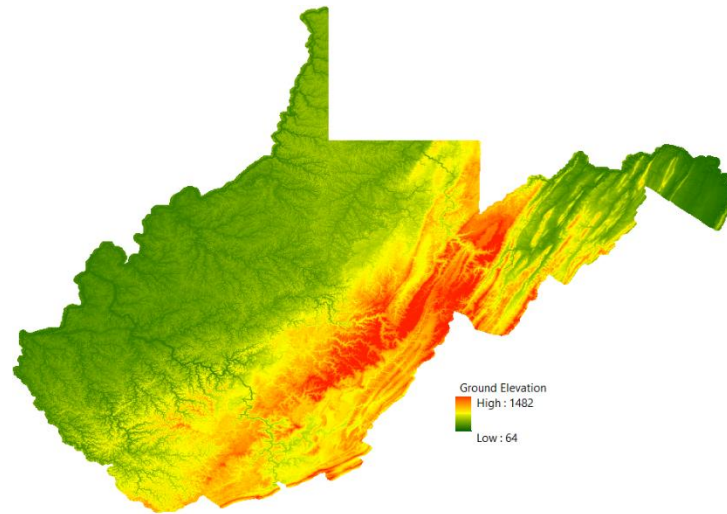
- **Socio-economic and demographic factors**

Rurality versus urbanity: The Geography Division at the US Census Bureau has classified US counties into metro and non-metro areas based on population, county size, and economic trends. According to the US Census Bureau, metropolitan areas contain at least one urbanized area with a population of 50,000 residents or more (any county that is not a part of a metropolitan area is considered rural). Such information is used to identify metro and non-metro counties in WV. This variable is used to investigate the possible relationship between urban/rural crashes and crash severity outcomes. Data were obtained from the website of the US Census Bureau.

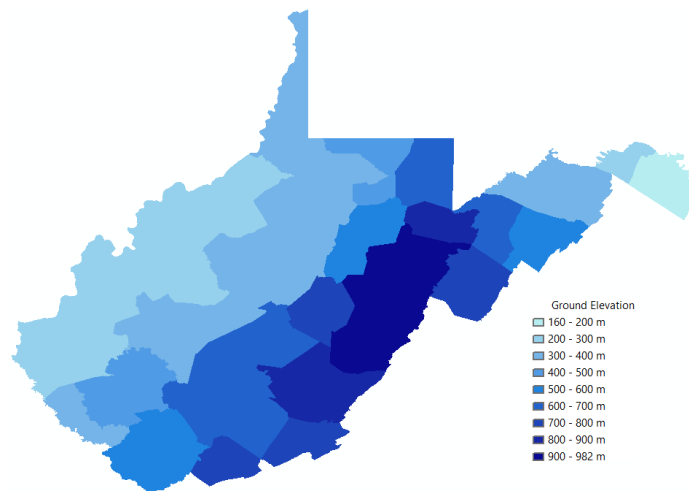
Average ground elevation: Elevation is used as a surrogate variable to account for the topography of each county (mountainous versus flat areas) in order to explore their impacts on traffic crashes. The average ground elevation was estimated using the digital elevation model (DEM) obtained from the West Virginia GIS Technical Center (see Figure 3.3).

Weather condition: The number of days with any measurable precipitation for each county is used as a surrogate for the climate condition. Because of lack of data, information for each county

was manually collected from the climate tab [in the website of Sperling's Best places](#) for the year 2016.



(a)



(b)

Figure 3.3 Map of (a) the ground elevation from DEM and (b) the average ground elevation per county

Figure 3.4 shows that all datasets from different resources are merged together for use in the macro-level data analysis. Table 3.1 presents the descriptive statistics of all data used in the macro-level data analysis.

TABLE 3.1 Descriptive statistics of the variables used in the macroscopic data analysis (330 observations)

Variable	Mean	St. Dev.	Min.	Max.
<i>Transport factors</i>				
Fatal crashes	5.19	4.69	0	29
Injury crashes	178.70	223.84	17	1558
PDO crashes	531.95	766.73	27	4791
Median travel time to work (minutes)	27.72	4.88	18.3	41.4
Mean time to the hospital (minutes)	20.16	15.36	2.99	69.84
Primary roadway density (miles per sq. mile)	0.06	0.09	0	0.37
Secondary roadway density (miles per sq. mile)	0.405	0.17	0.20	0.91
Intersection density	106.92	32.80	45.04	175.03
<i>Socio-economic factors</i>				
Labor force (%)	52.23	6.66	30	68
Median income (in thousands of dollars)	39.56	7.06	24.92	66.68
Median age (years)	43.00	3.12	30	49.10
Population with a B.S. degree or higher (%)	15.87	6.27	5.10	39.70
Population under poverty (%)	18.95	4.92	10.40	34.50
Male population (%)	49.88	1.90	45.10	59.70
Population under 19 (%)	22.54	1.64	17.62	25.93
Population between 20 to 24 (%)	5.96	1.90	4.55	16.30
Population over 65 (%)	16.16	2.06	9.09	20.99
Driving under influence (DUI) arrests per 10,000 drivers	47.29	29.30	5.10	227.48
School density (number of schools per area)	0.04	0.039	0.01	0.22
<i>Environment-related factors</i>				
Metro (1 if the county is a metropolitan area, 0 otherwise)	0.38	0.47	0	1
Mean ground elevation (in hundreds of meters)	4.46	2.02	1.61	9.81
Number of days with any measurable precipitation	90.11	10.90	70	122
<i>Exposure variable</i>				
County population	33,646	33,237	5,816	193,102

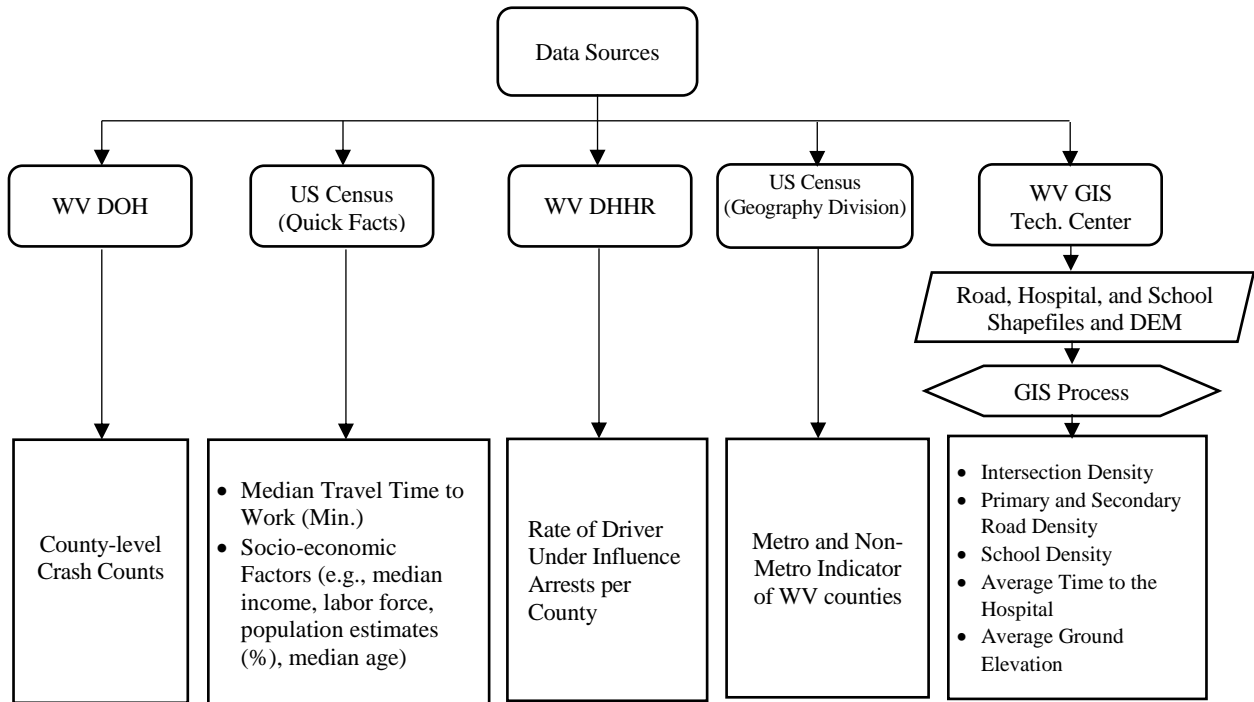


Figure 3.4 Area-level data collection and preparation process

3.2. Micro-level Data

The individual-level crash dataset for six years (2010–2015) for use in the hierarchical logit model was obtained from the FARS in the website of the National Highway Traffic Safety Administration. The database provides many attributes associated with each crash; these attributes can be categorized into four classes, which are occupant characteristics, crash information, vehicle characteristics, and roadway and environmental characteristics. The occupant attributes include age, gender, alcohol involvement, and a number of other attributes related to the profile of the occupant and/or the situational factors during the crash. Crash information refers to the characteristics of the crashes, such as severity. Roadway and environmental attributes include roadway surface condition, weather condition, and lighting condition. Table 3.2 shows the summary statistics of all variables used in the micro-level data analysis.

The micro-level crash data were analyzed in ArcGIS software in order to identify the census tracts in which they are located. In preparation for the modeling, all crash points within a census tract were assigned a census track unique ID. Figure 3.5 presents the framework used to relate the crash points to the corresponding census tracts.

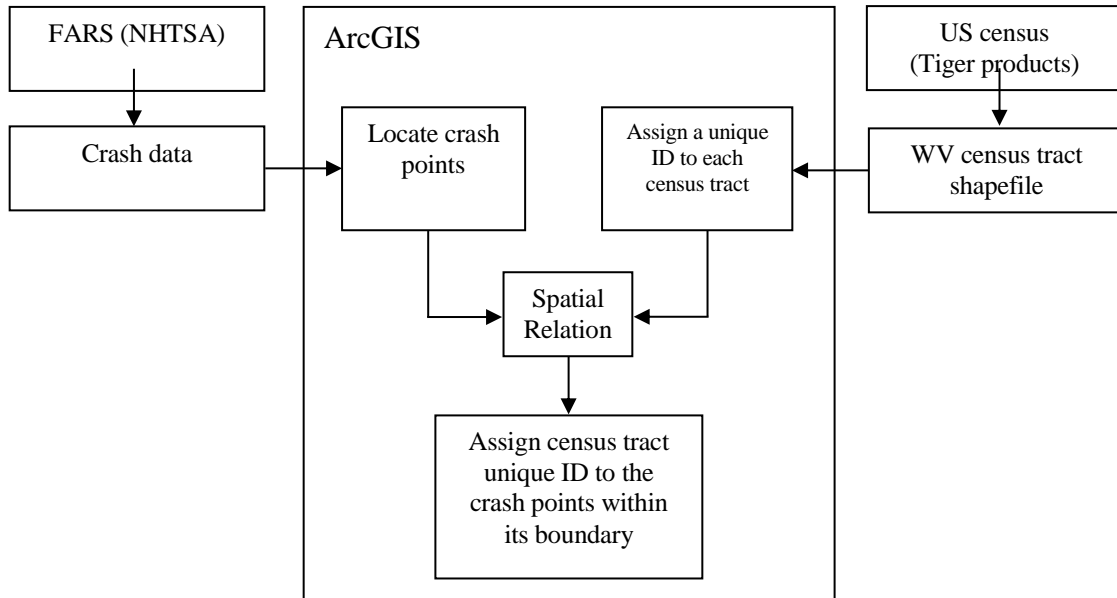


Figure 3.5 Procedure for relating crash points to census tracts

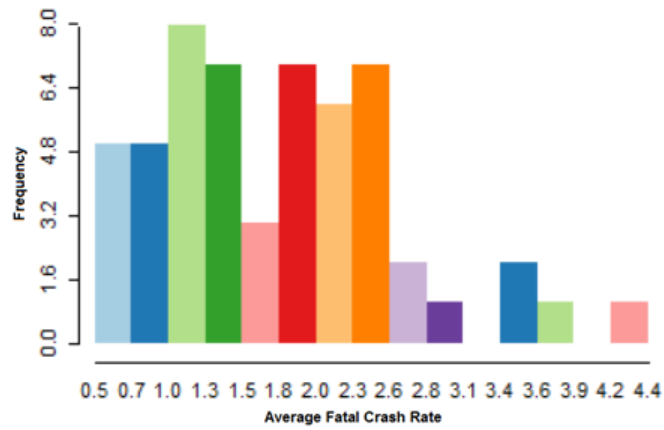
TABLE 3.2 Descriptive statistics of the variables used in the microscopic data analysis (3,724 observations)

Variables	Mean	SD	Min	Max
Occupant-level factors				
Severity (fatal=1, non-fatal=0)	0.462	0.499	0	1
Gender (male=1, female=0)	0.674	0.469	0	1
Age	39.378	20.333	0	99
Occupant type (driver=1, otherwise=0)	0.658	0.474	0	1
Ejection (occupant is ejected fully or partially=1, otherwise=0)	0.150	0.357	0	1
Vehicle-level factors				
Speeding (yes=1, No=0)	0.342	0.468	0	1
Alcohol involvement (yes=1, No=0)	0.222	0.415	0	1
Crash-level factors				
Surface condition (dry=1, otherwise=0)	0.790	0.407	0	1
Road alignment (straight=1, curve=0)	0.617	0.486	0	1
Road type (divided=1, otherwise=0)	0.724	0.447	0	1
Arterial	0.588	0.492	0	1
Collector	0.287	0.452	0	1
Weather (clear=1, otherwise=0)	0.664	0.472	0	1
Lighting (daylight=1, otherwise=0)	0.585	0.493	0	1

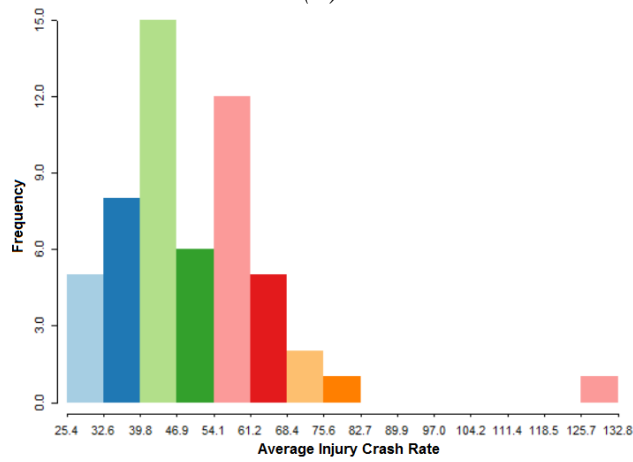
3.3. Exploratory Spatial Data Analysis

Exploratory data analysis (EDA) is the first step in analyzing the data characteristics, and it aims to identify the outliers and crash trends. This section presents some key results from the exploratory analysis of the macro-level data collected.

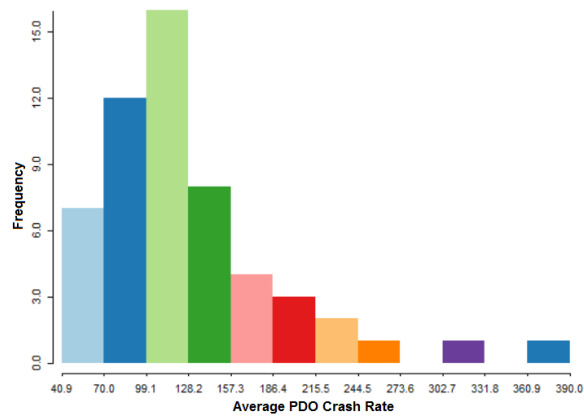
Figure 3.6 displays the histograms of the crash rates by severity, representing the density of the data. The figure shows that the crash rate distribution of all crash types tends to be skewed to the right, with a tail on the high end and taller bins on the low end, suggesting that more counties have a lower fatal crash rate in WV. An alternative way to present the distribution of spatial data is a boxmap. Figure 3.7 shows the boxmaps visualizing the distribution of the average crash rates per 10,000 population by severity based on six- year data (2010–2015).



(a)



(b)



(c)

Figure 3.6 Histograms of the average over the years 2010–2015: (a) fatal crash rate, (b) injury crash rate, and (c) PDO crash rate

Each map is a choropleth map, in which quantile classification of the data has been applied to reflect the data distribution and identify anomalous counties. Figure 3.7(a) shows that Pendleton County has the highest average fatal crash rate among WV counties. Moreover, WV eastern and southern counties tend to have higher fatal crashes. Figure 3.7(b) shows that Raleigh and McDowell Counties experienced more injury crashes than other counties did when their population is accounted for. Finally, as shown in Figure 3.7(c), Ohio, Lewis, Cabell, Kanawha, Raleigh, and McDowell Counties are among those with the highest PDO crash rates in WV.

Moran's I statistics can also be used to investigate the presence of spatial autocorrelation between traffic crashes in neighboring counties. It is defined as follows:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}, \quad (3.1)$$

where I is Moran's index value, N is the number of spatial units (counties), y_i and y_j are the crash rates related to targeted county i and neighboring county j , respectively, \bar{y} is the average crashes, and w_{ij} is an element of a matrix of spatial weights. A Moran's index value near +1.0 indicates clustering, an index value near -1.0 indicates dispersion, and a value close to zero indicates a random spatial pattern.

Queen contiguity was used to calculate the spatial weight; that is, counties that share an edge or have coincident boundaries are considered neighboring (the spatial weight for a neighboring county is 1, otherwise 0) and included in computations for the targeted county (see Figure 3.8). The global Moran's I statistics in each year from 2010 to 2015 are calculated using Geoda, and the results are summarized in Table 3.3. The results show no strong statistical evidence indicating that spatial autocorrelation is present in the crash data.

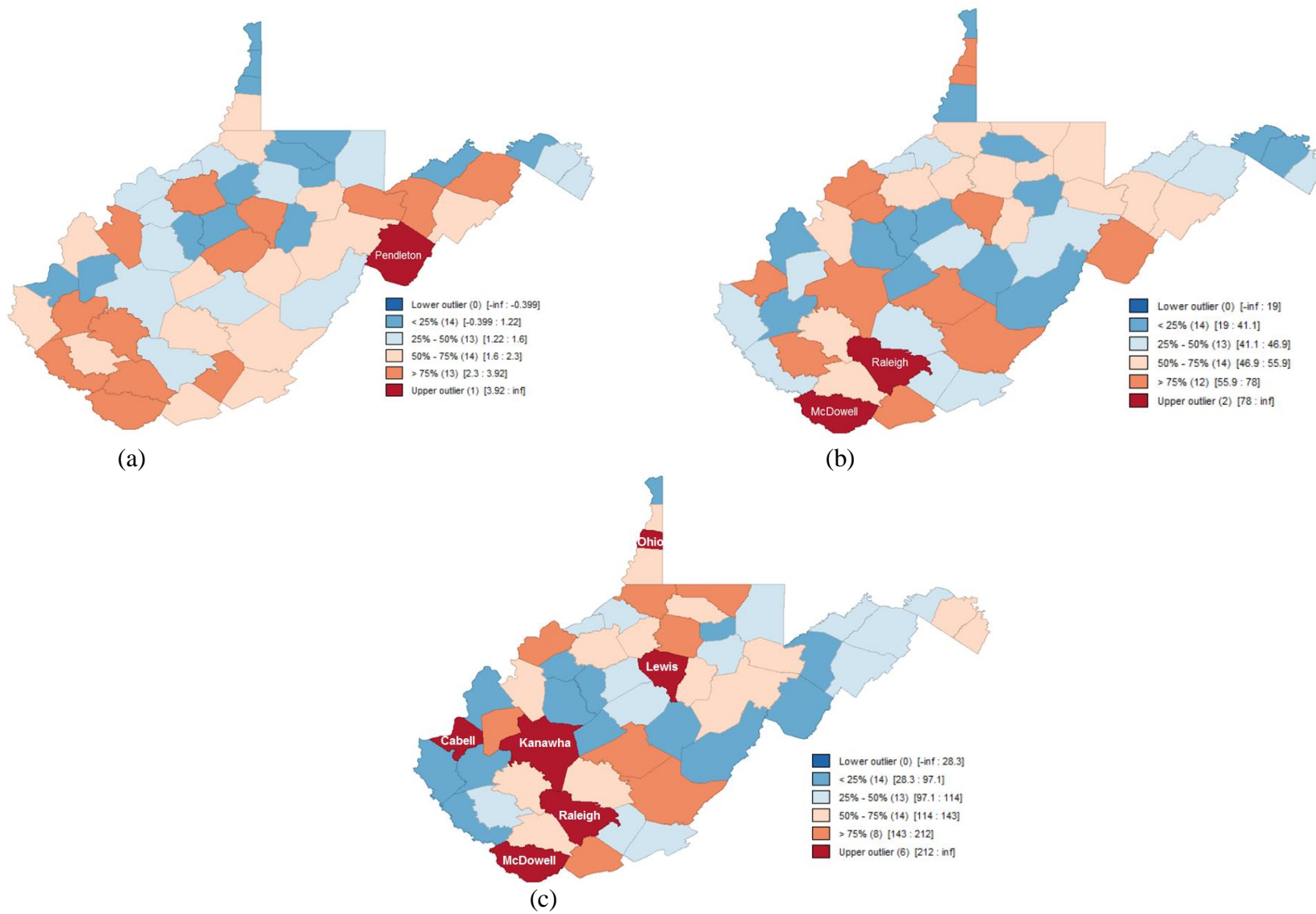


Figure 3.7 Boxmap of the average over the years 2010–2015: (a) fatal crash rate, (b) injury crash rate, and (c) PDO crash rate

TABLE 3.3 Global Moran's I statistics

Year	Fatal		Injury		PDO	
	Index	P-value	Index	P-value	Index	P-value
2015	-0.004	0.4	-0.07	0.3	-0.07	0.26
2014	-0.01	0.4	-0.06	0.3	-0.05	0.4
2013	0.04	0.2	-0.05	0.3	0.04	0.2
2012	0.05	0.2	0.08	0.13	-0.03	0.4
2011	0.03	0.3	-0.04	0.4	-0.01	0.4
2010	0.1	0.1	0.02	0.3	0.003	0.4

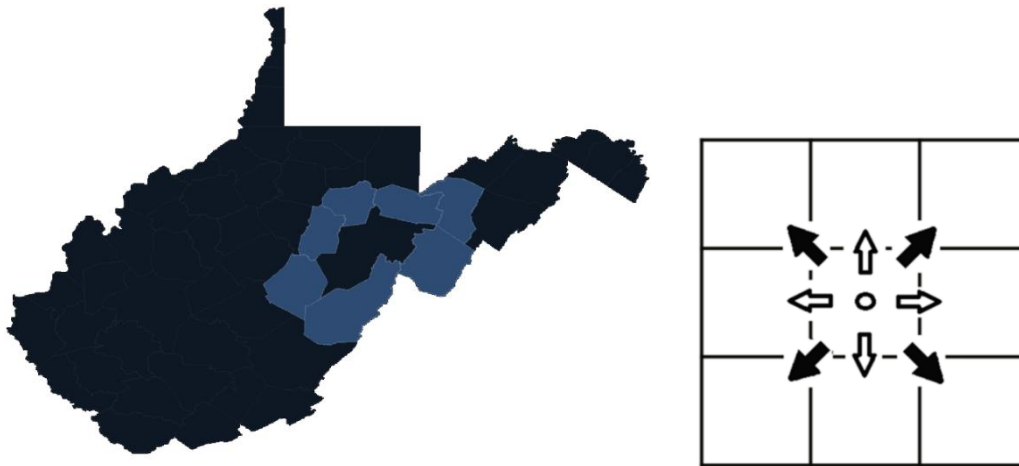
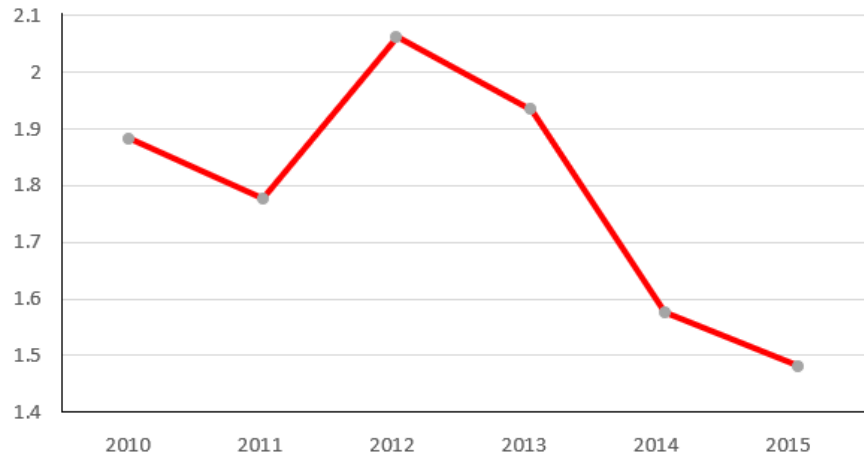


Figure 3.8 Queen's case contiguity

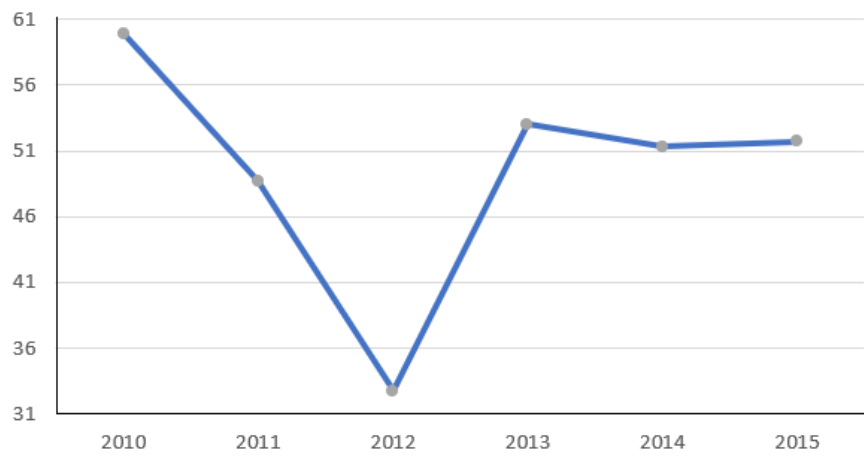
In addition, the Pearson correlation coefficients of fatal injury and PDO crashes were estimated and are summarized in Table 3.4. The results show that crashes of all crash severities are positively associated. For example, the locations with many fatal crashes will likely have many injury and PDO crashes. Figure 3.9 presents the average crash rates in 2010–2015 in WV. Figure 3.9(a) shows that the fatal crash rate decreased from 1.88 in 2010 to 1.77 in 2011, increased again

TABLE 3.4 Pearson correlation among crash severities

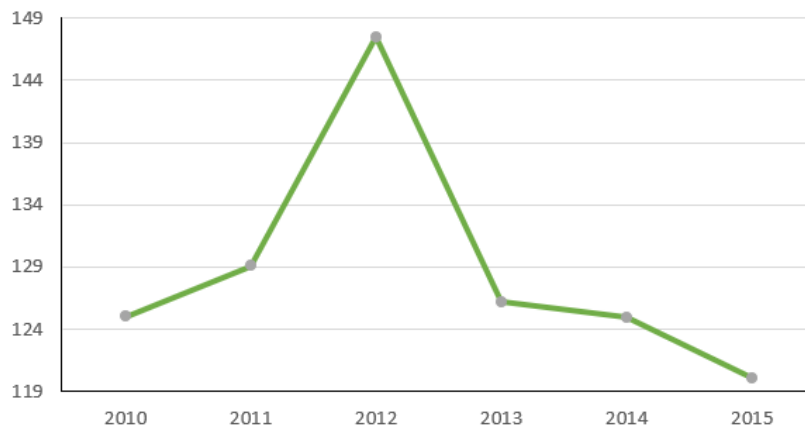
Crashes	Fatal	Injury	PDO
Fatal	1	0.71	0.78
Injury		1	0.87
PDO			1



(a)



(b)



(c)

Figure 3.9 Average (a) fatal crash rate, (b) injury crash rate, and (c) PDO crash rate in West Virginia

to 2.06 in 2012, but has been decreasing steadily since then. The injury crash rate was down from 59.9 in 2010 to 32.7 in 2012. However, it suddenly increased to 53.0 in 2013 and remained stable from 2013 to 2015 (see Figure 3.9(b)). The PDO crash rate had a downward trend that started in 2012 and continued to 2015 (see Figure 3.9 (c)). In general, the overview of traffic crashes shows an overall downward trend for fatal and PDO crashes.

Various scatter plots were created to investigate the potential relationship between the explanatory variables and crashes. Figure 3.10 shows a number of variables with positive, negative, and no relationship with the fatal crash rate (per 10,000 population). For instance, a negative, non-linear relationship seems to exist between the fatal crash rate and the median income, the labor force, the population with a BS degree or more, secondary roadway density, school density, the DUI rate, and the population with ages between 20 and 24. By contrast, the average travel time to the hospital is positively associated with the fatal crash rate. For the remaining variables, no obvious strong relationship seems to exist between them and the fatal crash rate, as the value of the fatal crash rate seems to change randomly. Additionally, according to the scatter plot of the fatal crash rate versus metropolitan areas, rural (i.e., nonmetropolitan) areas have a higher fatal crash rate.

Figure 3.11 shows that the median travel time and the population under 19 are negatively correlated with the injury crash rate, whereas the population with a BS degree, the population between 20 and 24, and primary roadway density are positively associated with the injury crash rate. Finally, an examination of the scatter plots in Figure 3.12 shows the presence of strong correlations between a number of variables and the PDO crash rate. For instance, the PDO crash rate is positively correlated with the labor force, the median income, the population with a BS degree or more, DUI, school density, and primary and secondary road density. Additionally, a significant negative correlation exists between the PDO crash rate and the median travel time to work.

