Accident Analysis and Prevention 42 (2010) 122–130

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Zero-state Markov switching count-data models: An empirical assessment

Nataliya V. Malyshkina*, Fred L. Mannering

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

ARTICLE INFO

Article history:
Received 26 April 2009
Received in revised form 14 July 2009
Accepted 16 July 2009

Keywords:
Accident frequency count-data models
Zero-inflated models
Negative binomial
Markov switching
Bayesian
Markov Chain Monte Carlo (MCMC)

ABSTRACT

In this study, a two-state Markov switching count-data model is proposed as an alternative to zero-inflated models to account for the preponderance of zeros sometimes observed in transportation count data, such as the number of accidents occurring on a roadway segment over some period of time. For this accident-frequency case, zero-inflated models assume the existence of two states: one of the states is a zero-accident count state, which has accident probabilities that are so low that they cannot be statistically distinguished from zero, and the other state is a normal-count state, in which counts can be non-negative integers that are generated by some counting process, for example, a Poisson or negative binomial. While zero-inflated models have come under some criticism with regard to accident-frequency applications – one fact is undeniable – in many applications they provide a statistically superior fit to the data. The Markov switching approach we propose seeks to overcome some of the criticism associated with the zero-accident state of the zero-inflated model by allowing individual roadway segments to switch between zero and normal-count states over time. An important advantage of this Markov switching approach is that it allows for the direct statistical estimation of the specific roadway-segment state (i.e., zero-accident or normal-count state) whereas traditional zero-inflated models do not. To demonstrate the applicability of this approach, a two-state Markov switching negative binomial model (estimated with Bayesian inference) and standard zero-inflated negative binomial models are estimated using five-year accident frequencies on Indiana interstate highway segments. It is shown that the Markov switching model is a viable alternative and results in a superior statistical fit relative to the zero-inflated models.

1. Introduction

The preponderance of zeros observed in many count-data applications has lead researchers to consider the possibility that two states exist; one state that is a "zero" state (where the probability of an event is so low that it cannot be statistically distinguished from zero) and the other that is a normal-count state that includes zeros and positive integers. This two-state assumption has led to the development of zero-inflated Poisson models and zero-inflated negative binomial models to account for possible overdispersion in the normal-count state. These zero-inflated models have been applied to a number of fields of study. For example, Lambert (1992) used a zero-inflated Poisson model to study manufacturing defects. Lambert argued that unobserved changes in the process caused manufacturing defects to move randomly between a state that was statistically near perfect (the zero state where defects were extremely rare) and an imperfect state where defects were possible but not inevitable (the normal-count state). Lambert's empirical assessment demonstrated that the zero-inflated modeling approach fit the data much better than the standard Poisson. In other work, van den Broek (1995) provided an application of the zero-inflated Poisson to the frequency of urinary tract infections in men diagnosed with the human immunodeficiency virus (HIV). In this case, it was postulated that a near zero-infection state existed for a portion of the patient population and that this state generated a large number of zeros in the frequency data, which was supported by the statistical findings. Also, Bohning et al. (1999) successfully applied the zero-inflated Poisson to study the frequency of dental decay in Portugal.

The frequency of vehicle accidents on a section of highway or at an intersection (over some time period) often exhibit excess zeros.[1] Similar to the literature discussed above, the excess of zeros observed in the data could potentially be explained by the existence of a two-state process for accident data generation (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002). In this case, roadway segments can belong to one of two states:

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

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.07.012

[1] Excess of zeros is relative to the number of zeros predicted by a fitted standard model (such as a Poisson or negative binomial model). A manifestation of excess zeros is a larger-than-predicted tabulated zero frequency (zero bar) in a histogram of accident frequencies (Miaou, 1994; Lord et al., 2005).

a zero-accident state (where zero accidents are expected) and a normal-count state, in which accidents can happen and accident frequencies are generated by some given counting process (Poisson or negative binomial). In the accident analysis arena, zero-inflated models have come under some criticism because many correctly point out that a zero-accident state cannot exist because no roadway is truly safe. However, if one looks through the literature on zero-inflated models and their numerous applications to a wide variety of fields, in no case does a true zero-state exist. In fact, the mechanics of zero-inflated models simply apply a two-state process in which one of the states has occurrence probabilities that are so low that they cannot be statistically distinguished from zero, and the other has a normal count distribution. It is unfortunate, in the accident analysis literature, that the zero state is often conveniently referred to as the "safe" state because this can be taken literally by readers implying the existence of some 100% safe road segments in perpetuity. However, this literal interpretation is not consistent with the underlying econometrics which has probabilistic assignments to the two states. Thus, the zero-accident state should be correctly viewed as a convenient and reasonable theoretical and empirical construct for the description of nearly safe states, in which accident rates are very low and accidents are extremely rare (over the considered time period).

To account for the two-state phenomena, the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models have been used in a number of roadway safety studies (Miaou, 1994; Shankar et al., 1997; Washington et al., 2003). These models explicitly account for an existence of the two states for accident data generation and allow modeling of the probabilities of being in these states. An application of ZIP and ZINB models was an empirical advance in statistical modeling of accident frequencies. And, what is undeniable, is that in many applications zero-inflated models fit accident data far better than traditional count models. However, although traditional zero-inflated models have become popular, they suffer from at least two important drawbacks. First, these models do not deal directly with the states of roadway segments, instead they consider probabilities of being in these states. As a result, zero-inflated models do not allow a direct statistical estimation of whether individual roadway segments are in the zero or normal-count state. For example, suppose a given roadway segment has zero accidents observed over a given time interval. This segment could truly be in a zero-accident state, or it may be in a normal-count state and just happened to have zero accidents over the considered time interval (Shankar et al., 1997). Distinguishing between these two possibilities is not straightforward in zero-inflated models. The second drawback of zero-inflated models is that, although they allow roadway segments to be in different states during different observation periods, zero-inflated models do not explicitly consider switching by the roadway segments between the states over time. It is reasonable to expect that this switching happens sometimes, for example, due to changes in regional or local traffic patterns (which can vary because of re-routing due to construction, changes in travel demands, urban development), changes in law enforcement amount and efficiency (which can vary together with changes in governmental policies), unobserved changes in some roadway characteristics (for example, changes in pavement quality and AADT may remain unobserved because they are measured infrequently), changes in weather conditions (for example, annual precipitation and snow fall amounts may change considerably from year to year), and changes in some other unobserved/unaccounted factors that influence roadway safety. The switching by the roadway segments between the states over time is also important from the theoretical point of view because it is unreasonable to expect any roadway segment to be in the zero state all the time and to have the long-term mean accident frequency equal to zero (Lord et al., 2005).

In this study, we propose a two-state Markov switching count-data model that considers a zero-accident state and a normal-count state. Similar to traditional zero-inflated models, the Markov switching model attempts to statistically account for the preponderance of zeros observed in accident count data. However, in contrast to traditional zero-inflated models, the Markov switching model that we propose allows a direct statistical estimation of the states roadway segments are in at specific points in time and explicitly considers changes in these states over time.

