Accident Analysis and Prevention 41 (2009) 217–226

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Markov switching negative binomial models: An application to vehicle accident frequencies

Nataliya V. Malyshkina*, Fred L. Mannering, Andrew P. Tarko

School of Civil Engineering, 550 Stadium Mall Drive, Purdue University, West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:
Received 13 August 2008
Received in revised form 28 October 2008
Accepted 7 November 2008

Keywords:
Accident frequency
Negative binomial
Count data model
Markov switching
Bayesian
MCMC

ABSTRACT

In this paper, two-state Markov switching models are proposed to study accident frequencies. These models assume that there are two unobserved states of roadway safety, and that roadway entities (roadway segments) can switch between these states over time. The states are distinct, in the sense that in the different states accident frequencies are generated by separate counting processes (by separate Poisson or negative binomial processes). To demonstrate the applicability of the approach presented herein, two-state Markov switching negative binomial models are estimated using five-year accident frequencies on Indiana interstate highway segments. Bayesian inference methods and Markov Chain Monte Carlo (MCMC) simulations are used for model estimation. The estimated Markov switching models result in a superior statistical fit relative to the standard (single-state) negative binomial model. It is found that the more frequent state is safer and it is correlated with better weather conditions. The less frequent state is found to be less safe and to be correlated with adverse weather conditions.

1. Introduction

Vehicle accidents place an incredible social and economic burden on society. As a result, considerable research has been conducted on understanding and predicting accident frequencies (the number of accidents occurring on roadway segments over a given time period). Because accident frequencies are non-negative integers, count data models are a reasonable statistical modeling approach (Washington et al., 2003). Simple modeling approaches include Poisson models and negative binomial (NB) models (Hadi et al., 1995; Shankar et al., 1995; Poch and Mannering, 1996; Miaou and Lord, 2003). These models assume a single process for accident data generation (a Poisson process or a negative binomial process) and involve a nonlinear regression of the observed accident frequencies on various roadway-segment characteristics (such as roadway geometric and environmental factors). Because a preponderance of zero-accident observations is often observed in empirical data, Miaou (1994), Shankar et al. (1997) and others have applied zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models for predicting accident frequencies. Zero-inflated models assume a two-state process for accident data generation: one state is assumed to be safe with zero accidents (over the duration of time being considered) and the other state is assumed to be unsafe with a possibility of nonzero accident frequencies. In zero-inflated models, individual roadway segments are assumed to be in either safe or unsafe state for each observation period. While the application of zero-inflated models often provides a better statistical fit of observed accident frequency data, the applicability of these models has been questioned by Lord et al. (2005, 2007). In particular, Lord et al. (2005, 2007) argue that zero-inflated models do not explicitly consider a likely possibility for roadway segments to change in time from one state to another.

In this paper, two-state Markov switching count data models are explored as a method for studying accident frequencies. These models assume Markov switching (over time) between two unobserved states of roadway safety.[1] There are a number of reasons to expect the existence of multiple states. First, the safety of roadway segments is likely to vary under different environmental conditions, driver reactions and other factors that may not necessarily be available to the analyst. For an illustration, consider Fig. 1, which shows weekly time series of the number of accidents on selected Indiana interstate segments during the 1995–1999 time interval. The figure shows that the number of accidents per week fluctuates widely over time. Thus, under different conditions, roads can become considerably more or less safe. As a result, it is reasonable to assume that there exist two or more states of roadway safety.

* Corresponding author.
E-mail addresses: nmalyshk@purdue.edu (N.V. Malyshkina), flm@ecn.purdue.edu (F.L. Mannering), tarko@ecn.purdue.edu (A.P. Tarko).

[1] In fact, there may be more than two states but, for illustration purposes, the two-state case will be considered in this paper. Extending the approach to the possibility of additional states would significantly complicate the model structure. However, once this extension were done, additional states could be empirically tested.

0001-4575/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2008.11.001

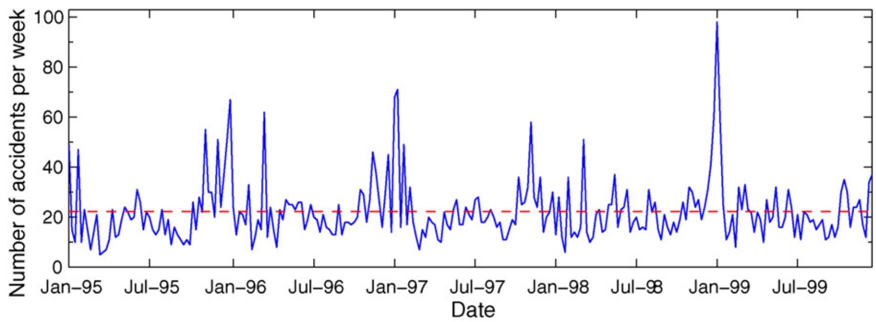

Fig. 1. Weekly accident frequencies on the sample of Indiana interstate segments from 1995 to 1999 (the horizontal dashed line shows the average value).

These states can help account for the existence of numerous unidentified and/or unobserved factors (unobserved heterogeneity) that influence roadway safety. Markov switching models are designed to account for unobserved states in a convenient and feasible way. In fact, these models have been successfully applied in other scientific fields. For example, two-state Markov switching autoregressive models have been used in economics, where the two states are usually identified as economic recession and expansion (Hamilton, 1989; McCulloch and Tsay, 1994; Tsay, 2002). In the context of finite mixture models of accident frequencies, unobserved states were considered by Park and Lord (2008).[2]

Another reason to expect the existence of multiple states is the empirical success of zero-inflated models, which are predicated on the existence of two-state process: a safe and an unsafe state (see Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002). Markov switching can be viewed as an extension of previous work on zero-inflated models, in that it relaxes the assumption that a safe state exists [which has been brought up as a concern, see Lord et al., 2005, 2007] and instead considers two significantly different unsafe states. In addition, in contrast to zero-inflated models, Markov switching explicitly considers the process of switching by roadway segments between the states over time.

2. Model specification

Our presentation of Markov switching negative binomial models is similar to that of Markov switching autoregressive models in econometrics (McCulloch and Tsay, 1994; Tsay, 2002). Markov switching models are parametric and can be fully specified by a likelihood function $f(\mathbf{Y}|\mathbf{\Theta}, \mathcal{M})$, which is the conditional probability distribution of the vector of all observations $\mathbf{Y}$, given the vector of all parameters $\mathbf{\Theta}$ of model $\mathcal{M}$. In our study, we observe the number of accidents $A_{t,n}$ that occur on the nth roadway segment during time period t. Thus $\mathbf{Y} = \{A_{t,n}\}$ includes all accidents observed on all roadway segments over all time periods ($n = 1, 2, \ldots, N_t$ and $t = 1, 2, \ldots, T$). Model $\mathcal{M} = \{M, \mathbf{X}_{t,n}\}$ includes the model's name M (for example, M = "negative binomial") and the vector $\mathbf{X}_{t,n}$ of all roadway segment characteristic variables (segment length, curve characteristics, grades, pavement properties, and so on).

To define the likelihood function, we first introduce an unobserved (latent) state variable s_t, which determines the state of all roadway segments during time period t. At each t, the state variable s_t can assume only two values: $s_t = 0$ corresponds to one state and $s_t = 1$ corresponds to the other state. The state variable s_t is assumed to follow a stationary two-state Markov chain process in time,[3] which can be specified by time-independent transition probabilities as

$$P(s_{t+1} = 1|s_t = 0) = p_{0\to1}, \qquad P(s_{t+1} = 0|s_t = 1) = p_{1\to0}. \tag{1}$$

Here, for example, $P(s_{t+1} = 1|s_t = 0)$ is the conditional probability of $s_{t+1} = 1$ at time $t + 1$, given that $s_t = 0$ at time t. Note that $P(s_{t+1} = 0|s_t = 0) = 1 - p_{0\to1}$ and $P(s_{t+1} = 1|s_t = 1) = 1 - p_{1\to0}$. Transition probabilities $p_{0\to1}$ and $p_{1\to0}$ are unknown parameters to be estimated from accident data. The stationary unconditional probabilities $\bar{p}_0$ and $\bar{p}_1$ of states $s_t = 0$ and $s_t = 1$ are[4]

$$\begin{aligned} \bar{p}_0 &= \frac{p_{1\to0}}{p_{0\to1} + p_{1\to0}} \quad \text{for state} \quad s_t = 0, \\ \bar{p}_1 &= \frac{p_{0\to1}}{p_{0\to1} + p_{1\to0}} \quad \text{for state} \quad s_t = 1. \end{aligned} \tag{2}$$

Without loss of generality, we assume that (on average) state $s_t = 0$ occurs more or equally frequently than state $s_t = 1$. Therefore, $\bar{p}_0 \geq \bar{p}_1$, and from Eq. (2) we obtain restriction[5]

$$p_{0\to1} \leq p_{1\to0}. \tag{3}$$

We refer to states $s_t = 0$ and $s_t = 1$ as "more frequent" and "less frequent" states respectively.