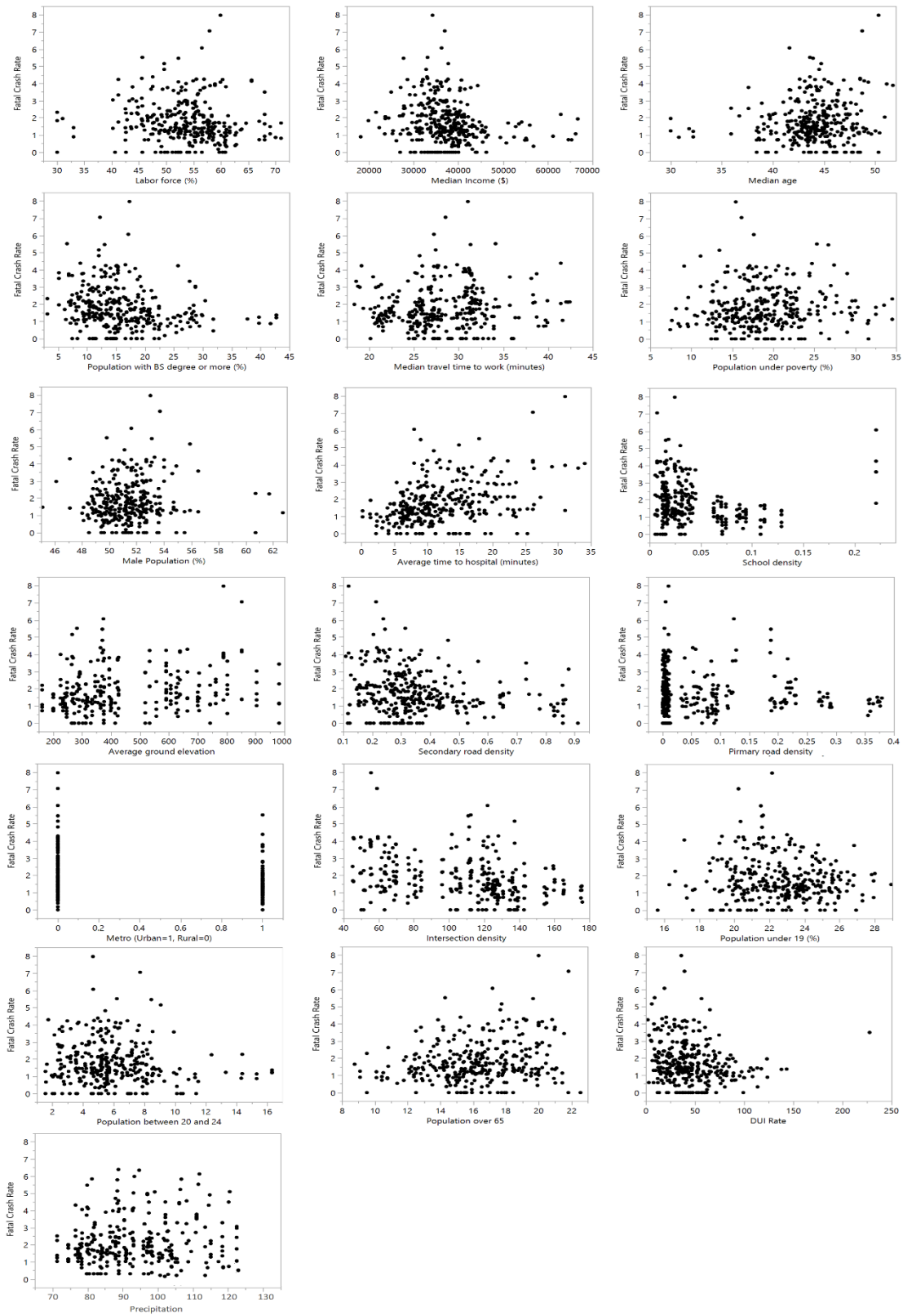


Figure 3.10 Scatter plot of the fatal crash rate (per 10,000 population) against covariates

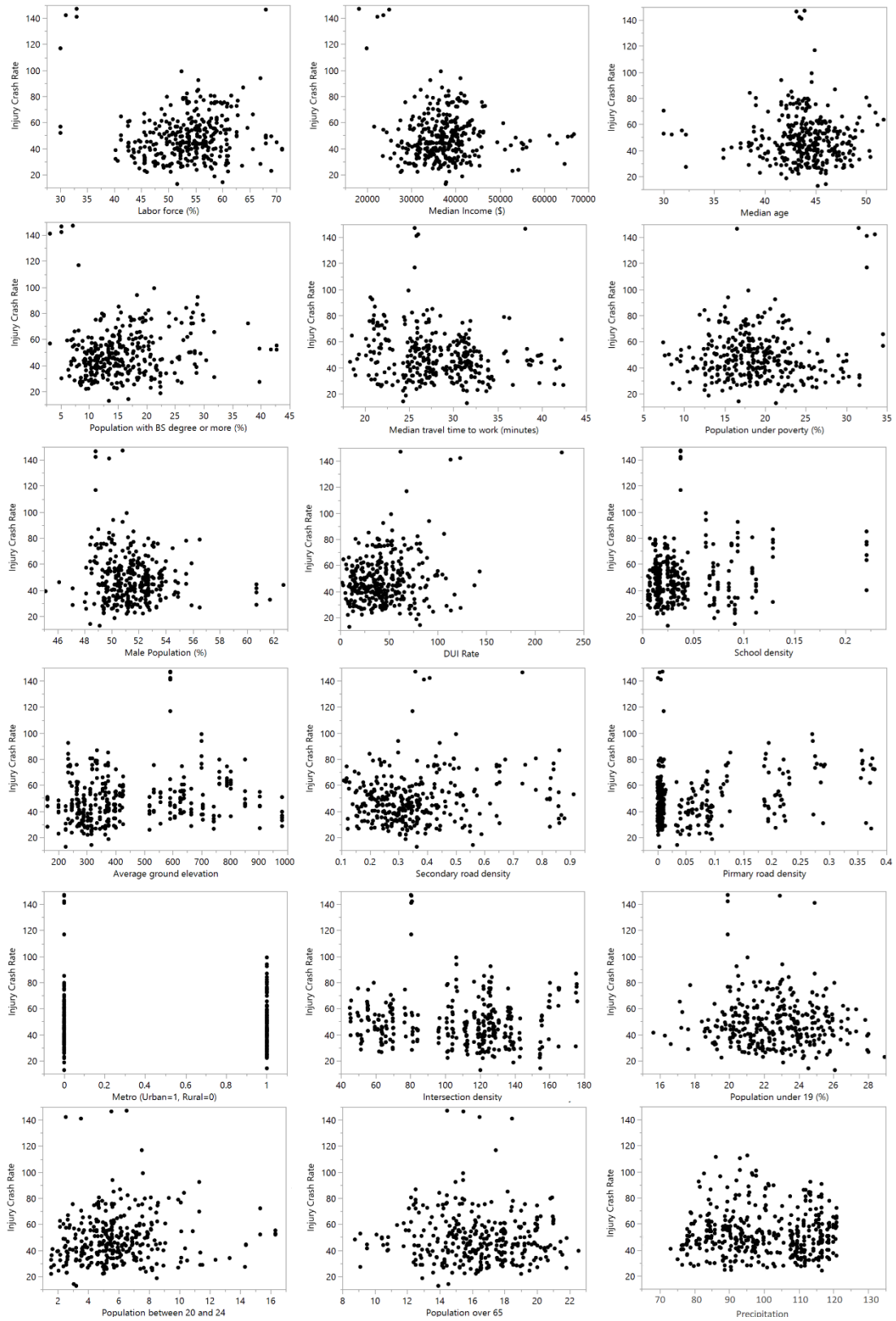


Figure 3.11 Scatter plot of the injury crash rate (per 10,000 population) against covariates

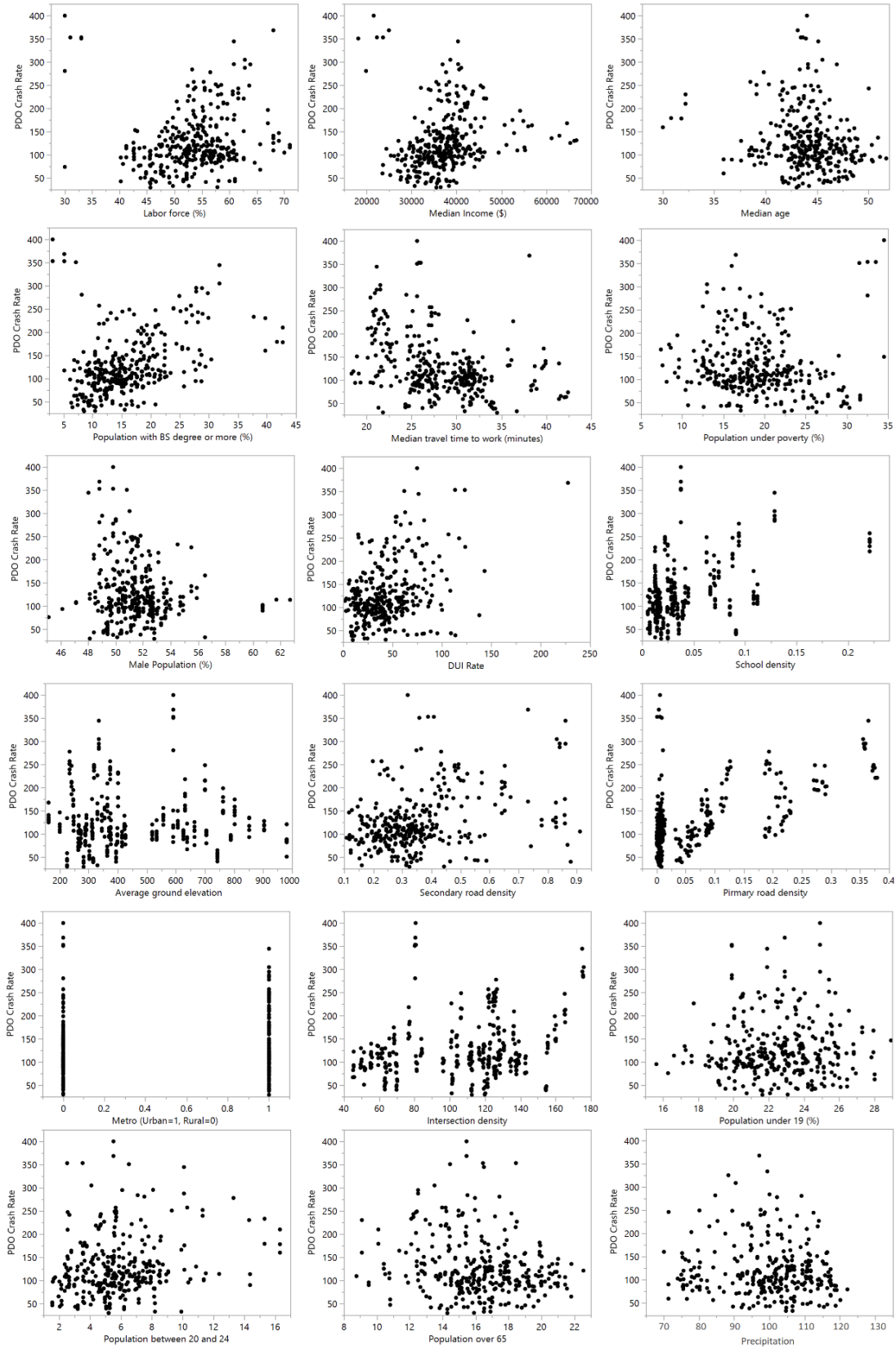


Figure 3.12 Scatter plot of the PDO crash rate (per 10,000 population) against covariates

3.4. Summary

In this section, various EDA techniques, such as histograms, boxplots, and 2D plots, have been utilized to identify high-risk locations in WV. The results of the analysis suggest that the distribution of traffic crashes at different severity levels is right skewed, suggesting that more counties have lower crash rates. Pendleton County has the highest fatal crash rates, whereas Raleigh and McDowell Counties have extreme (i.e., beyond the cut-off) injury crash rates. Moreover, counties located in eastern and southern WV tend to have higher fatal crash rates. However, injury and PDO crashes tend to be more randomly distributed across WV. From a temporal perspective, fatal and PDO crashes have had downward trends since 2012, whereas injury crashes have no clear pattern. The results of Moran's I test demonstrated that the number of crashes in a county is not significantly correlated with the number of crashes in neighboring counties. However, the Pearson correlation indicated that crashes of different severities within each county are positively correlated.

Investigation of a series of scatter plots showed that increases in median income, labor force, the population with a BS degree or more, secondary roadway density, school density, DUI rate, and population with ages between 20 and 24 will likely decrease the fatal crash rate. By contrast, increases in the average travel time to the hospital is positively correlated with the fatal crash rate. On the other hand, the median travel time and the population under 19 have a negative relationship with the injury crash rate, whereas the population with a BS or higher degree, the population between 20 and 24 years of age, and primary roadway density are positively associated with the injury crash rate. Lastly, a positive correlation exists between the PDO crash rate and the labor force, the median income, the population with a BS or higher degree, DUI, school density, and primary and secondary road density, whereas the PDO crash rate is negatively correlated with the median travel time to work.

CHAPTER 4: METHODOLOGY

The goal of this research is to develop an integrated screening method to detect hazardous locations. This section presents the methods developed to accomplish the research tasks discussed in Chapter 1, Section 1.2. Specifically, in the first part of this research, different full Bayesian Poisson lognormal models are used to predict crash rates as a function of macro-level data, such as socio-economic and transport-related factors. In the second part of this research, various crash severity/logit models are developed based on micro-level data and road characteristics to account for the hierarchical nature in the crash data using a Bayesian framework. The deviance information criterion (DIC) is used as the basis to select the models with the best fit, and the excess risk is defined as a measure to determine and prioritize the high-risk locations. Lastly, in the third part of this research, a GIS approach is used to combine the results from both the macro- and the micro-level data analyses in order to identify hazardous locations. Different random effect terms are explored in the models to improve their performance and prediction ability, which is critical in site ranking/prioritization. All models are illustrated for the case study of WV. Sections 4.1, 4.2, 4.3, 4.4, and 4.5 discuss the Bayesian framework, macro-level data analysis approaches, micro-level data analysis methods, model comparison, and the proposed integrated screening method, respectively.

4.1. Bayesian Framework

The Bayesian approach is extensively used in disease mapping and ecological analysis for developing spatial models because of its flexibility. Bayesian statistics consider parameters as random variables expressed in the form of probabilities. Based on this approach, a probability model of interest is fitted by incorporating prior information regarding the unknown parameters and the likelihood function of the observed data. To demonstrate this, let y represent a vector of n observations y_1, y_2, \dots, y_n , and let ψ represent a vector of m parameters $\psi_1, \psi_2, \dots, \psi_m$ on which the

distribution of the observations depends. Then, Bayes' theorem is based on the posterior distribution of the parameters given data $p(\psi|y)$, on a combination of the likelihood of observing the data given ψ , $P r[y/\psi]$ and our prior expectation/belief, expressed as the prior density $P r[\psi]$:

$$P r[\psi |y] \propto P r[y/ \psi] \cdot P r[\psi]. \quad (4.1)$$

The Bayesian approach shows numerous theoretical and practical advantages over classical likelihood-based inference methods, especially for hierarchical models and they will be the ones that will be estimated in this dissertation. For instance, the Bayesian approach accounts for sources of uncertainty/variability (individual- and group-level variation) and correlations among the response variables (e.g., severity levels).

4.2. Macro-level Data Analysis

4.2.1. Poisson Lognormal Model

Many researchers (see for example, Tunaru 2002; Ma and Kockelman, 2006; Aguerro-Valverde and Jovanis, 2009) have shown that crash frequencies are correlated across different severity levels because of the common unobserved factors at each road segment. To address this problem, this dissertation proposes a multivariate model to relate various county-level socio-economic and transport-related factors with crash occurrence while accounting for the possible temporal effects and spatial autocorrelations among adjacent counties. The proposed model is described below.

Let Obs_{ist} denote the observed crash count by severity level s in county i in year t . In this study, the following three severity levels are considered: fatal ($s=1$), injury ($s=2$), and PDO ($s=3$). The model can be expressed as follows:

$$Obs_{ist} \sim poisson (C_{ist}), \quad (4.2)$$

$$C_{ist} = \lambda_{ist} \cdot E_{it}, \quad (4.3)$$

where C_{ist} and λ_{ist} are, respectively, the expected number of crashes and crash rate by severity level ($s=1, 2, 3$) in county i in year t , and E_{it} refers to the exposure variable (population) in county i in year t . The first model to estimate the crash frequency for the three severity levels can be written as follows:

$$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it}, \quad (4.4)$$

where X_{it} denotes a vector of standardized covariates for county i at time t , β_{0s} is the intercept, and β_s is the vector of coefficients for the model of crash severity s . This model does not account for overdispersion, as well as for spatial and temporal effects. A set of models with random effects is developed to address these issues. Hence, the general random effects model is specified as follows:

$$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is} + \nu_{ts} + \delta_{ts} + \gamma_{its}, \quad (4.5)$$

where θ_{is} is the county-based unstructured random effects, which capture county-level heterogeneity or overdispersion, φ_{is} is the structured spatial random effect or spatial correlation term, ν_{ts} is the unstructured temporal random term, δ_{ts} is the structured temporal term that accounts for temporal correlation, and γ_{its} is the spatio-temporal interaction that captures unobserved heterogeneity over space and time. As other studies (Behnood and Mannering, 2015; Chiou and Chih-Wei, 2014) have established, unobserved effects are not spatially or temporally autocorrelated. Hence, the spatial and temporal correlation terms φ_{is} , and δ_{ts} may be dropped from Equation 5, and the model can be reduced. Therefore, nine models are established for the selection of an optimal model, as the full models might not be stable. These models are outlined in Table 4.1.

TABLE 4.1 Summary of the frequency models proposed in this study

No	Model
Basic (non-random effect model)	
1	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it}$
Full Models	
2	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is} + v_{ts} + \delta_{ts, MRW(1)} + \gamma_{it}$
3	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is} + v_{ts} + \delta_{ts, MRW(2)} + \gamma_{it}$
Spatial Models	
4	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is}$
5	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is}$
Temporal Models	
6	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + v_{ts} + \delta_{ts, MRW(1)}$
7	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + v_{ts} + \delta_{ts, MRW(2)}$
8	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + v_{ts}$
Spatio-Temporal Modes (without structured spatial and structured temporal terms)	
9	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + v_{ts} + \gamma_{it}$

Spatial effects: Spatial random effects follow the MCAR. With the CAR specifications proposed by Besag (1974) extended, the MCAR prior is defined as follows:

$$[\Phi_i | \Phi_j, i \neq j] \sim N \left(\bar{\Phi}_i, \frac{\sum_j \Phi_j}{\sum_j w_{ij}} \right), \quad (4.6)$$

$$\bar{\Phi}_i = (\bar{\varphi}_{i1}, \bar{\varphi}_{i2}, \bar{\varphi}_{i3}), \quad (4.7)$$

$$\bar{\varphi}_{is} = \frac{\sum_j \varphi_{js} w_{ij}}{\sum_j w_{ij}}, \quad (4.8)$$

$$\Sigma_{\Phi} = \begin{pmatrix} \sigma_{\varphi_1}^2 & r_{\varphi_{12}} \sigma_{\varphi_1}^2 \sigma_{\varphi_2}^2 & r_{\varphi_{13}} \sigma_{\varphi_1}^2 \sigma_{\varphi_3}^2 \\ r_{\varphi_{12}} \sigma_{\varphi_1}^2 \sigma_{\varphi_2}^2 & \sigma_{\varphi_2}^2 & r_{\varphi_{23}} \sigma_{\varphi_2}^2 \sigma_{\varphi_3}^2 \\ r_{\varphi_{13}} \sigma_{\varphi_1}^2 \sigma_{\varphi_3}^2 & r_{\varphi_{23}} \sigma_{\varphi_2}^2 \sigma_{\varphi_3}^2 & \sigma_{\varphi_3}^2 \end{pmatrix} \quad (4.9)$$

where $\Phi_i = (\varphi_{i1}, \varphi_{i2}, \varphi_{i3})$ is a 3D vector of spatial random effects; φ_{is} represents the spatial random effects in county i for the crash severity levels $s=1,2,3$; w_{ij} is an adjacency indicator (1 if counties i and j are adjacent and 0 otherwise); Σ_{Φ} is a covariance matrix with diagonal elements $\sigma_{\varphi_s}^2$

representing conditional variances for the severity levels; and $r_{\phi_{12}}$, $r_{\phi_{23}}$, and $r_{\phi_{13}}$ are the within-area conditional correlations of random effects for fatal injury crashes, injury PDO crashes, and fatal PDO crashes, respectively.

Unstructured spatial random effects follow a normal distribution, defined as follows:

$$\Theta_i \sim N(\mu_0, \Sigma_\Theta), \Theta_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}), \quad (4.10)$$

where θ_{is} represents the unstructured spatial random effect at different severity levels, μ_0 is a vector of zeroes $\mu_0 = (0, 0, 0)$, and Σ_Θ is a variance-covariance matrix:

$$\Sigma_\Theta = \begin{pmatrix} \sigma_{\theta_1}^2 & r_{\theta_{12}} \sigma_{\theta_1}^2 \sigma_{\theta_2}^2 & r_{\theta_{13}} \sigma_{\theta_1}^2 \sigma_{\theta_3}^2 \\ r_{\theta_{12}} \sigma_{\theta_1}^2 \sigma_{\theta_2}^2 & \sigma_{\theta_2}^2 & r_{\theta_{23}} \sigma_{\theta_2}^2 \sigma_{\theta_3}^2 \\ r_{\theta_{13}} \sigma_{\theta_1}^2 \sigma_{\theta_3}^2 & r_{\theta_{23}} \sigma_{\theta_2}^2 \sigma_{\theta_3}^2 & \sigma_{\theta_3}^2 \end{pmatrix}, \quad (4.11)$$

where $\sigma_{\theta_s}^2$ represents the conditional variances for the severity levels, and r_θ represents the within-area conditional correlation of random effects among the different severity levels.

Temporal effects: For the structured temporal component $\delta_{t,c}$, a multivariate random-walk prior of order 1 and 2, MRW(1) and MRW(2), are considered to account for the shared unmeasured risk factors across severities. As with the spatial severity model, the CAR prior is used to define the temporal neighboring points of t as $t-1$ and $t+1$. MRW(1) is defined as follows:

$$\Delta_t | \Delta_{-t} \sim \begin{cases} N(\Delta_{t+1}, \Sigma_\Delta) & \text{for } t=1 \\ N\left(\frac{\Delta_{t-1} + \Delta_{t+1}}{2}, \frac{\Sigma_\Delta}{2}\right) & \text{for } t=2, \dots, 5, \\ N(\Delta_{t-1}, \Sigma_\Delta) & \text{for } t=6 \end{cases} \quad (4.12)$$

MRW(2) is commonly used in epidemiology; it allows for flexibility in fitting the nonlinear effects of time and is defined as follows:

$$\Delta_t | \Delta_{-t} \sim \begin{cases} N(2\Delta_{t+1} - \Delta_{t+2}, \Sigma_{\Delta}) & \text{for } t = 1 \\ N\left(\frac{2\Delta_{t-1} + 4\Delta_{t+1} - \Delta_{t+2}}{5}, \frac{\Sigma_{\Delta}}{5}\right) & \text{for } t = 2 \\ N\left(\frac{4\Delta_{t-1} + 4\Delta_{t+1} - \Delta_{t+2} - \Delta_{t-2}}{6}, \frac{\Sigma_{\Delta}}{6}\right) & \text{for } t = 3, 4, \\ N\left(\frac{4\Delta_{t-1} + 2\Delta_{t+1} - \Delta_{t-2}}{5}, \frac{\Sigma_{\Delta}}{5}\right) & \text{for } t = 5 \\ N(-\Delta_{t-2} + 2\Delta_{t-1}, \Sigma_{\Delta}) & \text{for } t = 6 \end{cases} \quad (4.13)$$

where $\Delta_t = (\delta_{t1}, \delta_{t2}, \delta_{t3})$ is the 3D structured temporal effects, and Σ_{Δ} is the variance covariance matrix:

$$\Sigma_{\Delta} = \begin{pmatrix} \sigma_{\delta 1}^2 & r_{\delta 12} \sigma_{\delta 1}^2 \sigma_{\delta 2}^2 & r_{\delta 13} \sigma_{\delta 1}^2 \sigma_{\delta 3}^2 \\ r_{\delta 12} \sigma_{\delta 1}^2 \sigma_{\delta 2}^2 & \sigma_{\delta 2}^2 & r_{\delta 23} \sigma_{\delta 2}^2 \sigma_{\delta 3}^2 \\ r_{\delta 13} \sigma_{\delta 1}^2 \sigma_{\delta 3}^2 & r_{\delta 23} \sigma_{\delta 2}^2 \sigma_{\delta 3}^2 & r_{\delta 3}^2 \end{pmatrix}. \quad (4.14)$$

Unstructured temporal effects follow a multivariate normal distribution with a 3D variance covariance matrix Σ_v and is defined as follows:

$$V_t \sim N(\mu_0, \Sigma_v), V_t = (v_{t1}, v_{t2}, v_{t3}), \quad (4.15)$$

$$\Sigma_v = \begin{pmatrix} \sigma_{v1}^2 & r_{v12} \sigma_{v1}^2 \sigma_{v2}^2 & r_{v13} \sigma_{v1}^2 \sigma_{v3}^2 \\ r_{v12} \sigma_{v1}^2 \sigma_{v2}^2 & \sigma_{v2}^2 & r_{v23} \sigma_{v2}^2 \sigma_{v3}^2 \\ r_{v13} \sigma_{v1}^2 \sigma_{v3}^2 & r_{v23} \sigma_{v2}^2 \sigma_{v3}^2 & \sigma_{v3}^2 \end{pmatrix}, \quad (4.16)$$

Spatio-temporal effect: A spatio-temporal term is used to capture the unobserved factors over space and time across different crash severities, and it follows a multivariate normal distribution:

$$\Gamma_{it} \sim N(\mu_0, \Sigma_{\Gamma}), \Gamma_{it} = (\gamma_{it1}, \gamma_{it2}, \gamma_{it3}), \quad (4.17)$$

where Σ_{Γ} is the spatio-temporal variance-covariance matrix:

$$\Sigma_{\Gamma} = \begin{pmatrix} \sigma_{\gamma 1}^2 & \rho_{\gamma 12} \sigma_{\gamma 1}^2 \sigma_{\gamma 2}^2 & \rho_{\gamma 13} \sigma_{\gamma 1}^2 \sigma_{\gamma 3}^2 \\ \rho_{\gamma 12} \sigma_{\gamma 1}^2 \sigma_{\gamma 2}^2 & \sigma_{\gamma 2}^2 & \rho_{\gamma 23} \sigma_{\gamma 2}^2 \sigma_{\gamma 3}^2 \\ \rho_{\gamma 13} \sigma_{\gamma 1}^2 \sigma_{\gamma 3}^2 & \rho_{\gamma 23} \sigma_{\gamma 2}^2 \sigma_{\gamma 3}^2 & \sigma_{\gamma 3}^2 \end{pmatrix}. \quad (4.18)$$

Hyperprior: A normal prior $N(0, \tau_\beta)$ is assumed for β , where τ_β is the precision parameters that are the inverse of the variances and follow prior gamma (0.5, 0.0005) Wakefield et al. (2000). In case of multivariate updating, the variance-covariance matrix is defined as $\Sigma^{-1} \sim \text{Wishart}(R, d)$, where R and d are the scale matrix and the degree of freedom, respectively, and d is set to 3 for a weakly informative prior specification. As suggested by Gelman et al. (2014), R is defined as follows:

$$R = \begin{pmatrix} 0.1 & 0.005 & 0.005 \\ 0.005 & 0.1 & 0.005 \\ 0.005 & 0.005 & 0.1 \end{pmatrix}.$$

4.2.1.1. Macro-level Weighted Excess Risk

Safety is one of the five strategic goals identified by the Strategic Plan for fiscal years 2014–2018 (US Department of Transportation, 2014). In recent years, under the US Department of Transportation’s guidance, many state agencies have made significant efforts to incorporate traffic safety into transportation planning as a part of their long-term planning. Among other efforts, research and practice aim to find ways to identify and rank hotspot zones in terms of safety performance. In this study, the excess risk is defined as the ratio of the expected crash rate in one county to the average/expected crash rate in counties with similar characteristics (e.g., population, road density). The weighted average of the excess risk is chosen for the identification of areas with promise:

$$\varepsilon_i^{macro} = \frac{1}{6(C_1 + C_2 + C_3)} \sum_{t=1}^6 \sum_{s=1}^3 C_s er_{ist}, \quad (4.19)$$

$$er_{ist} = \frac{e^{[\beta X_{ist} + \theta_{is} + \varphi_{is} + \nu_{is} + \delta_{is} + \gamma_{it}]}}{e^{[\beta X_{ist}]}} = e^{\theta_{is} + \varphi_{is} + \nu_{is} + \delta_{is} + \gamma_{it}}, \quad (4.20)$$

where ε_i^{macro} is the weighted excess risk, er_{ist} and Obs_{ist} are, respectively, the excess risk and the observed crash risk for different severity levels, s , at time t ; and C_1 , C_2 , and C_3 are the costs/weights

associated with the fatal, injury, and PDO crashes. Counties with an weighted average excess risk higher than 1 are considered high-risk areas.

4.3. Micro-level Data Analysis

4.3.1. Bayesian Logit Model

In contrast to crash frequency models, crash severity models have normally been used to examine the effects of age, gender, road condition, and other factors on a severity outcome, which can be captured in discrete categories. As crash consequence or injury outcome is of a discrete nature, logit or discrete choice models are suitable in estimating a crash severity model.

Most of the crash data used in road safety studies are of a hierarchical nature, that is, the occupants are nested in vehicles, vehicles are within crashes, and crashes are in roadways and regions. Therefore, unobserved factors exist at each level, and these could contribute to an occupant's crash severity. Previous works attempted to model unobserved factors, but they did not account for all levels simultaneously.

Hence, in this section, we aim to fit a Bayesian logit model with various random effects to discrete data. The purposes are to explain the relationship between crash severity levels and the contributing factors of individuals involved in fatal crashes and to identify the distribution of areas with a high relative risk in WV.

Y_{ksit} is assumed to take two response values, 2 and 1, which refer to the fatal injury and non-fatal injury for person k involved in crash s in census tract i in year t , to define the principle of the multilevel logit model. Similar to the Poisson lognormal model, the logit model is a multi-stage model and can be expressed as follows:

$$Y_{ksit} \sim \text{Bernoulli}(P_{ksit}), \quad (4.21)$$

$$\text{Logit}(P_{ksit}) = \log\left[\frac{P_{ksit}}{1-P_{ksit}}\right] = \beta_0 + \beta X_{ksit} + \rho_c + \theta_i + \varphi_i + \delta_t + v_i + \gamma_{it}, \quad (4.22)$$

where P_{kcit} is the probability that person k in crash c in census tract i is killed, X_{kcit} is the deterministic part of the model, β is the vector of coefficients, ρ_c is a crash-level random effect term, and θ_i and φ_i stand for the uncorrelated and correlated heterogeneity, respectively. δ_t and v_t represent unstructured and structured spatial correlation, respectively. Additionally, the spatio-temporal effect, γ_{it} , was included in the model, as interaction is crucial; the trend of spatial heterogeneity may change over time and vice versa. β is assumed to have a normal prior $N(0, 10000)$, and the structured spatial and temporal terms follow the CAR specifications proposed by Besag (1974):

$$[\varphi_i | \varphi_j, i \neq j] \sim N\left(\mu_i, \frac{\sigma_\varphi^2}{\sum_j w_{ij}}\right), \quad (4.23)$$

$$\mu_i = \frac{\sum_j \varphi_j w_{ij}}{\sum_j w_{ij}}, \quad w_{ij}=1 \text{ if } i,j \text{ are adjacent or } 0 \text{ otherwise,} \quad (4.24)$$

where $\tau_\varphi = 1/\sigma_\varphi^2$ controls the variability of φ , and its prior is $\text{Ga}(0.5, 0.0005)$ (Wakefield et al., 2000).

The structured temporal component δ_t is defined using a univariate first-order random-walk prior,

RW(1):

$$\delta_t | \delta_{-t} \sim \begin{cases} N(\delta_{t+1,s}, \sigma_\delta) & \text{for } t = 1 \\ N\left(\frac{\delta_{t-1,s} + \delta_{t+1,s}}{2}, \frac{\sigma_\delta}{2}\right) & \text{for } t = 2, \dots, 5. \\ N(\delta_{t-1,s}, \sigma_\delta) & \text{for } t = 6 \end{cases} \quad (4.25)$$

Normal priors $N(0, \sigma_\rho^2)$, $N(0, \sigma_\theta^2)$, $N(0, \sigma_v^2)$, and $N(0, \sigma_\gamma^2)$ are assumed for ρ , θ , v , and γ , respectively, where $\tau_\rho = 1/\sigma_\rho^2$, $\tau_\theta = 1/\sigma_\theta^2$, $\tau_v = 1/\sigma_v^2$, and $\tau_\gamma = 1/\sigma_\gamma^2$ are the precision parameters that are the inverse of the variances. They follow prior gamma (0.5, 0.0005).