2. Model specification

Two-state Markov switching count-data models of accident frequencies were first presented in Malyshkina et al. (2009). Following that paper, we note that, although there are several major differences between Malyshkina et al. (2009) and this study, many ideas and statistical estimation methods developed in Malyshkina et al. (2009) apply in this study as well. In that paper, two states were assumed to exist but both were normal-count states (i.e., a zero-accident state did not exist). In the current paper, we take a different approach and consider the case where one of the states is a zero state and the other is a normal-count state and that individual roadway segments move between these two states over time. This differs from Malyshkina et al. (2009) in that their model assumes two normal-count states and that all roadway segments are in the same state at the same time.

To show this model, we note that Markov switching models are parametric and can be fully specified by a likelihood function. Let $\mathbf{Y}$ be the vector of all observations, and $\boldsymbol{\Theta}$ be the vector of all parameters of model $\mathcal{M}$. Then the likelihood function $f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})$ is the conditional probability distribution of $\mathbf{Y}$, given $\boldsymbol{\Theta}$ and $\mathcal{M}$. In this study, the observations are the accident numbers $A_{t,n}$ that are observed on roadway segments n during time periods t. Therefore, $\mathbf{Y} = \{A_{t,n}\}$, where $n = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, N is the total number of roadway segments observed (assumed to be constant over time), and T is the number of time periods. Model $\mathcal{M} = \{M, \mathbf{X}_{t,n}\}$, which includes the name M of the model (for example, M = "ZIP" *or* "ZINB"), as well as the vector $\mathbf{X}_{t,n}$ of all roadway segment characteristic variables (for example, number of lanes, ramps and bridges, outer and inner shoulder widths, AADT, median presence and type, segment length, and others).

The definition of a Markov switching model involves a concept of an unobserved (latent) state variable $s_{t,n}$. This variable determines the state of roadway segment n during time period t. In the present study, without loss of generality, we assume that $s_{t,n}$ can take on the following two values: $s_{t,n} = 0$ corresponds to the zero-accident state, and $s_{t,n} = 1$ corresponds to the normal-count state ($n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$). We further assume that, for each roadway segment n, the state variable $s_{t,n}$ follows a stationary two-state Markov chain process in time.[2] This stationary Markov process can be fully specified by time-independent transition probabilities as (Breiman, 1969)

$$P(s_{t+1,n} = 1|s_{t,n} = 0) = p^{(n)}_{0\to 1}, \qquad P(s_{t+1,n} = 0|s_{t,n} = 1) = p^{(n)}_{1\to 0}, \quad (1)$$

where, for example, $P(s_{t+1,n} = 1|s_{t,n} = 0)$ is the conditional probability of $s_{t+1,n} = 1$ at time $t + 1$, given that $s_{t,n} = 0$ at time t. Transition probabilities $p^{(n)}_{0\to 1}$ and $p^{(n)}_{1\to 0}$ are unknown model parameters that need to be estimated from the observed accident data ($n = 1, 2, \ldots, N$). Note that for all long-term averages, one needs to use the stationary unconditional probabilities of states $s_{t,n} =$

[2] Markov property implies that the probability distribution of $s_{t+1,n}$ depends only on the value $s_{t,n}$ at time t, but not on the previous history $s_{t-1,n}, s_{t-2,n}, s_{t-3,n}, \ldots$. Also note that stationarity of Markov process $\{s_{t,n}\}$ is in the statistical sense (Breiman, 1969).

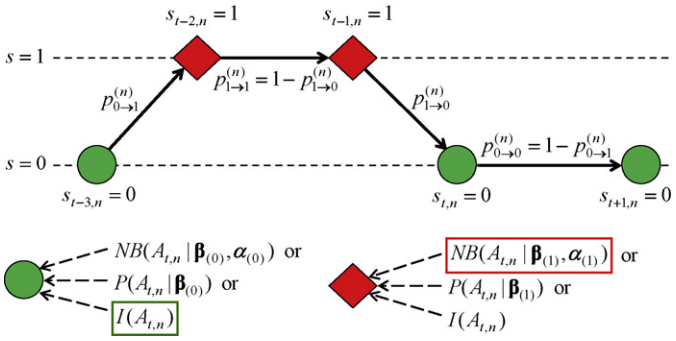

Fig. 1. Graphical demonstration of a two-state Markov switching model.

0 and $s_{t,n}=1$, which are $\tilde{p}_0^{(n)}=p_{1\to0}^{(n)}/(p_{0\to1}^{(n)}+p_{1\to0}^{(n)})$ and $\tilde{p}_1^{(n)}=p_{0\to1}^{(n)}/(p_{0\to1}^{(n)}+p_{1\to0}^{(n)})$, respectively.[3] If $p_{0\to1}^{(n)}<p_{1\to0}^{(n)}$, then $\tilde{p}_0^{(n)}>\tilde{p}_1^{(n)}$ and, on average (in the long-term), for roadway segment n state $s_{t,n}=0$ occurs more frequently than state $s_{t,n}=1$. If $p_{0\to1}^{(n)}>p_{1\to0}^{(n)}$, then state $s_{t,n}=1$ occurs more frequently for segment n.[4]

Now, let us consider a two-state Markov switching negative binomial (MSNB) model that assumes the standard negative binomial (NB) accident-generating process in the normal-count state $s_{t,n}=1$ and no accident generation in the zero-accident state. Thus, this model assumes that the probability of $A_{t,n}$ accidents occurring on the n th roadway segment during time period t is

$$P_{t,n}^{(A)}=\begin{cases}\mathcal{I}(A_{t,n}) & \text{if } s_{t,n}=0\\ \mathcal{NB}(A_{t,n}) & \text{if } s_{t,n}=1\end{cases}, \tag{2}$$

$$\mathcal{I}(A_{t,n})=\left\{1 \text{ if } A_{t,n}=0 \text{ and } 0 \text{ if } A_{t,n}>0\right\}, \tag{3}$$

$$\mathcal{NB}(A_{t,n})=\frac{\Gamma(A_{t,n}+1/\alpha)}{\Gamma(1/\alpha)A_{t,n}!}\left(\frac{1}{1+\alpha\lambda_{t,n}}\right)^{1/\alpha}\left(\frac{\alpha\lambda_{t,n}}{1+\alpha\lambda_{t,n}}\right)^{A_{t,n}}, \tag{4}$$

$$\lambda_{t,n}=\exp(\boldsymbol{\beta}'\mathbf{X}_{t,n}),\quad t=1,2,\ldots,T,\quad n=1,2,\ldots,N. \tag{5}$$

Here, Eq. (3) is the probability mass function that reflects the fact that accidents never happen in the zero-accident state $s_{t,n}=0$.[5] Eq. (4) is the standard negative binomial probability mass function, $\Gamma()$ is the gamma function, and prime means transpose (for example, $\boldsymbol{\beta}'$ is the transpose of $\boldsymbol{\beta}$). Parameter vector $\boldsymbol{\beta}$ and the over-dispersion parameter $\alpha\geq0$ are unknown estimable model parameters.[6] Scalars $\lambda_{t,n}$ are the accident rates in the normal-count state. The first component of $\mathbf{X}_{t,n}$ is set to unity, and, as a result, the first component of vector $\boldsymbol{\beta}$ is the intercept parameter.