Next, consider a two-state Markov switching negative binomial (MSNB) model that assumes negative binomial data-generating processes in each of the two states. With this, the probability of $A_{t,n}$ accidents occurring on roadway segment n during time period t is

$$P^{(A)}_{t,n} = \frac{\Gamma(A_{t,n} + 1/\alpha_t)}{\Gamma(1/\alpha_t)A_{t,n}!}\left(\frac{1}{1 + \alpha_t\lambda_{t,n}}\right)^{1/\alpha_t}\left(\frac{\alpha_t\lambda_{t,n}}{1 + \alpha_t\lambda_{t,n}}\right)^{A_{t,n}}, \tag{4}$$

$$\begin{aligned} \lambda_{t,n} &= \begin{cases} \exp(\boldsymbol{\beta}'_{(0)}\mathbf{X}_{t,n}) & \text{if} \quad s_t = 0 \\ \exp(\boldsymbol{\beta}'_{(1)}\mathbf{X}_{t,n}) & \text{if} \quad s_t = 1 \end{cases}, \\ \alpha_t &= \begin{cases} \alpha_{(0)} & \text{if} \quad s_t = 0 \\ \alpha_{(1)} & \text{if} \quad s_t = 1 \end{cases}, \\ t &= 1, 2, \ldots, T, \quad n = 1, 2, \ldots, N_t. \end{aligned} \tag{5}$$

Here, Eq. (4) is the standard negative binomial probability mass function (Washington et al., 2003), $\Gamma(\ )$ is the gamma function, prime means transpose (so $\boldsymbol{\beta}'_{(0)}$ is the transpose of $\boldsymbol{\beta}_{(0)}$), N_t is the number of roadway segments observed during time period t, and T is the total number of time periods. Parameter vectors $\boldsymbol{\beta}_{(0)}$ and

[2] Recently, Anastasopoulos and Mannering (submitted for publication) applied random parameters count models to the analysis of accident frequencies. Random parameter models can potentially define unique parameters for each roadway segment, but they still assume a single state for each segment. This single-state assumption would also be true for count models with random effects (see Shankar et al., 1998).

[3] Markov property means that the probability distribution of s_{t+1} depends only on the value s_t at time t, but not on the previous history $s_{t-1}, s_{t-2}, \ldots$ (Breiman, 1969). Stationarity of $\{s_t\}$ is in the statistical sense. Below we will relax the assumption of stationarity and discuss a case of time-dependent transition probabilities.

[4] These can be found from stationarity conditions $\bar{p}_0 = (1 - p_{0\to1})\bar{p}_0 + p_{1\to0}\bar{p}_1$, $\bar{p}_1 = p_{0\to1}\bar{p}_0 + (1 - p_{1\to0})\bar{p}_1$ and $\bar{p}_0 + \bar{p}_1 = 1$ (Breiman, 1969).

[5] Without any loss of generality, restriction (3) is introduced for the purpose of avoiding the problem of state label switching $0 \leftrightarrow 1$. This problem would otherwise arise because of the symmetry of Eqs. (1)–(7) under the label switching.

$\boldsymbol{\beta}_{(1)}$, and over-dispersion parameters $\alpha_{(0)} \geq 0$ and $\alpha_{(1)} \geq 0$ are the unknown estimable parameters of negative binomial models in the two states, $s_t = 0$ and $s_t = 1$ respectively.[6] We set the first component of $\mathbf{X}_{t,n}$ to unity, and, therefore, the first components of $\boldsymbol{\beta}_{(0)}$ and $\boldsymbol{\beta}_{(1)}$ are the intercepts in the two states.

If accident events are assumed to be independent, the likelihood function is

$$f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M}) = \prod_{t=1}^{T}\prod_{n=1}^{N_t} P_{t,n}^{(A)}. \tag{6}$$

Here, because the state variables s_t are unobservable, the vector of all estimable parameters $\boldsymbol{\Theta}$ must include all states, in addition to all model parameters (β, α) and transition probabilities. Thus,

$$\boldsymbol{\Theta} = [\boldsymbol{\beta}'_{(0)}, \alpha_{(0)}, \boldsymbol{\beta}'_{(1)}, \alpha_{(1)}, p_{0\to1}, p_{1\to0}, \mathbf{S}']', \quad \mathbf{S}' = [s_1, \ldots, s_T]. \tag{7}$$

Vector $\mathbf{S}$ has length T and contains all state values. Eqs. (1)–(7) define the two-state Markov switching negative binomial (MSNB) models considered in this study.

3. Model estimation methods

Statistical estimation of Markov switching models is complicated by unobservability of the state variables s_t.[7] As a result, the traditional maximum likelihood estimation (MLE) procedure is of very limited use for Markov switching models. Instead, a Bayesian inference approach is used. Given a model $\mathcal{M}$ with likelihood function $f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})$, the Bayes formula is

$$f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M}) = \frac{f(\mathbf{Y}, \boldsymbol{\Theta}|\mathcal{M})}{f(\mathbf{Y}|\mathcal{M})} = \frac{f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})\pi(\boldsymbol{\Theta}|\mathcal{M})}{\int f(\mathbf{Y}, \boldsymbol{\Theta}|\mathcal{M})\,d\boldsymbol{\Theta}}. \tag{8}$$

Here $f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M})$ is the posterior probability distribution of model parameters $\boldsymbol{\Theta}$ conditional on the observed data $\mathbf{Y}$ and model $\mathcal{M}$. Function $f(\mathbf{Y}, \boldsymbol{\Theta}|\mathcal{M})$ is the joint probability distribution of $\mathbf{Y}$ and $\boldsymbol{\Theta}$ given model $\mathcal{M}$. Function $f(\mathbf{Y}|\mathcal{M})$ is the marginal likelihood function—the probability distribution of data $\mathbf{Y}$ given model $\mathcal{M}$. Function $\pi(\boldsymbol{\Theta}|\mathcal{M})$ is the prior probability distribution of parameters that reflects prior knowledge about $\boldsymbol{\Theta}$. The intuition behind Eq. (8) is straightforward: given model $\mathcal{M}$, the posterior distribution accounts for both the observations $\mathbf{Y}$ and our prior knowledge of $\boldsymbol{\Theta}$. We use the harmonic mean formula to calculate the marginal likelihood $f(\mathbf{Y}|\mathcal{M})$ of data $\mathbf{Y}$ (see Kass and Raftery, 1995) as,

$$f(\mathbf{Y}|\mathcal{M})^{-1} = \int \frac{f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M})}{f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})}\,d\boldsymbol{\Theta} = E[f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})^{-1}|\mathbf{Y}], \tag{9}$$

where $E(\ldots|\mathbf{Y})$ is the posterior expectation (which is calculated by using the posterior distribution).

A full specification of Bayesian methodology and model estimation requires a specification of the prior distribution. We choose the prior distribution $\pi(\boldsymbol{\Theta}|\mathcal{M})$ of model parameters $\boldsymbol{\Theta}$, given by Eq. (7), as

$$\pi(\boldsymbol{\Theta}|\mathcal{M}) = f(\mathbf{S}|p_{0\to1}, p_{1\to0})\pi(p_{0\to1}, p_{1\to0})\prod_{s=0}^{1}\left[\pi(\alpha_{(s)})\prod_{k}\pi(\beta_{(s),k})\right], \tag{10}$$

$$\pi(p_{0\to1}, p_{1\to0}) \propto \pi(p_{0\to1})\pi(p_{1\to0})I(p_{0\to1} \leq p_{1\to0}), \tag{11}$$

$$\begin{aligned} f(\mathbf{S}|p_{0\to1}, p_{1\to0}) &= P(s_1)\prod_{t=2}^{T}P(s_t|s_{t-1}) \propto \prod_{t=2}^{T}P(s_t|s_{t-1}) \\ &= (p_{0\to1})^{n_{0\to1}}(1-p_{0\to1})^{n_{0\to0}}(p_{1\to0})^{n_{1\to0}} \\ &\quad \times (1-p_{1\to0})^{n_{1\to1}} \end{aligned} \tag{12}$$

Here $\beta_{(s),k}$ is the kth component of vector $\boldsymbol{\beta}_{(s)}$. The priors of $\beta_{(s),k}$ are chosen to be normal distributions, $\pi(\beta_{(s),k}) = \mathcal{N}(\mu_k, \Sigma_k)$, and the priors of $\alpha_{(s)}$ are chosen to be log-normal, $\pi(\ln[\alpha_{(s)}]) = \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$. The joint prior of $p_{0\to1}$ and $p_{1\to0}$ is given by Eq. (11), where $\pi(p_{0\to1}) = \mathcal{B}eta(\upsilon_0, \nu_0)$ and $\pi(p_{1\to0}) = \mathcal{B}eta(\upsilon_1, \nu_1)$ are beta distributions, and function $I(p_{0\to1} \leq p_{1\to0})$ is equal to unity if the restriction given by Eq. (3) is satisfied and to zero otherwise. We disregard distribution $P(s_1)$ in Eq. (12) because its contribution is negligible when T is large (alternatively, we can assume $P(s_1 = 0) = P(s_1 = 1) = 1/2$). Number $n_{i\to j}$ is the total number of state transitions $i \to j$ (where $i, j = 0, 1$). Parameters that enter the prior distributions are called hyper-parameters. For these, the means μ_α and μ_k are chosen to be equal to the maximum likelihood estimation (MLE) values of $\ln(\alpha)$ and β_k for the standard NB model. The variances Σ_α and Σ_k are chosen to be ten times larger than the maximum between the MLE values of $\ln(\alpha)$ and β_k squared, and the MLE variances of $\ln(\alpha)$ and β_k for the standard NB model. Thus, the prior distributions for α and β are chosen to be very wide and essentially flat.[8] We choose $\upsilon_0 = \nu_0 = \upsilon_1 = \nu_1 = 1$, in which case the beta distributions become the uniform distribution between zero and one. Note that formula (12) for distribution $f(\mathbf{S}|p_{0\to1}, p_{1\to0})$ follows from the Markov process property, specified by Eq. (1). In other words, we a priori specify that the state variable s_t follows a Markov process in time, and this must be and is reflected in the prior distribution.