The crash severity outcome can also be considered ordinal in nature (from a less severe, i.e., PDO, to a more severe, i.e., fatal, outcome). Therefore, considering an ordered response model like ordered logit is reasonable, as it has the advantage of making full use of ranked data. The ordered logit model can be derived using an unobserved injury severity level or a latent variable model specified as follows:

$$Y_{kcit}^* = \beta_0 + \beta X_{kcit} + \rho_c + \theta_i + \varphi_i + \delta_t + \nu_t + \gamma_{it}, \quad (4.26)$$

$$Y_{kcit} = \begin{cases} 1 & \text{if } Y_{kcit}^* \leq c \\ 2 & \text{if } c < Y_{kcit}^* \end{cases}, \quad (4.27)$$

where Y_{kcit}^* is a latent variable that measures the crash severity ranging from $-\infty$ to $+\infty$, and c is the threshold to be estimated. Hence, the odds of fatality for an occupant involved in a fatal crash can be estimated as follows:

$$\frac{P_{kcit}}{1 - P_{kcit}} = \exp(\beta_0 + \beta X_{kcit} + \rho_c + \theta_i + \varphi_i + \delta_t + \nu_t + \gamma_{it} - c) \quad (4.28)$$

4.3.1.1. Micro-level Weighted Excess Risk

In micro-level data analysis, the weighted excess risk can be formulated as a ratio of the expected odds of fatality (based on spatial and temporal components) to the average odds of fatality:

$$\varepsilon_i^{micro} = \frac{1}{6} \sum_{t=1}^6 er_{it}, \quad (4.29)$$

$$er_{it} = \frac{\exp(\beta_0 + \beta X_{ksit} + \theta_i + \varphi_i + \delta_t + \nu_t + \gamma_{it})}{\exp(\beta_0 + \beta X_{kcit})} = \exp(\theta_i + \varphi_i + \delta_t + \nu_t + \gamma_{it}), \quad (4.30)$$

This measure, in fact, represents the contribution of spatial and temporal effects to an occupant's odds of fatality. If the weighted excess risk is greater than 1, the occupant's odds of fatality would be higher than the average.

4.4. Marginal Effects

Marginal effects (MEs) are also calculated to enable the interpretation of the results. In the case of the crash frequency model, the ME of a continuous covariate $x_{it}^{(n)}$ is the partial derivative of the expected value of the response variable $E[Y_{ist}]$ with respect to $x_{it}^{(n)}$, and it refers to the expected change in the response variable as a function of the change in $x_{it}^{(n)}$, with the other explanatory variables being fixed. The MEs for this study were computed at the median of the explanatory variables with the following equation:

$$\frac{\partial E[Y_{ist}]}{\partial x_{it}^{(n)}} = \beta_s^{(n)} \cdot E_{it} \exp(\beta_s^{(0)} + \beta_s^{(1)} \cdot x_{it}^{(1)} + \dots + \beta_s^{(n)} \cdot x_{it}^{(n)}) \quad (4.31)$$

Similarly, for the crash severity model, the ME of a continuous variable z_{kit} would be $\frac{\partial P_{ksit}}{\partial z_{kit}}$

. All explanatory variables are held at their mean values to estimate the ME. Additionally, for the dummy variables, the ME is calculated as the discrete change of dummy variables from 0 to 1.

4.5. Model Comparison

The deviance information criterion (DIC) will be used to compare the models. The final model, which has the smallest DIC among candidate models, is of our interest. The DIC is calculated using the following equation (Spiegelhalter et al., 2002):

$$DIC = 2\bar{D} - \hat{D}, \quad (4.32)$$

where

\bar{D} = posterior mean of deviance

$\hat{D} = 2p(y | \psi)$

$\bar{\psi}$ = posterior mean of ψ

(Miaou et al., 2003) proposed a *pseudo-R²* measure that attempted to standardize the DIC as follows:

$$Pseudo R^2 = 1 - \frac{DIC_{model}}{DIC_{null model}}. \quad (4.33)$$

The mean absolute error (MAE) is used as a measure to compare the performance of Bayesian and NB models, and it is as follows:

$$MAE = \frac{|C_{ist} - Obs_{ist}|}{n}. \quad (4.34)$$

The unweighted mean of squares is utilized to assess the goodness of fit for both the Bayesian and classical logit models. It is defined as follows:

$$UMS = \frac{|Y_{kcit} - P_{kcit}|}{N}. \quad (4.35)$$

4.6. Integration of Macro- and Micro-level Data Analyses

The goal of this part of the framework is to combine the results from both the macro- and micro-level data analyses to provide a comprehensive and effective measure for hotspot identification. The advantage of this approach is that it overcomes the limitations of the macroscopic (e.g., it ignores engineering issues on the roadway) and microscopic (e.g., it does not consider system-wide effects) safety analyses because it simultaneously considers the observed and unobserved effects affecting traffic crashes at the macro and micro levels.

Figure 4.1 represents the schematic framework proposed to combine the macro- and micro-level metrics for the development of the combined or integrated metric. This metric can be used to identify areas that can be considered hazardous based on the analysis of both the macro and micro levels. Specifically, the census tracts with high values in this integrated metric are tracts associated

with both higher expected crash costs because of the expected fatal, injury, and PDO crashes and with higher odds of a vehicle occupant involved in a fatal crash being killed (versus surviving).

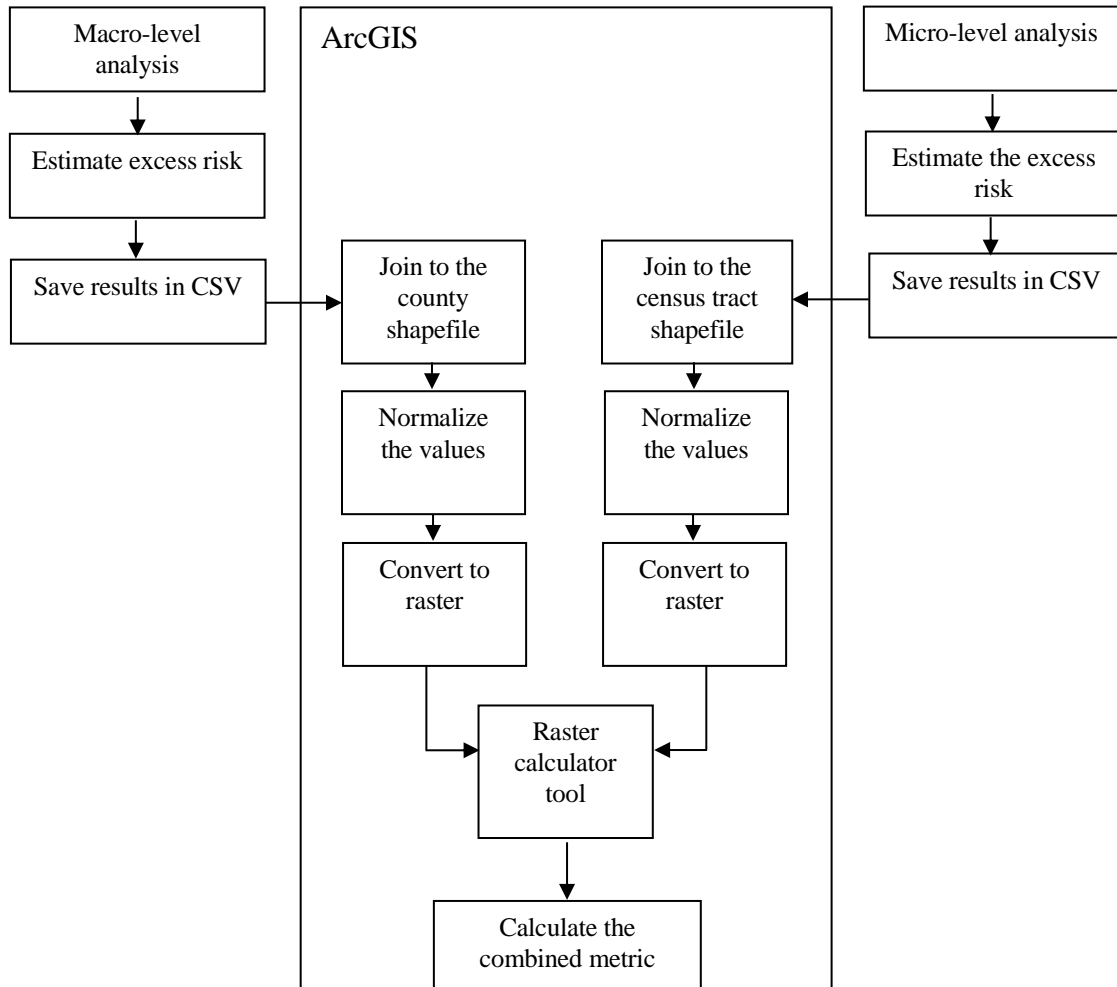


Figure 4.1 Schematic framework for calculating the combined metrics

The following steps are followed to construct the combined metrics:

1. Store the two metrics/measures (the macro- and micro-level weighted excess risks discussed in Sections 4.2.1.2 and 4.3.1.1, respectively) in two separate Excel files in CSV format. Each Excel file has two columns. The first column contains the unique code for each county/census tract, and the second column contains the corresponding weighted excess risk value.

2. Import the excel files into ArcMap and join them to the corresponding shapefiles by using the unique codes associated with each zone.
3. Assign the macro-level excess risk value (ε_1^{macro}) for each county to all of its sub-census tracts, that is, all sub-census tracts in that county will have the same value (refer to Figure 4.2).
4. As the two metrics/measures have different distributions (i.e., their standard deviations follow inverse Wishart and Gamma distribution, as explained in Sections 4.2.1. and 4.3.1, respectively), scaling both the macro- and micro-level metrics between 0 and 1 by using a min–max normalization strategy is necessary.

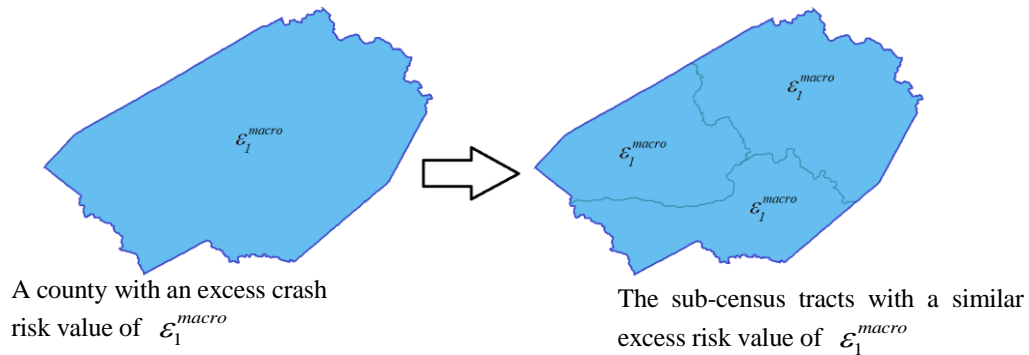


Figure 4.2 Example of assigning the county's weighted excess risk value to its sub-census tracts.

5. Convert the shapefiles to two raster files to represent the normalized weighted excess risk at the macro and micro levels, $N(\varepsilon_i^{macro})$ and $N(\varepsilon_i^{micro})$.

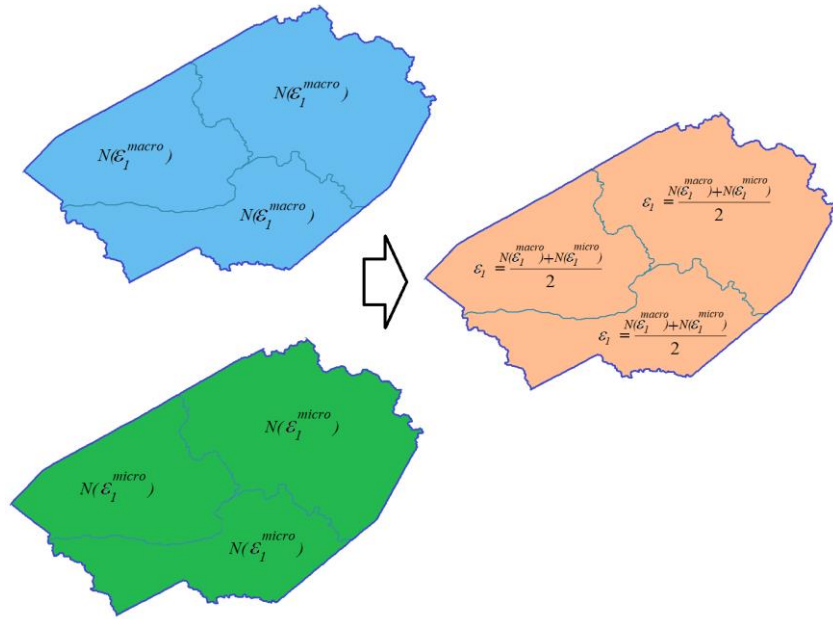


Figure 4.3 Example of estimating the combined metrics for the census tracts in Braxton County.

6. Utilize the *raster calculator* tool to take the average of the normalized values for each census tract. These values correspond to the combined metrics and will be the basis for identifying and ranking the zones with promise (refer to Figure 4.3). The resulting values will then be classified using the natural break (Jenks) classification method, as it reduces the variances within each group and maximizes the variance between groups.

$$N(\varepsilon_i) = \frac{N(\varepsilon_i^{macro}) + N(\varepsilon_i^{micro})}{2} . \quad (4.36)$$

4.7. Summary

This chapter provided a discussion of the methodology to be followed in this dissertation, including statistical methods for macro and micro level data analyses as well as random effect terms that address unobserved factors over space and time. Macro-level analysis aims to establish the relationship between county-level crash frequency and area-wide factors, such as socio-

economic factors and environmental characteristics, whereas micro-level analysis, explores the relationship between crash severity and a range of factors at different levels (occupant-level, crash-level and spatial level). This chapter has also explained various measures to compare Bayesian and classical methods for model validation purpose, including MAE and UMS.

This chapter has also described a novel metric that identifies and prioritizes high-risk areas by capturing the effects of missing influential factors. The proposed metric/measure combines the results from both crash frequency and crash severity models; it has advantages over traditional ranking methods, such as the crash rate and EB methods, because it addresses both system-wide and engineering issues. Therefore, it can be used by transport policy makers to prioritize SWiPs and effectively allocate funding and resources.

CHAPTER 5: RESULTS

5.1. Macro-Level Data Analysis

Transportation agencies are required to incorporate traffic safety into long-term transport planning, which is typically implemented at the aggregate level. Crash models developed at a similar level of aggregation, that is, zonal crash models, can be used to establish the association between observed crashes and a set of variables in each zone to inform such long-term plans and implementations. The aim of this section is to present the results of crash frequency models and examine the macro-level factors affecting county-level traffic crashes, on the basis of the research framework presented in Chapter 4, for the case study of WV. Specifically, this section presents the estimates of various Poisson lognormal models that include various random effect terms. The details of the development of the macro-level models proposed in this dissertation and the empirical results are discussed in this section. In addition, the reliability and the contributions of the proposed methodology are given.

5.1.1. Model Development

Nine models were estimated in this study using the WINBUGS software (Spiegelhalter et al. 2003). The potential explanatory variables to be included in the Bayesian hierarchical models were screened by estimating the variance inflation factor (VIF) to avoid potential misspecification issues. The estimated VIF confirms the absence of multicollinearity in that the values were found to be smaller than 10 for all variables.

As the full models were found to be unstable and some variables did not converge, identifying and eliminating random effect terms that did not significantly contribute to the model were necessary. Inclusion of spatial and temporal correlation terms in the model when they are not significant may deteriorate the precision of crash estimation, so these terms should be dropped from

the model. Different spatial and temporal models were therefore examined to evaluate the significance of spatial and temporal correlation terms by using structured and unstructured spatial and temporal effects.

The first model developed (Model 1) is a non-spatial model that does not account for spatial or temporal effects. The second and third models developed (Models 2 and 3) are full models with structured and unstructured spatial and temporal random effects, as well as a spatio-temporal interaction term. The temporal correlation terms in Model 2 and 3 are assumed to follow MRW(1) and MRW(2) priors, respectively.

Two spatial models (Model 4 and 5) were also considered to assess the significance of the spatial correlation terms. Model 4 contained both unstructured and structured spatial effects, whereas Model 5 included only an unstructured spatial effect term.

Similarly, three temporal models (Model 6, Model 7 and 8) were defined to assess the presence of temporal correlation. Models 6 and 7 contain an unstructured term, along with a structured component with MRW(1) and MRW(2) priors, respectively. However, Model 8 included only an unstructured component. Finally, Model 9 was a spatio-temporal model with both unstructured spatial and temporal components, along with a spatio-temporal interaction term. Table 5.1 summarizes the goodness of fit measures (DIC values and R^2_{DIC}) for all models.

For the models, the first 10,000 iterations were discarded as burn-in, and the 200,000 iterations that followed were used to obtain the summary statistics of the posterior distribution of the parameters. All covariates in the models were standardized (centered around the mean and then divided by the standard deviation) to speed up the convergence. The convergence was evaluated by visual inspection of the Markov chains for the parameters in the history plots. The plots of the final model can be found in the Appendix, Figure 1. Furthermore, the Monte Carlo errors for the

parameters were found to be less than 5% of the posterior standard deviation, which indicates satisfactory convergence. Note that the Monte Carlo error is the standard deviation divided by the number of sampling iterations, and it generally corresponds to an indication of the increasing precision of the estimates as the number of sampling iterations increases.

TABLE 5.1 DIC values associated with the models

No	Model	DIC	R ² _{DIC}
Basic (non-random effect model)			
1	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it}$	21,556	0.54
Full Models			
2	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is} + \nu_{ts} + \delta_{ts, MRW(1)} + \gamma_{it}$	9,838	0.80
3	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is} + \nu_{ts} + \delta_{ts, MRW(2)} + \gamma_{it}$	9,874	0.79
Spatial Models			
4	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \varphi_{is}$	12,119	0.74
5	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is}$	11,993	0.75
Temporal Models			
6	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \nu_{ts} + \delta_{ts, MRW(1)}$	20,543	0.57
7	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \nu_{ts} + \delta_{ts, MRW(2)}$	20,607	0.56
8	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \nu_{ts}$	18,094	0.61
Spatio-Temporal Modes (without structured spatial and temporal terms)			
9	$\log(C_{ist}) = \log(E_{it}) + \beta_{0s} + \beta_s X_{it} + \theta_{is} + \nu_{ts} + \gamma_{it}$	8,487	0.82

From the spatial and temporal models, the models with only unstructured spatial or temporal terms (Model 5 and 8) performed better than their counterparts, suggesting that structured spatial and temporal correlation terms are negligible and can be eliminated from the models. This result confirms that the covariates are less likely to be spatially and temporally correlated, but unobserved heterogeneity resulting from unstructured components seems to be present. This finding further corroborated the preliminary findings from the EDA. Furthermore, the DIC for Model 9 is considerably smaller, which further confirms the existence of unstructured spatial and unstructured temporal effects, as well as their interaction.

In addition to DIC, another goodness of fit measure proposed by Miaou et al., (2003), R^2_{DIC} , was estimated to assess the improvement offered by the fitted models over the intercept model. The results show that R^2_{DIC} increased from 0.54 for Model 1 to 0.82 for Model 9. This result shows the necessity of accounting the random effects in crash analysis. The results of both estimated goodness of fit measures suggest that Model 9 best fits the data for the case study of WV. Figures 1 and 2 in the Appendix show the estimates for the posterior distribution, as well as the density curves, which confirm that the variables are converged for this model. The fixed effect parameters have normal curves, which is consistent with the prior assumption of the distribution of these parameters.

The results of the models are presented in Tables 5.2 to 5.10. As the tables show, after the random effect terms are accounted for, some significant variables (e.g., median age and median travel time to work) in Model 1 are no longer significant in Model 9. An explanation for this is that such variables significantly varied across space and time in a systematic way. Therefore, the use of random effect terms in the model captured their variations used to estimate the fixed effect parameters in the basic model (Model 1). As a result, those variables became insignificant. This result indicates that the estimated parameters in non-random effects may not be efficient and reliable, as they are unable to capture the variability over space and/or time.

Although a model that includes a smaller number of significant variables might seem less useful for planning and practice purposes, it is important to note that the insights that the inclusion of random effects provide are valuable. First, targeting factors for policy and planning interventions based on variables that appear to be significant because of misspecifications will likely not provide the expected results in crash reductions. Second, significant spatial and/or temporal components essentially capture unobserved characteristics that vary across space and time, so their existence highlights the need for the further identification of meaningful factors that can be used for planning

TABLE 5.2 Parameter estimates for Model 1

Crash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.576	-8.654	-8.500	-5.312	-5.327	-5.297	-4.383	-4.393	-4.374
Median travel time to work	0.114	0.026	0.201	-0.047	-0.063	-0.031	-0.094	-0.104	-0.085
Mean travel time to the hospital	0.251	0.162	0.338	N/A			N/A		
Primary road density	0.134	0.056	0.211	0.154	0.140	0.168	0.202	0.194	0.211
Secondary road density	0.046	-0.037	0.127	0.104	0.090	0.118	0.051	0.043	0.060
Intersection density	-0.159	-0.250	-0.067	-0.153	-0.169	-0.136	-0.144	-0.155	-0.133
Labor force	-0.051	-0.146	0.045	-0.075	-0.093	-0.058	-0.147	-0.158	-0.136
Median income	-0.033	-0.125	0.058	0.072	0.054	0.089	0.148	0.137	0.159
Median age	-0.106	-0.207	-0.001	-0.022	-0.045	0.000	-0.019	-0.034	-0.003
Population with a B.S. degree or higher (%)	-0.084	-0.217	0.048	-0.043	-0.068	-0.018	0.058	0.043	0.074
Population under poverty (%)	-0.089	-0.232	0.055	0.076	0.048	0.103	0.094	0.076	0.111
Male population (%)	-0.156	-0.285	-0.027	-0.017	-0.041	0.007	-0.004	-0.020	0.011
Population under 19 (%)	-0.042	-0.148	0.067	-0.001	-0.022	0.019	0.059	0.047	0.072
Population 20 to 24 (%)	-0.042	-0.160	0.076	0.029	0.006	0.051	0.056	0.042	0.070
Population over 65 (%)	-0.083	-0.226	0.062	0.048	0.020	0.076	0.150	0.132	0.168
DUI	-0.034	-0.093	0.024	0.005	-0.005	0.015	0.058	0.053	0.064
School density	0.065	-0.002	0.130	0.086	0.074	0.097	0.135	0.128	0.141
Metro	-0.307	-0.442	-0.172	-0.072	-0.097	-0.047	-0.097	-0.113	-0.082
Mean ground elevation	0.046	-0.035	0.128	-0.004	-0.018	0.010	-0.067	-0.076	-0.058
Number of days with any measurable precipitation	0.048	-0.024	0.121	-0.003	-0.016	0.011	0.023	0.014	0.031
DIC	21556								
R ² _{DIC}	0.54								

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.3 Parameter estimates for Model 2

Crash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-7.270	-9.947	-4.419	-5.173	-5.921	-4.557	-4.615	-4.995	-4.294
Median travel time to work	0.035	-0.185	0.256	-0.106	-0.507	0.144	-0.011	-0.270	0.294
Mean travel time to the hospital	0.274	-0.060	0.560	N/A			N/A		
Primary road density	0.214	0.008	0.629	0.146	-0.213	0.377	0.332	0.163	0.710
Secondary road density	-0.012	-0.296	0.383	-0.074	-0.490	0.293	0.228	0.048	0.493
Intersection density	-0.161	-0.399	0.059	-0.099	-0.389	0.234	-0.173	-0.654	0.137
Labor force (%)	-0.008	-0.316	0.279	0.020	-0.204	0.259	-0.303	-0.577	-0.052
Median income	-0.013	-0.256	0.230	-0.013	-0.256	0.230	0.126	-0.218	0.340
Median age	-0.166	-0.441	0.073	-0.011	-0.323	0.257	0.015	-0.279	0.322
Population with a B.S. degree or higher (%)	-0.131	-0.439	0.161	-0.130	-0.457	0.195	-0.308	-0.543	-0.029
Population under poverty (%)	-0.049	-0.370	0.254	-0.004	-0.290	0.320	-0.295	-0.571	0.099
Male population (%)	-0.151	-0.422	0.183	0.065	-0.289	0.655	-0.159	-0.462	0.095
Population under 19 (%)	-0.013	-0.249	0.228	-0.025	-0.284	0.218	-0.060	-0.368	0.271
Population 20 to 24 (%)	-0.120	-0.484	0.169	0.158	-0.166	0.599	0.184	-0.048	0.459
Population over 65 (%)	-0.105	-0.511	0.250	0.104	-0.373	0.574	-0.113	-0.410	0.189
DUI	-0.019	-0.100	0.061	-0.007	-0.021	0.007	0.008	-0.001	0.016
School density	0.066	-0.097	0.248	0.008	-0.304	0.186	0.200	0.012	0.397
Metro	-0.256	-0.784	0.257	-0.299	-1.075	0.229	0.044	-0.349	0.418
Mean ground elevation	0.047	-0.227	0.268	0.009	-0.262	0.313	0.020	-0.247	0.301
Number of days with any measurable precipitation	0.048	-0.121	0.224	-0.004	-0.103	0.093	-0.001	-0.138	0.140
σ_v	2.8			1.8			0.28		
σ_δ	0.33			0.35			0.37		
σ_θ	0.33			0.52			0.46		
σ_ϕ	0.82			0.17			0.13		
σ_γ	0.2			1.73			0.29		
DIC	9,838								

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.4 Parameter estimates for Model 3

Cash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.028	-12.870	-4.817	-5.181	-5.864	-4.523	-4.370	-4.620	-4.035
Median travel time to work	0.018	-0.260	0.220	-0.047	-0.408	0.225	-0.076	-0.263	0.121
Mean travel time to the hospital	0.284	0.091	0.494	N/A			N/A		
Primary road density	0.183	-0.018	0.445	0.276	0.080	0.471	0.331	0.153	0.522
Secondary road density	-0.042	-0.275	0.159	0.075	-0.273	0.328	0.105	-0.088	0.286
Intersection density	-0.154	-0.392	0.068	-0.160	-0.687	0.161	-0.129	-0.348	0.061
Labor force (%)	-0.022	-0.262	0.260	-0.108	-0.447	0.258	-0.276	-0.563	-0.090
Median income	0.004	-0.210	0.218	0.004	-0.210	0.218	0.155	-0.012	0.331
Median age	-0.140	-0.397	0.105	-0.044	-0.427	0.227	-0.127	-0.428	0.066
Population with a B.S. degree or higher (%)	-0.136	-0.464	0.158	-0.338	-0.832	-0.070	-0.170	-0.390	0.066
Population under poverty (%)	-0.055	-0.353	0.248	-0.222	-0.615	0.141	-0.195	-0.441	0.038
Male population (%)	-0.127	-0.431	0.321	-0.075	-0.575	0.460	-0.173	-0.566	0.185
Population under 19 (%)	-0.018	-0.252	0.237	-0.082	-0.512	0.178	-0.080	-0.349	0.179
Population 20 to 24 (%)	-0.071	-0.337	0.199	0.108	-0.192	0.329	0.079	-0.164	0.293
Population over 65 (%)	-0.066	-0.396	0.322	-0.079	-0.541	0.422	-0.029	-0.225	0.183
DUI	-0.022	-0.102	0.056	-0.007	-0.021	0.007	0.008	-0.001	0.016
School density	0.079	-0.085	0.336	0.089	-0.115	0.465	0.088	-0.140	0.230
Metro	-0.252	-0.600	0.099	-0.117	-0.552	0.545	-0.135	-0.533	0.178
Mean ground elevation	0.071	-0.149	0.329	0.133	-0.111	0.510	-0.006	-0.233	0.207
Number of days with any measurable precipitation	0.051	-0.113	0.232	0.002	-0.095	0.098	0.001	-0.130	0.145
σ_v	2.95			1.74			0.61		
σ_δ	0.11			0.13			0.04		
σ_θ	0.25			0.26			0.27		
σ_ϕ	0.09			0.07			0.06		
σ_γ	0.21			0.29			0.23		
DIC	10								

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.5 Parameter estimates for Model 4

Crash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.583	-8.718	-8.432	-5.173	-5.379	-4.966	-4.328	-4.478	-3.854
Median travel time to work	0.078	-0.077	0.232	-0.027	-0.158	0.103	-0.032	-0.247	0.139
Mean travel time to the hospital	0.282	0.130	0.429	N/A			N/A		
Primary road density	0.170	0.029	0.319	0.160	0.030	0.289	0.262	0.078	0.447
Secondary road density	0.009	-0.141	0.158	0.111	-0.005	0.228	0.100	-0.080	0.281
Intersection density	-0.155	-0.321	0.011	-0.095	-0.230	0.040	-0.086	-0.284	0.111
Labor force	-0.027	-0.201	0.148	-0.044	-0.183	0.095	-0.159	-0.354	0.036
Median income	0.003	-0.156	0.161	0.014	-0.108	0.136	0.099	-0.064	0.262
Median age	-0.108	-0.287	0.071	0.002	-0.133	0.136	-0.043	-0.224	0.139
Population with a B.S. degree or higher (%)	-0.107	-0.336	0.122	-0.111	-0.356	0.135	-0.125	-0.490	0.240
Population under poverty (%)	-0.071	-0.314	0.172	-0.062	-0.296	0.171	-0.184	-0.512	0.145
Male population (%)	-0.151	-0.349	0.047	-0.050	-0.204	0.104	-0.061	-0.263	0.142
Population under 19 (%)	-0.025	-0.203	0.154	-0.019	-0.161	0.122	0.001	-0.188	0.189
Population 20 to 24 (%)	-0.063	-0.273	0.147	0.060	-0.113	0.233	0.121	-0.120	0.361
Population over 65 (%)	-0.079	-0.330	0.173	-0.081	-0.278	0.116	-0.013	-0.276	0.251
DUI	-0.004	-0.080	0.072	-0.041	-0.056	-0.026	0.025	0.015	0.035
School density	0.072	-0.046	0.189	0.068	-0.027	0.164	0.128	0.001	0.254
Metro	-0.291	-0.540	-0.043	-0.121	-0.346	0.103	-0.222	-0.551	0.106
Mean ground elevation	0.041	-0.119	0.201	0.058	-0.086	0.201	0.057	-0.130	0.244
Number of days with any measurable precipitation	0.061	-0.062	0.183	-0.014	-0.111	0.084	-0.008	-0.146	0.129
σ_θ	0.28			0.23			0.36		
σ_ϕ	0.17			0.19			0.19		
Fraction	0.62			0.55			0.65		
DIC	12417								
R^2_{DIC}	0.67								

Notes: Shaded cells indicate significant coefficients at $\alpha=5\%$ level. Bolded text indicates significant coefficients at $\alpha=10\%$ level.

TABLE 5.6 Parameter estimates for Model 5

Crash model	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.587	-8.724	-8.435	-5.272	-5.386	-4.974	-4.330	-4.488	-3.857
Median travel time to work	0.076	-0.081	0.226	-0.020	-0.168	0.097	-0.036	-0.252	0.134
Mean travel time to the hospital	0.272	0.121	0.426	N/A			N/A		
Primary road density	0.166	0.021	0.313	0.163	0.029	0.284	0.264	0.072	0.445
Secondary road density	0.005	-0.148	0.154	0.116	-0.003	0.241	0.086	-0.089	0.275
Intersection density	-0.162	-0.327	0.001	-0.100	-0.231	0.035	-0.090	-0.286	0.102
Labor force	-0.037	-0.210	0.142	-0.047	-0.184	0.094	-0.174	-0.363	0.030
Median income	0.000	-0.158	0.160	0.013	-0.109	0.134	0.092	-0.068	0.259
Median age	-0.114	-0.294	0.064	0.001	-0.134	0.130	-0.045	-0.229	0.131
Population with a B.S. degree or higher (%)	-0.113	-0.345	0.119	-0.137	-0.323	0.142	-0.128	-0.393	0.250
Population under poverty (%)	-0.079	-0.319	0.165	-0.077	-0.270	0.161	-0.192	-0.450	0.157
Male population (%)	-0.157	-0.354	0.037	-0.068	-0.207	0.103	-0.074	-0.272	0.169
Population under 19 (%)	-0.028	-0.205	0.149	-0.029	-0.165	0.115	-0.010	-0.189	0.188
Population 20 to 24 (%)	-0.062	-0.276	0.145	0.070	-0.121	0.230	0.141	-0.124	0.351
Population over 65 (%)	-0.084	-0.340	0.168	-0.101	-0.281	0.111	-0.039	-0.280	0.247
DUI	-0.008	-0.082	0.065	-0.043	-0.057	-0.029	0.024	0.016	0.032
School density	0.068	-0.049	0.183	0.067	-0.034	0.162	0.124	-0.016	0.253
Metro	-0.295	-0.547	-0.051	-0.115	-0.349	0.100	-0.199	-0.558	0.096
Mean ground elevation	0.038	-0.124	0.197	0.064	-0.093	0.192	0.060	-0.179	0.238
Number of days with any measurable precipitation	0.052	-0.071	0.174	-0.013	-0.118	0.080	-0.005	-0.154	0.119
σ_0	0.26			0.26			0.35		
DIC	11993								
R ² _{DIC}	0.75								

Notes: Shaded cells indicate significant coefficients at $\alpha=5\%$ level. Bolded text indicates significant coefficients at $\alpha=10\%$ level.