A two-state Markov switching model of accident frequencies is graphically demonstrated in Fig. 1. In the two roadway safety states $s=0$ and $s=1$ shown in the figure, the accident frequency data are generated by two different processes, shown by the circles (for state $s=0$) and the diamonds (for $s=1$). In this study, we assume that accident frequency is generated according to the zero-accident distribution $\mathcal{I}(A_{t,n})$ in state $s=0$, and according to the standard negative binomial distribution $\mathcal{NB}(A_{t,n})$ in state $s=1$ (these two distributions are outlined by the boxes in Fig. 1). The state variable $s_{t,n}$ follows a Markov process over time, with transition probabilities $p_{0\to0}^{(n)}$, $p_{0\to1}^{(n)}$, $p_{1\to0}^{(n)}$ and $p_{1\to1}^{(n)}$, as shown in Fig. 1.

Accident events are assumed to be independent. As a result, the likelihood function is

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

In this formula, because the state variables $s_{t,n}$ are not unobserved, the vector of all estimable parameters $\boldsymbol{\Theta}$ must include all state values, in addition to all other model parameters (β, α) and all transition probabilities. Thus, $\boldsymbol{\Theta}=[\boldsymbol{\beta}',\alpha,p_{0\to1}^{(1)},\ldots,p_{0\to1}^{(N)},p_{1\to0}^{(1)},\ldots,p_{1\to0}^{(N)},\mathbf{S}']'$, where vector $\mathbf{S}=[(s_{1,1},\ldots,s_{T,1}),\ldots,(s_{1,N},\ldots,s_{T,N})]'$ has length $T\times N$ and contains all values of the state variables.

Eqs. (1)–(6) define the two-state Markov switching negative binomial (MSNB) model considered here. Note that in this model the estimable state variables $s_{t,n}$ explicitly specify the states of all roadway segments $n=1,2,\ldots,N$ during all time periods $t=1,2,\ldots,T$.

In this study, in addition to the MSNB model, we also consider the standard zero-inflated negative binomial (ZINB) models. In this case, the probability of $A_{t,n}$ accidents occurring is (Washington et al., 2003)

$$P_{t,n}^{(A)}=q_{t,n}\mathcal{I}(A_{t,n})+(1-q_{t,n})\mathcal{NB}(A_{t,n}), \tag{7}$$

$$q_{t,n}=\frac{1}{1+e^{-\tau\log\lambda_{t,n}}}, \tag{8}$$

$$q_{t,n}=\frac{1}{1+e^{-\boldsymbol{\gamma}'\mathbf{X}_{t,n}}}, \tag{9}$$

where we use two different specifications for the probability $q_{t,n}$ that the n th roadway segment is in the zero-accident state during time period t. The right-hand-side of Eq. (7) is a mixture of zero-accident distribution $\mathcal{I}(A_{t,n})$ given by Eq. (3) and negative binomial distribution $\mathcal{NB}(A_{t,n})$ given by Eq. (4). Scalar τ and vector $\boldsymbol{\gamma}$ are estimable model parameters. Accident rate $\lambda_{t,n}$ is given by Eq. (5). We call "ZINB- τ" the model specified by Eqs. (7) and (8). We call "ZINB- γ" the model specified by Eqs. (7) and (9). Note that $q_{t,n}$ depends on the estimable model parameters and gives the probability of being in the zero-accident state $s_{t,n}=0$, but it is not an estimable parameter by itself and does not explicitly specify the state value $s_{t,n}$.

3. Model estimation methods

Markov switching models are difficult to estimate statistically because the state variables $s_{t,n}$ are unobservable.[7] As a result, we cannot use the conventional maximum likelihood estimation (MLE) method in this study. Instead, we use a Bayesian inference approach, which is based on the Bayes formula (Robert, 2001; Congdon, 2007)

$$f(\boldsymbol{\Theta}|\mathbf{Y},\mathcal{M})=\frac{f(\mathbf{Y},\boldsymbol{\Theta}|\mathcal{M})}{f(\mathbf{Y}|\mathcal{M})}=\frac{f(\mathbf{Y}|\boldsymbol{\Theta},\mathcal{M})\pi(\boldsymbol{\Theta}|\mathcal{M})}{\int f(\mathbf{Y},\boldsymbol{\Theta}|\mathcal{M})\,d\boldsymbol{\Theta}}. \tag{10}$$

Here $f(\mathbf{Y}|\boldsymbol{\Theta},\mathcal{M})$ is the likelihood function for model $\mathcal{M}$; $f(\boldsymbol{\Theta}|\mathbf{Y},\mathcal{M})$ is the posterior probability distribution of model parameters $\boldsymbol{\Theta}$ conditional on the observed data $\mathbf{Y}$ and model $\mathcal{M}$; $f(\mathbf{Y},\boldsymbol{\Theta}|\mathcal{M})$ is the joint probability distribution of $\mathbf{Y}$ and $\boldsymbol{\Theta}$ given $\mathcal{M}$; $f(\mathbf{Y}|\mathcal{M})$ is the marginal likelihood function that is the probability distribution of $\mathbf{Y}$ given $\mathcal{M}$;

[3] These follow from stationarity conditions $\tilde{p}_0^{(n)}=[1-p_{0\to1}^{(n)}]\tilde{p}_0^{(n)}+p_{1\to0}^{(n)}\tilde{p}_1^{(n)}$, $\tilde{p}_1^{(n)}=p_{0\to1}^{(n)}\tilde{p}_0^{(n)}+[1-p_{1\to0}^{(n)}]\tilde{p}_1^{(n)}$ and $\tilde{p}_0^{(n)}+\tilde{p}_1^{(n)}=1$ (Breiman, 1969).

[4] Here, Eq. (1) is a significant departure from Malyshkina et al. (2009) in that individual roadway segments can be in different states at the same time (i.e., the state variable is subscripted by roadway segment n). Also, in contrast to Malyshkina et al. (2009), here we do not restrict state $s_{t,n}=0$ to be more frequent than state $s_{t,n}=1$.

[5] Although Eq. (3) formally assumes $s_{t,n}=0$ to be a zero-accident state, in which accidents do not happen, this state can be viewed as an approximation for a nearly safe state, in which the average accident rate is negligible ($\lambda_{t,n}\ll1$) and accidents are extremely rare (over the considered time period).

[6] To ensure that α is non-negative, we estimate its logarithm.

[7] For example, below we will have 335 roadway segments ($N=335$) and five time periods ($T=5$), in which case there are $2^{TN}=2^{1675}$ possible combinations for value of vector $\mathbf{S}=[(s_{1,1},\ldots,s_{T,1}),\ldots,(s_{1,N},\ldots,s_{T,N})]'$.

finally, $\pi(\boldsymbol{\Theta}|\mathcal{M})$ is the prior probability distribution of parameters $\boldsymbol{\Theta}$. While the prior distribution reflects our prior knowledge about model parameters $\boldsymbol{\Theta}$, the posterior distribution, given by Eq. (10), reflects our improved knowledge that accounts for the observed data $\mathbf{Y}$.