In our study (and in most practical studies), the direct application of Eq. (8) is not feasible because the parameter vector $\boldsymbol{\Theta}$ contains too many components, making integration over $\boldsymbol{\Theta}$ in Eq. (8) extremely difficult. However, the posterior distribution $f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M})$ in Eq. (8) is known up to its normalization constant, $f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M}) \propto f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})\pi(\boldsymbol{\Theta}|\mathcal{M})$. As a result, we use Markov Chain Monte Carlo (MCMC) simulations, which provide a convenient and practical computational methodology for sampling from a probability distribution known up to a constant (the posterior distribution in our case). Given a large enough posterior sample of parameter vector $\boldsymbol{\Theta}$, any posterior expectation and variance can be found and Bayesian inference can be readily applied. For the MCMC simulations in this paper, special numerical code was written in the MATLAB programming language and tested on artificial accident data sets. The test procedure included a generation of artificial data with a known model. Then these data were used to estimate the underlying model by means of our simulation code. With this procedure we found that the MSNB models, used to generate the artificial data, were reproduced successfully with our estimation code.

By using the hybrid Gibbs sampler MCMC algorithm, briefly described in the Appendix A, we obtain a Markov chain of posterior draws $\{\boldsymbol{\Theta}^{(g)}\}$, where $g = 1, 2, \ldots, G_{bi}, G_{bi}+1, \ldots, G$. We discard the first G_{bi} "burn-in" draws because they can depend on the initial choice $\boldsymbol{\Theta}^{(0)}$. Of the remaining $G - G_{bi}$ draws, we typically store every third or every tenth draw in the computer memory. We use these draws for Bayesian inference. Our typical choice is $G_{bi} = 3 \times 10^5$ and $G = 3 \times 10^6$ (in which case, one MCMC simulation run typically

[6] To ensure that $\alpha_{(0)}$ and $\alpha_{(1)}$ are non-negative, their logarithms are used in estimation.

[7] Below we will have 260 time periods ($T = 260$). In this case, there are 2^{260} possible combinations for value of vector $\mathbf{S} = [s_1, \ldots, s_T]'$.

[8] Eq. (8) shows that for nearly flat prior distributions, when $\pi(\boldsymbol{\Theta}|\mathcal{M})$ is approximately constant around the peak of the likelihood function, the posterior distribution only weakly depends on the exact choice of the prior. We have verified this during our test MCMC runs. Using the MLE estimates for the hyper-parameters might introduce a possible bias, but this bias is small for nearly flat priors.

Table 1
Estimation results for negative binomial models of accident frequency (the superscript and subscript numbers to the right of individualparameter posterior/MLE estimates are 95% confidence/credible intervals—see text for further explanation).

Variable	NB-by-MLE [a]	NB-by-MCMC [b]	Restricted MSNB [c]		Full MSNB [d]	
			State $s=0$	State $s=1$	State $s=0$	State $s=1$
Intercept (constant term)	$-21.3^{-18.7}_{-23.9}$	$-20.6^{-18.5}_{-22.7}$	$-20.9^{-18.7}_{-23.0}$	$-19.9^{-17.8}_{-22.1}$	$-20.7^{-18.7}_{-22.8}$	$-20.7^{-18.7}_{-22.8}$
Accident occurring on interstates I-70 or I-164 (dummy)	$-.655^{-.562}_{-.748}$	$-.657^{-.565}_{-.750}$	$-.656^{-.564}_{-.748}$	$-.656^{-.564}_{-.748}$	$-.660^{-.568}_{-.752}$	$-.660^{-.568}_{-.752}$
Pavement quality index (PQI) average [e]	$-.0132^{-.00581}_{-.0205}$	$-.0189^{-.0134}_{-.0244}$	$-.0195^{-.0141}_{-.0248}$	$-.0195^{-.0141}_{-.0248}$	$-.0220^{-.0166}_{-.0273}$	$-.0125^{-.00700}_{-.0180}$
Road segment length (in miles)	$.0512^{.0809}_{.0215}$	$.0546^{.0826}_{.0266}$	$.0538^{.0812}_{.0264}$	$.0538^{.0812}_{.0264}$	$.0395^{.0625}_{.0165}$	$.0395^{.0625}_{.0165}$
Logarithm of road segment length (in miles)	$.909^{.974}_{.845}$	$.903^{.964}_{.842}$	$.900^{.961}_{.840}$	$.900^{.961}_{.840}$	$.913^{.973}_{.853}$	$.913^{.973}_{.853}$
Total number of ramps on the road viewing and opposite sides	$-.0172^{-.00174}_{-.0327}$	$-.021^{-.00624}_{-.0358}$	$-.0187^{-.00423}_{-.0331}$	$-.0187^{-.00423}_{-.0331}$	–	$-.0264^{-.00656}_{-.0464}$
Number of ramps on the viewing side per lane per mile	$.394^{.479}_{.309}$	$.400^{.479}_{.319}$	$.397^{.475}_{.317}$	$.397^{.475}_{.317}$	$.359^{.429}_{.289}$	$.359^{.429}_{.289}$
Median configuration is depressed (dummy)	$.210^{.314}_{.106}$	$.214^{.318}_{.111}$	$.211^{.315}_{.108}$	$.211^{.315}_{.108}$	$.209^{.313}_{.107}$	$.209^{.313}_{.107}$
Median barrier presence (dummy)	$-3.02^{-2.38}_{-3.67}$	$-2.99^{-2.40}_{-3.67}$	$-3.01^{-2.42}_{-3.69}$	$-3.01^{-2.42}_{-3.69}$	$-3.01^{-2.42}_{-3.69}$	$-3.01^{-2.42}_{-3.69}$
Interior shoulder presence (dummy)	$-1.15^{-.486}_{-1.81}$	$-1.06^{.135}_{-2.26}$	$-1.02^{.148}_{-2.23}$	$-1.02^{.148}_{-2.23}$	$-1.16^{-.523}_{-1.87}$	$-1.16^{-.523}_{-1.87}$
Width of the interior shoulder is less that 5 ft (dummy)	$.373^{.477}_{.270}$	$.384^{.491}_{.279}$	$.386^{.492}_{.281}$	$.386^{.492}_{.281}$	$.380^{.486}_{.275}$	$.380^{.486}_{.275}$
Interior rumble strips presence (dummy)	$-.166^{-.0382}_{-.293}$	$-.142^{.857}_{-1.16}$	$-.163^{.836}_{-1.14}$	$-.163^{.836}_{-1.14}$	–	–
Width of the outside shoulder is less that 12 ft (dummy)	$.281^{.380}_{.182}$	$.272^{.370}_{.174}$	$.268^{.366}_{.170}$	$.268^{.366}_{.170}$	$.267^{.365}_{.170}$	$.267^{.365}_{.170}$
Outside barrier is absent (dummy)	$-.249^{-.139}_{-.358}$	$-.255^{-.142}_{-.366}$	$-.255^{-.142}_{-.366}$	$-.255^{-.142}_{-.366}$	$-.251^{-.140}_{-.362}$	$-.251^{-.140}_{-.362}$
Average annual daily traffic (AADT)	$-4.09^{-3.04}_{-5.15}\times 10^{-5}$	$-4.09^{-3.24}_{-4.95}\times 10^{-5}$	$-4.07^{-3.22}_{-4.94}\times 10^{-5}$	$-4.07^{-3.22}_{-4.94}\times 10^{-5}$	$-3.90^{-3.11}_{-4.72}\times 10^{-5}$	$-4.53^{-3.61}_{-5.48}\times 10^{-5}$
Logarithm of average annual daily traffic	$2.08^{2.36}_{1.80}$	$2.06^{2.30}_{1.83}$	$2.07^{2.30}_{1.83}$	$2.07^{2.30}_{1.83}$	$2.07^{2.30}_{1.84}$	$2.07^{2.30}_{1.84}$
Posted speed limit (in mph)	$.0154^{.0244}_{.00643}$	$.0150^{.0241}_{.00589}$	$.0161^{.0251}_{.00697}$	$.0161^{.0251}_{.00697}$	$.0161^{.0252}_{.00712}$	$.0161^{.0252}_{.00712}$
Number of bridges per mile	$-.0213^{-.00187}_{-.0407}$	$-.0241^{-.00721}_{-.0419}$	$-.0233^{-.00648}_{-.0410}$	$-.0233^{-.00648}_{-.0410}$	–	$-.0607^{-.0232}_{-.102}$
Maximum of reciprocal values of horizontal curve radii (in 1/mile)	$-.182^{-.122}_{-.242}$	$-.179^{-.118}_{-.241}$	$-.178^{-.117}_{-.239}$	$-.178^{-.117}_{-.239}$	$-.175^{-.114}_{-.237}$	$-.175^{-.114}_{-.237}$
Maximum of reciprocal values of vertical curve radii (in 1/mile)	$.0191^{.0285}_{.00972}$	$.0177^{.027}_{.00843}$	$.0183^{.0275}_{.00917}$	$.0183^{.0275}_{.00917}$	$.0184^{.0274}_{.00925}$	$.0184^{.0274}_{.00925}$
Number of vertical curves per mile	$-.0535^{-.0180}_{-.0889}$	$-.057^{-.0233}_{-.0924}$	$-.0586^{-.0249}_{-.0940}$	$-.0586^{-.0249}_{-.0940}$	$-.0565^{-.0231}_{-.0917}$	$-.0565^{-.0231}_{-.0917}$
Percentage of single unit trucks (daily average)	$1.38^{1.88}_{.886}$	$1.25^{1.75}_{.758}$	$1.19^{1.68}_{.701}$	$1.19^{1.68}_{.701}$	$.726^{1.28}_{.171}$	$2.57^{3.39}_{1.77}$
Winter season (dummy)	$.148^{.226}_{.0698}$	$.148^{.226}_{.0689}$	$-.116^{.0563}_{-.261}$	$-.116^{.0563}_{-.261}$	$-.159^{-.0494}_{-.269}$	–
Spring season (dummy)	$-.173^{-.0878}_{-.258}$	$-.173^{-.0899}_{-.257}$	$-.0932^{.0547}_{-.209}$	$-.0932^{.0547}_{-.209}$	–	–
Summer season (dummy)	$-.179^{-.0921}_{-.266}$	$-.180^{-.0963}_{-.263}$	$-.0332^{.111}_{-.146}$	$-.0332^{.111}_{-.146}$	–	$-.549^{-.293}_{-.883}$
Over-dispersion parameter α in NB models	$.957^{1.07}_{.845}$	$.968^{1.09}_{.849}$	$.537^{.677}_{.392}$	$1.24^{1.51}_{.986}$	$.443^{.595}_{.300}$	$1.16^{1.39}_{.945}$
Mean accident rate ($\lambda_{t,n}$ for NB), averaged over all values of $\mathbf{X}_{t,n}$	–	.0663	.0558	.1440	.0533	.1130
Standard deviation of accident rate ($\sqrt{\lambda_{t,n}(1+\alpha\lambda_{t,n})}$ for NB), averaged over all values of explanatory variables $\mathbf{X}_{t,n}$	–	.2050	.1810	.3350	.1760	.2820
Markov transition probability of jump $0\to 1$ ($p_{0\to 1}$)	–	–	$.0933^{.147}_{.0531}$		$.158^{.225}_{.100}$	
Markov transition probability of jump $1\to 0$ ($p_{1\to 0}$)	–	–	$.651^{.820}_{.463}$		$.627^{.773}_{.474}$	
Unconditional probabilities of states 0 and 1 ($\bar{p}_0$ and $\bar{p}_1$)	–	–	$.873^{.929}_{.797}$ and $.127^{.203}_{.0713}$		$.798^{.868}_{.718}$ and $.202^{.282}_{.132}$	
Total number of free model parameters (β and α)	26	26	28		28	
Posterior average of the log-likelihood (LL)	–	$-16097.2^{-16091.3}_{-16105.0}$	$-15821.8^{-15807.9}_{-15835.2}$		$-15778.0^{-15672.9}_{-15794.9}$	