TABLE 5.7 Parameter estimates for Model 6

Crash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.663	-10.250	-7.524	-5.325	-5.614	-5.027	-4.284	-4.541	-4.006
Median travel time to work	0.090	0.008	0.173	-0.048	-0.063	-0.033	-0.102	-0.112	-0.093
Mean travel time to the hospital	0.238	0.151	0.325	N/A			N/A		
Primary road density	0.127	0.050	0.205	0.153	0.139	0.166	0.201	0.192	0.209
Secondary road density	0.032	-0.049	0.113	0.101	0.087	0.114	0.048	0.040	0.056
Intersection density	-0.144	-0.233	-0.055	-0.147	-0.163	-0.130	-0.142	-0.152	-0.131
Labor force (%)	-0.050	-0.147	0.046	-0.078	-0.095	-0.060	-0.148	-0.159	-0.138
Median income	-0.020	-0.110	0.069	-0.020	-0.110	0.069	0.154	0.143	0.164
Median age	-0.105	-0.206	-0.001	-0.025	-0.047	-0.003	-0.016	-0.031	-0.001
Population with a B.S. degree or higher (%)	-0.071	-0.201	0.061	-0.046	-0.071	-0.022	0.066	0.051	0.081
Population under poverty (%)	-0.054	-0.192	0.084	0.072	0.046	0.098	0.108	0.092	0.124
Male population (%)	-0.128	-0.247	-0.008	-0.012	-0.035	0.011	0.007	-0.007	0.022
Population under 19 (%)	-0.035	-0.142	0.072	0.006	-0.014	0.027	0.056	0.043	0.068
Population 20 to 24 (%)	-0.050	-0.168	0.068	0.031	0.008	0.053	0.051	0.036	0.065
Population over 65 (%)	-0.073	-0.211	0.068	0.063	0.035	0.090	0.148	0.132	0.165
DUI	-0.048	-0.107	0.011	0.026	0.016	0.036	0.050	0.044	0.056
School density	0.055	-0.012	0.120	0.086	0.075	0.098	0.131	0.125	0.138
Metro	-0.289	-0.424	-0.156	-0.072	-0.097	-0.048	-0.095	-0.110	-0.080
Mean ground elevation	0.049	-0.032	0.130	-0.008	-0.022	0.007	-0.066	-0.075	-0.058
Number of days with any measurable precipitation	0.048	-0.115	0.233	-0.002	-0.098	0.095	0.004	-0.134	0.148
σ_v	1.12			1.34			1.91		
σ_δ	0.16			0.12			0.33		
DIC	20,543								

Notes: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.8 Parameter estimates for Model 7

Crash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-11.810	-17.320	-8.208	-5.267	-6.022	-4.520	-4.181	-4.623	-3.765
Median travel time to work	0.091	0.009	0.172	-0.048	-0.063	-0.033	-0.102	-0.111	-0.093
Mean travel time to the hospital	0.238	0.151	0.325	N/A			N/A		
Primary road density	0.127	0.050	0.206	0.152	0.138	0.167	0.200	0.192	0.209
Secondary road density	0.032	-0.047	0.113	0.101	0.087	0.114	0.048	0.040	0.056
Intersection density	-0.145	-0.235	-0.055	-0.147	-0.163	-0.130	-0.142	-0.152	-0.131
Labor force (%)	-0.051	-0.148	0.045	-0.078	-0.095	-0.060	-0.148	-0.159	-0.138
Median income	-0.021	-0.111	0.070	-0.021	-0.111	0.070	0.154	0.143	0.164
Median age	-0.105	-0.206	-0.002	-0.025	-0.048	-0.003	-0.016	-0.031	0.000
Population with a B.S. degree or higher (%)	-0.071	-0.202	0.060	-0.046	-0.071	-0.021	0.066	0.050	0.081
Population under poverty (%)	-0.054	-0.191	0.079	0.072	0.046	0.099	0.108	0.092	0.124
Male population (%)	-0.128	-0.249	-0.010	-0.012	-0.035	0.011	0.007	-0.007	0.022
Population under 19 (%)	-0.036	-0.143	0.072	0.006	-0.014	0.027	0.056	0.043	0.069
Population 20 to 24 (%)	-0.050	-0.169	0.066	0.030	0.008	0.054	0.051	0.036	0.065
Population over 65 (%)	-0.073	-0.216	0.070	0.063	0.036	0.090	0.148	0.132	0.166
DUI	-0.048	-0.107	0.010	0.026	0.016	0.036	0.050	0.044	0.056
School density	0.055	-0.011	0.120	0.086	0.075	0.098	0.131	0.125	0.138
Metro	-0.290	-0.427	-0.155	-0.072	-0.097	-0.048	-0.095	-0.111	-0.080
Mean ground elevation	0.049	-0.034	0.130	-0.008	-0.022	0.007	-0.066	-0.075	-0.058
Number of days with any measurable precipitation	0.051	-0.111	0.233	-0.001	-0.098	0.096	0.006	-0.131	0.148
σ_v	1.28			1.74			1.79		
σ_δ	0.22			0.24			0.12		
DIC	20607								

Notes: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.9 Parameter estimates for Model 8

Crash model Parameter	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.703	-8.990	-8.403	-4.101	-5.010	-3.270	-3.484	-3.877	-3.044
Median travel time to work	0.091	0.008	0.173	-0.048	-0.063	-0.033	-0.102	-0.112	-0.093
Mean travel time to the hospital	0.239	0.151	0.327	N/A			N/A		
Primary road density	0.127	0.049	0.206	0.153	0.139	0.166	0.201	0.192	0.209
Secondary road density	0.032	-0.048	0.111	0.101	0.087	0.114	0.048	0.040	0.056
Intersection density	-0.144	-0.234	-0.056	-0.147	-0.164	-0.130	-0.142	-0.153	-0.132
Labor force (%)	-0.054	-0.151	0.043	-0.078	-0.095	-0.060	-0.148	-0.159	-0.138
Median income	-0.020	-0.110	0.070	-0.020	-0.110	0.070	0.154	0.143	0.165
Median age	-0.106	-0.209	-0.001	-0.025	-0.048	-0.002	-0.016	-0.031	0.000
Population with a B.S. degree or higher (%)	-0.072	-0.207	0.058	-0.046	-0.071	-0.022	0.066	0.050	0.080
Population under poverty (%)	-0.056	-0.196	0.083	0.072	0.046	0.098	0.108	0.092	0.124
Male population (%)	-0.128	-0.248	-0.006	-0.012	-0.035	0.010	0.007	-0.007	0.022
Population under 19 (%)	-0.035	-0.142	0.075	0.006	-0.015	0.027	0.056	0.043	0.069
Population 20 to 24 (%)	-0.050	-0.170	0.070	0.031	0.009	0.053	0.051	0.037	0.065
Population over 65 (%)	-0.070	-0.213	0.075	0.063	0.034	0.091	0.148	0.131	0.166
DUI	-0.041	-0.101	0.018	0.026	0.016	0.036	0.050	0.043	0.056
School density	0.056	-0.012	0.120	0.086	0.075	0.098	0.131	0.125	0.138
Metro	-0.290	-0.428	-0.151	-0.073	-0.098	-0.048	-0.095	-0.111	-0.080
Mean ground elevation	0.048	-0.035	0.129	-0.008	-0.022	0.007	-0.067	-0.076	-0.058
Number of days with any measurable precipitation	0.044	-0.113	0.229	0.001	-0.093	0.101	0.008	-0.129	0.141
σ_v	1.21			1.49			1.03		
DIC	18,094								

Notes: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.10 Parameter estimates for Model 9

Crash model	Fatal			Injury			PDO		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Intercept	-8.639	-8.860	-8.419	-5.362	-5.640	-5.085	-4.395	-4.598	-4.199
Median travel time to work	0.049	-0.113	0.206	-0.013	-0.131	0.100	-0.021	-0.170	0.133
Mean travel time to the hospital	0.255	0.092	0.418	N/A			N/A		
Primary road density	0.160	0.002	0.320	0.184	0.064	0.309	0.304	0.163	0.476
Secondary road density	-0.021	-0.183	0.139	0.075	-0.042	0.197	0.061	-0.097	0.233
Intersection density	-0.142	-0.311	0.028	-0.136	-0.257	-0.018	-0.126	-0.280	0.028
Labor force	-0.040	-0.231	0.150	-0.074	-0.215	0.068	-0.200	-0.378	-0.006
Median income	0.017	-0.154	0.190	0.042	-0.082	0.164	0.108	-0.060	0.267
Median age	-0.111	-0.306	0.085	0.020	-0.119	0.167	-0.023	-0.204	0.162
Population with a B.S. degree or higher (%)	-0.113	-0.371	0.141	-0.148	-0.351	0.042	-0.160	-0.410	0.084
Population under poverty (%)	-0.049	-0.304	0.208	-0.073	-0.251	0.109	-0.212	-0.433	0.007
Male population (%)	-0.136	-0.334	0.062	-0.026	-0.152	0.093	-0.061	-0.225	0.086
Population under 19(%)	-0.013	-0.197	0.175	-0.001	-0.130	0.123	-0.003	-0.176	0.159
Population 20 to 24 (%)	-0.060	-0.290	0.169	0.099	-0.078	0.268	0.187	-0.045	0.408
Population over 65 (%)	-0.071	-0.348	0.205	-0.024	-0.213	0.161	-0.007	-0.244	0.221
DUI	-0.019	-0.095	0.057	-0.007	-0.021	0.007	0.008	-0.001	0.016
School density	0.060	-0.066	0.186	0.056	-0.041	0.153	0.121	-0.011	0.253
Metro	-0.245	-0.510	0.025	-0.059	-0.251	0.135	-0.145	-0.388	0.115
Mean ground elevation	0.054	-0.118	0.228	0.043	-0.081	0.170	0.055	-0.108	0.216
Number of days with any measurable precipitation	0.051	-0.079	0.178	-0.005	-0.103	0.092	-0.001	-0.14	0.136
σ_θ	0.21			0.19			0.25		
σ_ν	0.21			0.33			0.17		
σ_γ	0.21			0.20			0.28		
Fraction: $\sigma_\theta / (\sigma_\theta + \sigma_\nu + \sigma_\gamma)$	0.33			0.26			0.36		
DIC	8,487								

NOTES: Shaded cells with bold text indicate significant coefficients at $\alpha=5\%$ level. Bolded text only indicates significant coefficients at $\alpha=10\%$ level.

and policy purposes. To determine such potential factors, one can turn to the findings of previous literature and explore the factors that can, in principle, have a spatial and/or temporal component. For example, in previous works (Wang and Kockelman 2013; Lee et al. 2015), many spatial-related factors, such as land use, bus stop density, side walk density, and number of traffic signals, were found to contribute to traffic crashes. These factors are not routinely collected in WV. For the purpose of crash prediction, the random effect terms suggested in these models can be used to account for the effects of those omitted factors. However, for planning and policy purposes, the findings of the case study suggest a significant spatial random effect, so collecting and exploring the effects of similar factors in crashes in WV would be advisable.

5.1.2. Model Validation

Prior to the exploration of the model parameters, confirming that the model is an appropriate fit to the data is prudent. One possible model validation procedure is to examine how well the preferred Bayesian model (Model 9) fits the data compared with classical count models. In this section, a set of the most suitable classical count models is estimated for the fatal, injury, and PDO crashes, and they are compared with the preferred Bayesian model.

Classical count methods, such as the NB, Poisson, and zero-inflated models, have been popular in crash frequency modeling. In general, the choice of the model depends on the nature of the crash data. For example, crash data tend to be overdispersed, that is, the variance exceeds the mean (Karlaftis and Golias, 2002). As the Poisson model assumes equality of the mean and variance, using a Poisson model when the data are overdispersed can underestimate the variance of the parameter estimates (Abdel-Aty and Radwan, 2000). Hence, many safety practitioners have used the negative binomial model in lieu of the Poisson model to account for the over-dispersion issue. An NB model can be reduced to a Poisson model when overdispersion is not present in the model.

It should be noted that in some cases, overdispersion might be caused by the presence of excess zeros in a dataset, which cannot be handled properly by neither a Poisson nor an NB model. In these cases, the zero-inflated model should be used to model count data. The zero-inflated model involves two zero-generating processes: one for the binary response model that generates structural zeros and one for the frequency model that generates counts and is conditioned on the binary response (Farewell et al. 2017). Mamun (2014) recommended the use of the zero-inflated model when more than 30% of the counts are zeros. The histograms of the dependent variables are first studied to explore the suitability of a Poisson, NB, or zero-inflated model (see Figure 5.1).

In this case study, the distribution of fatal, injury, and PDO crash data displays signs of overdispersion, that is, the variances are higher than the means (see Figure 5.1). However, the injury and PDO crash data do not contain any zero values, and the fatal crash data contain a small portion of zeros (around 8%). Hence, three separate NB models for each crash severity outcome are independently estimated to compare the performance of the proposed full Bayesian model (Model 9) and the classical model. The parameter estimates of the models can be found in Table 5.11.

As Table 5.11 shows, the estimated log likelihood ratio test of $\alpha=0$ for the fatal, injury, and PDO models is associated with a p-value of 0.018, a p-value smaller than 0.0001, and a p-value smaller than 0.0001, respectively, suggesting that for all cases, overdispersion is present and that NB models should be estimated.

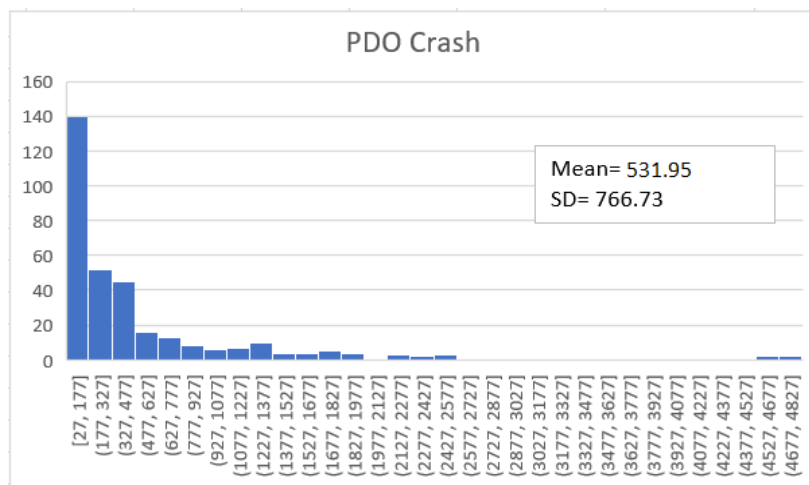
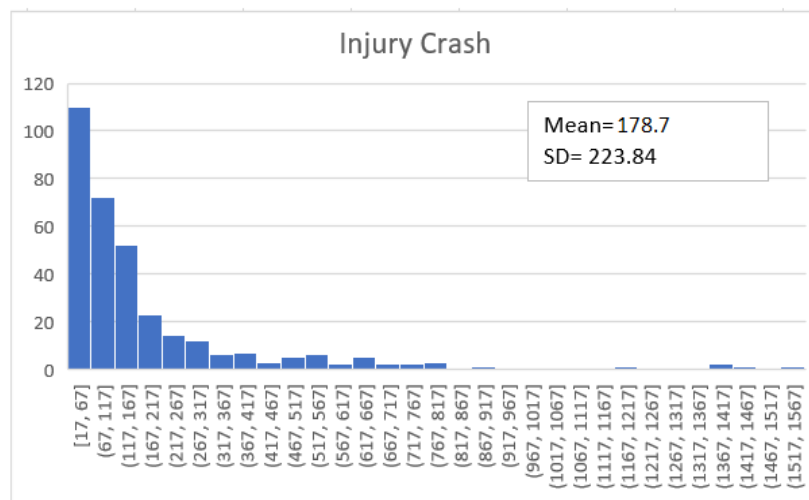
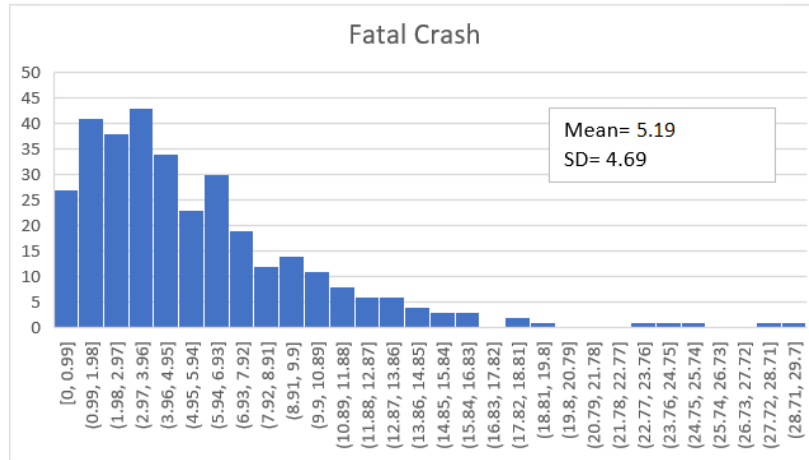


Figure 5.1 Distribution of traffic crashes by severity

TABLE 5.11 Negative binomial models' parameter estimates

	Fatal		Injury		PDO	
	Coef.	p-value	Coef.	p-value	Coef.	p-value
Intercept	-8.672	<0.0001	-2.440	<0.0001	-2.160	<0.0001
Median travel time to work	0.059	0.082	-0.002	0.928	-0.055	0.021
Mean travel time to the hospital	0.184	<0.0001	NA	NA	NA	NA
Primary road density	0.128	<0.0001	0.133	<0.0001	0.201	<0.0001
Secondary road density	0.039	0.292	0.052	0.024	0.027	0.311
Intersection density	-0.159	<0.0001	-0.067	0.012	-0.044	0.143
Labor force (%)	-0.038	0.300	-0.043	0.058	-0.082	0.002
Median income	-0.017	0.660	-0.014	0.554	0.057	0.036
Median age	-0.049	0.135	-0.010	0.639	-0.016	0.503
Population with a B.S. degree or higher	-0.058	0.175	-0.039	0.166	0.009	0.775
Population under poverty (%)	-0.047	0.202	0.022	0.350	0.028	0.311
Male population (%)	-0.031	0.415	-0.016	0.486	0.014	0.556
Population under 19 (%)	-0.026	0.394	0.012	0.528	0.007	0.728
Population 20 to 24 (%)	-0.033	0.332	0.023	0.286	0.032	0.170
Population over 65 (%)	0.007	0.832	0.003	0.892	0.032	0.208
DUI	-0.048	0.128	0.056	0.003	0.111	<0.0001
School density	0.042	0.217	0.078	<0.0001	0.140	<0.0001
Metro	-0.237	0.001	-0.043	0.338	-0.119	0.022
Mean ground elevation	0.0450	0.871	0.0480	0.061	-0.04	0.164
Number of days with any measurable precipitation	0.370	0.212	-0.001	0.176	0.019	0.331
Alpha	0.031	<0.0001	0.087	<0.0001	0.115	<0.0001
Log likelihood ratio test of alpha=0	$\chi^2=4.370$ (p-value=0.018)		$\chi^2=4,402.900$ (p-value<0.0001)		$\chi^2=11,000.000$ (p-value<0.0001)	
Log likelihood	-711		-1641		-1976	

With regard to the comparison of the Bayesian model with the classical models, the MAE is used as a measure to demonstrate how well the models fit the data. Table 5.12 shows that for the fatal model, the full Bayesian approach performs slightly better than the NB model. However, for the injury and the PDO models, the predictive ability of the Bayesian approach is markedly stronger than that of the NB model. This result illustrates the necessity of accounting for the variation caused

by random effects in crash severity estimation, in addition to the risk factors, especially for a predictive model.

TABLE 5.12 Comparison of Bayesian and NB models

Model	MAE
Bayesian fatal model (Model 9)	1.71
NB fatal model	1.74
Bayesian injury model (Model 9)	21.52
NB injury model	43
Bayesian PDO model (Model 9)	67.76
NB PDO model	113

Although the inclusion of random effect terms in the Bayesian model led to a more precise crash estimation compared with the NB model (as the MAE revealed), such an improvement could also be due to more precise parameter estimates in the Bayesian model (as it incorporates prior information with data) compared with the NB model. Thus, comparing the significance and magnitude of the parameters in both models is also critical. Additionally, for future directions in the transport planning field and to help researchers better decide between the Bayesian model's high predictive power and the classical model's low runtime, knowing to what extent the parameter estimates for low (fatal crashes) and high (injury and PDO crashes) frequency counts differ in both Bayesian and classical models is valuable.

The parameter estimates for the Bayesian and classical fatal models show that the variables that are significant in the Bayesian models are also significant in the NB models and that the signs and magnitudes of the coefficients are consistent. However, in the injury model, the variables secondary roadway density, DUI, and school density were only significant in the NB model but not in the Bayesian model (Model 9). Finally, in the PDO model, the variables median travel time to work and median income were only significant in the NB model but not in the Bayesian model (Model 9). This result could be due to differences in the nature of the Bayesian and classical models.

Unlike the NB model, which that assumes that there is one true fixed parameter estimate within a confidence interval whose bounds are fixed, the Bayesian model provides a random parameter within a credible interval whose bounds are fixed and is based on prior distribution (Van de Schoot et al. 2014). Because of this fact, the credible interval and confidence interval may not coincide; as a result, some variables that are found significant using one approach may not be found significant in the other.

Along the same lines, the magnitude of parameter estimates for fatal models in both Bayesian and NB models is very close, but this does not apply for the injury and PDO models. This difference can be explained by the fact that compared with the NB model, the Bayesian model provides more precise parameter estimates for high-frequency crashes (e.g., injury and PDO). The fact that the results of the fatal model are closer is expected because the difference between MAEs associated with the Bayesian and NB fatal models is negligible, whereas for injury and PDO crashes, the difference is significant (see Table 5.11).

5.1.3. Interpretation of Parameter Estimates

As discussed earlier, previous studies that have utilized a full Bayesian framework similar to that proposed in this dissertation (see for example, Boulieri et al., 2016; Liu and Sharma, 2018) have explored few covariates. Furthermore, from the ones explored, few were found to be significant in their models. For example, in the work of Liu and Sharma (2018), only vehicle miles traveled, which is essentially an exposure variable, was found to affect the number of fatal crashes. This dissertation examines several potential macro-level factors that may affect crash severities to address such a problem. The effects of some of these factors, such as road and intersection density (Aguero-Valverde and Jovanis, 2006; Lee et al. 2015), as well as income and climate conditions (Liu and Sharma, 2018), have been evaluated in previous works but were found insignificant. Nevertheless,

it should be noted that their impacts on traffic crashes may vary from one state to another because of differences in road design criteria, travel and pedestrian behaviors, and other underlying factors. Hence, the use of such variables in this research is critical. In addition to the factors that emerged from previous studies, some important variables that have not been investigated before, such as access time to the hospital and urbanity/rurality, are considered in this study.

With regard to the case study results, Table 5.10 shows that primary roadway density is positively correlated with all crash severities, which is consistent with the findings of previous studies (Aguero-Valverde and Jovanis, 2006a; Lee et al., 2015). This result might be due to the higher traffic speed and speed limit in primary roadways than in local roadways, thus increasing the risk of traffic crashes. Likewise, Wang and Kockelman (2013) reported that the freeway and arterial density are positively associated with severe and non-severe pedestrian crashes.

Intersection density is found to be negatively associated with injury crashes. This result means that an increase in intersection density will likely decrease the injury crashes in every county. This finding is not consistent with that of Guevara et al. (2004), who explored crashes in Tucson, Arizona. Two main reasons explain this contradiction. First, driver and pedestrian behaviors might vary from state to state. Second, intersection type and the portion of signalized and stopped control intersections in WV may be different from those in other states, such as Arizona. In principle, an increase in intersection density in WV could potentially be associated with fewer injury crashes; people may have to drive with caution and at a lower traffic speed near intersections, as WV is among top five states with strict penalties on reckless drivers (Comoreanu, 2018).

Labor force was only significant in PDO models. Its negative sign indicates that areas with higher labor force participation rates may have fewer PDO crashes. With the assumption that areas with a higher labor force participation rate are likely to be more congested because of more

commuter trips, this finding is consistent with the literature suggesting that congested roads can be safer because drivers tend to drive with greater caution (Zhou and Sisiopiku, 1997).

School density and DUI rates are the factors that were insignificant in the literature, but they are significant in the present study. School density is positively associated with PDO crashes, whereas it is not correlated with fatal and injury models. This result is possibly caused by the increase in traffic volumes and the lower speed limit near school zones. Furthermore, the positive sign of DUI suggests that an increase in DUI rates will likely increase PDO crashes. The reason is that alcohol and drug impairment can worsen a driver's ability to control a vehicle, thus increasing the risk of PDO crashes in WV.

Lastly, mean travel time to the hospital and urbanity/rurality are investigated in this study only. The mean travel time to the hospital was only evaluated in the fatal crash model because severe injuries may become fatal as the travel time to the hospital increases, and it was found to be significant. The positive coefficient of this variable implies that counties with lower accessibility to the hospital will likely have more fatal crashes. Furthermore, the negative coefficient of metro implies that counties in metro areas have a lower risk of fatal crashes, possibly caused by the reduction in traffic speed as a result of congestion.

Generally, in the interpretation of a regression model, the MEs often provide more information than the coefficients do. Table 5.13 presents the MEs of significant variables on crash counts by severity on the basis of Model 9.

As standardized covariates were used in the model, the interpretation of the ME results would be as follows. An increase in the mean travel time to the hospital by one standard deviation (which is 15.36 minutes, as shown in Table 3.1) will likely increase the number of fatal crashes in every county by 1.38. A one standard deviation increase in primary roadway density (s.d.=0.09 miles per

sq.) is associated with an increase of 0.85, 14.8, and 38.01, on average, in the number of fatal, injury, and PDO crashes, respectively. Furthermore, a one standard deviation increase in intersection density (s.d.=32.8 intersections/county area) and labor force participation rate (s.d.=6.6%) will likely reduce the number of injury and PDO crashes by 9.04 and 25.79, respectively. An increase in DUI rate (s.d.=29.30) and school density (s.d.=0.039) will likely change the number of PDO crashes by 1.36 and 17.65, respectively. Lastly, with respect to the estimated ME for metro, on average, the number of fatal crashes is 1.32 greater in rural counties than in urban counties.

TABLE 5.13 Marginal effects of significant covariates in Model 9

Parameter	Fatal	Injury	PDO
Mean travel time to hospital	1.38	-	-
Primary road density	0.85	14.8	38.01
Intersection density	-	-9.04	-
Labor force	-	-	-25.79
DUI	-	-	1.36
School density	-	-	17.65
Metro (1 if urban, 0 otherwise)	-1.32	-	-

Note: For the calculation of MEs, explanatory variables are set equal to their medians in the dataset.

The extent to which these factors can reduce or increase the number of traffic crashes could give some insights to policy makers into prioritizing engineering countermeasures or any corrective actions by performing cost-benefit analyses. For example, various medical clinics in WV do not have any trauma centers and thus cannot offer emergency services to crash victims. If setting up a level-2 trauma center in a county reduces the average access time by 15 minutes, it will likely reduce the number of fatal crashes by 1.38. As a comparison, Ashley et al., (2017) stated that the average readiness cost of level-2 trauma centers in Georgia in 2010 was US\$2,333,113, whereas that of a fatal injury in WV in 2012 was US\$5,289,928 (Harmon et al., 2018). Readiness costs are associated with administrative, clinical medical staff, operating room, and education/outreach costs, which are

involved with maintaining the essential infrastructure that can facilitate emergent services on a continuous basis.

5.1.4. Interpretation of Random Effect Terms

Generally, the development of a model with random effect terms would be beneficial, as random effect terms can capture the impacts of unmeasured factors, such as economic condition and travel behavior, on traffic crashes. Additionally, they can represent some other factors (number of traffic signals, housing units per acre, number of workers who commute by walking) that were found to affect traffic crashes in previous work (Lee et al. 2015), whereas these factors cannot be measured or are not readily available in WV.

This section aims to discuss the spatial and temporal random effect terms used in Model 9. As Table 5.10 shows, the fractions of random effects for fatal, injury, and PDO crashes obtained from Model 9 are 33%, 26%, and 36%, respectively. This result demonstrates that all three random effect terms have equally contributed to the fatal model, whereas in the injury model, temporal (σ_v) and spatio-temporal (σ_γ) terms explained more variation than the spatial (σ_θ) term did, and in the PDO model, spatio-temporal terms, followed by the spatial term, had the highest effect.

With regard to the temporal component, it is structured across time, accounting for potential state-wide time effects. As Figure 5.2 shows, the temporal effects, which are on the exponential scale, exhibit a slight downward trend in posterior fatal and PDO crash rates from 2011 and 2012, respectively. However, for injury crashes, no clear pattern across the years can be observed. The sudden decrease and increase in 2012 and 2013 might be due to the hand-held cell phone/texting ban introduced in 2012 and the decrease in seatbelt use observed in 2013, respectively (for more information, please refer to the West Virginia Highway Safety Performance Plan FY 2017 and the West Virginia Observational Survey of Seat Belt Use 2013).

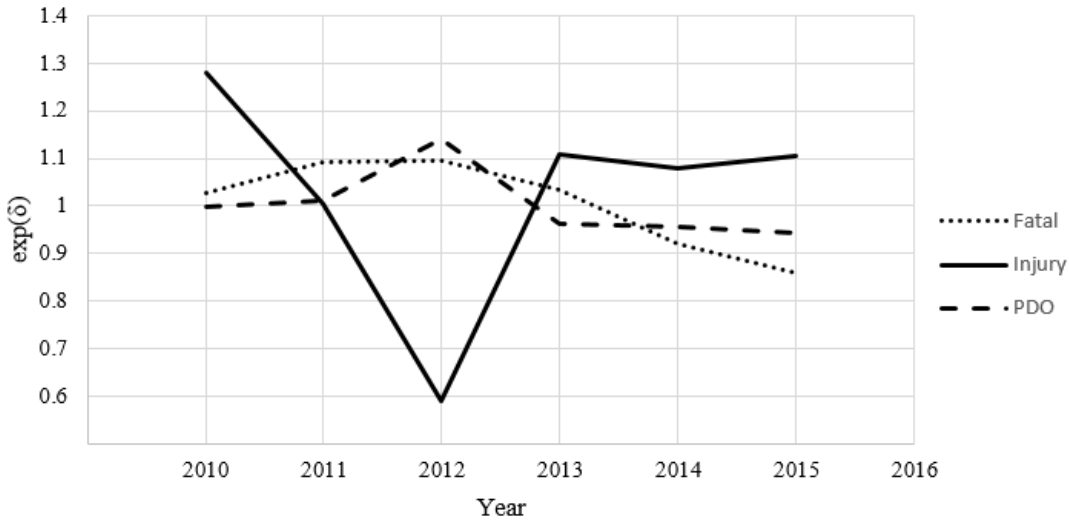


Figure 5.2 Temporal effects on the exponential scale by crash severity under Model 9

Another aspect related to the random effects that the proposed model allows researchers to explore is the correlation among crash severity levels across space and time. It has been speculated that crash severities are not independent, that is, a change in the number of crashes in one crash severity may be associated with a change in the number of crashes in another crash severity. This assumption may pose a potential dilemma for decision makers. The benefit of reducing less severe crashes may be offset by more severe crashes. Hence, exploring the possible correlation between crash severities is critical before making any corrective actions. Variance-covariance matrices can be estimated to facilitate this exploration.