Because the parameter vector $\boldsymbol{\Theta}$ has many components (refer to footnote 7), the integration over $\boldsymbol{\Theta}$ in Eq. (10) is not feasible. As a result, the direct application of Eq. (10) is not possible. Instead, we calculate the posterior distribution $f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M})$ in Eq. (10) up to its normalization constant, $f(\boldsymbol{\Theta}|\mathbf{Y}, \mathcal{M}) \propto f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})\pi(\boldsymbol{\Theta}|\mathcal{M})$, and use Markov Chain Monte Carlo (MCMC) simulations (Robert, 2001; Tsay, 2002; Congdon, 2007). MCMC simulations is a standard practical computational methodology for sampling from a probability distribution known up to a constant (the posterior distribution in our study). We use MCMC simulations to draw a large enough posterior sample of parameter vector $\boldsymbol{\Theta}$ from the posterior distribution, and then we use this sample for Bayesian statistical inference (for example, for calculation of posterior expectations and 95% credible intervals of $\boldsymbol{\Theta}$). A reader interested in details is referred to Malyshkina (2008), where we comprehensively describe our choice of the prior distribution $\pi(\boldsymbol{\Theta}|\mathcal{M})$ and the MCMC simulation algorithm.[8] We used MATLAB language for programming and running the MCMC simulations.

We rely on a formal Bayesian approach for comparison of two different statistical models $\mathcal{M}_1$ and $\mathcal{M}_2$ that have parameter vectors $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$. Under condition of equal preferences of these models, the prior probabilities of these models are the same, $\pi(\mathcal{M}_1) = \pi(\mathcal{M}_2) = 1/2$. In this case, the ratio of the models' posterior probabilities, $P(\mathcal{M}_1|\mathbf{Y})$ and $P(\mathcal{M}_2|\mathbf{Y})$, is (Kass and Raftery, 1995)

$$\frac{P(\mathcal{M}_2|\mathbf{Y})}{P(\mathcal{M}_1|\mathbf{Y})} = \frac{f(\mathcal{M}_2, \mathbf{Y})/f(\mathbf{Y})}{f(\mathcal{M}_1, \mathbf{Y})/f(\mathbf{Y})} = \frac{f(\mathbf{Y}|\mathcal{M}_2)\pi(\mathcal{M}_2)}{f(\mathbf{Y}|\mathcal{M}_1)\pi(\mathcal{M}_1)} = \frac{f(\mathbf{Y}|\mathcal{M}_2)}{f(\mathbf{Y}|\mathcal{M}_1)}. \quad (11)$$

Here $f(\mathcal{M}_1, \mathbf{Y})$ and $f(\mathcal{M}_2, \mathbf{Y})$ are the joint distributions of the models and the data; $f(\mathbf{Y}|\mathcal{M}_1)$ and $f(\mathbf{Y}|\mathcal{M}_2)$ are the marginal likelihoods of the models; $f(\mathbf{Y})$ is the unconditional probability distribution of the data $\mathbf{Y}$. The right-hand-side of Eq. (11) is called the Bayes factor (Kass and Raftery, 1995; Robert, 2001). As in Malyshkina et al. (2009), to calculate the marginal likelihoods, we use the harmonic mean formula $f(\mathbf{Y}|\mathcal{M})^{-1} = E\left[f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})^{-1}|\mathbf{Y}\right]$, where $E(\ldots|\mathbf{Y})$ means posterior expectation calculated by using the posterior distribution. If the ratio in Eq. (11) 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. It is important to note that the Bayesian approach to model comparison guards against overfitting due to inclusion of too many model parameters.

For evaluation of the performance of model $\{\mathcal{M}, \boldsymbol{\Theta}\}$ in fitting the observed data $\mathbf{Y}$, we use a χ^2 goodness-of-fit test (Maher and Summersgill, 1996; Cowan, 1998; Wood, 2002; Press et al., 2007). We perform this test by Monte Carlo simulations to find the distribution of the χ^2 quantity, which measures the discrepancy between the observations and the model predictions (Cowan, 1998). Then we use this distribution to find the goodness-of-fit p-value, which is the probability that χ^2 exceeds the observed value of χ^2 under the hypothesis that the model is true (the observed value of χ^2 is calculated by using the observed data $\mathbf{Y}$). For additional details, please see Malyshkina (2008).

4. Empirical results

Data are used from 5769 accidents that were observed on 335 interstate highway segments in Indiana in 1995–1999. The highways were I-64, I-65, I-70, I-74 and I-164. The segments were chosen as non-overlapping highway segments with homogeneous pavement type and geometric characteristics (number of lanes, median type and width, median barrier presence, inside and outside shoulder presence and width, rumble strips presence, and speed limit). At least one of these characteristic varied between adjacent segments. We use annual time periods, $t = 1, 2, 3, 4$, $T = 5$ in total.[9] Thus, for each roadway segment $n = 1, 2, \ldots, N = 335$ the state $s_{t,n}$ can change every year.[10] The histogram of annual accident frequencies is shown in Fig. 2. Four types of accident frequency models are estimated:

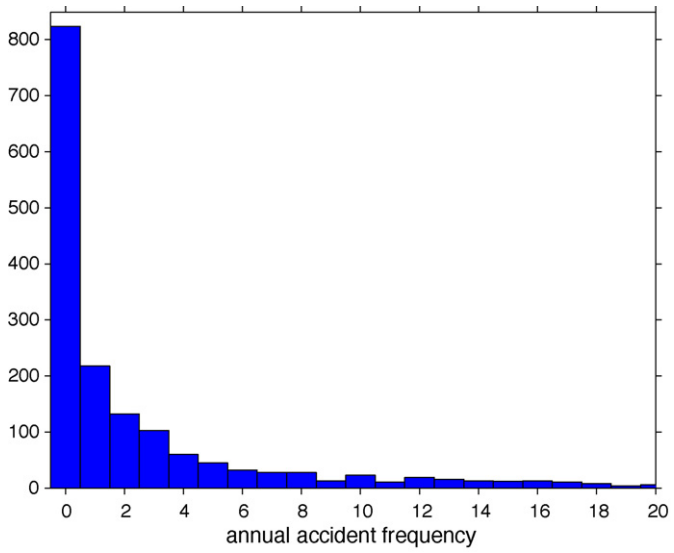

Fig. 2. Histogram of annual accident frequencies.

(1) First, for the purpose of explanatory variable selection, we estimate an auxiliary standard negative binomial (NB) model, which is not reported here. We estimate this model by maximum likelihood estimation (MLE). To obtain a standard NB model, we choose 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). In order to make a comparison of explanatory variable effects in different models straightforward, in all other models, described below, we use only those explanatory variables that enter the standard NB model.[12]

[8] Our priors for α, β, $p_{0\to1}$ and $p_{1\to0}$ are flat or nearly flat, while the prior for the states $\mathbf{S}$ reflects the Markov process property, specified by Eq. (1).

[9] We also considered quarterly time periods and obtained qualitatively similar results (not reported here).

[10] Note that Malyshkina et al. (2009) used weekly accident frequencies with a large number of zeros which would possibly have made it a candidate for zero-inflated switching. However, recall that the current paper is a generalization of Malyshkina et al. (2009) in that individual segments can switch between states from one time period to the next (Malyshkina et al., 2009 had the simplifying assumption that all segments must switch at the same time). This generalization would make estimation using weekly data extremely difficult because even with the annual data that we use the length of the state vector $\mathbf{S}$ is 1675 (see footnote 7), as opposed to only 260 in Malyshkina et al. (2009). The longer vector $\mathbf{S}$ is, the slower MCMC convergence becomes, and extensive supercomputer access would be needed to apply this model to weekly data. Again, our intent in this paper is to demonstrate methodology and the annual accident data works best for this purpose in this case.