Max(LL): estimated maximum value of log-likelihood (LL) for MLE; maximum observed LL during MCMC simulations for Bayesian estimation	−16081.2 (MLE)	−16086.3 (observed)	−15786.6 (observed)	−15744.8 (observed)
Logarithm of marginal likelihood of data ($\ln[f(\mathbf{Y}\mid\mathcal{M})]$)	–	$-16108.6^{-16105.7}_{-16110.7}$	$-15850.2^{-15840.1}_{-15849.5}$	$-15809.4^{-15801.7}_{-15811.9}$
Goodness-of-fit p-value	–	0.701	0.729	0.647
Maximum of the potential scale reduction factors (PSRF) [f]	–	1.00874	1.00754	1.00939
Multivariate potential scale reduction factor (MPSRF) [f]	–	1.00928	1.00925	1.01002

[a] Standard (conventional) negative binomial estimated by maximum likelihood estimation (MLE).
[b] Standard negative binomial estimated by Markov Chain Monte Carlo (MCMC) simulations.
[c] Restricted two-state Markov switching negative binomial (MSNB) model with only the intercept and over-dispersion parameters allowed to vary between states.
[d] Full two-state Markov switching negative binomial (MSNB) model with all parameters allowed to vary between states.
[e] The pavement quality index (PQI) is a composite measure of overall pavement quality evaluated on a 0–100 scale.
[f] PSRF/MPSRF are calculated separately/jointly for all continuous model parameters. PSRF and MPSRF are close to 1 for converged MCMC chains.

takes few days on a single computer CPU). We usually consider eight choices of the initial parameter vector $\mathbf{\Theta}^{(0)}$. Thus, we obtain eight Markov chains of $\mathbf{\Theta}$, and use them for the Brooks–Gelman–Rubin diagnostic of convergence of our MCMC simulations (Brooks and Gelman, 1998). The resulting potential scale reduction factors (PSRF) and multivariate potential scale reduction factors (MPSRF) are close to unity (see Table 1), which indicates good convergence of the MCMC chains. We also check convergence by monitoring the likelihood $f(\mathbf{Y}|\mathbf{\Theta}^{(g)})$ and the joint distribution $f(\mathbf{Y},\mathbf{\Theta}^{(g)})$. For all details on MCMC simulations, see Malyshkina (2008), available at http://arxiv.org/abs/0808.1448.

For comparison of different models we use a formal Bayesian approach. Let there be two models $\mathcal{M}_1$ and $\mathcal{M}_2$ with parameter vectors $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ respectively. Assuming that we have equal preferences of these models, their prior probabilities are $\pi(\mathcal{M}_1)=\pi(\mathcal{M}_2)=1/2$. In this case, the ratio of the models' posterior probabilities, $P(\mathcal{M}_1|\mathbf{Y})$ and $P(\mathcal{M}_2|\mathbf{Y})$, is equal to the Bayes factor. The later is defined as the ratio of the models' marginal likelihoods (Kass and Raftery, 1995). Thus, we have

$$\frac{P(\mathcal{M}_2|\mathbf{Y})}{P(\mathcal{M}_1|\mathbf{Y})}=\frac{f(\mathcal{M}_2,\mathbf{Y})/f(\mathbf{Y})}{f(\mathcal{M}_1,\mathbf{Y})/f(\mathbf{Y})}=\frac{f(\mathbf{Y}|\mathcal{M}_2)\pi(\mathcal{M}_2)}{f(\mathbf{Y}|\mathcal{M}_1)\pi(\mathcal{M}_1)}=\frac{f(\mathbf{Y}|\mathcal{M}_2)}{f(\mathbf{Y}|\mathcal{M}_1)}, \tag{13}$$

where $f(\mathcal{M}_1,\mathbf{Y})$ and $f(\mathcal{M}_2,\mathbf{Y})$ are the joint distributions of the models and the data, $f(\mathbf{Y})$ is the unconditional distribution of the data, and the marginal likelihoods $f(\mathbf{Y}|\mathcal{M}_1)$ and $f(\mathbf{Y}|\mathcal{M}_2)$ are given by Eq. (9). If the ratio in Eq. (13) is larger than one, then model $\mathcal{M}_2$ is favored, if the ratio is less than one, then model $\mathcal{M}_1$ is favored. An advantage of the use of Bayes factors is that it has an inherent penalty for including too many parameters in the model and guards against overfitting.

To evaluate the performance of model $\{\mathcal{M},\mathbf{\Theta}\}$ in fitting the observed data $\mathbf{Y}$, we carry out a χ^2 goodness-of-fit test (Maher and Summersgill, 1996; Cowan, 1998; Wood, 2002; Press et al., 2007). Quantity χ^2 is[9]

$$\chi^2=\sum_{t=1}^{T}\sum_{n=1}^{N_t}\frac{[Y_{t,n}-E(Y_{t,n}|\mathbf{\Theta},\mathcal{M})]^2}{var(Y_{t,n}|\mathbf{\Theta},\mathcal{M})}, \tag{14}$$