For this case study, Table 5.13 demonstrates the variance-covariance matrices associated with the random effect terms for the preferred model. The results show that all off-diagonal elements of the variance-covariance matrix attributed to the spatial component are significant. This finding confirms that crash severities in WV are not independent. However, their signs are positive, suggesting that crash severities are positively correlated over space, that is, unobserved factors that increase the number of fatal crashes in one county will likely increase the number of injury and PDO crashes in the same county. Nevertheless, none of the off-diagonal elements of the temporal

variance-covariance matrix is significant, which suggests that little correlation exists across crash severities over the years. In other words, in an individual year, there is no state-wide unobserved factor that simultaneously affects the number of different crash severities.

TABLE 5.14 Estimated variance-covariance matrices for Model 9

Unstructured Spatial Effects			
	Fatal	Injury	PDO
Fatal	0.047 (0.013, 0.096)	0.021 (0.008, 0.0493)	0.030 (0.012, 0.065)
Injury		0.037 (0.010, 0.07)	0.034 (0.003, 0.080)
PDO			0.065 (0.014, 0.151)
Unstructured Temporal Effects			
	Fatal	Injury	PDO
Fatal	0.043 (0.01, 0.13)	-0.017 (-0.109, 0.044)	0.005 (-0.030, 0.050)
Injury		0.110 (0.032, 0.34)	-0.017 (-0.092, 0.037)
PDO			0.031 (0.008, 0.100)
Spatio-Temporal Effects			
	Fatal	Injury	PDO
Fatal	0.046 (0.013, 0.109)	0.001 (-0.030, 0.035)	-0.0005 (-0.045, 0.044)
Injury		0.039 (0.011, 0.087)	0.038 (0.003, 0.084)
PDO			0.071 (0.016, 0.151)

Additionally, according to the spatio-temporal variance-covariance matrix, a significant positive correlation exists between injury and PDO crashes. This result implies that there might be unobserved factors in a county in an individual year that simultaneously increase both PDO and injury crashes.

Lastly, the most important component in terms of implications is the spatial random effect term, which captures the variation across neighboring zones; insights from this component could be widely used by safety practitioners. Specifically, one of the major strengths of the model with a spatial random effect term is that it provides ranked areas based on the average (mean) crash rates. As the spatial random effect terms reflect the unique characteristics of each county, these terms can be used to identify the counties whose crash rate are higher than the average. In other words, they

can be used to estimate the excess crashes associated with each county for the purpose of identifying those counties with a high crash risk, providing decision makers with a high level of information to draw priority plans for actions.

For the case study, the results of Model 9 were plugged into Equation 4.20 to rank and highlight WV counties using the weighted average excess risk. According to the information provided by (Harmon et al., 2018), the average costs per casualty related to fatal, injury (including incapacitating, non-incapacitating, and possible injury crashes), and PDO crashes in WV in 2012 were US\$5,289,928, US\$149,505, and US\$9,764, respectively. In view of these costs, the weighted average excess risk for all counties was estimated. The spatial distribution of the weighted excess risk from 2010 to 2015 for all WV counties is shown in Figure 5.3.

The results show that the hotspots (high-risk counties) tend to be locally clustered across the study areas, whereas cold spots (low-risk counties) tend to be clustered in central and southwestern WV. Overall, Marion, Doddridge, and Raleigh Counties are associated with the lowest excess risks. On the other hand, Marshall, Braxton, Lewis and Jefferson Counties have the highest excess risk and, therefore, should be given more attention during the planning stage.

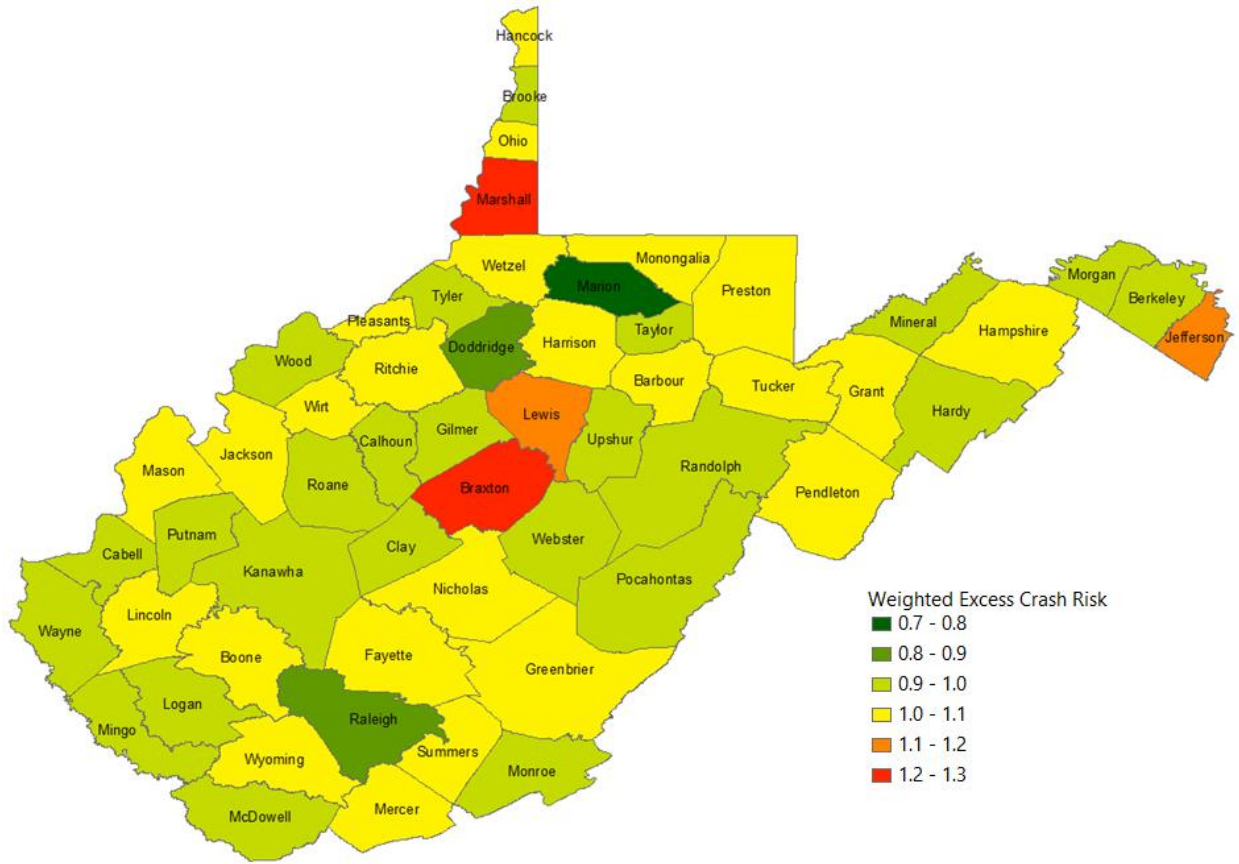


Figure 5.3 Spatial distribution of the weighted excess risk for counties in 2010–2015

5.2. Micro-level Data Analysis

As explained in Section 4.3.1, most of the crash data used in road safety studies are of a hierarchical nature, that is, the occupants are nested in vehicles, the vehicles are within crashes, and the crashes are in roadways and regions. Additionally, because the crash consequence or injury outcome is of a discrete nature, logit or discrete choice models are suitable for estimating a crash severity model. Logit models include ordered or nominal (un-ordered) response models. Each of these models has its own advantages. For example, the ordered response model considers the ordinal nature of the crash severity levels, whereas the nominal response model is known to have consistent parameter estimates with under-reported data (Savolainen and Mannering, 2007) .

In this section, four models, including two ordered response models and two nominal response models, have been developed to examine the relationship between various factors, such as age, gender, and crash-level factors (e.g., road condition), and the odds of a vehicle occupant involved in a fatal crash being killed (as opposed to surviving). The first model is a basic (non-random) nominal response model, and the second model is a full nominal response model with crash-level, spatial, temporal, and spatio-temporal random effect terms. The third model is a basic ordered logit model, and the fourth model is a full ordered logit model with the same random effect terms utilized in Model 2. The injury severity of a vehicle occupant involved in a fatal crash is coded as follows: 2 for an occupant's fatal injury and 1 for any other outcomes (non-fatal injury or no injury). The variance inflation factor (VIF) was estimated for all variables to measure potential multicollinearity; the VIF did not exceed 2.74, suggesting that no evidence of multicollinearity exists. Nominal and ordered response models are explained in Sections 5.2.1.1 and 5.2.1.2, respectively.

5.2.1. Model Development

5.2.1.1. Nominal Response Models

Two full Bayesian nominal response models are discussed in this section: one basic or non-random effect model and one full model with all random effect terms to account for the unobserved effects at the crash level over space and time. The models were computed with the MCMC technique using WINBUGS software. A 95% Bayesian credible interval was used to assess the significance of the covariates, and the DIC was used as a model diagnostic. The first 1,000 iterations were removed as burn-ins to ensure the convergence of all parameters, and the next 100,000 were used to establish the model. MCMC chain and history plots were chosen to monitor the MCMC chains and the convergence of the parameters. Table 5.15 shows the DIC

values for both proposed models. The full model (Model 2) had a lower DIC compared with the basic model (Model 1). This implies that the inclusion of random effect terms explained the unobserved factors affecting the occupants' injury severity and contributed to the model's accuracy. The model estimation results are summarized in Tables 5.16 and 5.17. The estimation results of the two models are similar in terms of the set of statistically significant variables, as well as the magnitude of the coefficients and their signs.

TABLE 5.15 DIC values for the nominal response models

No	Model	DIC
1	$\beta_0 + \beta X_{ksit}$	4,256
2	$\beta_0 + \beta X_{ksit} + \rho_s + \theta_i + \varphi_i + \delta_t + \nu_t + \gamma_{st}$	4,242

TABLE 5.16 Parameter estimates for the basic nominal response model (Model 1)

Parameters	Mean	Odds Ratio	SD	25% CI	97% CI
Intercept	-1.956	0.141	0.196	-2.353	-1.584
Occupant-level factors					
Gender (male=1, female=0)	-0.190	0.827	0.083	-0.350	-0.027
Age	0.023	1.024	0.002	0.020	0.027
Occupant type (driver=1, otherwise=0)	1.053	2.866	0.088	0.881	1.224
Ejection (occupant is ejected fully or partially=1, otherwise=0)	2.128	8.398	0.133	1.871	2.396
Vehicle-level factors					
Speeding (yes=1, no=0)	0.191	1.211	0.083	0.027	0.353
Alcohol involvement (yes=1, no=0)	0.867	2.380	0.101	0.669	1.068
Crash-level factors					
Surface condition (dry=1, otherwise=0)	-0.142	0.867	0.115	-0.367	0.084
Road alignment (straight=1, curve=0)	-0.314	0.730	0.079	-0.471	-0.159
Road type (divided=1, undivided=0)	0.039	1.040	0.095	-0.147	0.230
Arterial	-0.185	0.831	0.129	-0.435	0.068
Collector	0.091	1.095	0.132	-0.166	0.350
Weather (clear=1, otherwise=0)	0.119	1.127	0.101	-0.078	0.317
Lighting (daylight=1, otherwise=0)	0.050	1.052	0.081	-0.107	0.212
DIC	4,256				

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.17 Parameter estimates for the full nominal response model (Model 2)

Parameters	Mean	Odds Ratio	SD	25% CI	97% CI
Intercept	-1.966	0.140	0.207	-2.390	-1.560
Occupant-level factors					
Gender (male=1, female=0)	-0.193	0.825	0.086	-0.364	-0.026
Age	0.024	1.024	0.002	0.020	0.028
Occupant type (driver=1, otherwise=0)	1.053	2.866	0.089	0.878	1.227
Ejection (occupant is ejected fully or partially=1, otherwise=0)	2.135	8.457	0.134	1.881	2.399
Vehicle-level factors					
Speeding (yes=1, no=0)	0.203	1.224	0.083	0.041	0.367
Alcohol involvement (yes=1, no=0)	0.873	2.393	0.103	0.671	1.073
Crash-level factors					
Surface condition (dry=1, otherwise=0)	-0.133	0.875	0.114	-0.359	0.097
Road alignment (straight=1, curve=0)	-0.315	0.730	0.078	-0.467	-0.165
Road type (divided=1, undivided=0)	0.038	1.039	0.097	-0.152	0.226
Arterial	-0.181	0.834	0.130	-0.434	0.071
Collector	0.085	1.088	0.132	-0.171	0.344
Weather (clear=1, otherwise=0)	0.115	1.122	0.102	-0.084	0.317
Lighting (daylight=1, otherwise=0)	0.045	1.046	0.084	-0.121	0.210
Standard deviations of random effect terms					
σ_v	0.609	-	2.072	0.05927	4.919
σ_δ	0.813	-	3.717	0.061	4.309
σ_θ	0.252	-	0.182	0.060	0.754
σ_ϕ	0.138	-	0.049	0.101	0.304
σ_γ	0.482	-	0.430	0.085	1.648
σ_p	0.319	-	0.274	0.083	1.065
DIC	4,242				

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

5.2.1.2. Ordered Response Model

Because crash severities are ordinal in nature, two ordered logit models were also established to account for the full data hierarchy: a basic ordered model (Model 3) and a full ordered logit model with all random effect terms (Model 4). One chain was simulated with an initial 1,000 iterations discarded as burn-ins, and it was followed by 100,000 iterations to achieve convergence. Figure 3 in the Appendix shows the MCMC output for the parameters, indicating that the variables are converged. Additionally, as shown in Figure 4, the fixed effect parameters have normal curves, which is consistent with the prior assumption of the distribution of these

parameters. According to the DIC values listed in Tables 5.17 and 5.18, the full ordered model (Model 4) fits the dataset the best among all the models; its DIC value is slightly lower than that of the full nominal response model. Therefore, the full ordered logit model is considered the preferred model to use for site ranking. The model estimation results are summarized in Tables 5.19 to 5.20.

TABLE 5.18 DIC values for the ordered response models

No.	Model	DIC
3	$\beta_0 + \beta X_{ksit}$	4,255
4	$\beta_0 + \beta X_{ksit} + \rho_s + \theta_i + \varphi_i + \delta_t + \nu_t + \gamma_{st}$	4,240

TABLE 5.19 Parameter estimates for the basic ordered logit model (Model 3)

Parameters	Mean	Odds Ratio	SD	25% CI	97% CI
Occupant-level factors					
Gender (male=1, female=0)	-0.188	0.829	0.084	-0.351	-0.024
Age	0.024	1.024	0.002	0.020	0.027
Occupant type (driver=1, otherwise=0)	1.054	2.869	0.093	0.881	1.232
Ejection (occupant is ejected fully or partially=1, otherwise=0)	2.127	8.390	0.133	1.868	2.394
Vehicle-level factors					
Speeding (yes=1, no=0)	0.190	1.209	0.083	0.026	0.350
Alcohol involvement (yes=1, no=0)	0.870	2.387	0.102	0.669	1.070
Crash-level factors					
Surface condition (dry=1, otherwise=0)	-0.142	0.867	0.114	-0.359	0.088
Road alignment (straight=1, curve=0)	-0.314	0.730	0.079	-0.472	-0.161
Road type (divided=1, undivided=0)	0.037	1.037	0.096	-0.161	0.223
Arterial	-0.181	0.834	0.124	-0.431	0.062
Collector	0.095	1.099	0.128	-0.160	0.342
Weather (clear=1, otherwise=0)	0.122	1.129	0.100	-0.075	0.318
Lighting (daylight=1, otherwise=0)	0.050	1.051	0.081	-0.107	0.209
Threshold value	1.998		0.201	1.636	2.443
DIC			4,255		

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

TABLE 5.20 Parameter estimates for the full ordered logit model (Model 4)

Parameters	Mean	Odds Ratio	SD	25% CI	97% CI
Occupant-level factors					
Gender (male=1, female=0)	-0.204	0.815	0.085	-0.373	-0.042
Age	0.024	1.024	0.002	0.020	0.028
Occupant type (driver=1, otherwise=0)	1.050	2.858	0.089	0.868	1.230
Ejection (occupant is ejected fully or partially=1, otherwise=0)	2.161	8.680	0.136	1.901	2.432
Vehicle-level factors					
Speeding (yes=1, no=0)	0.212	1.236	0.087	0.043	0.381
Alcohol involvement (yes=1, no=0)	0.890	2.434	0.106	0.686	1.095
Crash-level factors					
Surface condition (dry=1, otherwise=0)	-0.133	0.876	0.118	-0.372	0.089
Road alignment (straight=1, curve=0)	-0.308	0.735	0.083	-0.473	-0.144
Road type (divided=1, undivided=0)	0.027	1.027	0.103	-0.177	0.226
Arterial	-0.194	0.823	0.137	-0.455	0.071
Collector	0.067	1.069	0.138	-0.192	0.348
Weather (clear=1, otherwise=0)	0.120	1.127	0.106	-0.081	0.335
Lighting (daylight=1, otherwise=0)	0.041	1.042	0.086	-0.129	0.210
Threshold value	1.967		0.240	1.514	2.426
Standard deviations of random effect terms					
σ_v	0.160	-	0.114	0.055	0.467
σ_δ	0.124	-	0.062	0.054	0.279
σ_θ	0.172	-	0.032	0.117	0.241
σ_ϕ	0.205	-	0.046	0.129	0.307
σ_γ	0.112	-	0.034	0.059	0.188
σ_ρ	0.170	-	0.030	0.120	0.236
DIC	4,240				

Note: Shaded cells indicate significant coefficients at $\alpha=5\%$ level.

5.2.2. Model Validation

As discussed before, the validation of a model is important before interpreting the parameter estimates. In the macro-level data analysis section, the validation of a Bayesian model was performed by comparing its goodness of fit with those of the corresponding classical models using MAE, which is the difference between the expected number of crashes estimated by the model and the observed number of crashes over the number of observations. Following the same line of thought, a simple ordered logit model was estimated (see Table 5.21) to compare the

accuracy of the full Bayesian ordered logit model (Model 4) and its classical counterpart. However, instead of MAE, which is only applicable for the count models, the unweighted mean of squares (UMS) is considered as a measure to understand how well the models fit the data. In this application, if the estimated probability of being killed for an individual is close to its observed crash severity outcome (1 if the outcome is fatal injury, 0 otherwise), then the UMS would be smaller, and the model better fits our data. Table 5.22 shows that the estimated UMS for the Bayesian approach is smaller than that for the classical model, suggesting that the Bayesian logit model provides a better fit to the data. As the parameter estimates in both models are similar, this difference in UMS is due to the use of the random effect terms in the Bayesian model, which captured the unobserved effects not explained in the classical model.

Table 5.21 Classical ordered logit model parameter estimates

Parameters	Mean	SD	Z-value	P> Z
Gender (male=1, female=0)	-0.189	0.083	-2.270	0.023
Speeding (yes=1, no=0)	0.190	0.083	2.290	0.022
Alcohol involvement (yes=1, no=0)	0.864	0.101	8.540	<0.0001
Ejection (occupant is ejected fully or partially=1, otherwise=0)	2.116	0.132	16.040	<0.0001
Surface condition (dry=1, otherwise=0)	-0.142	0.118	-1.210	0.227
Age	0.023	0.002	11.640	<0.0001
Road alignment (straight=1, curve=0)	-0.313	0.079	-3.970	<0.0001
Road type (divided=1, undivided=0)	0.038	0.094	0.400	0.688
Arterial	-0.189	0.129	-1.470	0.142
Collector	0.088	0.131	0.670	0.502
Weather (clear=1, otherwise=0)	0.119	0.102	1.160	0.245
Lighting (daylight=1, otherwise=0)	0.051	0.082	0.620	0.536
Occupant type (driver=1, otherwise=0)	1.049	0.087	12.000	<0.0001
Threshold value	1.945	0.203		
Log likelihood	-2114			

TABLE 5.22 Comparison of Bayesian and classical ordered logit models

Model	UMS
Bayesian ordered logit	0.38
Classical ordered logit	0.61

5.2.3. Interpretation of Parameter Estimates

The variables that were found to be significant in Model 4 are the same as the significant variables in Models 1 to 3. Additionally, the sign and magnitude of the parameters are consistent.

In terms of occupant-level factors, the estimated odds ratios demonstrate that conditional on fatal crashes, a male occupant had 19% lower odds of fatality than a female one. This finding is in line with those of previous studies (see for example, Shaheed et al., 2016; Chen and Chen, 2011; Romano et al., 2008) that reported a lower injury propensity for male occupants than for female ones.

In addition, an occupant who was ejected was nine times as likely to be killed compared with an occupant who was not ejected. This finding is consistent with those of previous research (Shaheed et al., 2016) reporting that occupants are likely to sustain more severe injury when they are ejected from a vehicle. Furthermore, older occupants had higher odds of fatality compared with younger occupants. Some other researchers (Eluru et al., 2010; Kahane, 2013) reported similar results. This finding can be justified by this fact that older occupants are more vulnerable to thoracic injuries, especially multiple rib fractures (Kahane, 2013). Lastly, a driver was 2.86 times more likely to be killed in a fatal crash than a passenger was. This result contradicts that of Shaheed et al. (2016) who found that driver is associated with lower odds of fatality, probably because of the less severe injuries sustained by passengers in back seats in the case of frontal impacts.

With regard to vehicle-level factors, the odds ratio associated with speeding implies that speeding increased the odds of fatality by 24%. Other researchers used speed from different aspects, such as speed limit (Usman et al. 2016) and average vehicle speed (Meng et al., 2017). In both cases, an increase in speed was found to increase the odds of fatality or major injury. Furthermore, a vehicle occupant was found to be 2.43 times more likely to be killed when the

vehicle's driver was alcohol impaired than when the driver was not. This finding is consistent with that reported in the literature (Eluru et al., 2010; Yuan et al., 2017).

Driving in a straight alignment was found to reduce the odds of fatality by 27% compared with driving in a curved alignment. In other words, the presence of a curve on a road segment is associated with higher odds of fatality in a fatal crash, as the driver may not have a clear sight of the distance or may not be able to control the vehicle. However, Usman et al. (2016) found that the presence of a curve on a hilly road is only associated with an increase in the odds of minor injury.

The MEs of the statistically significant variables on the probability of a fatal injury were calculated and are summarized in Table 5.23 to further assess the parameter estimates of the preferred model (Model 4). Here, ME gives instantaneous effect that a change in a particular explanatory variable has on an occupant likelihood of being killed in a fatal crash.

TABLE 5.23 Marginal effects of significant variables in the full ordered logit model

Variables	ME
Gender (male=1, female=0)	-0.036
Occupant type (driver=1, otherwise=0)	0.184
Age	0.006
Speeding (yes=1, no=0)	0.018
Alcohol involvement (yes=1, no=0)	0.052
Ejection (occupant is ejected fully or partially=1, otherwise=0)	0.086
Road alignment (straight=1, curve=0)	-0.046

As Table 5.23 shows, on average, being in the driver's seat will likely increase the likelihood of being killed by 0.184. Similarly, being a male occupant will likely reduce the probability of being killed by 0.036. According to the parameter estimate for age, which is a continuous variable, an increase in occupant's age by one year will likely increase the probability of being killed by 0.006. Additionally, alcohol impairment, speeding, and full or partial ejection will likely increase the

likelihood of fatal injury in a fatal crash by 0.018, 0.052, and 0.086, respectively. Lastly, driving in a straight road alignment decreases the probability of being killed by 0.046.

5.2.4. Interpretation of Random Effect Terms

As discussed before, in the macro-level data analysis, we dealt with crash counts that were aggregated at the county level. Therefore, they were modeled by only incorporating spatial and temporal components. However, at the micro level, we deal with multilevel data, which implies that there might be unobserved factors at each level that can affect crash severity. This section focuses on interpreting random effect terms, especially the spatial and temporal components used in the full Bayesian logit model (Model 4). As Table 5.17 shows, the standard deviations of the five random effect terms were found to be significant. The standard deviation associated with both structures (σ_ϕ) was found to be slightly larger and explained more variation than the other random effect terms did.

With regard to the crash-level random effect term (ρ_c), it captures the variation across traffic crashes. In other words, it accounts for crash-level unobserved factors, such as the impact speed between vehicles and road geometric features. As discussed by previous researchers (Shaheed et al., 2016; Vanlaar, 2005), ignoring the crash-level variation is equivalent to ignoring the clustering nature of data, that is, all observations within a cluster are independent, which can result in underestimated standard errors.

From the temporal aspect, the structured (δ) and unstructured (ν) temporal effects nearly played the same roles for the crash severity model. Hence, the findings suggest that there are both unobserved factors that were temporally correlated and unobserved factors that were present in individual years. Figure 5.4 plots the exponential of the sum of the posterior means of structured

and unstructured temporal effects from 2010 to 2015. Temporal effects have a decreasing trend in 2010 and 2012, suggesting that the unobserved factors (e.g., decline in vehicle ownership or increased use of the public transport system) over these years tended to reduce the odds of fatality in WV. However, the trend line has been increasing since 2012, which may represent the decline in seatbelt use. In principle, not too many global unobserved factors have state-wide effects; examples could be changes in economic status, changes in transit ridership, and traffic safety laws. Therefore, exploring the evolution of temporal effects can provide authorities with some insights into large-scale and state-wide transportation and safety planning and can help design new policies (e.g., banning the use of wireless electronic devices, installation of an ignition interlock device on a vehicle for DUI offenders) that target the mitigation of injury severity.

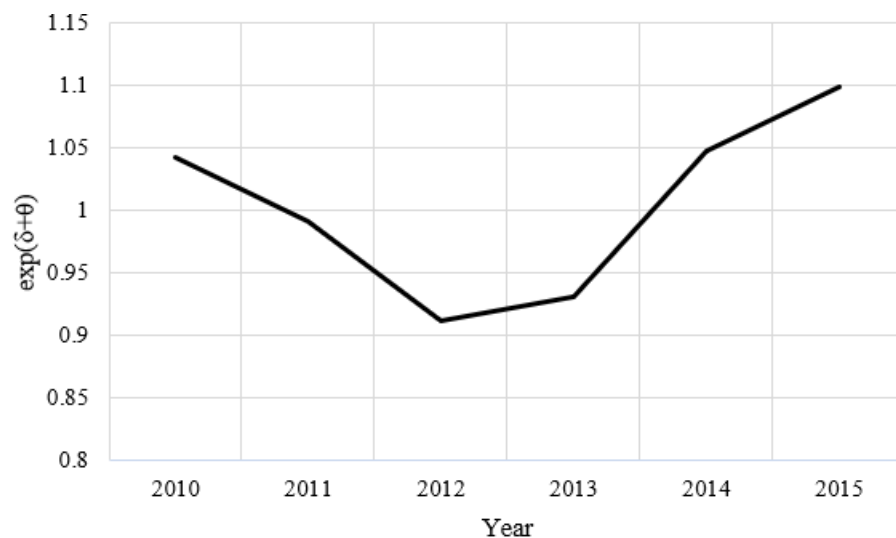


Figure 5.4 Exponential of the temporal effects derived from the Bayesian logit model (Model 4)

The last components discussed in this section are the structured (ϕ) and unstructured (θ) spatial random effect terms, which capture spatial correlation and census-tract-level unobserved factors, respectively. Similar to those in macro-level data analysis, the spatial random effect terms

allow researchers to identify high-risk zones. Again, the findings from this analysis can help identify areas in need of interventions and/or provide insights into the prioritization of sites.

However, the two applications have some major differences. First, the spatial random effect terms at the macro level aim to represent the areas with system-wide issues on the basis of the average crash rate, whereas the spatial random effect terms at the micro level tend to show the sites with engineering problems on the basis of the average occupants' odds of fatality. Second, macro-level spatial random effect terms represent the factors that contribute to the occurrence of traffic crashes, whereas micro-level spatial random effect terms may only represent the factors that contribute to crash severity. Figure 5.5 represents the estimated weighted excess risk of a fatality, which is the exponential of the random effect terms across the census tracts on the basis of Model 4.

As Figure 5.5 shows, the majority of the census tracts have a value between 0.9 and 1.1. The census tracts in Braxton and Wood Counties were identified as SWiPs, whose odds of fatality are higher than the average. The remaining high-risk census tracts tend to be concentrated in the central part. However, the majority of low-risk areas are located in the northern and southwestern areas.

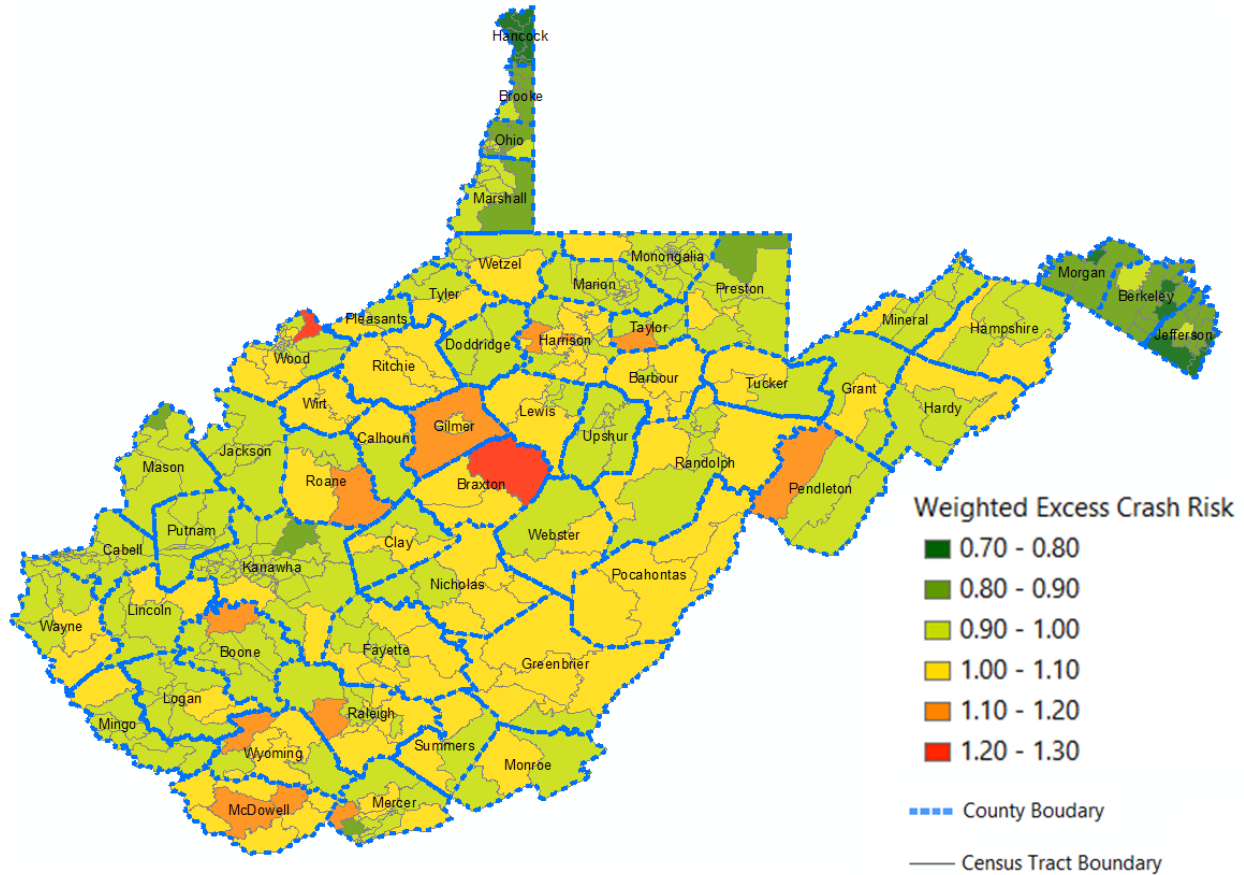


Figure 5.5 Spatial distribution of the weighted excess risk of fatality for the census tracts in 2010–2015

5.3. Integration of Macro- and Micro-Level Data Analyses

This section presents an application of the crash prediction models discussed in this dissertation, which aims to identify SWiPs (i.e., areas with excess risk) for further investigation and safety treatment. It demonstrates the results from a GIS-based approach that combines both accident frequency and severity models for site ranking. As discussed previously, the combination of macro- and micro-level analyses has many advantages over classical models, as it incorporates both individual and area-wide factors that affect traffic crashes.