[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 (11). Unfortunately, because MCMC

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

Variable	ZINB- τ [a]		ZINB- γ [b]		MSNB [c]
	by MLE	by MCMC	by MLE	by MCMC	by MCMC
β- and α-parameters in Eq. (5)					
Intercept (constant term)	$-15.0^{-12.5}_{-17.5}$	$-15.2^{-13.0}_{-17.4}$	$-11.6^{-8.32}_{-14.8}$	$-11.6^{-8.29}_{-14.6}$	$-17.3^{-13.0}_{-21.3}$
Accident occurring on interstates I-70 or I-164 (dummy)	$-.683^{-.570}_{-.797}$	$-.685^{-.575}_{-.794}$	$-.715^{-.602}_{-.829}$	$-.715^{-.593}_{-.836}$	$-.734^{-.617}_{-.850}$
Pavement quality index (PQI) average [d]	$-.0122^{-.0189}_{-.00550}$	$-.0122^{-.00562}_{-.0188}$	$-.0140^{-.00627}_{-.0217}$	$-.0143^{-.00643}_{-.0221}$	$-.0163^{-.00850}_{-.0240}$
Logarithm of road segment length (in miles)	$.791^{.832}_{.751}$	$.791^{.829}_{.754}$	$.929^{.978}_{.880}$	$.939^{.993}_{.886}$	$.887^{.929}_{.845}$
Number of ramps on the viewing side per lane per mile	$.226^{.300}_{.153}$	$.227^{.306}_{.149}$	$.298^{.387}_{.209}$	$.304^{.394}_{.214}$	$.317^{.404}_{.230}$
Number of lanes on a roadway	–	–	–	–	$1.19^{2.04}_{.386}$
Median configuration is depressed (dummy)	$.184^{.288}_{.0795}$	$.183^{.282}_{.0839}$	$.201^{.319}_{.0820}$	$.202^{.325}_{.0781}$	–
Median barrier presence (dummy)	$-1.43^{-1.22}_{-1.64}$	$-1.43^{-1.14}_{-1.72}$	–	–	$-1.69^{-1.00}_{-2.46}$
Width of the interior shoulder is less that 5 ft. (dummy)	$.323^{.443}_{.202}$	$.323^{.434}_{.211}$	$.435^{.572}_{.297}$	$.437^{.569}_{.307}$	$.374^{.505}_{.243}$
Outside shoulder width (in feet)	$-.0480^{-.0196}_{-.0764}$	$-.0478^{-.0207}_{-.0749}$	$-.0532^{-.0176}_{-.0887}$	$-.0532^{-.020}_{-.0867}$	$-.0537^{-.0214}_{-.0862}$
Outside barrier is absent (dummy)	–	–	$-.245^{-.117}_{-.373}$	$-.245^{-.101}_{-.389}$	$-.264^{-.124}_{-.403}$
Average annual daily traffic (AADT)	$-4.07^{-3.17}_{-4.97} \times 10^{-5}$	$-4.14^{-3.31}_{-5.04} \times 10^{-5}$	$-1.93^{-3.21}_{-6.50} \times 10^{-5}$	$-1.91^{-3.16}_{-5.83} \times 10^{-5}$	$-3.78^{-2.02}_{-5.26} \times 10^{-5}$
Logarithm of average annual daily traffic	$1.89^{2.17}_{1.61}$	$1.91^{2.16}_{1.67}$	$1.52^{1.88}_{1.15}$	$1.52^{1.86}_{1.15}$	$1.95^{2.34}_{1.49}$
Number of bridges per mile	–	–	–	–	$-.0214^{-.00164}_{-.0428}$
Maximum of reciprocal values of horizontal curve radii (in 1/mile)	$-.140^{-.0710}_{-.209}$	$-.141^{-.0734}_{-.208}$	$-.134^{-.0559}_{-.213}$	$-.138^{-.0593}_{-.217}$	$-.106^{-.0289}_{-.183}$
Percentage of single unit trucks (daily average)	$1.23^{1.84}_{.624}$	$1.23^{1.82}_{.646}$	$1.32^{1.96}_{.693}$	$1.32^{1.96}_{.691}$	$1.29^{1.90}_{.688}$
Number of changes per vertical profile along a roadway segment	$.0555^{.0930}_{.0180}$	$.0562^{.0903}_{.0226}$	–	–	–
Over-dispersion parameter α in NB models	$.144^{.183}_{.105}$	$.150^{.192}_{.114}$	$.130^{.168}_{.0925}$	$.142^{.185}_{.105}$	$.114^{.147}_{.0847}$
τ- and γ-parameters in Eqs. (8) and (9)					
The model parameter τ in Eq. (8)	$-1.72^{-1.45}_{-2.00}$	$-1.73^{-1.50}_{-1.98}$	–	–	–
Intercept (constant term)	–	–	$23.1^{41.3}_{4.99}$	$26.5^{47.0}_{10.9}$	–
Logarithm of road segment length (in miles)	–	–	$-1.34^{-.942}_{-1.73}$	$-1.4^{-1.03}_{-1.83}$	–
Median barrier presence (dummy)	–	–	$3.97^{4.86}_{3.08}$	$4.16^{5.20}_{3.27}$	–
Average annual daily traffic (AADT)	–	–	$9.23^{15.1}_{3.35} \times 10^{-5}$	$10.5^{17.4}_{5.72} \times 10^{-5}$	–
Logarithm of average annual daily traffic	–	–	$-2.88^{-.901}_{-4.86}$	$-3.28^{-1.59}_{-5.57}$	–
Mean accident rate ($\lambda_{t,n}$ for NB), averaged over all values of $X_{t,n}$	–	3.38	–	3.42	3.88
Standard deviation of accident rate ($\sqrt{\lambda_{t,n}(1+\alpha\lambda_{t,n})}$ for NB), averaged over all values of explanatory variables $X_{t,n}$	–	2.14	–	2.15	2.13
Total number of free model parameters (β-s, γ-s, α and τ)	16	16	19	19	16
Posterior average of the log-likelihood (LL)	–	$-2510.68^{-2506.13}_{-2517.12}$	––	$-2436.34^{-2431.12}_{-2443.54}$	$-2124.82^{-2096.30}_{-2153.91}$
Max(LL) : estimated max. value of log-likelihood (LL) for MLE; maximum observed value of LL for Bayesian-MCMC	−2502.67(MLE)	−2503.21(observed)	−2426.54(MLE)	−2427.41(observed)	−2049.45(observed)
Logarithm of marginal likelihood of data ($\ln[f(\mathbf{Y}\|\mathcal{M})]$)	–	$-2519.90^{-2516.95}_{-2521.59}$	–	$-2447.33^{-2443.93}_{-2448.86}$	$-2184.21^{-2186.70}_{-2169.56}$
Goodness-of-fit p-value	–	0.005	–	0.177	0.191
Maximum of the potential scale reduction factors (PSRF) [e]	–	1.01006	–	1.02200	1.02117
Multivariate potential scale reduction factor (MPSRF) [e]	–	1.01023	–	1.02302	1.02189

[a] Standard (conventional) ZINB- τ model estimated by maximum likelihood estimation (MLE) and Markov Chain Monte Carlo (MCMC) simulations.
[b] Standard ZINB- γ model estimated by maximum likelihood estimation (MLE) and Markov Chain Monte Carlo (MCMC) simulations.
[c] Two-state Markov switching negative binomial (MSNB) model where all reported parameters are for the normal-count state $s = 1$.
[d] The pavement quality index (PQI) is a composite measure of overall pavement quality evaluated on a 0–100 scale.
[e] PSRF/MPSRF are calculated separately/jointly for all continuous model parameters. PSRF and MPSRF are close to 1 for converged MCMC chains.