where $E(Y_{t,n}|\mathbf{\Theta},\mathcal{M})$ and $var(Y_{t,n}|\mathbf{\Theta},\mathcal{M})$ are the expectations and variances of the observations $Y_{t,n}$. In our study, the observations are the accident frequencies, $Y_{t,n}=A_{t,n}$. From Eqs. (4)–(5) for the MSNB model we find (unconditional of state) expectations and variances $E(Y_{t,n}|\mathbf{\Theta},\mathcal{M})=\bar{p}_0\lambda^{(0)}_{t,n}+\bar{p}_1\lambda^{(1)}_{t,n}$ and $\mathrm{var}(Y_{t,n}|\mathbf{\Theta},\mathcal{M})=\bar{p}_0\lambda^{(0)}_{t,n}(1+\alpha_{(0)}\lambda^{(0)}_{t,n})+\bar{p}_1\lambda^{(1)}_{t,n}(1+\alpha_{(1)}\lambda^{(1)}_{t,n})+\bar{p}_0\bar{p}_1(\lambda^{(1)}_{t,n}-\lambda^{(0)}_{t,n})^2$, where $\lambda^{(0)}_{t,n}=\exp(\boldsymbol{\beta}'_{(0)}\mathbf{X}_{t,n})$ and $\lambda^{(1)}_{t,n}=\exp(\boldsymbol{\beta}'_{(1)}\mathbf{X}_{t,n})$ are the mean accident rates in the states $s_t=0$ and $s_t=1$ respectively. In the limit of asymptotically normal distribution of large accident frequencies, χ^2 has the chi-square distribution with degrees of freedom equal to the number of observations minus the number of model parameters (Wood, 2002). Because weekly accident frequencies are typically small, in this study, we do not rely on the assumption of their asymptotic normality. Instead, we carry out Monte Carlo simulations to find the distribution of χ^2 (Cowan, 1998). This is done by generating a large number of artificial data sets under the hypothesis that the model $\{\mathcal{M},\mathbf{\Theta}\}$ is true, computing and recording the χ^2 value for each data set, and then using these values to find the distribution of χ^2. This distribution is then used to find the goodness-of-fit p-value, equal to the probability that χ^2 exceeds the observed value of χ^2 (the later is calculated by using the observed data $\mathbf{Y}$).

[9] Note that for a standard Poisson distribution, the variances are equal to the means, $var(Y_{t,n}|\mathbf{\Theta},\mathcal{M})=E(Y_{t,n}|\mathbf{\Theta},\mathcal{M})$, and Eq. (14) reduces to the Pearson's χ^2.

4. Model estimation results

Data are used from 5769 accidents that were observed on 335 interstate highway segments in Indiana in 1995–1999. We use weekly time periods, $t = 1, 2, 3, \ldots, T = 260$ in total.[10] Thus, in the present study the state (s_t) is the same for all roadway segments and can change every week. Four types of accident frequency models are estimated:

- First, we estimate a standard (single-state) negative binomial (NB) model without Markov switching by maximum likelihood estimation (MLE) method with the help of the LIMDEP software package. To obtain a parsimonious standard NB model, estimated by MLE, we choose the explanatory variables and their dummies by using the Akaike Information Criterion (AIC)[11] and the 5% statistical significance level for the two-tailed t-test (for details on our variable selection methods, see Malyshkina, 2006). We refer to this model as "NB-by-MLE".
- Second, we estimate the same standard negative binomial model by the Bayesian inference approach and the MCMC simulations. We refer to this model as "NB-by-MCMC". As one expects, for our choice of a nearly flat prior distribution, the estimated NB-by-MCMC model turned out to be very similar to the NB-by-MLE model.
- Third, we estimate a restricted two-state Markov switching negative binomial (MSNB) model. In this restricted switching model only the intercept in the model parameters vector $\boldsymbol{\beta}$ and the over-dispersion parameter α are allowed to switch between the two states of roadway safety. In other words, in Eq. (5) only the first components of vectors $\boldsymbol{\beta}_{(0)}$ and $\boldsymbol{\beta}_{(1)}$ may differ, while the remaining components are restricted to be the same. In this case, the two states can have different average accident rates, given by Eq. (5), but the rates have the same dependence on the explanatory variables. We refer to this model as "restricted MSNB"; it is estimated by the Bayesian-MCMC methods. Note that, in order to make comparison of explanatory variable effects in different models straightforward, in all MSNB models we use only those explanatory variables that enter the standard NB model.[12]
- Fourth, we estimate a full two-state Markov switching negative binomial (MSNB) model. In this model all estimable model parameters (β and α) are allowed to switch between the two states of roadway safety. To obtain the final full MSNB model reported here, we consecutively construct and use 60%, 85% and 95% Bayesian credible intervals for evaluation of the statistical significance of each β-parameter. As a result, in the final model some components of $\boldsymbol{\beta}_{(0)}$ and $\boldsymbol{\beta}_{(1)}$ are restricted to zero or restricted to be the same in the two states.[13] We do not impose any restrictions on over-dispersion parameters (αs). We refer to the final full MSNB model as "full MSNB"; it is estimated by the Bayesian-MCMC methods.

Note that the two states, and thus the MSNB models, do not have to exist. For example, they will not exist if all estimated model parameters turn out to be statistically the same in the two states, $\boldsymbol{\beta}_{(0)} = \boldsymbol{\beta}_{(1)}$, (which suggests the two states are identical and the MSNB models reduce to the standard NB model). Also, the two states will not exist if all estimated state variables s_t turn out to be close to zero, resulting in $p_{0\to1} \ll p_{1\to0}$ (compare to Eq. (3)), then the less frequent state $s_t = 1$ is not realized and the process stays in state $s_t = 0$.

The model estimation results for accident frequencies are given in Table 1. Posterior (or MLE) estimates of all continuous model parameters (β, α, $p_{0\to1}$ and $p_{1\to0}$) are given together with the corresponding 95% confidence intervals for MLE models and 95% credible intervals for Bayesian-MCMC models (refer to the superscript and subscript numbers adjacent to parameter posterior/MLE estimates in Table 1).[14] Table 2 gives summary statistics of all roadway segment characteristic variables $\mathbf{X}_{t,n}$ (except the intercept).

To visually see how the model tracks the data, consider Fig. 2. The top plot in Fig. 2 shows the weekly accident frequencies in our data as given in Fig. 1. The bottom plot in Fig. 2 shows corresponding weekly posterior probabilities $P(s_t = 1|\mathbf{Y})$ of the less frequent state $s_t = 1$ for the full MSNB model. These probabilities are equal to the posterior expectations of s_t, $P(s_t = 1|\mathbf{Y}) = 1 \times P(s_t = 1|\mathbf{Y}) + 0 \times P(s_t = 0|\mathbf{Y}) = E(s_t|\mathbf{Y})$. Weekly values of $P(s_t = 1|\mathbf{Y})$ for the restricted MSNB model are very similar to those given on the top plot in Fig. 2, and, as a result, are not shown on a separate plot. Indeed, the time-correlation[15] between $P(s_t = 1|\mathbf{Y})$ for the two MSNB models is about 99.5%.

Turning to the estimation results, the findings show that two states exist and Markov switching models are strongly favored by the empirical data. In particular, in the restricted MSNB model we over 99.9% confident that the difference in values of β-intercept in the two states is non-zero.[16] To compare the Markov switching models (restricted and full MSNB) with the corresponding standard non-switching model (NB), we calculate and use Bayes factors given by Eq. (13). We use Eq. (9) for calculation of the values and the 95% confidence intervals of the logarithms of the marginal likelihoods given in Table 1. The confidence intervals are found by bootstrap simulations.[17] The log-marginal-likelihoods are -16108.6, -15850.2 and -15809.4 for the NB, restricted MSNB and full MSNB models respectively. Therefore, the restricted and full MSNB models provide considerable (258.4 and 299.2) improvements of the log-marginal-likelihoods of the data as compared to the corresponding standard non-switching NB model. Thus, given the accident data, the posterior probabilities of the restricted and full MSNB models are larger than

[10] A week is from Sunday to Saturday, there are 260 full weeks in the 1995–1999 time interval. We also considered daily time periods and obtained qualitatively similar results (not reported here).

[11] Minimization of AIC $= 2K - 2LL$, were K is the number of free continuous model parameters and LL is the log-likelihood, ensures an optimal choice of explanatory variables in a model and avoids overfitting (Tsay, 2002; Washington et al., 2003).

[12] A formal Bayesian approach to model variable selection is based on evaluation of model's marginal likelihood and the Bayes factor (13). Unfortunately, because MCMC simulations are computationally expensive, evaluation of marginal likelihoods for a large number of trial models is not feasible in our study.

[13] A β-parameter is restricted to zero if it is statistically insignificant. A β-parameter is restricted to be the same in the two states if the difference of its values in the two states is statistically insignificant. A $(1 - a)$ credible interval is chosen in such way that the posterior probabilities of being below and above it are both equal to $a/2$ (we use significance levels $a = 40\%, 15\%, 5\%$).

[14] Note that MLE assumes asymptotic normality of the estimates, resulting in confidence intervals being symmetric around the means (a 95% confidence interval is ±1.96 standard deviations around the mean). In contrast, Bayesian estimation does not require this assumption, and posterior distributions of parameters and Bayesian credible intervals are usually non-symmetric.

[15] Here and below we calculate weighted correlation coefficients. For variable $P(s_t = 1|\mathbf{Y}) \equiv E(s_t|\mathbf{Y})$ we use weights w_t inversely proportional to the posterior standard deviations of s_t. That is $w_t \propto \min\left\{1/\mathrm{std}(s_t|\mathbf{Y}), \mathrm{median}[1/\mathrm{std}(s_t|\mathbf{Y})]\right\}$.

[16] The difference of the intercept values is statistically non-zero despite the fact that the 95% credible intervals for these values overlap (see the "Intercept" line and the "Restricted MSNB" columns in Table 1). The reason is that the posterior draws of the intercepts are correlated. The statistical test of whether the intercept values differ, must be based on evaluation of their difference.