As discussed in Chapter 4, the weighted excess risk is defined as a metric in order to identify high-risk locations. For the macro-level analysis, the weighted excess risk was defined as the ratio of the expected number of crashes in a county to the average number of crashes in similar counties. For the micro-level analysis, the excess risk was defined as the ratio of the expected odds of fatality to the mean odds of fatality in similar census tracts. Both metrics highlight the increased risk resulting from unobserved spatial and temporal effects.

As these two metrics follow different distributions, they were normalized using the min–max normalization strategy (the normalized values range from 0 to 1. The larger the value is, the more hazardous the area is) and were stored in separate files in CSV format. Later, the CSV files were added to the county and census tract shapefiles by using unique codes assigned to each county and census tract. Next, as shown in Figure 5.6, the Feature to Raster tool was utilized to convert the shapefiles to raster files representing the corresponding normalized weighted excess risk values. In Figure 5.6, county and Census tract shapefiles are selected from the “input features” dropdown menu. Then, “field” menu was set to “ratio,” which contains the normalized excess risk values. Once the two raster files were created, the raster calculator tool was used to combine the generated raster files. This tool allows the execution of algebra expressions (taking the average of two normalized excess risks from the macro- and micro-level parts) (see Figure 5.7). The produced map is shown in Figure 5.8.

The results of the integrated metric can be used to identify areas that can be considered hazardous based on the analysis of both the macro and micro levels. As seen in Figure 5.8, these areas are randomly located throughout the study area. The Census tracts in Marshal County in northern WV, the majority of the Census tracts in Lewis County, and all Census tracts in Braxton County in central WV, one Census tract in Pendleton County in the eastern side, and one Census

tract in Wyoming and Mercer Counties have been identified as hazardous areas. In other word, these areas have two main characteristics: (1) they have higher expected crash costs because of expected fatal, injury, and PDO crashes, and (2) they are associated with a higher vehicle occupant's odd of fatality. Hence, these areas require further safety examination and treatment. Additionally, the top-ranked Census tracts should be given more attention and higher priorities for safety improvement when limited funds and resources are available.

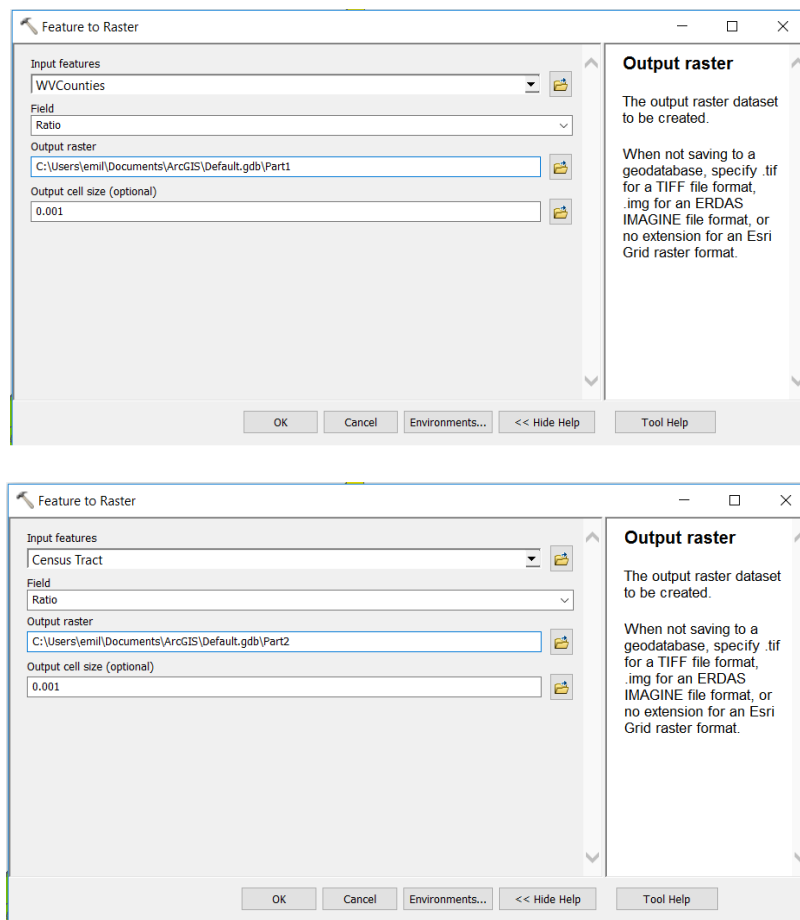


Figure 5.6 Conversion of county and census tract shapefiles to raster

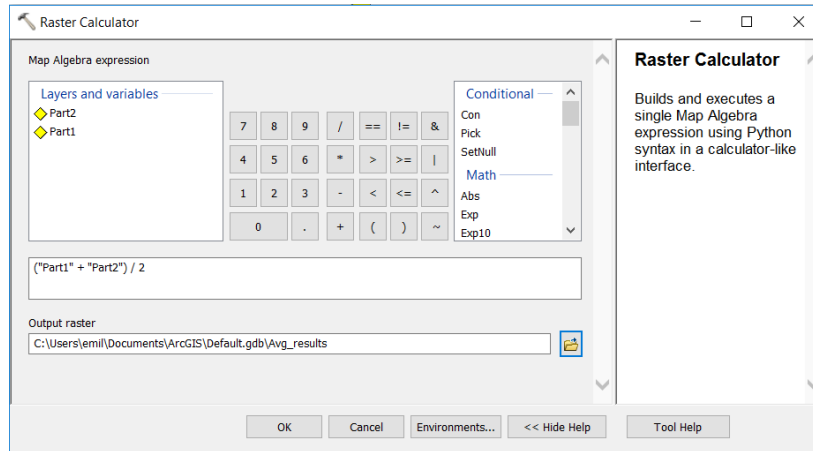


Figure 5.7 Estimating the combined metrics through the raster calculator tool

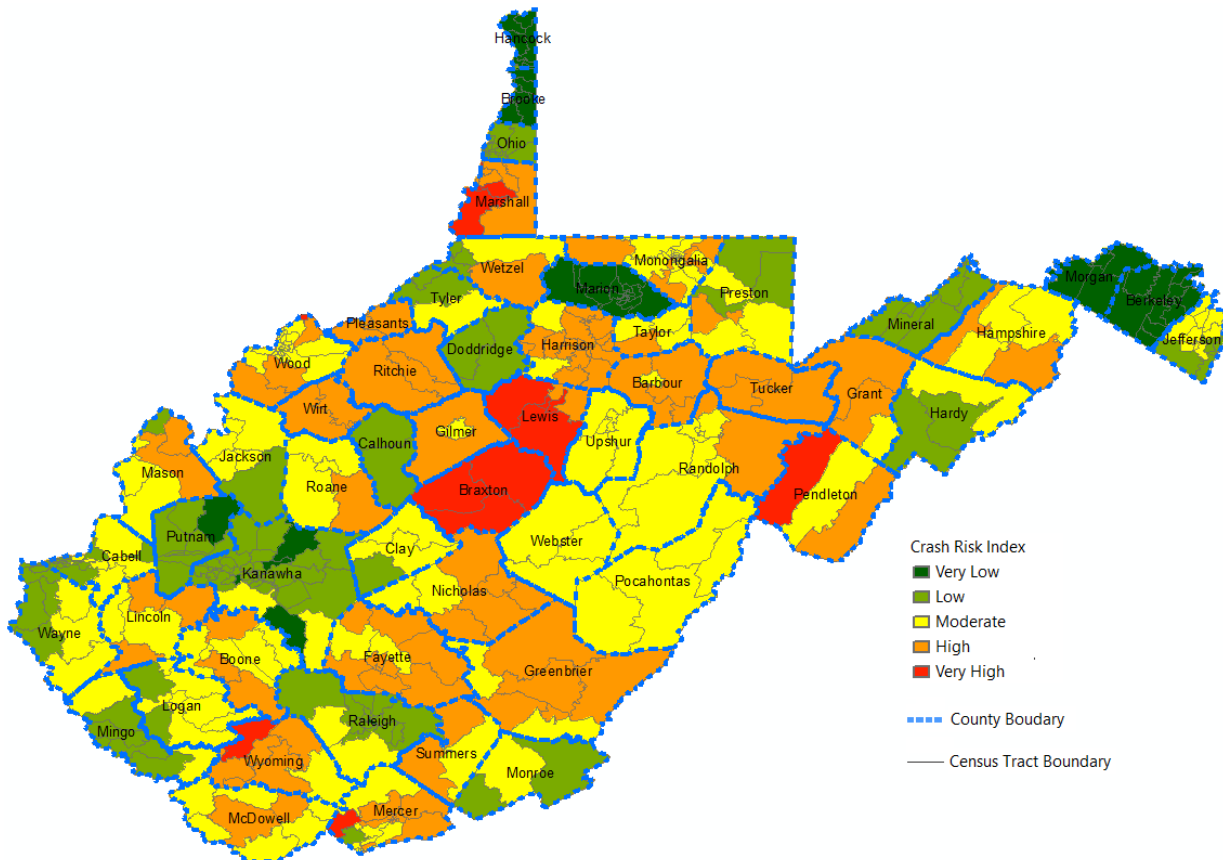


Figure 5.8 Final map from the combination of the macro- and micro-level results

5.4. Summary

Generally, many Bayesian spatial models have been developed for crash analyses in the literature. However, most models ignored either temporal effects or correlations among crash severities, which may lead to biased parameter estimates.

In this study, in the case of macro-level data analysis, various multivariate models were established to simultaneously analyze county-level fatal, injury, and PDO crashes, which was not done before. Various socio-economic, transport-related, and environmental factors were examined. The results demonstrated strong correlations among crash severities across WV counties, which cannot be ignored. However, little correlation exists among crash severities over time.

As with other studies (Boulieri et al., 2016; Huang et al., 2010; Liu and Sharma, 2018), unstructured spatial and temporal components, as well as spatio-temporal effects, were found to play important roles in model performance. Nevertheless, in this study, the structured spatial and temporal random effect terms were found to be negligible. From temporal viewpoints, MRW(1) was also shown to perform better than MRW(2) for crashes at different severity levels.

The variables found to be significant in the preferred model were reasonable and explainable. According to the parameter estimates of the final model, primary roadway density is the common significant variable that contributes to all crash severities, and it is positively associated with fatal, injury, and PDO crash risk. Some variables are solely significant for one severity type. Intersection density is only significant in the injury crash model. Moreover, the sign of the coefficient is negative, suggesting that an increase in intersection density will likely decrease injury crashes. The mean travel time to the hospital was only evaluated in the fatal crash model because severe injuries may become fatal as the travel time to the hospital increases, and it was found to be significant. The positive coefficient of this variable implies that people with severe injuries are less likely to survive in

counties with longer access times to hospitals. The metro and labor force variables were only found to be significant in the fatal and PDO models, respectively. Their negative coefficients indicate that urbanized areas and areas with higher labor force participation rates may have fewer fatal and PDO crashes. With the assumption that metropolitan areas and areas with higher labor force participation rates are likely to be more congested, this finding is consistent with the literature, suggesting that congested roads can be safer because drivers tend to drive with greater caution (Zhou and Sisiopiku, 1997). The DUI and school density variables were positive and significant in the PDO crash model. This result implies that counties with more DUI arrests and a higher school density will likely have a higher number of PDO crashes.

According to the weighted excess crash risk values, counties with a high crash risk are locally clustered across WV. Additionally, they are evenly distributed in metro and non-metro areas. All crashes, except injury crash, have downward trends from 2012 to 2015.

In the case of micro-level data analysis, ordered and nominal response models were established by accounting for crash level, spatial, temporal, and spatio-temporal random effect terms. The results demonstrated that the ordered logit model with all random effect terms provided a better fit.

With regard to the estimated odds ratios in the full ordered response model, conditional on fatal crashes, a male occupant had lower odds of fatality than a female one. The odds ratio associated with speeding implies that speeding increased the odds of fatality. Furthermore, an occupant was more likely to be killed when the vehicle's driver was alcohol impaired than when the driver was not. Occupants who were ejected were more likely to be killed compared with those who were not ejected. Furthermore, older occupants had higher odds of fatality than their younger counterparts.

Driving in a straight alignment was also found to reduce the odds of fatality. Finally, a driver was more likely to be killed than a passenger.

CHAPTER 6: CONCLUSION

6.1. Summary

This dissertation explored the relationship between traffic crashes and several factors that might affect the number and severity of traffic crashes both at the micro and macro levels. Various suitable Bayesian models were used for both crash frequency and severity analyses. Finally, a novel method was proposed to combine the macro- and micro-level results in order to identify and rank high-risk areas. This chapter aims to provide an overview of this dissertation and further discusses the conclusions of this study based on the analysis results from Chapter 5. The contributions and policy implications of this research will also be discussed.

As discussed in the literature review (Chapter 2), many Bayesian models have been developed for crash analyses using both macroscopic and microscopic analyses. Macro-level analysis typically divides the study area into a number of smaller areas, such as counties or districts, and it aims to establish the relationship between crash frequency and area-wide factors, such as socio-economic factors and environmental characteristics. Micro-level analysis, on the other hand, explores the relationship between crash frequency or crash severity and different factors, such as number of lanes, lane width, and individual factors (e.g., age, gender).

However, the literature has some major limitations. First, most models that were proposed did not examine unobserved effects over space and time as well as correlations between crash severities, which may lead to biased parameter estimates. Second, the effects of important factors, such as urbanity/rurality and access time to the hospital, on crash severities were not explored. Third, little efforts were made to combine macro- and micro-level analyses for site ranking purposes.

6.2. Research Contributions

6.2.1. Empirical Contributions and Policy and Planning Implications

In the macro-level data analysis in this dissertation, nine Bayesian models with and without random effects were estimated and tested for the case study of WV. The results demonstrated that unstructured spatial random effects, as well as unstructured temporal effects, play important roles in the model performance, as the estimated goodness of fit measures (DIC and R^2_{DIC}) suggest. Moreover, evidence of strong within-area correlations among crash severities was found. The comparison of the Poisson lognormal and negative binomial models also demonstrated that the parameter estimates for the fatal models are very close, whereas those for the injury and PDO models are slightly different. This finding implies that unobserved factors explained less variation in the fatal model than in the injury and PDO models.

The variables that were found to be significant in the preferred model were reasonable and explainable. The variable primary roadway density was positively correlated with all crash severities, which is consistent with the findings of previous studies. However, some of the variables used in this study were not consistent with the results of other research. For example, intersection density was found to have a negative impact on injury crashes in this study, whereas in a previous work (such as that of Guevara et al. 2004), it was positively associated with injury crashes. School density and DUI rates, which were found to be insignificant in the literature, were established to be significant and positively associated with PDO crashes. Furthermore, some variables, such as the mean travel time to the hospital and urbanity/rurality, were only investigated in the present study, and they were found to have significant positive and negative impacts on fatal crashes, respectively.

In the micro-level data analysis, different nominal response and ordered logit models were established and tested for the case study of WV. In all models, age, gender, occupant type, occupant's

ejection, speeding, alcohol involvement, and road alignment were significant. At the occupant level, older occupants, female occupants, and ejected occupants, respectively, had higher odds of fatality than young drivers, male occupants, and unejected occupants. These findings are in line with those of previous studies. On the other hand, drivers were more likely to be killed in a fatal crash than passengers were, which contradicts the finding of Shaheed et al. (2015). At the vehicle level, a driver's alcohol involvement and speeding contributed to higher odds of fatality for the occupant, a finding that is consistent with those of previous works. Finally, at the crash level, a curved segment was associated with higher odds of fatality. This result is not consistent with those of previous research (see for example, Usman et al., 2016), which reported that the presence of curves was only associated with an increase in the odds of minor injury. The comparison of the ordered and nominal response models suggested that the full ordered logit model performed better than the other models. Additionally, the comparison of the preferred model and the classical logit model demonstrated that the former predicted the crash probabilities more accurately. Analysis of the model results showed that unobserved factors at the crash level and over space and time exist. Furthermore, the spatial correlation term explained more variations than the other components did.

In terms of site ranking, the results from the two levels were normalized and combined using a GIS-based approach to provide more accurate information. Very high-risk census tracts were found in Marshal, Lewis, Braxton, Pendleton, Wyoming, and Mercer Counties. Furthermore, both very-high-risk and high-risk areas were distributed throughout the study area. This finding can be used by state agencies and corresponding decision makers to effectively allocate resources and funds in order to mitigate safety issues.

In addition to the identification of high-risk areas, exploring the potential of various policies and strategies that can reinforce traffic safety in the community based on the modeling results can

also be of value to policy makers. Several policy implementations can be proposed to improve road safety. The rest of this section discusses some of these approaches that can be effective for WV according to the research findings.

Effective speed management countermeasures: As discussed in Chapter 5, WV counties with a higher primary road density will likely experience more crashes, potentially because of their higher traffic speed and speed limits in primary roadways. Furthermore, speeding has been identified as one of the key risk factors that influence the risk of fatality in WV. Considering these two factors together indicates that adopting effective speed management strategies in primary roadways can create a safer driving environment. Some examples of engineering speed management countermeasures (Federal Highway Administration, 2014a) that can be effective are the implementation of speed humps, reduction of the posted speed limit in expressways, use of transverse rumble strips, and enhancement of road delineation. Additionally, from the research findings in rural or non-metro areas, higher traffic crashes are expected on rural arterial roads and for vehicles traveling at higher speeds and substantial distances. Thus, speed limit signs should be placed at least every 5 km (3.11 miles) along rural roads, where the default limit does not apply and the conditions are reasonably consistent (World Health Organization, 2008).

Establishment of hospitals with trauma centers in deprived counties: Travel time to the hospital is among the factors that were found to directly affect the risk of fatalities. Additionally, on the basis of the policies and treatment protocol developed by the West Virginia Department of Health and Human Resource, traffic crash victims who are injured seriously should be transported to the highest-level trauma center within 30 minutes (West Virginia Department of Health and Human Resources, 2018). According to the GIS analysis performed in this study, 10 counties in the eastern side have a more than 30-minute average travel time to hospitals with trauma centers. In light of

these findings, the establishment of trauma centers or air ambulance services in these counties is recommended.

Roadway departure countermeasures: As found in this research, horizontal curves are associated with higher odds of fatality than straight roadways are. The majority of crashes (more than 80%) in horizontal curves are roadway departure crashes in which a vehicle crosses an edge line or leaves the traveled way (Federal Highway Administration, 2014b). The Federal Highway Administration suggests some effective countermeasures to mitigate injury severity as a result of roadway departure crashes on horizontal curves. Some examples are the use of longitudinal barriers in locations where a high potential for crossover crashes exists, the addition of clear zones and roadside terrains to allow a driver to stop safely or regain control of a vehicle, and the use of crash cushions and high-friction pavements on horizontal curves (Federal Highway Administration, 2014b).

Deployment of in-vehicle safety technology: As the fatality risk increases with the age of vehicle occupants, including the drivers, the use of in-vehicle systems, such as collision warning systems by car manufacturers, can compensate for an old driver's poor reaction time. Additionally, as a driver's alcohol impairment is found to increase the odds of fatality, implementing an ignition interlock device would be helpful in reducing alcohol-impaired driving. Lastly, compared with other variables, the magnitude of odds ratios showed that an occupant's ejection had a major influence on crash severities. Therefore, car manufacturers should place a high degree of emphasis on the development of advanced restraint systems to prevent occupants from being ejected.

6.2.2. Methodological Contributions

This dissertation has made a number of methodological contributions, which are discussed in this section.

To improve crash estimation, this research proposed a Bayesian multivariate Poisson lognormal model that simultaneously analyzed three crash severities (fatal, injury, and PDO crashes) and accounted for spatial and temporal effects. This novel model has not been fully exploited in the transportation safety literature. Specifically, the proposed model estimates crash frequencies (especially injury and PDO crash frequencies) more precisely than do popular classical models, such as the NB model, as confirmed by the validation of the model. This finding implies that the proposed model has a better ability to identify true/actual high-risk areas. One of the main reasons why the proposed model provides more accurate estimates is its inclusion of spatial and temporal components. Additionally, the simultaneous analysis of crash severities demonstrated that crash severities are correlated. Accounting for this correlation among crash severities further improves the accuracy of the results.

Aside from improved accuracy, the proposed macro-level model has additional capabilities that can make it more useful for planning and policy purposes compared with traditionally used models. For example, exploring the significance of the correlations across space and time can further provide insights into policy and planning implementations. As an illustration, the results of the case study implied that the unobserved factors increasing the number of fatal crashes in one county will likely increase the number of PDO and injury crashes in the same county. This finding underlies the assumption that traffic crashes are independent and can provide some insights to policy makers in targeting some of the potential shared unobserved factors (e.g., transit ridership) in order to effectively reduce fatal, injury, and PDO crashes simultaneously. In addition, one of the major

strengths of the proposed model with a spatial random effect term is that it can be used to rank areas based on the excess crash risk that can be estimated using the average/expected crash rates.

At the micro level, this dissertation extended previous works in this research topic by establishing comprehensive Bayesian hierarchical logit models; it did so by simultaneously capturing unobserved factors at the crash level, as well as spatial, temporal, and spatio-temporal effects, to obtain robust parameter estimates. The major advantage of this method is that it reveals the covariates that can contribute to crash severity (e.g., occupant's ejection, age, gender) and/or crash occurrence (e.g., presence of curves on the roadway, speeding, alcohol involvement). These findings can help authorities identify and prioritize engineering countermeasures.

Finally, this research developed a novel metric that identifies high-risk areas requiring safety improvement. The proposed metric/measure combines the results from both crash frequency and crash severity models; it has advantages over traditional ranking methods, such as the crash rate and EB methods, because it addresses both system-wide and engineering issues. Therefore, it can be used by transport policy makers to prioritize SWiPs and effectively allocate funding and resources. It should be noted that in most State Departments of Transportation, the distribution of federal Highway Safety Improvement Program funding to each county is generally based on the number of county-level crashes. That is, counties with a higher number of fatal and serious crashes will receive a higher percentage of the funding available. This method may not be accurate or efficient because crash frequency does not account for exposure and, therefore, is not a good indicator of high-risk locations. Even when crash rates that can account for exposure based on VMT, population, and other factors are used instead of crash frequency, the proposed ranking method is still more suitable and has a better ability to identify true high-risk areas, as it measures safety within the area not only based on population but also with respect to additional factors (which may generally reflect the

characteristics of that area) that could affect crashes and/or risk of fatalities. The model can also identify and prioritize high-risk areas by capturing the effects of missing influential factors

6.3. Generalizability and Transferability

The spatial transferability of crash models for the identification of SWiPs is of safety practitioners' interest, as it can help save time and costs associated with model development. However, as discussed by Sikder et al. (2013), from a practical point of view, models cannot be completely transferred with the same specification between states or regions. Nevertheless, the proposed methods and overall framework can be transferred with some modifications, and some states or regions may need to use the estimated models of other places because of data limitations.

Sikder et al. (2013) discussed the theoretical considerations for the transferability of travel forecasting models, which was the basis for defining three aspects of crash models at which transferability should be assessed. These three aspects are model structure, model specification, and parameter estimates; each of these three aspects is discussed herein to provide insights into the potential transferability of the proposed models and the overall framework.

Model structure: The needs of State organizations may vary across regions, depending on the regions' priorities and problems; for example, depending on whether the goal is to allocate funds at the county level or to identify roadway segments that might be in need of interventions, the focus can be in different spatial units, such as county or traffic analysis zones. Therefore, the model structure proposed in this dissertation may not be suitable to use without modifications for any applications. With that said, given the flexibility of Bayesian methods, the basic Bayesian framework that accounts for spatial and temporal effects is scalable. For example, the methods (Poisson lognormal or NB) used to model crash frequency at the county level will likely not be suitable to model crash frequency at the traffic analysis zone level or census tract level, as many areas there

have zero crashes. Therefore, the choice of a more suitable model, such as the Bayesian zero-inflated model, should be explored.

In addition, some States may only need to model winter-related crashes in order to identify a high-risk location during the winter season, or some other States may need to pay explicit attention to pedestrian crashes or crashes associated with different transport modes. Therefore, the distribution of traffic crashes in a different context may not be similar, as these might be overdispersed or might have excess zeros. Therefore, the transferability of the model structure should also be assessed when the base and application context are different.

Model specification: Model specification refers to the explanatory variables included in the model. The transferability of the model specification to different regions may introduce additional approximations because of missing influential factors and unobserved effects. For example, the factors that affect traffic crashes in one State might not have been included in the model developed for WV either because of data limitations or because these data might be irrelevant in the case of WV and, therefore, would potentially introduce irrelevant variable bias. In classical models, ignoring such unobserved influential factors may introduce new errors that can deteriorate the estimation accuracy of the model. However, this concern has been resolved using the framework proposed in this study by accommodating random effect terms that can address unobserved factors at different levels. Therefore, although it is suggested that the proposed specification of the models be assessed for relevance and be enhanced to explore any additional factors that might be considered important in other states when transferred, the models are fully transferable in terms of model specification.

Parameter estimates: Generally, parameter estimates may not be easily transformed to make them suitable for use in other regions because of differences in sampling errors, variable measurement errors, and resident and other relevant regional characteristics. However, in the case of

data limitations for a region, the parameters already estimated from another area may be used by applying various model enhancement techniques. For example, in the Bayesian updating approach Atherton and Ben-Akiva (1976), the prior distribution of parameters in the base region can be incorporated with the sample distribution of the parameter estimates from a small sample in the intended region in order to obtain the posterior estimates.

In addition, states with similar empirical settings could benefit from the empirical results of this dissertation's case study, including the parameter estimates. In other words, the results of the case study could be generalizable to states that are similar with WV in terms of topography, climate condition, and demography. Such states could be, for example, Pennsylvania, Virginia, and Kentucky.

6.4. Limitations and Future Work

This research has some limitations, mostly related to data, which should be noted. In the case of the crash frequency model, population estimates per county were utilized as an exposure variable because of the lack of alternative data. Similar research can be conducted using VMT to represent crash risk exposure more accurately. The intersection density variable used in this study was calculated from a road network shapefile that did not differentiate between signalized and stopped control intersections. Thus, the separate use of signalized and stopped control intersection density in a model, subject to data availability, would help better understand the effect of signalized intersections on traffic crashes. In the case of the crash severity model, the use of a comprehensive crash dataset that includes information pertaining to crashes of any severity (instead of a fatal crash dataset only) is recommended. Using such data, one could estimate the odds of a fatality, conditional on being involved in a crash, and, therefore, provide additional insights into the factors affecting

fatalities. Unfortunately, the West Virginia Department of Highways does not have an accessible crash database, and even most crashes do not have GPS locations.

Another extension of this research could be the development of a joint model that includes both micro- and macro-level data at the census tract level or traffic analysis zones to identify the SWiPs and compare the results with those of the present study in order to make valid inferences. This can be done by including additional factors, such as transit stop density, land use type, and trip production and attractions. Finally, additional data could be collected from states with similar empirical settings, and the results can then be compared so that the transferability of the parameter estimates can be explored.

REFERENCES

- Abdel-Aty, M. A., Chen, C. L., & Schott, J. R. (1998). An assessment of the effect of driver age on traffic accident involvement using log-linear models. *Accident Analysis & Prevention*, 30(6), 851–861.
- Abdel-Aty, M., & Radwan, A. E. (1998). *DEMOGRAPHIC FACTORS AND TRAFFIC CRASHES: PART II-GIS ANALYSIS*. Retrieved from <https://trid.trb.org/view.aspx?id=536571>
- Aguero-Valverde, J. (2013). Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. *Accident Analysis & Prevention*, 59, 365–373.
- Aguero-Valverde, J., & Jovanis, P. (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record: Journal of the Transportation Research Board*, (2136), 82–91.
- Aguero-Valverde, J., & Jovanis, P. (2010). Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. *Transportation Research Record: Journal of the Transportation Research Board*, (2165), 21–32.
- Aguero-Valverde, J., & Jovanis, P. P. (2006a). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, 38(3), 618–625.
- Aguero-Valverde, J., & Jovanis, P. P. (2006b). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, 38(3), 618–625.
- Ahmed, M., Huang, H., Abdel-Aty, M., & Guevara, B. (2011). Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. *Accident Analysis & Prevention*, 43(4), 1581–1589.
- Anselin, L. (2013). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.
- Ashley, D. W., Mullins, R. F., Dente, C. J., Garlow, L., Medeiros, R. S., Atkins, E. V., ... Ferdinand, C. H. (2017). What Are the Costs of Trauma Center Readiness? Defining and Standardizing Readiness Costs for Trauma Centers Statewide. *The American Surgeon*, 83(9), 979–990.
- Atherton, T. J., & Ben-Akiva, M. (1976). *Transferability and updating of disaggregate travel demand models*.
- Banerjee, Sudipto. (2009). Retrieved from <http://www.unc.edu/~rls/s940/Areal.pdf>
- Barua, S., El-Basyouny, K., & Islam, M. T. (2014). A full Bayesian multivariate count data model of collision severity with spatial correlation. *Analytic Methods in Accident Research*, 3, 28–43.

- Bedard, M., Guyatt, G. H., Stones, M. J., & Hirdes, J. P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accident Analysis & Prevention*, 34(6), 717–727.
- Behnood, A., & Mannering, F. L. (2015). The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: some empirical evidence. *Analytic Methods in Accident Research*, 8, 7–32.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.
- Bhat, C. R., Astroza, S., & Lavieri, P. S. (2017). A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. *Analytic Methods in Accident Research*, 16, 1–22.
- Boulieri, A., Liverani, S., Hoogh, K., & Blangiardo, M. (2016). A space–time multivariate Bayesian model to analyse road traffic accidents by severity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/rssa.12178/pdf>
- CDC. (n.d.). Injury Prevention & Control: Data & Statistics (WISQARSTM). Retrieved January 31, 2017, from https://www.cdc.gov/injury/wisqars/overview/key_data.html
- Chen, F., & Chen, S. (2011). Injury severities of truck drivers in single-and multi-vehicle accidents on rural highways. *Accident Analysis & Prevention*, 43(5), 1677–1688.
- Cheng, W., Gill, G. S., Ensich, J. L., Kwong, J., & Jia, X. (2018). Multimodal crash frequency modeling: multivariate space-time models with alternate spatiotemporal interactions. *Accident Analysis & Prevention*, 113, 159–170.
- Chin, H. C., & Quddus, M. A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention*, 35(2), 253–259.
- Chiou, Y.-C., Fu, C., & Chih-Wei, H. (2014). Incorporating spatial dependence in simultaneously modeling crash frequency and severity. *Analytic Methods in Accident Research*, 2, 1–11.
- Comoreanu, A. (2018). Strictest And Most Lenient States on Speeding and Reckless Driving. Retrieved July 9, 2018, from <https://wallethub.com/edu/strictest-and-most-lenient-states-on-speeding/14211/>
- Darwiche, A. (2009). A GIS safety study and a county-level spatial analysis of crashes in the state of Florida.
- Davis, G. (2000). Estimating traffic accident rates while accounting for traffic-volume estimation error: A Gibbs sampling approach. *Transportation Research Record: Journal of the Transportation Research Board*, (1717), 94–101.