(2) We estimate the standard ZINB- τ model, specified by Eqs. (6)–(8). First, we estimate this model by maximum likelihood estimation (MLE) and use the 5% statistical significance level for evaluation of the statistical significance of each β-parameter. Second, we estimate the same ZINB- τ model by the Bayesian inference approach and MCMC simulations. As one expects, the Bayesian–MCMC estimation results turned out to be similar to the MLE estimation results for the ZINB- τ model.

(3) We estimate the standard ZINB- γ model, specified by Eqs. (6), (7) and (9). First, we estimate this model by MLE and use the 5% statistical significance level for evaluation of the statistical significance of each β-parameter. Second, we estimate the same ZINB- γ model by the Bayesian inference approach and MCMC simulations. The Bayesian–MCMC and the MLE estimation results for the ZINB- γ model turned out to be similar.

(4) We estimate the two-state Markov switching negative binomial (MSNB) model, specified by Eqs. (1)–(6), by the Bayesian–MCMC methods. We consecutively construct and use 60%, 85% and 95% Bayesian credible intervals for evaluation of the statistical significance of each β-parameter in the MSNB

simulations are computationally expensive, evaluation of marginal likelihoods for a large number of trial models is not feasible in our study.

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
Logarithm of road segment length (in miles)	−.901	1.22	−4.71	−1.03	2.44
Number of ramps on the viewing side per lane per mile	.138	.408	0	0	3.27
Number of lanes on a roadway	2.09	.286	2.00	2.00	3.00
Median configuration is depressed (dummy)	.630	.484	0	1.00	1.00
Median barrier presence (dummy)	.161	.368	0	0	1
Width of the interior shoulder is less that 5 ft. (dummy)	.696	.461	0	1.00	1.00
Outside shoulder width (in feet)	11.3	1.74	6.20	11.2	21.8
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
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
Percentage of single unit trucks (daily average)	.0859	.0678	.00975	.0683	.322
Number of changes per vertical profile along a roadway segment	.522	.908	0	0	6.00

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

model. As a result, in the final MSNB model some components of $\boldsymbol{\beta}$ are restricted to zero.[13] No restriction is imposed on the over-dispersion parameter α, which turns out to be significant anyway.

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

The estimation results show that the MSNB model is strongly favored by the empirical data, as compared to the standard ZINB models.[15] Indeed, from Table 1 we see that the MSNB model provides considerable, 335.69 and 263.12, improvements of the logarithm of the marginal likelihood of the data as compared to the ZINB- τ and ZINB- γ models.[16] Thus, from Eq. (11), we find that, given the accident data, the posterior probability of the MSNB model is larger than the probabilities of the ZINB- τ and ZINB- γ models by $e^{335.69}$ and $e^{263.12}$, respectively.[17]

Let us now consider the maximum likelihood estimation (MLE) of the standard ZINB- τ and ZINB- γ models and an imaginary MLE estimation of the MSNB model. Referring to Table 1, the MLE gave maximum log-likelihood values −2502.67 and −2426.54 for the ZINB- τ and ZINB- γ models. The maximum log-likelihood value observed during our MCMC simulations for the MSNB model is equal to −2049.45. An imaginary MLE, at its convergence, would give MSNB log-likelihood value that would be even larger than this observed value. Therefore, the MSNB model, if estimated by the MLE, would provide very large, at least 453.22 and 377.09, improvements in the maximum log-likelihood value over the ZINB- τ and ZINB- γ models. These improvements would come with no increase or a decrease in the number of free continuous model parameters $(\beta, \alpha, \tau, \gamma)$ that enter the likelihood function.

To evaluate the goodness-of-fit for a model, we use the posterior (or MLE) estimates of all continuous model parameters $(\beta, \alpha, p^{(n)}_{0\to1}, p^{(n)}_{1\to0})$ and generate 10^4 artificial data sets under the hypothesis that the model is true.[18] We find the distribution of χ^2 and calculate the goodness-of-fit p-value for the observed value of χ^2. For details, see (Malyshkina et al., 2009). The resulting p-values for our models are given in Table 1. For the ZINB- γ and MSNB models the p-values are sufficiently large, around 20%, which indicates that these models fit the data reasonably well. At the same time, for the ZINB- τ model the goodness-of-fit p-value is only around 0.5%, which indicates a much poorer fit.[19]

The estimation results also show that the over-dispersion parameter α is higher for the ZINB- τ and ZINB- γ models, as compared to the MSNB model (refer Table 1). 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 of roadway safety.

Now, refer to Fig. 3, made for the case of the MSNB model. The four plots in this figure show five-year time series of the posterior probabilities $P(s_{t,n} = 1|\mathbf{Y})$ of the normal-count state for four selected roadway segments. These plots represent the following four categories of roadway segments:

(1) For roadway segments from the first category we have $P(s_{t,n} = 1|\mathbf{Y}) = 1$ for all $t = 1, 2, 3, 4, 5$. Thus, we can say with absolute certainty that these segments were always in the normal-count state $s_{t,n} = 1$ during the considered five-year time interval. A

[13] A β-parameter is restricted to zero if it 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] The ZINB models are, in turn, strongly favored over a simple NB model, see Malyshkina (2008). The logarithm of the marginal likelihood of the data for the NB model is about −2554.16. As a result, the posterior probability of the MSNB model is about $e^{369.95}$ larger than the posterior probability of the NB model.

[16] We use the harmonic mean formula to calculatethe values and the 95% confidence intervals of the log-marginal-likelihoods given in Table 1. The confidence intervals are calculated by bootstrap simulations. For details, see Malyshkina et al. (2009) or Malyshkina (2008).

[17] There are other frequently used model comparison criteria, for example, the deviance information criterion, $\mathrm{DIC} = 2E[D(\boldsymbol{\Theta})|\mathbf{Y}] - D(E[\boldsymbol{\Theta}|\mathbf{Y}])$, where deviance $D(\boldsymbol{\Theta}) \equiv -2\ln[f(\mathbf{Y}|\boldsymbol{\Theta}, \mathcal{M})]$ (Robert, 2001). Models with smaller DIC are favored to models with larger DIC. We find DIC values 5037.3, 4891.4, 4261.5 for the ZINB- τ, ZINB- γ and MSNB models, respectively. This means that the MSNB model is favored over the standard ZINB models. 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. (11).

[18] Note that the state values $\mathbf{S}$ are generated by using $p^{(n)}_{0\to1}$ and $p^{(n)}_{1\to0}$.

[19] It is worth to mention that for the auxiliary standard negative binomial (NB) model, which we do not report here, the goodness-of-fit p-value was also very poor, ≈ 0.3%. This is an expected result because of a preponderance of zeros in the data, not accounted for in the NB model.