[17] During bootstrap simulations we repeatedly draw, with replacement, posterior values of $\boldsymbol{\Theta}$ to calculate the posterior expectation in Eq. (9). In each of 10^5 bootstrap draws that we make, the number of $\boldsymbol{\Theta}$ values drawn is 1/100 of the total number of all posterior $\boldsymbol{\Theta}$ values available from MCMC simulations. The bootstrap simulations show that Eq. (9) gives sufficiently accurate answers, and that the expectation in Eq. (9) is not dominated too much by just few posterior values of $\boldsymbol{\Theta}$, at which the likelihood function happens to be extremely small.

Table 2
Summary statistics of roadway segment characteristic variables.

Variable	Mean	Standard deviation	Minimum	Median	Maximum
Accident occurring on interstates I-70 or I-164 (dummy)	.155	.363	0	0	1.00
Pavement quality index (PQI) average [a]	88.6	5.96	69.0	90.3	98.5
Road segment length (in miles)	.886	1.48	.00900	.356	11.5
Logarithm of road segment length (in miles)	−.901	1.22	−4.71	−1.03	2.44
Total number of ramps on the road viewing and opposite sides	.725	1.79	0	0	16
Number of ramps on the viewing side per lane per mile	.138	.408	0	0	3.27
Median configuration is depressed (dummy)	.630	.484	0	1.00	1.00
Median barrier presence (dummy)	.161	.368	0	0	1
Interior shoulder presence (dummy)	.928	.258	0	1	1
Width of the interior shoulder is less that 5 ft (dummy)	.696	.461	0	1.00	1.00
Interior rumble strips presence (dummy)	.722	.448	0	1.00	1.00
Width of the outside shoulder is less that 12 ft (dummy)	.752	.432	0	1.00	1.00
Outside barrier absence (dummy)	.830	.376	0	1.00	1.00
Average annual daily traffic (AADT)	3.03×10^4	2.89×10^4	$.944 \times 10^4$	1.65×10^4	14.3×10^4
Logarithm of average annual daily traffic	10.0	.623	9.15	9.71	11.9
Posted speed limit (in mph)	63.1	3.89	50.0	65.0	65.0
Number of bridges per mile	1.76	8.14	0	0	124
Maximum of reciprocal values of horizontal curve radii (in 1/mile)	.650	.632	0	.589	2.26
Maximum of reciprocal values of vertical curve radii (in 1/mile)	2.38	3.59	0	0	14.9
Number of vertical curves per mile	1.50	4.03	0	0	50.0
Percentage of single unit trucks (daily average)	.0859	.0678	.00975	.0683	.322
Winter season (dummy)	.242	.428	0	0	1.00
Spring season (dummy)	.254	.435	0	0	1.00
Summer season (dummy)	.254	.435	0	0	1.00

[a] The pavement quality index (PQI) is a composite measure of overall pavement quality evaluated on a 0–100 scale.

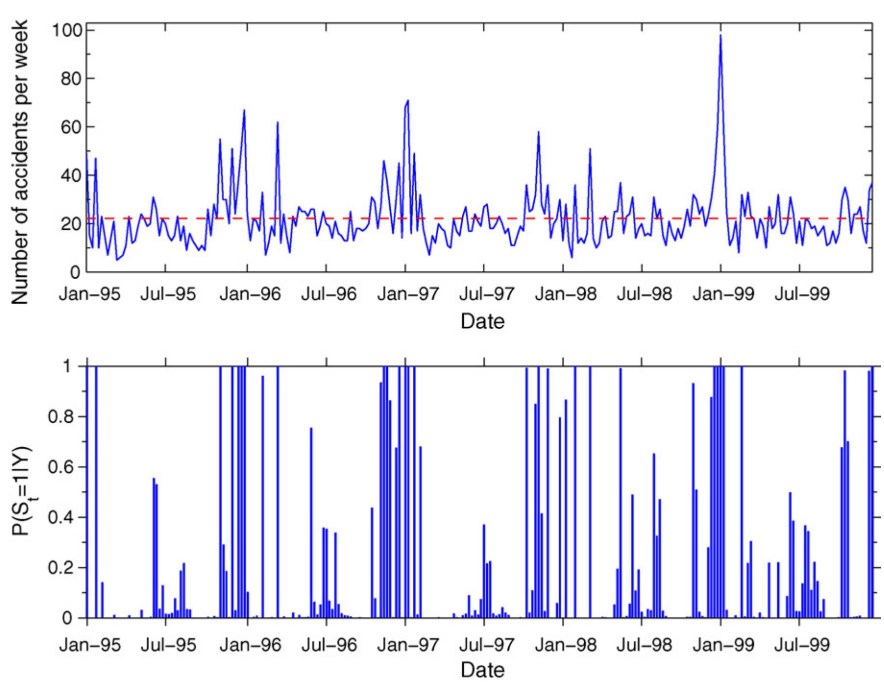

Fig. 2. The top plot is the same as Fig. 1. The bottom plot shows weekly posterior probabilities $P(s_t = 1|\mathbf{Y})$ for the full MSNB model.

the probability of the standard NB model by $e^{258.4}$ and $e^{299.2}$ respectively.[18]

[18] There are other frequently used model comparison criteria, for example, the deviance information criterion, $\text{DIC} = 2E[D(\mathbf{\Theta})|\mathbf{Y}] - D(E[\mathbf{\Theta}|\mathbf{Y}])$, where deviance $D(\mathbf{\Theta}) \equiv -2\ln[f(\mathbf{Y}|\mathbf{\Theta}, \mathcal{M})]$ (Robert, 2001). Models with smaller DIC are favored to models with larger DIC. We find DIC values 32,219, 31,662, 31,577 for the NB, restricted MSNB and full MSNB models respectively. This means that the MSNB models are favored over the standard NB model (the full MSNB is favored most). However, DIC is theoretically based on the assumption of asymptotic multivariate normality of the posterior distribution, in which case DIC reduces to AIC (Spiegelhalter et al., 2002). As a result, we prefer to rely on a mathematically rigorous and formal Bayes factor approach to model selection, as given by Eq. (13).

To evaluate the goodness-of-fit for a model, we use the posterior (or MLE) estimates of all continuous model parameters (βs, α, $p_{0\to1}$, $p_{1\to0}$) and generate 10^4 artificial data sets under the hypothesis that the model is true.[19] We find the distribution of χ^2, given by Eq. (14), and calculate the goodness-of-fit p-value for the observed value of χ^2. The resulting p-values for our models are given in Table 1. These p-values are around 65–70%. Therefore, all models fit the data well.

Focusing on the full MSNB model, which is statistically superior, its estimation results show that the less frequent state $s_t = 1$ is about four times as rare as the more frequent state $s_t = 0$ (refer

[19] Note that the state values $\mathbf{S}$ are generated by using $p_{0\to1}$ and $p_{1\to0}$.

Table 3
Correlations of the posterior probabilities $P(s_t = 1|\mathbf{Y})$ with weather-condition variables for the full MSNB model.

	All year	Winter (November–March)	Summer (May–September)
Precipitation (in.)	.031	–	.144
Temperature (F)	−.518	−.591	.201
Snowfall (in.)	.602	.577	–
> .2 (dummy)	.651	.638	–
Fog/frost (dummy)	.223	(frost) .539	(fog) .051
Visibility distance (mile)	−.221	−.232	−.126

to the estimated values of the unconditional probabilities $\bar{p}_0$ and $\bar{p}_1$ of states 0 and 1, which are given in the "Full MSNB" columns in Table 1).

Also, the findings show that the less frequent state $s_t = 1$ is considerably less safe than the more frequent state $s_t = 0$. This result follows from the values of the mean weekly accident rate $\lambda_{t,n}$ [given by Eq. (5) with model parameters β taken to be equal to their posterior means in the two states], averaged over all values of the explanatory variables $\mathbf{X}_{t,n}$ observed in the data sample (see "mean accident rate" in Table 1). For the full MSNB model, on average, state $s_t = 1$ has about two times more accidents per week than state $s_t = 0$ has.[20] Therefore, it is not a surprise, that in Fig. 2 the weekly number of accidents (shown on the bottom plot) is larger when the posterior probability $P(s_t = 1|\mathbf{Y})$ of the state $s_t = 1$ (shown on the top plot) is higher.

Note that the long-term unconditional mean of accident rate $A_{t,n}$ is $E(A_{t,n}) = \bar{p}_0\lambda^{(0)}_{t,n} + \bar{p}_1\lambda^{(1)}_{t,n}$, where $\lambda^{(0)}_{t,n} = \exp(\boldsymbol{\beta}'_{(0)}\mathbf{X}_{t,n})$ and $\lambda^{(1)}_{t,n} = \exp(\boldsymbol{\beta}'_{(1)}\mathbf{X}_{t,n})$ are the mean accident rates in the states $s_t = 0$ and $s_t = 1$ respectively [see Eq. (5)]. The unconditional expectation $E(A_{t,n})$ should be used in all predictions of long-term averaged accident rates. In the formula for this expectation, the mean accident rate $\lambda_{t,n}$ is averaged over the two states by using the stationary unconditional probabilities $\bar{p}_0$ and $\bar{p}_1$ given by Eq. (2)(see the "unconditional probabilities of states 0 and 1" in Table 1).