- Davis, G., & Yang, S. (2001). Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. *Transportation Research Record: Journal of the Transportation Research Board*, (1746), 84–89.
- Eluru, N., Paleti, R., Pendyala, R., & Bhat, C. (2010). Modeling injury severity of multiple occupants of vehicles: Copula-based multivariate approach. *Transportation Research Record: Journal of the Transportation Research Board*, (2165), 1–11.
- Farewell, V. T., Long, D. L., Tom, B. D. M., Yiu, S., & Su, L. (2017). Two-part and related regression models for longitudinal data. *Annual Review of Statistics and Its Application*, 4, 283–315.
- Federal Highway Administration. (2014a). Engineering Speed Management Countermeasures. Retrieved July 12, 2018, from https://safety.fhwa.dot.gov/speedmgt/ref_mats/eng_count/2014/reducing_crashes.cfm
- Federal Highway Administration. (2014b). Roadway Departure Countermeasures. Retrieved July 13, 2018, from https://safety.fhwa.dot.gov/roadway_dept/rdctrm.cfm
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Chapman & Hall/CRC Boca Raton, FL, USA. Retrieved from <http://amstat.tandfonline.com/doi/full/10.1080/01621459.2014.963405>
- Gross, F., Persaud, B., & Lyon, C. (2010). *A guide to developing quality crash modification factors*.
- Guo, F., Wang, X., & Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis & Prevention*, 42(1), 84–92.
- Hadayeghi, A., Shalaby, A., & Persaud, B. (2003). Macrolevel accident prediction models for evaluating safety of urban transportation systems. *Transportation Research Record: Journal of the Transportation Research Board*, (1840), 87–95.
- Hadi, M. A., Aruldas, J., Chow, L.-F., & Wattleworth, J. A. (1995). Estimating safety effects of cross-section design for various highway types using negative binomial regression. *Transportation Research Record*, 1500, 169.
- Harmon, T., Bahar, G., & Gross, F. (2018). *Crash Costs for Highway Safety Analysis*. United States. Federal Highway Administration.
- Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis & Prevention*, 18(1), 1–12.
- Hauer, E. (1996). Identification of sites with promise. *Transportation Research Record: Journal of the Transportation Research Board*, (1542), 54–60.
- Hauer, E., Harwood, D., Council, F., & Griffith, M. (2002). Estimating safety by the empirical Bayes method: a tutorial. *Transportation Research Record: Journal of the Transportation Research Board*, (1784), 126–131.

- Hauer, E., & Persaud, B. N. (1987). *How to estimate the safety of rail-highway grade crossings and the safety effects of warning devices.*
- Higle, J. L., & Witkowski, J. M. (1988). *Bayesian Identification of Hazardous Locations (with Discussion and Closure).*
- Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention, 42*(6), 1556–1565.
- Huang, H., Abdel-Aty, M., & Darwiche, A. (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record: Journal of the Transportation Research Board, (2148)*, 27–37.
- Huang, H. F., Schneider, R. J., Zegeer, C. V., Amerlynck, A. J., & Lacy, J. K. (2002). *Identification of Severe Crash Factors and Countermeasures in North Carolina-Final Report.* FHWA/NC/2001-03, University of North Carolina.
- Jang, H., Lee, S., & Kim, S. W. (2010). Bayesian analysis for zero-inflated regression models with the power prior: Applications to road safety countermeasures. *Accident Analysis & Prevention, 42*(2), 540–547.
- Jonathan, A.-V., Wu, K.-F. K., & Donnell, E. T. (2016). A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accident Analysis & Prevention, 87*, 8–16.
- Kahane, C. J. (2013). *Injury vulnerability and effectiveness of occupant protection technologies for older occupants and women.*
- Kant, E. J. (2008). The Florida Safety Circuit Rider Program: Successes and Challenges. In *ITE 2008 Technical Conference and Exhibit*. Retrieved from <https://trid.trb.org/view.aspx?id=868742>
- Khan, G., Qin, X., & Noyce, D. A. (2008). Spatial analysis of weather crash patterns. *Journal of Transportation Engineering, 134*(5), 191–202.
- Khattak, A. J., Schneider, R. J., & Targa, F. (2003). Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions. *Transportation Research Record, 40*. Retrieved from http://www.ltrc.lsu.edu/TRB_82/TRB2003-000331.pdf
- Kim, K., Brunner, I., & Yamashita, E. (2006). Influence of land use, population, employment, and economic activity on accidents. *Transportation Research Record: Journal of the Transportation Research Board, (1953)*, 56–64.
- Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis & Prevention, 34*(3), 313–321.
- Krull, K., Khattak, A., & Council, F. (2000). Injury effects of rollovers and events sequence in single-vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board, (1717)*, 46–54.

- Labadie, M. J., & Barbaresso, J. C. (1982). *Development of Priority Program for Roadside Hazard Abatement*.
- Ladron de Guevara, F., Washington, S., & Oh, J. (2004). Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, (1897), 191–199.
- LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis & Prevention*, 32(5), 651–658.
- Lee, C., Hellinga, B., & Saccomanno, F. (2006). Evaluation of variable speed limits to improve traffic safety. *Transportation Research Part C: Emerging Technologies*, 14(3), 213–228.
- Lee, J., Abdel-Aty, M., & Choi, K. (2014). Analysis of residence characteristics of at-fault drivers in traffic crashes. *Safety Science*, 68, 6–13.
- Lee, J., Abdel-Aty, M., Choi, K., & Huang, H. (2015). Multi-level hot zone identification for pedestrian safety. *Accident Analysis & Prevention*, 76, 64–73.
- Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis & Prevention*, 27(5), 663–674.
- Liu, C., & Sharma, A. (2018). Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Analytic Methods in Accident Research*, 17, 14–31.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291–305.
- Lord, D., Smiley, A., & Haroun, A. (1998). Pedestrian accidents with left-turning traffic at signalized intersections: Characteristics, human factors, and unconsidered issues. In *77th Annual Transportation Research Board Meeting, Washington, DC*.
- 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.
- Loukaitou-Sideris, A., Liggett, R., & Sung, H.-G. (2007). Death on the crosswalk A study of pedestrian-automobile collisions in los angeles. *Journal of Planning Education and Research*, 26(3), 338–351.
- Ma, J., & Kockelman, K. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record: Journal of the Transportation Research Board*, (1950), 24–34.

- Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention, 40*(3), 964–975.
- Mamun, M. A. A. (2014). Zero-inflated regression models for count data: an application to under-5 deaths.
- 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.
- Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research, 11*, 1–16.
- Meng, F., Xu, P., Wong, S. C., Huang, H., & Li, Y. C. (2017). Occupant-level injury severity analyses for taxis in Hong Kong: A Bayesian space-time logistic model. *Accident Analysis & Prevention, 108*, 297–307.
- Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention, 26*(4), 471–482.
- Miaou, S.-P., Hu, P. S., Wright, T., Rathi, A. K., & Davis, S. C. (1992). Relationship between truck accidents and highway geometric design: a Poisson regression approach. *Transportation Research Record, (1376)*.
- 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*, 33–58.
- Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway accident severities and the mixed logit model: an exploratory empirical analysis. *Accident Analysis & Prevention, 40*(1), 260–266.
- National Highway Traffic Safety Administration. (2016). Fatality Analysis Reporting System Encyclopedia. Retrieved July 15, 2018, from <https://www-fars.nhtsa.dot.gov/QueryTool/QuerySection/SelectCriteria.aspx>
- National Safety Council. (2016). Retrieved from <http://www.nsc.org/NewsDocuments/2017/12-month-estimates.pdf>
- Noland, R. B., & Oh, L. (2004a). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. *Accident Analysis & Prevention, 36*(4), 525–532.
- Noland, R. B., & Quddus, M. A. (2004). A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention, 36*(6), 973–984.
- O'donnell, C. J., & Connor, D. H. (1996). Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. *Accident Analysis & Prevention, 28*(6), 739–753.

- Pawlovich, M. D. (2003). Evaluating traffic safety network screening: An initial framework utilizing the hierarchical Bayesian philosophy.
- Persaud, B., Lyon, C., & Nguyen, T. (1999). Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. *Transportation Research Record: Journal of the Transportation Research Board*, (1665), 7–12.
- Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, 122(2), 105–113.
- Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. *Accident Analysis & Prevention*, 40(4), 1486–1497.
- Rhee, K.-A., Kim, J.-K., Lee, Y., & Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis & Prevention*, 91, 190–199.
- Richardson, S. (1992). *Statistical methods for geographical correlation studies*. Oxford: University Press.
- Romano, E., Kelley-Baker, T., & Voas, R. B. (2008). Female involvement in fatal crashes: Increasingly riskier or increasingly exposed? *Accident Analysis & Prevention*, 40(5), 1781–1788.
- Savolainen, P., & Mannering, F. (2007). Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. *Accident Analysis & Prevention*, 39(5), 955–963.
- Shaheed, M. S., Gkritza, K., Carriquiry, A. L., & Hallmark, S. L. (2016). Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach. *Analytic Methods in Accident Research*, 11, 33–47.
- Shankar, V., Milton, J., & Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. *Accident Analysis & Prevention*, 29(6), 829–837.
- Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382–391.
- Sikder, S., Pinjari, A. R., Srinivasan, S., & Nowrouzian, R. (2013). Spatial transferability of travel forecasting models: a review and synthesis. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, 5(2–3), 104–128.
- Smeed, R. J. (1949). Some statistical aspects of road safety research. *Journal of the Royal Statistical Society. Series A (General)*, 112(1), 1–34.
- Song, J. J., Ghosh, M., Miaou, S., & Mallick, B. (2006). Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis*, 97(1), 246–273.

- Soro, W. L., Zhou, Y., & Wayoro, D. (2017). Crash rates analysis in China using a spatial panel model. *IATSS Research*, 41(3), 123–128.
- 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 (Statistical Methodology)*, 64(4), 583–639.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS user manual*. version.
- Spring, G. S., & Hummer, J. (1995). Identification of hazardous highway locations using knowledge-based GIS: a case study. *Transportation Research Record*, 1497, 83–90.
- Srinivasan, K. (2002). Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. *Transportation Research Record: Journal of the Transportation Research Board*, (1784), 132–141.
- Tay, R., Choi, J., Kattan, L., & Khan, A. (2011). A multinomial logit model of pedestrian–vehicle crash severity. *International Journal of Sustainable Transportation*, 5(4), 233–249.
- Taylor, J. I., & Thompson, H. T. (1977). *Identification of hazardous locations* (No. FHWA-RD-77-83).
- Truong, L. T., Kieu, L.-M., & Vu, T. A. (2016). Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. *Accident Analysis & Prevention*, 94, 153–161.
- Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. *Austrian Journal of Statistics*, 31(3), 221–229.
- U.S. Census Bureau. (2015). 2010-2015 American Community Survey 6-Year Estimates. American Factfinder
- U.S. Department of Transportation. (2014). *Transportation for a new generation: Strategic Plan / Fiscal Years 2014-18* (FY 2014-2018 Strategic Plan). Retrieved from <http://www.dot.gov/administrations/office-policy/fy-2014-2018-strategic-plan>
- U.S. Government Accountability Office. (2003). *Research Continues on a Variety of Factors That Contribute to Motor Vehicle Crashes*. Report GAO-03-436. US General Accounting Office. Retrieved from <http://161.203.16.70/assets/240/237711.pdf>
- Usman, T., Fu, L., & Miranda-Moreno, L. F. (2016). Injury severity analysis: comparison of multilevel logistic regression models and effects of collision data aggregation. *Journal of Modern Transportation*, 24(1), 73–87.
- Vanlaar, W. (2005). Multilevel modeling in traffic safety research: two empirical examples illustrating the consequences of ignoring hierarchies. *Traffic Injury Prevention*, 6(4), 311–316.

- Wakefield, J. C., Best, N. G., & Waller, L. (2000). Bayesian approaches to disease mapping. *Spatial Epidemiology: Methods and Applications*, 104–127.
- Wang, J., & Huang, H. (2016). Road network safety evaluation using Bayesian hierarchical joint model. *Accident Analysis & Prevention*, 90, 152–158.
- Wang, X., & Abdel-Aty, M. (2006). Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis & Prevention*, 38(6), 1137–1150.
- Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, 60, 71–84.
- Washington, S. (2006). *Incorporating safety into long-range transportation planning*. Transportation Research Board.
- West Virginia Department of Health and Human Resources. (2016). 2016 County Epidemiological Profiles. Retrieved July 15, 2018, from <https://dhhr.wv.gov/bhhf/Pages/2016-County-Epidemiological-Profiles-.aspx>
- West Virginia Department of Health and Human Resources. (2018). *Paramedic Treatment Protocols*. Retrieved from <https://www.wvoems.org/media/355789/paramedic%20protocols%202018.pdf>
- WV GIS Technical Center: GIS Data Clearinghouse. (2016). Retrieved June 15, 2017, from <http://wvgis.wvu.edu/data/data.php>
- Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, 41(1), 137–145.
- World Health Organization. (2008). *Speed management: a road safety manual for decision-makers and practitioners*.
- Yuan, Q., Lu, M., Theofilatos, A., & Li, Y.-B. (2017). Investigation on occupant injury severity in rear-end crashes involving trucks as the front vehicle in Beijing area, China. *Chinese Journal of Traumatology*, 20(1), 20–26.
- Zhou, M., & Sisiopiku, V. (1997). Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record: Journal of the Transportation Research Board*, (1581), 47–52.

APPENDIX

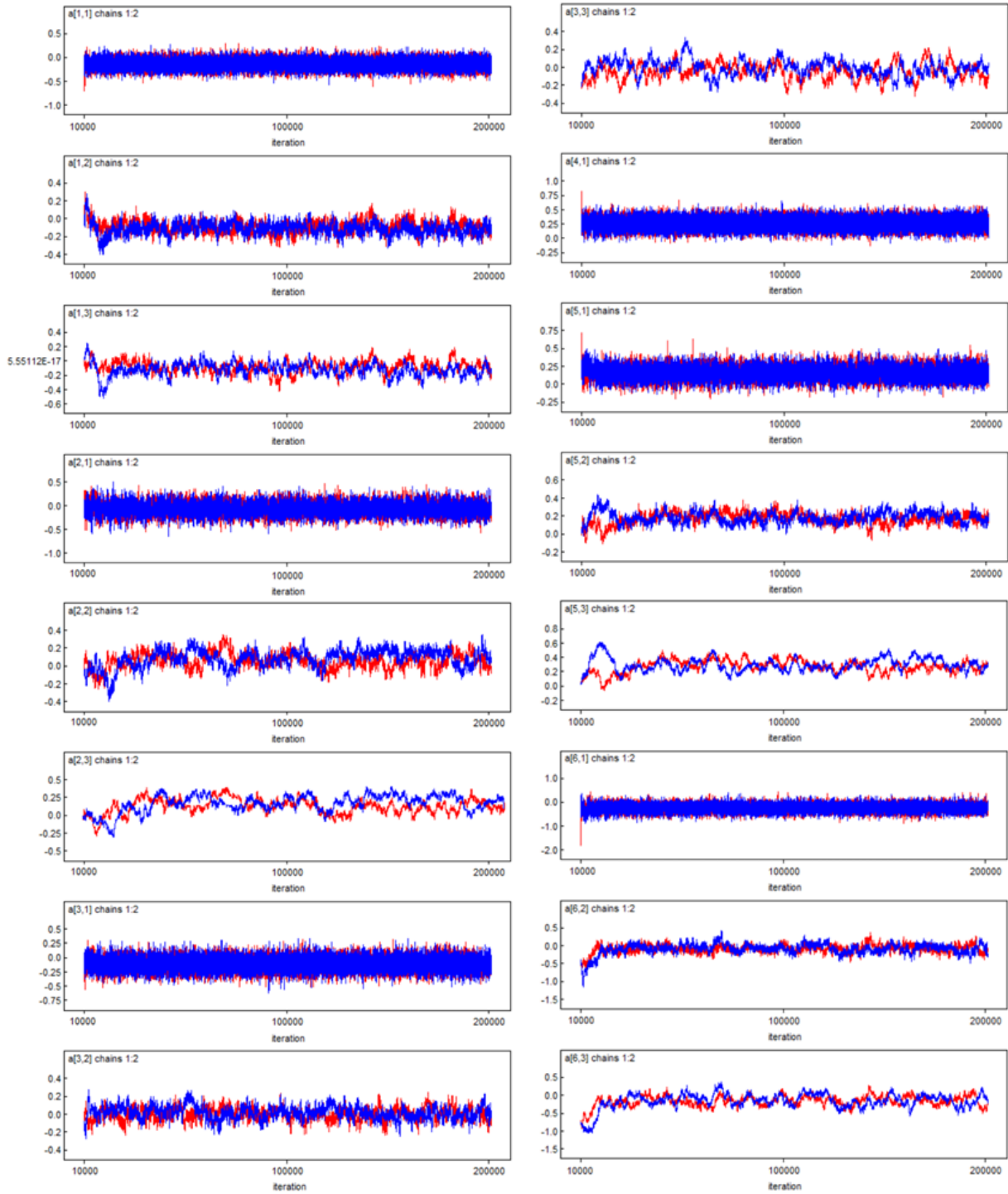


FIGURE 1 MCMC output for parameters under Model 9, macro-level analysis

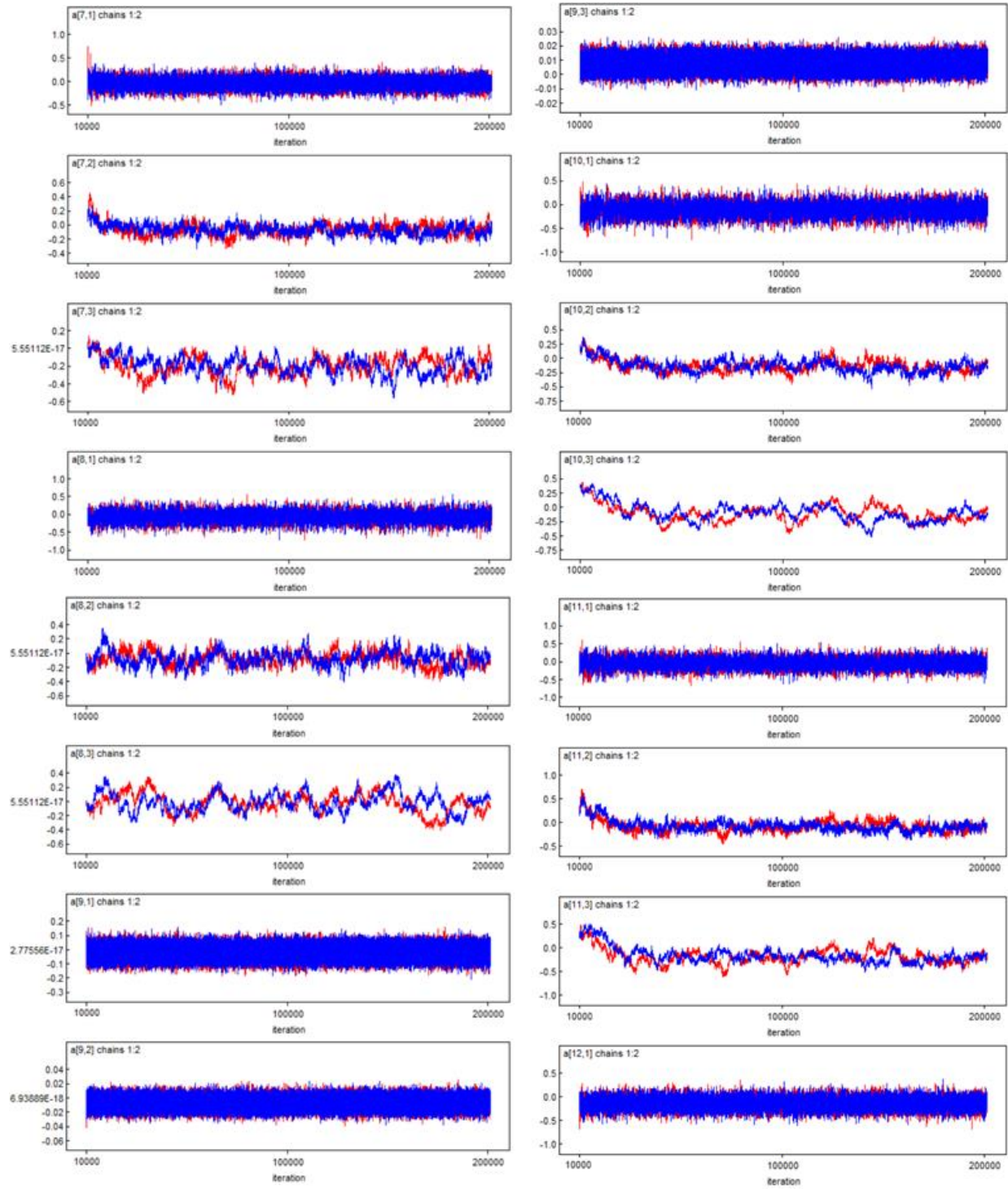


FIGURE 1 MCMC output for parameters under Model 9, macro-level analysis (continued)

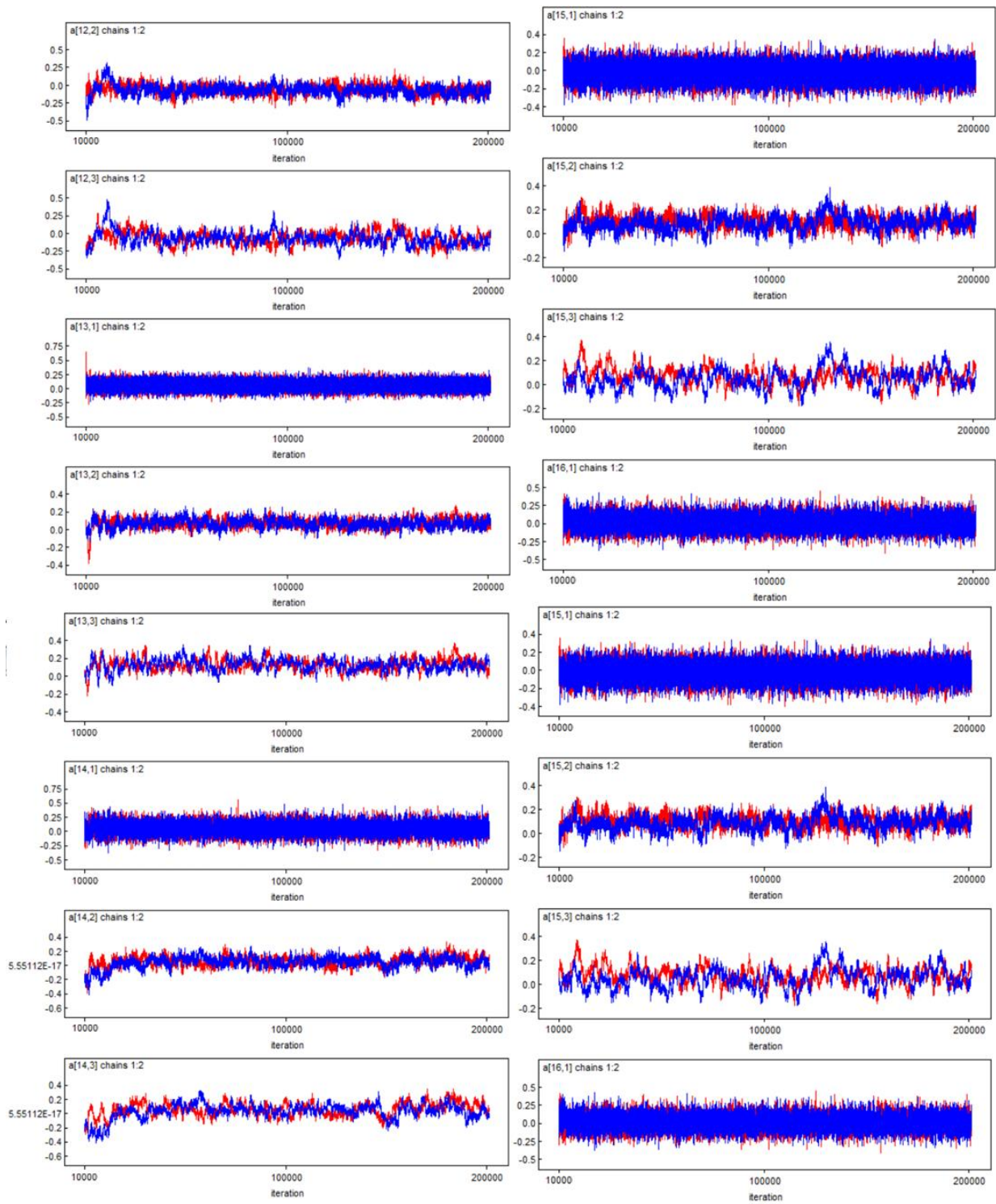


FIGURE 1 MCMC output for parameters under Model 9, macro-level analysis (continued)

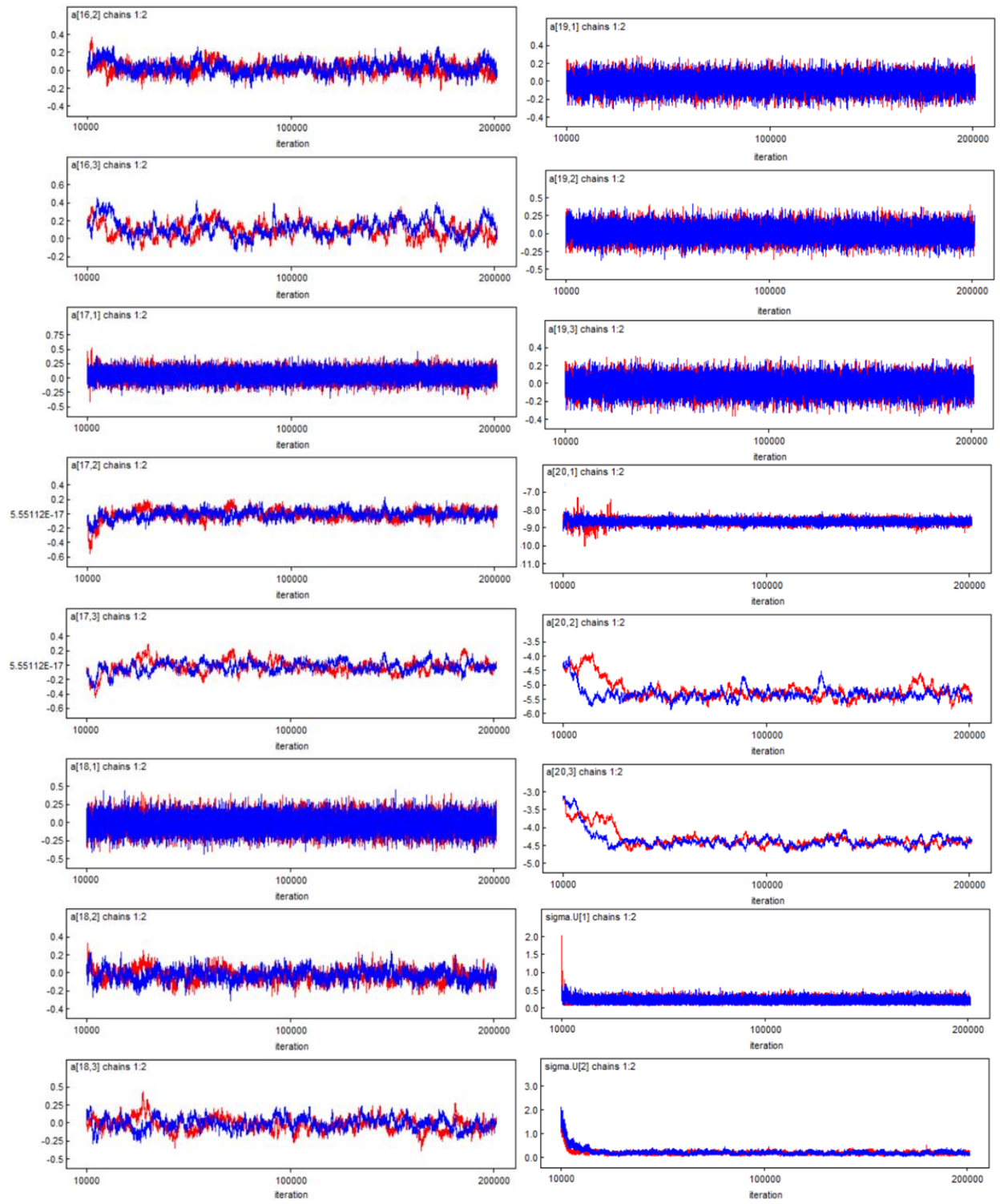


FIGURE 1 MCMC output for parameters under Model 9, macro-level analysis (continued)

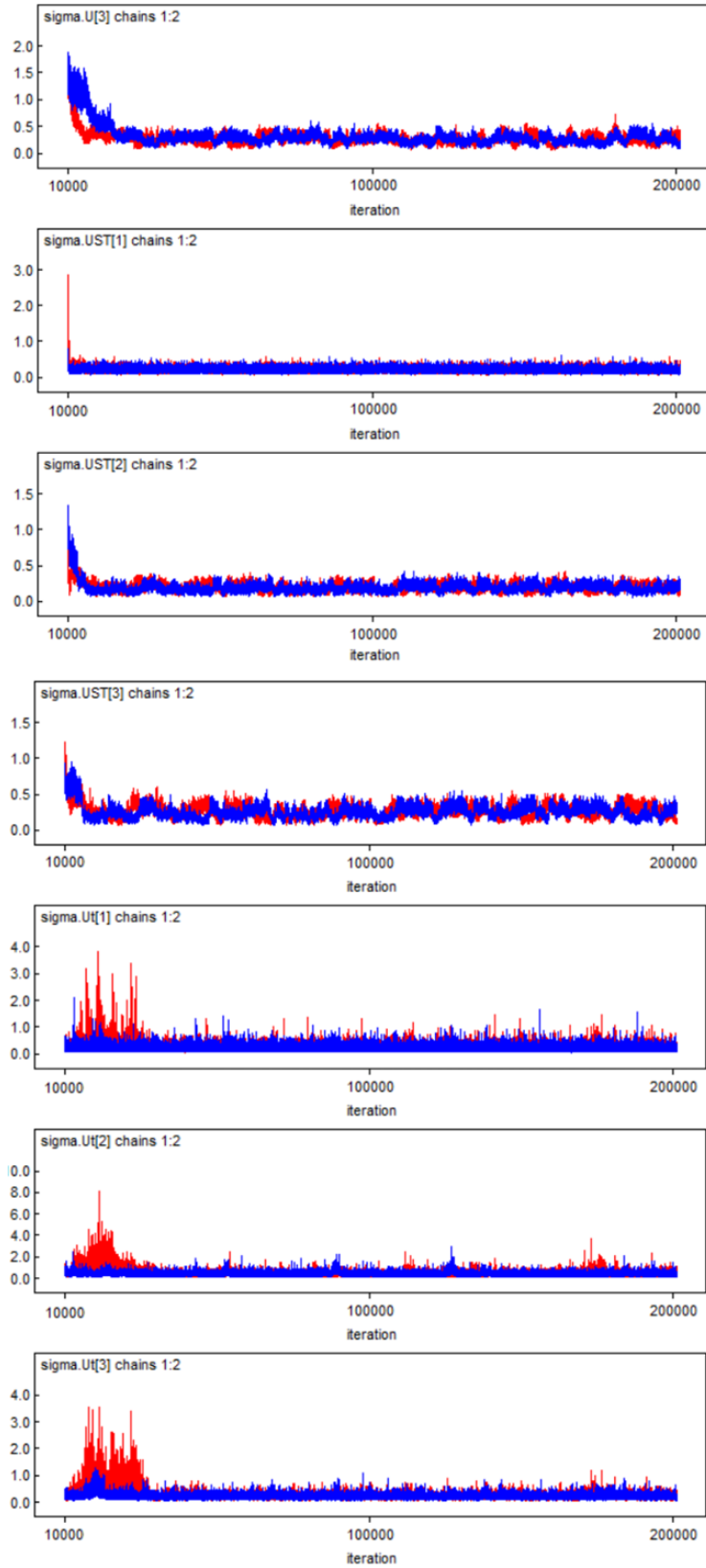


FIGURE 1 MCMC output for parameters under Model 9, macro-level analysis (continued)