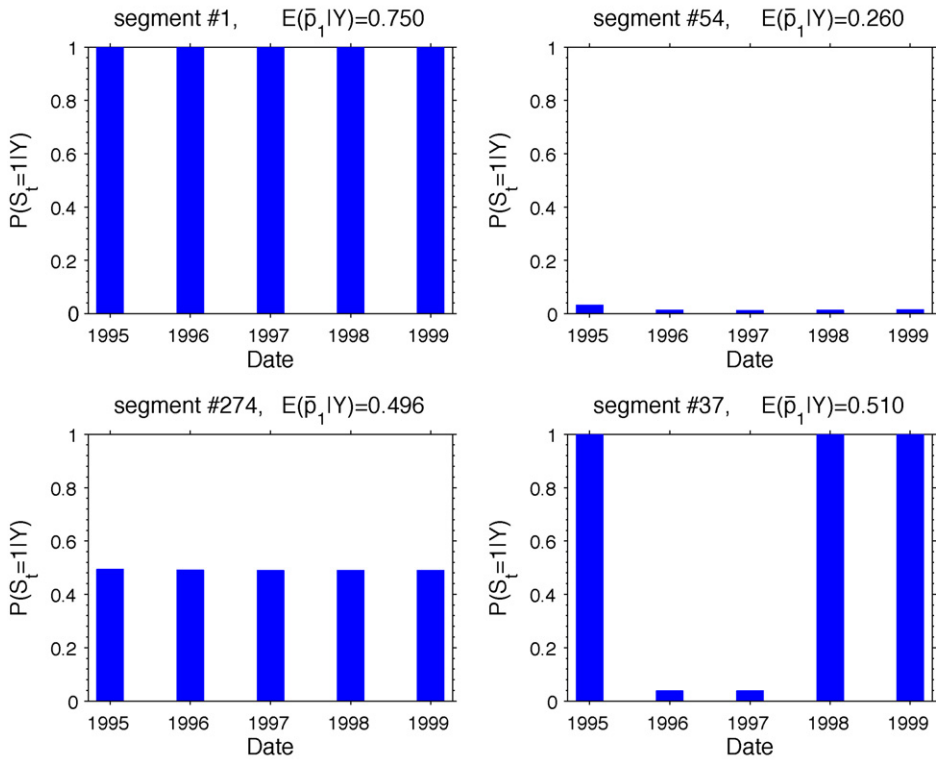

Fig. 3. Five-year time series of the posterior probabilities $P(s_{t,n} = 1|\mathbf{Y})$ of the normal-count state $s_{t,n} = 1$ for four selected roadway segments ($t = 1, 2, 3, 4, 5$).

roadway segment belongs to this category if and only if it had at least one accident during each year ($t = 1, 2, 3, 4, 5$). An example of such roadway segment is given in the top-left plot in Fig. 3. For this segment the posterior expectation of the long-term unconditional probability $\bar{p}_1$ of being in the normal-count state is large, $E(\bar{p}_1|Y) = 0.750$.

(2) For roadway segments from the second category $P(s_{t,n} = 1|\mathbf{Y}) \ll 1$ for all $t = 1, 2, 3, 4, 5$. Thus, we can say with high degree of certainty that these segments were always in the zero-accident state $s_{t,n} = 0$ during the considered five-year time interval. A roadway segment n belongs to this category if it had no accidents observed over the five-year interval despite the accident rates given by Eq. (5) were large, $\lambda_{t,n} \gg 1$ for all $t = 1, 2, 3, 4, 5$. Clearly this segment would be unlikely to have zero accidents observed, if it were not in the zero-accident state all the time.[20] An example of such roadway segment is given in the top-right plot in Fig. 3. For this segment $E(\bar{p}_1|Y) = 0.260$ is small.

(3) For roadway segments from the third category $P(s_{t,n} = 1|\mathbf{Y})$ is neither one nor close to zero for all $t = 1, 2, 3, 4, 5$.[21] For these segments we cannot determine with high certainty what states these segments were in during years $t = 1, 2, 3, 4, 5$. A roadway segment n belongs to this category if it had no any accidents observed over the considered five-year time interval and the accident rates were not large, $\lambda_{t,n} \lesssim 1$ for all $t = 1, 2, 3, 4, 5$. In fact, when $\lambda_{t,n} \ll 1$, the posterior probabilities of the two states are close to one-half, $P(s_{t,n} = 1|\mathbf{Y}) \approx P(s_{t,n} = 0|\mathbf{Y}) \approx 0.5$, and no inference about the value of the state variable $s_{t,n}$ can be made. In this case of small accident frequencies, the observation of zero accidents is perfectly consistent with both states $s_{t,n} = 0$ and $s_{t,n} = 1$. An example of a roadway segment from the third category is given in the bottom-left plot in Fig. 3. For this segment $E(\bar{p}_1|Y) = 0.496$ is about one-half.

(4) Finally, the fourth category is a mixture of the three categories described above. Roadway segments from this fourth category have posterior probabilities $P(s_{t,n} = 1|\mathbf{Y})$ that change in time between the three possibilities given above. In particular, for some roadway segments we can say with high certainty that they changed their states in time from the zero-accident state $s_{t,n} = 0$ to the normal-count state $s_{t,n} = 1$ or vice versa. An example of a roadway segment from the fourth category is given in the bottom-right plot in Fig. 3. For this segment $E(\bar{p}_1|Y) = 0.510$ is about one-half. Thus we find a direct empirical evidence that some roadway segments do change their states over time.

Next, it is useful to consider roadway segment statistics by state of roadway safety. Referring to Fig. 4, a case is made for the MSNB model. The top plot in this figure shows the histogram of the posterior probabilities $P(s_{t,n} = 1|\mathbf{Y})$ for all $N = 335$ roadway segments during all $T = 5$ years (1675 values of $s_{t,n}$ in total). For example, we find that during five years roadway segments had $P(s_{t,n} = 1|\mathbf{Y}) = 1$ and were normal-count in 851 cases, and they had $P(s_{t,n} = 1|\mathbf{Y}) < 0.2$ and were likely to be zero-accident in 212 cases. The bottom plot in Fig. 4 shows the histogram of the posterior expectations $E[\bar{p}_1^{(n)}|\mathbf{Y}]$, where $\bar{p}_1^{(n)} = p_{0\to1}^{(n)}/(p_{0\to1}^{(n)} + p_{1\to0}^{(n)})$ are the stationary unconditional probabilities of the normal-count state (see Section 2). We find that $0.2 \le E[\bar{p}_1^{(n)}|\mathbf{Y}] \le 0.8$ for all segments $n = 1, 2, \ldots, 335$. This means

[20] Note that the zero-accident state may exist due to under-reporting of minor, low-severity accidents (Shankar et al., 1997).

[21] If there were no Markov switching, which introduces time-dependence of states via Eqs. (1), then, assuming non-informative priors $\pi(s_{t,n} = 0) = \pi(s_{t,n} = 1) = 1/2$ for states $s_{t,n}$, the posterior probabilities $P(s_{t,n} = 1|\mathbf{Y})$ would be either exactly equal to 1 (when $A_{t,n} > 0$) or necessarily below 1/2 (when $A_{t,n} = 0$). In other words, we would have $P(s_{t,n} = 1|\mathbf{Y}) \notin [0.5, 1)$ for any t and n. Even with Markov switching existent, in this study we have never found any $P(s_{t,n} = 1|\mathbf{Y})$ close but not equal to 1, refer to the top plot in Fig. 4.