It is also noteworthy that the number of accidents is more volatile in the less frequent and less-safe state ($s_t = 1$). This is reflected in the fact that the standard deviation of the accident rate ($\text{std}_{t,n} = \sqrt{\lambda_{t,n}(1 + \alpha\lambda_{t,n})}$ for NB distribution), averaged over all values of explanatory variables $\mathbf{X}_{t,n}$, is higher in state $s_t = 1$ than in state $s_t = 0$ (refer to Table 1). Moreover, for the full MSNB model the over-dispersion parameter α is higher in state $s_t = 1$ ($\alpha = 0.443$ in state $s_t = 0$ and $\alpha = 1.16$ in state $s_t = 1$). Because state $s_t = 1$ is relatively rare, this suggests that over-dispersed volatility of accident frequencies, which is often observed in empirical data, could be in part due to the latent switching between the states, and in part due to high accident volatility in the less frequent and less safe state $s_t = 1$.

To study the effect of weather (which is usually unobserved heterogeneity in most data bases) on states, Table 3 gives time-correlation coefficients between posterior probabilities $P(s_t = 1|\mathbf{Y})$ for the full MSNB model and weather-condition variables. These correlations were found by using daily and hourly historical weather data in Indiana, available at the Indiana State Climate Office at Purdue University (www.agry.purdue.edu/climate). For these correlations, the precipitation and snowfall amounts are daily amounts in inches averaged over the week and across several weather observation stations that are located close to the roadway segments.[21] The temperature variable is the mean daily air temperature (F) averaged over the week and across the weather stations. The effect of fog/frost is captured by a dummy variable that is equal to one if and only if the difference between air and dewpoint temperatures does not exceed 5 F (in this case frost can form if the dewpoint is below the freezing point 32 F, and fog can form otherwise). The fog/frost dummies are calculated for every hour and are averaged over the week and across the weather stations. Finally, visibility distance variable is the harmonic mean of hourly visibility distances, which are measured in miles every hour and are averaged over the week and across the weather stations.[22]

Table 3 shows that the less frequent and less safe state $s_t = 1$ is positively correlated with extreme temperatures (low during winter and high during summer), rain precipitations and snowfalls, fogs and frosts, low visibility distances. It is reasonable to expect that during bad weather, roads can become significantly less safe, resulting in a change of the state of roadway safety. As a useful test of the switching between the two states, all weather variables, listed in Table 3, were added into our full MSNB model. However, when doing this, the two states did not disappear and the posterior probabilities $P(s_t = 1|\mathbf{Y})$ did not changed substantially (the correlation between the new and the old probabilities was around 90%). As another test, we modified the standard single-state NB model by adding the weather variables into it. As a result, the marginal likelihood for this model improved noticeably, but the modified single-state NB model was still strongly disfavored by the data as compared to the restricted and full MSNB models. This result emphasizes the importance of the two-state approach.

Let us give a brief summary of the effects of explanatory variables on accident rates. We will focus on those variables that are significantly different between the two states in the full MSNB model. Table 1 shows that parameter estimates for pavement quality index, total number of ramps on the road viewing and opposite sides, average annual daily traffic (AADT), number of bridges per mile, percentage of single unit trucks, and season dummy variables are all significantly different between the two states. All these differences are reasonable and could be explained by adverse weather/pavement conditions in the less-safe state $s_t = 1$, and by the resulting lighter than usual traffic and more alert/defensive driving in this state. In particular, as compared to variable effects in the safe state $s_t = 0$, in the less safe state $s_t = 1$ an improvement of pavement quality leads to a smaller reduction of the accident rate, an increase in percentage of single unit trucks results in a larger increase of the accident rate, and an increase in AADT leads to a smaller increase of the accident rate (note that the effects of AADT and its logarithm should be considered simultaneously). An increase in number of ramps and bridges, and the summer season indicator significantly reduce the accident rate only in the less-safe state $s_t = 1$. The winter season indicator reduces the accident rate only in the safe state $s_t = 0$ (this result, which might look counter-intuitive, could be explained by an increase in cases of over-confident, reckless driving during good weather/pavement conditions, unless there is a winter).

In addition to the MSNB models, we estimated two-state Markov switching Poisson (MSP) models, which have the Poisson likelihood function instead of the NB likelihood function in Eq. (4). Our findings for the MSP models are very similar to those for the MSNB models (Malyshkina, 2008). Also, because the time series in Fig. 2 seems to

[20] Note that accident frequency rates can easily be converted from one time period to another (for example, weekly rates can be converted to annual rates). Because accident events are independent, the conversion is done by a summation of moment-generating (or characteristic) functions. The sum of Poisson variates is Poisson. The sum of NB variates is also NB if all explanatory variables do not depend on time ($\mathbf{X}_{t,n} = \mathbf{X}_n$).

[21] Snowfall and precipitation amounts are weakly related with each other because snow density (g/cm^3) can vary by more than a factor of ten.

[22] The harmonic mean $\bar{d}$ of distances d_n is calculated as $\bar{d}^{-1} = (1/N)\sum_{n=1}^{N} d_n^{-1}$, assuming $d_n = 0.25$ miles if $d_n \leq 0.25$ miles.

exhibit a seasonal pattern (roads appear to be less safe and $P(s_t = 1|\mathbf{Y})$ appears to be higher during winters), we estimated MSNB and MSP models in which the transition probabilities $p_{0\to 1}$ and $p_{1\to 0}$ are not constant (allowing each of them to assume two different values: one during winters and the other during all remaining seasons). However, these models did not perform as well as the MSNB and MSP models with constant transition probabilities [as judged by the Bayes factors, see Eq. (13)].[23]

5. Summary and conclusions

The empirical finding that two states exist and that these states are correlated with weather conditions has important implications. The findings suggest that multiple states of roadway safety can exist due to slow and/or inadequate adjustment by drivers (and possibly by roadway maintenance services) to adverse conditions and other unpredictable, unidentified, and/or unobservable variables that influence roadway safety. All these variables are likely to interact and change over time, resulting in transitions from one state to the next.

As discussed earlier, the empirical findings show that the less frequent state is significantly less safe than the other, more frequent state. The full MSNB model results show that explanatory variables $\mathbf{X}_{t,n}$, other than the intercept, exert different influences on roadway safety in different states as indicated by the fact that some of the parameter estimates for the two states of the full MSNB model are significantly different. Thus, the states not only differ by average accident frequencies, but also differ in the magnitude and/or direction of the effects that various variables exert on accident frequencies. This again underscores the importance of the two-state approach.

The Markov switching models presented in this study are similar to zero-inflated count data models (which have been previously applied in accident frequency research) in the sense that they are also two-state models (see Shankar et al., 1997; Lord et al., 2007). However, in contrast to zero-inflated models, the models presented herein allow for switching between the two states over time.[24] In addition, in this study, a "safe" state is not assumed and accident frequencies can be nonzero in both states.

In terms of future work on Markov switching models for roadway safety, additional empirical studies (for other accident data samples), and multi-state models (with more than two states of roadway safety) are two areas that would further demonstrate the potential of the approach.

Acknowledgments

We thank Dominique Lord, Byung-Jung Park and Srinivas Geedipally for their interest in this study and for useful discussions.

Appendix A. MCMC simulation algorithm

For brevity, in this Appendix A we omit model notation $\mathcal{M}$ in all equations. For example, here we write the posterior distribution, given by Eq. (8), as $f(\mathbf{\Theta}|\mathbf{Y})$.

To obtain draws from a posterior distribution, we use the hybrid Gibbs sampler, which is an MCMC simulation algorithm that involves both Gibbs and Metropolis–Hasting sampling (McCulloch and Tsay, 1994; Tsay, 2002; SAS Institute Inc., 2006). Assume that $\mathbf{\Theta}$ is composed of K components: $\mathbf{\Theta} = [\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \ldots, \boldsymbol{\theta}_K']'$, where $\boldsymbol{\theta}_k$ can be scalars or vectors, $k = 1, 2, \ldots, K$. Then, the hybrid Gibbs sampler works as follows:

(1) Choose an arbitrary initial value of the parameter vector, $\mathbf{\Theta} = \mathbf{\Theta}^{(0)}$, such that $f(\mathbf{Y}, \mathbf{\Theta}^{(0)}) > 0$.

(2) For each $g = 1, 2, 3, \ldots$, parameter vector $\mathbf{\Theta}^{(g)}$ is generated component-by-component from $\mathbf{\Theta}^{(g-1)}$ by the following procedure:

(a) First, draw $\boldsymbol{\theta}_1^{(g)}$ from the conditional posterior probability distribution $f(\boldsymbol{\theta}_1^{(g)}|\mathbf{Y}, \boldsymbol{\theta}_2^{(g-1)}, \ldots, \boldsymbol{\theta}_K^{(g-1)})$. If this distribution is exactly known in a closed analytical form, then we draw $\boldsymbol{\theta}_1^{(g)}$ directly from it. This is Gibbs sampling. If the conditional posterior distribution is known up to an unknown normalization constant, then we draw $\boldsymbol{\theta}_1^{(g)}$ by using the Metropolis–Hasting (M–H) algorithm. This is M–H sampling.