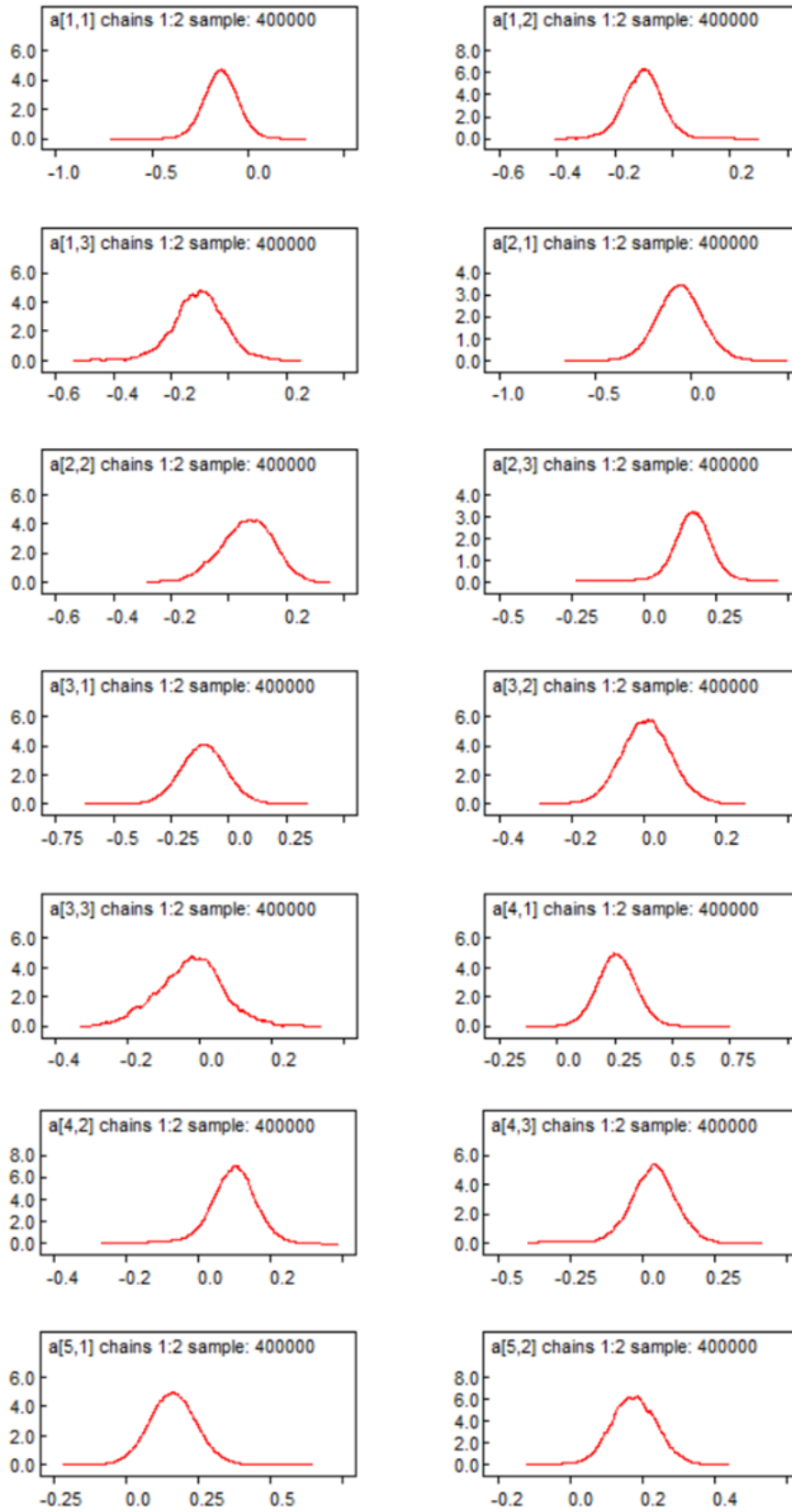


FIGURE 2 Posterior density curves of parameters under Model 9 macro-level analysis

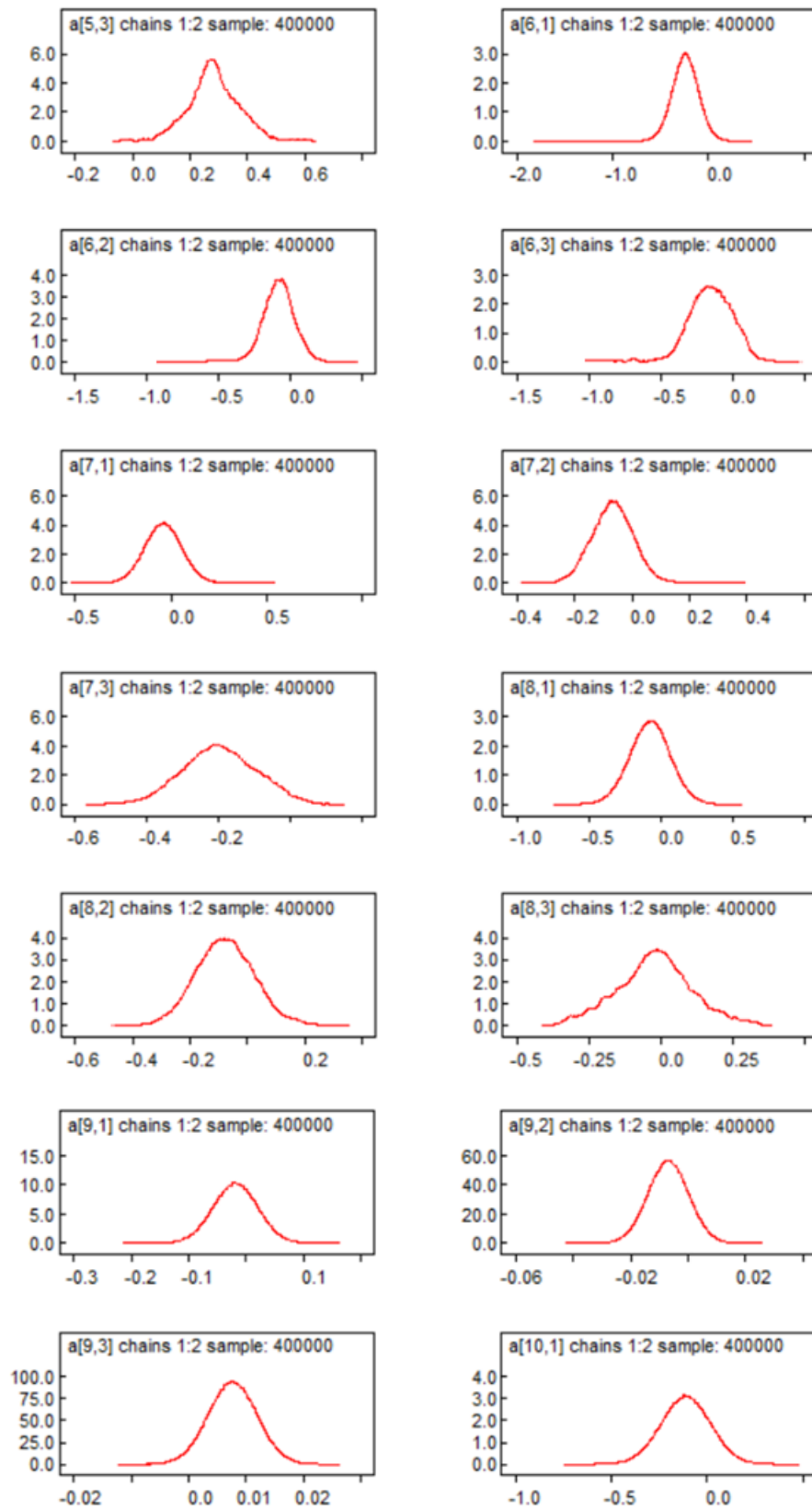


FIGURE 2 Posterior density curves of parameters under Model 9 macro-level analysis (continued)

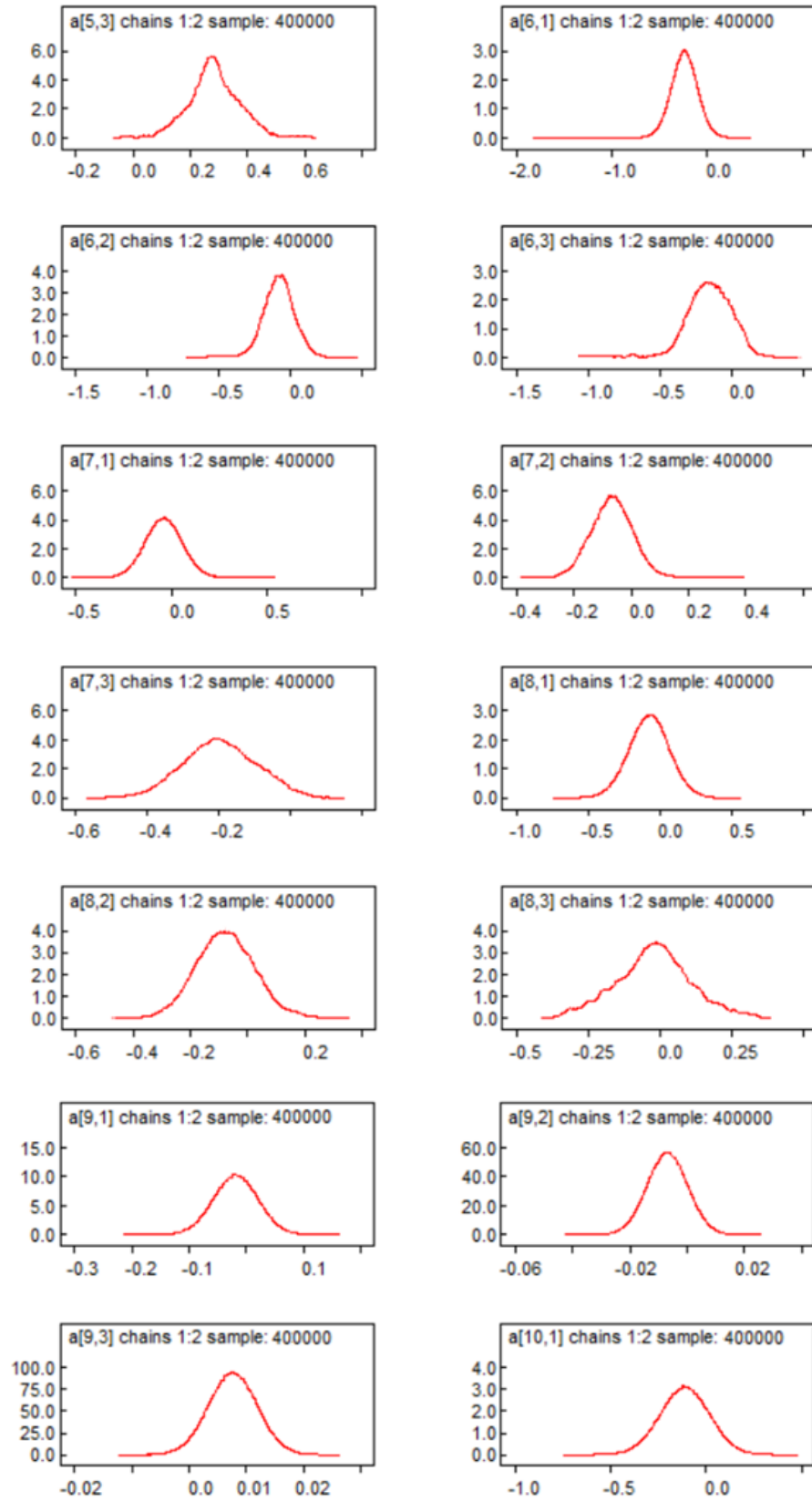


FIGURE 2 Posterior density curves of parameters under Model 9 macro-level analysis (continued)

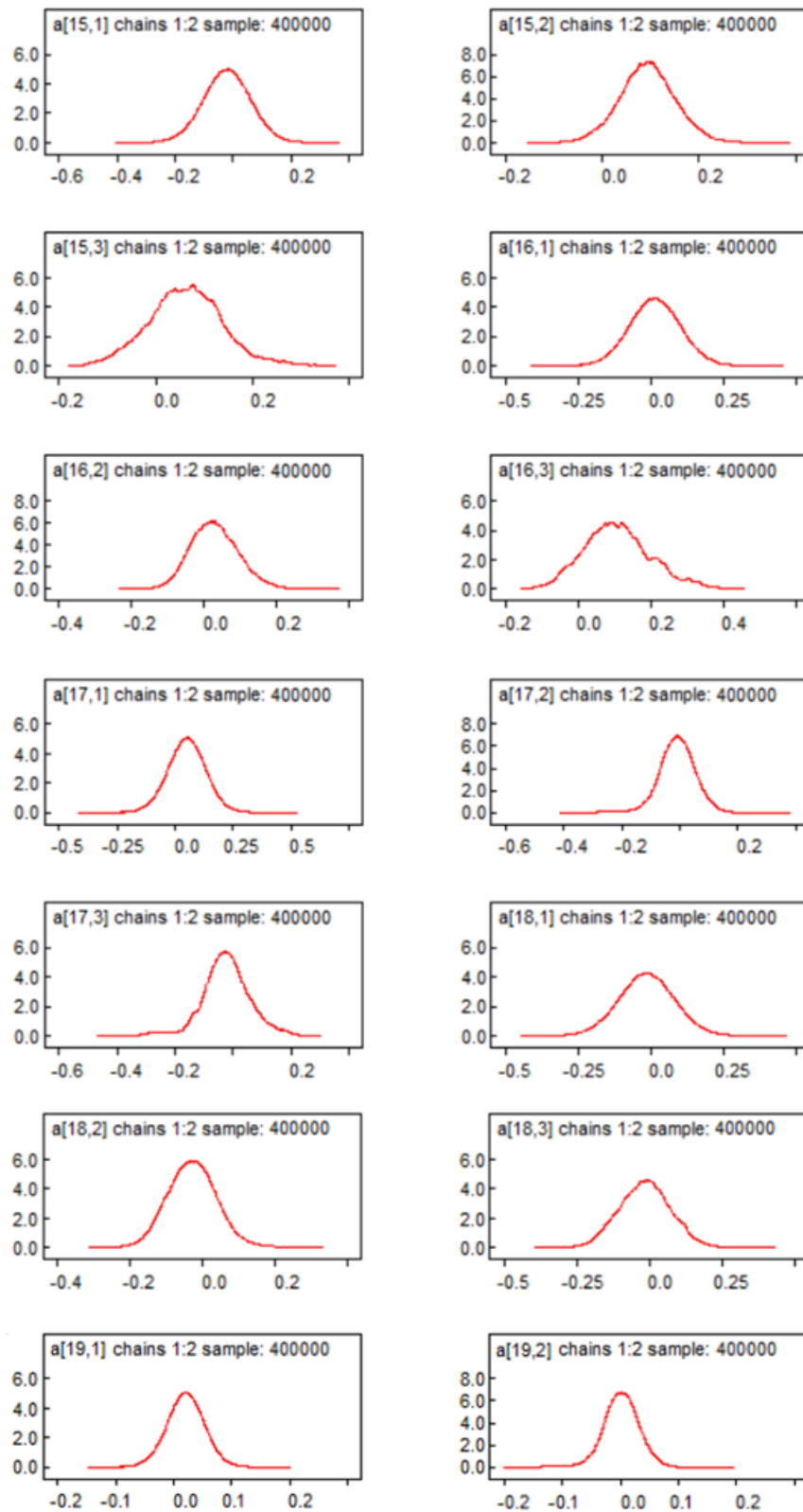


FIGURE 2 Posterior density curves of parameters under Model 9 macro-level analysis (continued)

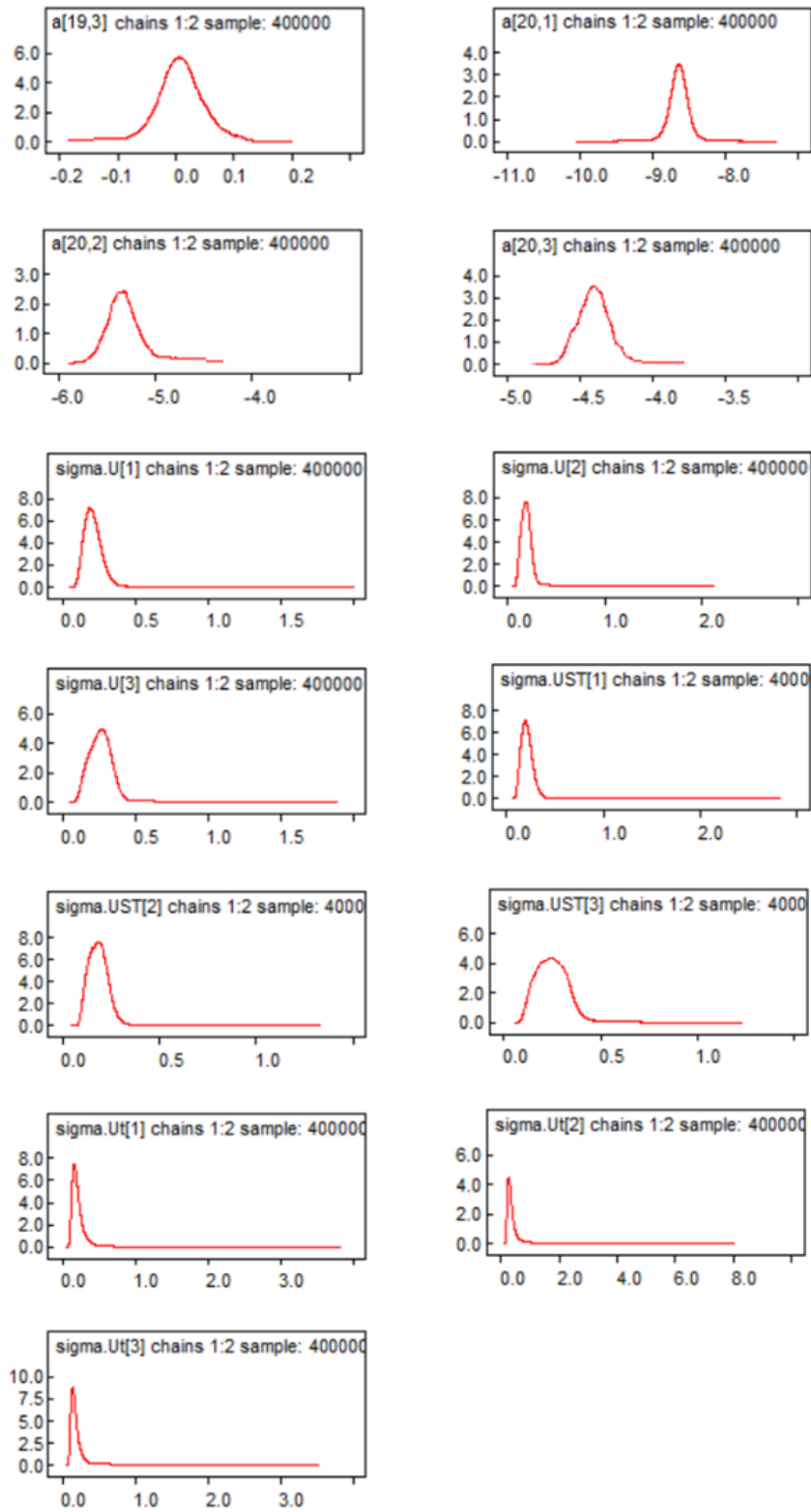


FIGURE 2 Posterior density curves of parameters under Model 9 macro-level analysis (continued)

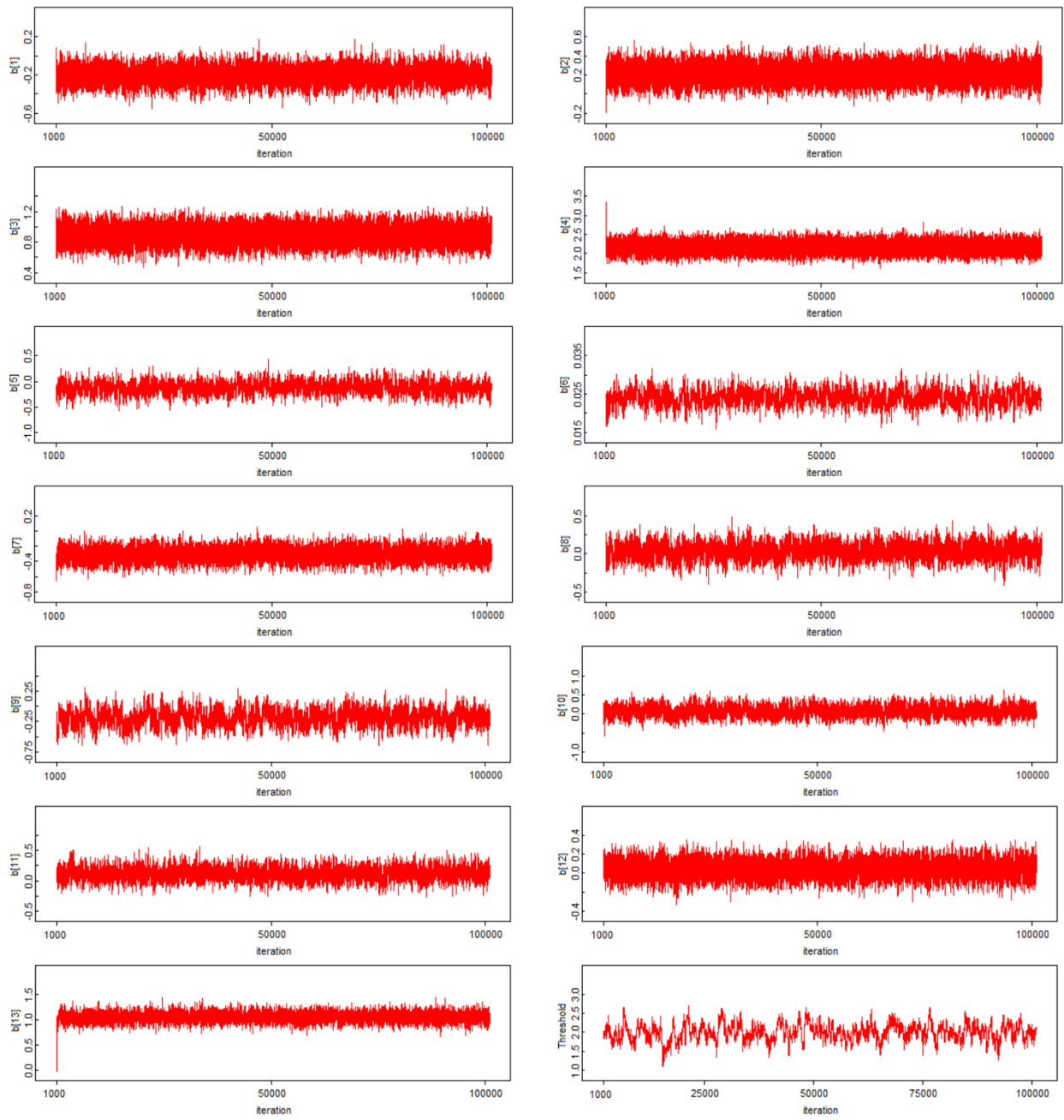


Figure 3 MCMC output for parameters in full ordered logit model, micro-level analysis

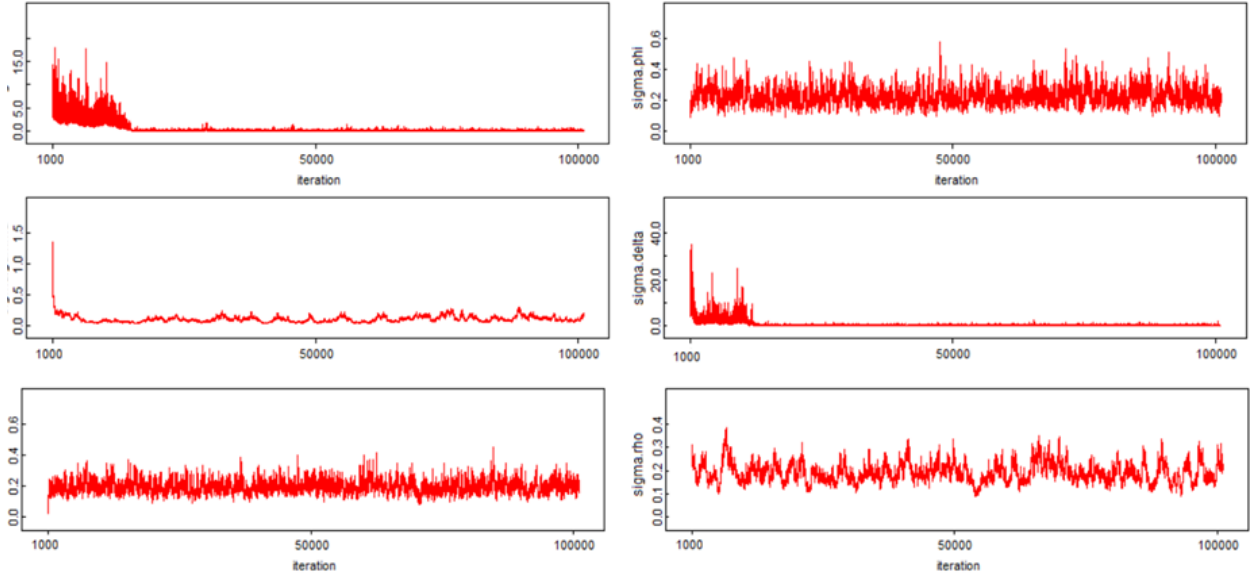


Figure 3 MCMC output for parameters in full ordered logit model, micro-level analysis (continued)

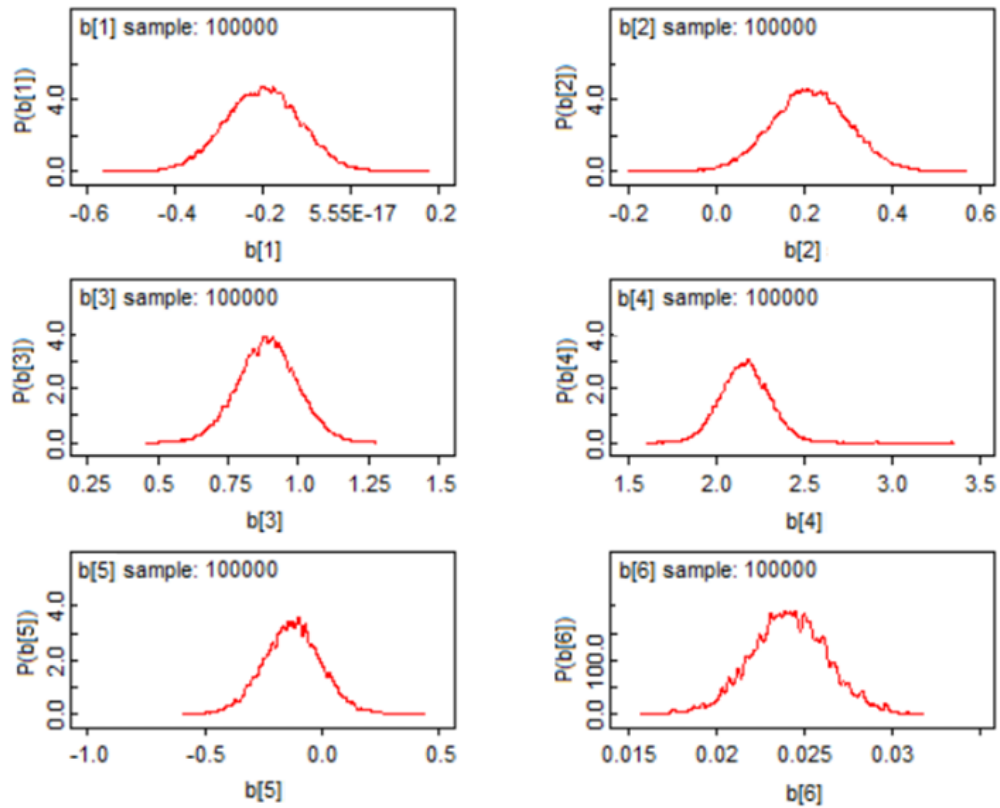


FIGURE 4 Posterior density curves of parameters under Model 4, micro-level analysis

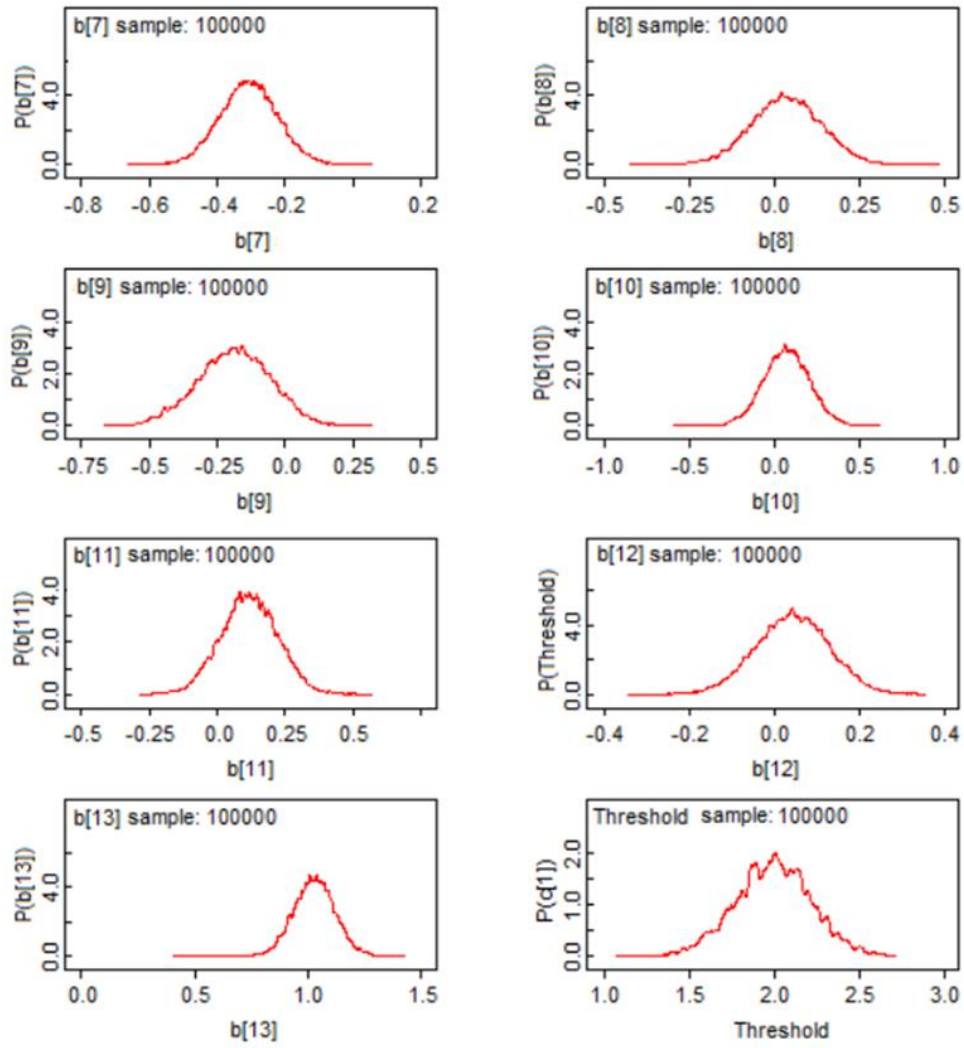


FIGURE 4 Posterior density curves of parameters under Model 4, micro-level analysis (continued)