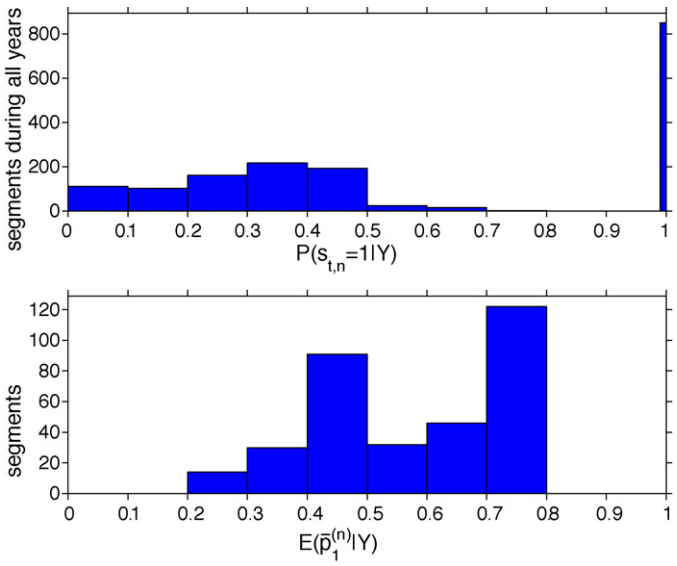

Fig. 4. Histograms of the posterior probabilities $P(s_{t,n} = 1|\mathbf{Y})$ (the top plot) and of the posterior expectations $E[\bar{p}_1^{(n)}|\mathbf{Y}]$ (the bottom plot). Here $t = 1, 2, 3, 4, 5$ and $n = 1, 2, \ldots, 335$.

that in the long run, all roadway segments have significant probabilities of visiting both the zero-accident and the normal-count states.

To save space and because this paper is methodologically oriented, we do not discuss the safety effects of all explanatory variables that enter Table 1. However, it is noteworthy to consider the effects of AADT (averaged annual daily traffic) and roadway segment length. To be specific, we focus on the MSNB model, reported in the far right column in Table 1. Refer to this column and to Eq. (5). We find that the average accident rate $\lambda_{t,n}$ depends on both AADT and the logarithm of AADT, namely, $\lambda_{t,n} \propto \exp[-.0000378\,\mathrm{AADT} + 1.95\,\log(\mathrm{AADT})] = \mathrm{AADT}^{1.95}\exp(-.0000378\,\mathrm{AADT})$. The $\mathrm{AADT}^{1.95}$ power-law term in this formula implies a faster-than-linear growth in accident rate with an increase of AADT. This is a reasonable result, and its likely explanation is that vehicles interfere with each other while sharing the road (for example, when AADT increases, driving behavior and traffic patterns may change adversely). On the other hand, this adverse interference effect is mitigated by the $\exp(-.0000378\,\mathrm{AADT})$ exponential term, which decreases with AADT. Moreover, when AADT exceeds approximately 51,600, the exponential term starts to dominate over the power-law term, resulting in a decrease of the average accident rate with a further increase of AADT. Apparently, this decrease can take place in a very heavy traffic (traffic jams), in which driving speeds decline considerably below speed limit values. However, one has to look at this prediction with caution, it may be not statistically very robust because in our data sample there are only 16.1% observations with AADT above 51,600 (in addition, underreporting of minor property-damage-only accidents can be an issue in slow-speed high AADT traffic flows). As far as the roadway segment length effect is concerned, from the results of Table 1 we see that the average accident rate $\lambda_{t,n}$ is proportional to the segment length in the power less than one. This finding is likely due to segment-boundary effects since segment lengths are defined by changing roadway geometrics, and thus changing geometrics (number of lanes, shoulder widths, etc.) may be associated with accident clustering. This same segment-length effect was also found in the study conducted by Anastasopoulos and Mannering (2009).

Finally, it is also worth mentioning that, in addition to negative binomial models, we estimated Poisson models for the same accident data and obtained similar results (Malyshkina, 2008). In particular, we found that a two-state Markov switching Poisson (MSP) model, which has the Poisson likelihood function instead of the NB likelihood function in Eq. (4), is strongly favored by the empirical data as compared to standard zero-inflated Poisson models.

5. Conclusions

A number of important observations can be made with regard to our empirical findings. First, Markov switching count-data models provide a superior statistical fit for accident frequencies relative to standard zero-inflated models. Second, Markov switching models, which explicitly consider transitions between the zero-accident state and the normal-count state over time, permit a direct empirical estimation of what states roadway segments are in at different time periods. This estimation can help in a detailed study of individual roadway segments, which can lead to a better understanding of why some segments are considerably safer than others during some time periods. The estimation results also provide evidence that some roadway segments changed their states over time (see the bottom-right plot in Fig. 3). Third, the Markov switching models avoid a theoretically implausible assumption that some roadway segments are always in the zero-accident state because, in these models, every segment has a non-zero probability of being in the normal-count state. Indeed, the long-term unconditional mean of the accident rate for the n th roadway segment is equal to $\bar{p}_1^{(n)}\langle\lambda_{t,n}\rangle_t$, where $\bar{p}_1^{(n)} = p_{0\to1}^{(n)}/(p_{0\to1}^{(n)} + p_{1\to0}^{(n)})$ is the stationary probability of being in the normal-count state $s_{t,n} = 1$ and $\langle\lambda_{t,n}\rangle_t$ is the time average of the accident rate in the normal-count state [refer to Eq. (5)]. This long-term mean is always above zero (see the bottom plot in Fig. 4), even for segments that were likely to be in the zero-accident state over the whole observed five-year time interval. Note that it is the unconditional expectation $\bar{p}_1^{(n)}\langle\lambda_{t,n}\rangle_t$ that should be used for long-term predictions, safety decision making and efforts to reduce long-term averaged accident rates in practical applications of the models. Finally, we conclude that two-state Markov switching count-data models are likely to be a better alternative to zero-inflated models, in order to account for excess of zeros often observed in accident-frequency data.[22]

References

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle-accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.
Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., Kirchner, U., 1999. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. J. R. Stat. Soc. A 162 (2), 195–209.
Breiman, L., 1969. Probability and Stochastic Processes with a View Toward Applications. Houghton Mifflin Co., Boston.
van den Broek, J., 1995. A score test for zero-inflation in a Poisson distribution. Biometrics 51 (2), 738–743.
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.
Congdon, P., 2007. Bayesian Statistical Modelling. John Wiley & Sons, Inc.
Cowan, G., 1998. Statistical Data Analysis. Clarendon Press, Oxford Univ. Press, USA.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795.
Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 1–14.
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.

[22] Overdispersion errors in a gamma-Poisson (negative binomial) distribution can be correlated over time (and/or between adjacent roadway segments). These correlations can be accounted for by extension of a Markov switching model to include random effects (similar to inclusion of random effects into the standard NB model). To be sure, past research has found such time and spatial correlation to be statistically insignificant (see footnote 4 in Anastasopoulos and Mannering, 2009). Still, an extension of our model to include the possibility of random effects would be an important direction for future work.

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.
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.
Malyshkina, N.V., Mannering, F.L., Tarko, A.P., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accid. Anal. Prev. 41 (2), 217–226.
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.
Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes 3rd edition: The Art of Scientific Computing. Cambridge Univ. Press, UK.
Robert, C.P., 2001. The Bayesian Choice: From Decision-theoretic Foundations to Computational Implementation. Springer-Verlag, New York.
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.