(b) Second, for all $k = 2, 3, \ldots, K-1$, draw $\boldsymbol{\theta}_k^{(g)}$ from the conditional posterior distribution $f(\boldsymbol{\theta}_k^{(g)}|\mathbf{Y}, \boldsymbol{\theta}_1^{(g)}, \ldots, \boldsymbol{\theta}_{k-1}^{(g)}, \boldsymbol{\theta}_{k+1}^{(g-1)}, \ldots, \boldsymbol{\theta}_K^{(g-1)})$ by using either Gibbs sampling (if the distribution is known exactly) or M–H sampling (if the distribution is known up to a constant).

(c) Finally, draw $\boldsymbol{\theta}_K^{(g)}$ from conditional posterior probability distribution $f(\boldsymbol{\theta}_K^{(g)}|\mathbf{Y}, \boldsymbol{\theta}_1^{(g)}, \ldots, \boldsymbol{\theta}_{K-1}^{(g)})$ by using either Gibbs or M–H sampling.

(3) The resulting Markov chain $\{\mathbf{\Theta}^{(g)}\}$ converges to the true posterior distribution $f(\mathbf{\Theta}|\mathbf{Y})$ as $g \to \infty$.

For a description of the Metropolis–Hasting (M–H) algorithm, used to draw from a probability distribution known up to a constant, see Tsay (2002), SAS Institute Inc. (2006), or a standard textbook in Bayesian statistics. Note that all conditional posterior distributions are proportional to the joint distribution $f(\mathbf{Y}, \mathbf{\Theta}) = f(\mathbf{Y}|\mathbf{\Theta})\pi(\mathbf{\Theta})$, where the likelihood $f(\mathbf{Y}|\mathbf{\Theta})$ is given by Eq. (6) and the prior $\pi(\mathbf{\Theta})$ is given by Eq. (10).

In this study $\mathbf{\Theta}$ is given by Eq. (7), and the hybrid Gibbs sampler generates draws $\mathbf{\Theta}^{(g)}$ from $\mathbf{\Theta}^{(g-1)}$ as follows (for brevity, below we drop g indexing):

(a) We draw vector $\boldsymbol{\beta}_{(0)}$ component-by-component by using the M–H algorithm. For each component $\beta_{(0),k}$ of $\boldsymbol{\beta}_{(0)}$ we use a normal jumping distribution $J(\hat{\beta}_{(0),k}|\beta_{(0),k}) = \mathcal{N}(\beta_{(0),k}, \sigma^2_{(0),k})$ for the M–H algorithm. Variances $\sigma^2_{(0),k}$ are adjusted during the burn-in sampling ($g = 1, 2, \ldots, G_{bi}$) to have approximately 30% acceptance rate during the M–H sampling.[25] The conditional posterior distribution of $\beta_{(0),k}$ is

$$
\begin{aligned}
f(\beta_{(0),k}|\mathbf{Y}, \mathbf{\Theta} \setminus \beta_{(0),k}) &\propto f(\mathbf{Y}, \mathbf{\Theta}) \\
&= f(\mathbf{Y}|\mathbf{\Theta})\pi(\mathbf{\Theta}) \propto f(\mathbf{Y}|\mathbf{\Theta})\pi(\beta_{(0),k}).
\end{aligned}
$$

(b) We draw $\alpha_{(0)}$ first, all components of $\boldsymbol{\beta}_{(1)}$ second, and $\alpha_{(1)}$ third, from their conditional posterior distributions by using the M–H algorithm in a way very similar to the drawing the components of $\boldsymbol{\beta}_{(0)}$. In all cases, we use normal jumping distributions with variances chosen to have ≈ 30% acceptance rates.

(c) By using Gibbs sampling, we draw, first, $p_{0\to 1}$ and, second, $p_{1\to 0}$ from their conditional posterior distributions, which are

[23] We have only five winter periods in our five-year data. MSNB and MSP with seasonally changing transition probabilities could perform better for an accident data that covers a longer time interval.

[24] One might also consider a threshold model in which the state value is a function of explanatory variables [similar to threshold autoregressive models used in econometrics (Tsay, 2002)]. This interesting possibility is beyond the scope of this study.

[25] We also tried Cauchy jumping distributions and obtained similar results.

truncated beta distributions,

$$f(p_{0\to1}|\mathbf{Y},\mathbf{\Theta}\setminus p_{0\to1}) \propto f(\mathbf{Y},\mathbf{\Theta}) \propto f(\mathbf{S}|p_{0\to1},p_{1\to0})\pi(p_{0\to1},p_{1\to0}) \propto \mathcal{B}eta(\upsilon_0+n_{0\to1},\nu_0+n_{0\to0})I(p_{0\to1}\le p_{1\to0}),$$
$$f(p_{1\to0}|\mathbf{Y},\mathbf{\Theta}\setminus p_{1\to0}) \propto \mathcal{B}eta(\upsilon_1+n_{1\to0},\nu_1+n_{1\to1})I(p_{0\to1}\le p_{1\to0}). \tag{A.1}$$

(d) Finally, we draw components of $\mathbf{S}=[s_1,s_2,\ldots,s_T]'$ by Gibbs sampling. Neighboring components of $\mathbf{S}$ can be strongly (anti-)correlated. Therefore, to speed up MCMC convergence in this case, we draw subsections $\mathbf{S}_{t,\tau}=[s_t,s_{t+1},\ldots,s_{t+\tau-1}]'$ of $\mathbf{S}$ at a time. The conditional posterior distribution of $\mathbf{S}_{t,\tau}$ is

$$f(\mathbf{S}_{t,\tau}|\mathbf{Y},\mathbf{\Theta}\setminus\mathbf{S}_{t,\tau}) \propto f(\mathbf{Y},\mathbf{\Theta}) \propto f(\mathbf{Y}|\mathbf{\Theta})f(\mathbf{S}|p_{0\to1},p_{1\to0}). \tag{A.2}$$

Vector $\mathbf{S}_{t,\tau}$ has length τ and can assume 2^τ possible values. By choosing τ small enough, we can compute the right-hand-side of Eq. (A.3) for each of these values and find the normalization constant of $f(\mathbf{S}_{t,\tau}|\mathbf{Y},\mathbf{\Theta}\setminus\mathbf{S}_{t,\tau})$. This allows us to make Gibbs sampling of $\mathbf{S}_{t,\tau}$. Our typical choice of τ is from 10 to 14. We draw all subsections $\mathbf{S}_{t,\tau}$ one after another.

A detailed description of MCMC simulation algorithms and their implementation in the context of accident modeling can be found in Malyshkina (2008), available at http://arxiv.org/abs/0808.1448.

References

Anastasopoulos, P.Ch., Mannering, F.L. A note on modelingvehicle-accident frequencies with random-parameters count models. Accid. Anal. Prev., in press.
Breiman, L., 1969. Probability and Stochastic Processes with a View Toward Applications. Houghton Mifflin Co., Boston.
Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graphic. Stat. 7 (4), 434–455.
Carson, J., Mannering, F.L., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accid. Anal. Prev. 33 (1), 99–109.
Cowan, G., 1998. Statistical Data Analysis. Clarendon Press, Oxford University Press, USA.
Hadi, M.A., Aruldhas, J., Lee-Fang, C., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transport. Res. Rec. 1500, 169–177.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795.
Lee, J., Mannering, F.L., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prev. 34 (2), 149–161.
Lord, D., Washington, S., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37 (1), 35–46.
Lord, D., Washington, S., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accid. Anal. Prev. 39 (1), 53–55.
Maher, M.J., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accid. Anal. Prev. 28 (3), 281–296.
Malyshkina, N.V., 2006. Influence of speed limit on roadway safety in Indiana. MS thesis, Purdue University. http://arxiv.org/abs/0803.3436.
Malyshkina, N.V., 2008. Markov Switching Models: An Application to Roadway Safety. Ph.D. thesis, Purdue University. http://arxiv.org/abs/0808.1448.
McCulloch, R.E., Tsay, R.S., 1994. Statistical analysis of economic time series via Markov switching models. J. Time Series Anal. 15 (5), 523–539.
Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.
Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Transport. Res. Rec. 1840, 31–40.
Park, B.-J., Lord, D., 2008. Application of Finite Mixture Models for Vehicle Crash Data Analysis. Texas A&M University, unpublished.
Poch, M., Mannering, F.L., 1996. Negative binomial analysis of intersection accident frequency. J. Transport. Eng. 122 (2), 105–113.
Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press, UK.
Robert, C.P., 2001. The Bayesian Choice: from Decision-Theoretic Foundations to Computational Implementation. Springer-Verlag, New York.
SAS Institute Inc., 2006. Preliminary Capabilities for Bayesian Analysis in SAS/STAT Software. SAS Institute Inc., Cary, NC. http://support.sas.com/rnd/app/papers/bayesian.pdf.
Shankar, V., Albin, R., Milton, J., Mannering, F., 1998. Evaluating median cross-over likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transport. Res. Rec. 1635, 44–48.
Shankar, V.N., Mannering, F.L., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.
Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. B 64, 583–639.
Tsay, R.S., 2002. Analysis of Financial Time Series: Financial Econometrics. John Wiley & Sons, Inc.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC.
Wood, G.R., 2002. Generalised linear accident models and goodness of fit testing. Accid. Anal. Prev. 34, 417–427.