Analytic Methods in Accident Research 1 (2014) 86–95

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Modeling safety of highway work zones with random parameters and random effects models

Erdong Chen *, Andrew P. Tarko [1]

Center for Road Safety, School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, USA

A R T I C L E I N F O

Article history:
Received 30 May 2013
Received in revised form
30 October 2013
Accepted 30 October 2013
Available online 2 December 2013

Keywords:
Highway safety
Work zone safety
Crash frequency modeling
Random parameters
Negative binomial model

A B S T R A C T

This paper presents an investigation of traffic safety in highway work zones using detailed data obtained from the results of a survey of project engineers and existing datasets. The observations were organized in monthly clusters that correspond to individual work zones; and a two-level random parameters negative binomial model that reflected the structure of the observations was estimated. The safety effects of various work zone design and traffic management features were identified, including lane shift, lane split, and detour, whose safety effects have not been evaluated in past research. This new insight into highway work zone safety was accomplished thanks to the better data acquired and the improved statistical model. A fixed parameters negative binomial model with random effects then was estimated to check its viability as an alternative to the random parameters model when the sample's large size makes estimation of the latter challenging. From a practical standpoint, the marginal effects on crash frequency computed from the model with random effects were quite similar to those computed from the random parameters model. This result indicates that, at least in some cases, convenient fixed parameters models may be a practical alternative to random parameters models. Utilization of an entire sample to estimate these conventional models may further compensate a less advanced model specification. The obtained negative binomial model with random effects has become useful for programming police enforcement in highway work zones in Indiana.

1. Introduction

Statistical models play an important role in acquiring knowledge and attaining a better understanding of highway safety, its factors, and effective countermeasures. Mannering and Bhat (this issue) point out the recent fast advancement in the analytic methodology of safety analysis. They also note the widening gap between the "frontier" models and modeling practice in highway safety. One point must be emphasized though. Not all applications of early models are missed opportunities for applying a better methodology if the selection of a model type considers not only the state-of-the art but also the purpose of developing the model.

An example of safety management comes to mind here. Safety management includes identification of roads that exhibit excessive frequency and severity of crashes; and, typically, only traffic volumes and key road characteristics are available for

* Corresponding author. Tel.: +1 765 494 9821; fax: +1 765 494 0395.
E-mail addresses: chen608@purdue.edu, chenerdong007@gmail.com (E. Chen), tarko@purdue.edu (A.P. Tarko).
[1] Tel.: +1 765 494 5027; fax: +1 765 494 0395.

2213-6657/$ - see front matter
http://dx.doi.org/10.1016/j.amar.2013.10.003

all the roads in large road networks. Thus, a model with a limited number of variables is fitted to data that represent the population of roads, and then the model is used to estimate the expected number of crashes on these roads. A road with a large positive residual (excessive crash frequency) is a good candidate for a detailed site investigation. Even if the model parameters reflect other variables omitted from the model (e.g., the traffic volume parameter may reflect better winter maintenance of busy roads), the model may still produce reasonable estimates of the crash frequency if applied to the same population. The parsimony of the model can sometimes be based on necessity and, more importantly, purpose. Nevertheless, applicable modeling advancements, such as accounting for a spatial correlation, should be utilized even in this case. On the other hand, if individual safety effects are to be estimated and this knowledge is to be applied outside of the studied population or sample, then the model must be carefully specified and the benefits offered by the state-of-the-art models utilized as discussed by Mannering and Bhat (this issue).

This paper discusses two statistical models developed to analyze work zone safety: a random parameters model and a random effects model. The benefit of using the random parameters model in providing additional insight into the safety factors of work zones will be emphasized. Furthermore, the justification of using a model with random effects, which might be considered by some researchers as less advanced among the two discussed, will be shown based on its performance and intended use.

First, the existing literature relevant to this paper's topic is summarized by pointing out representative examples of published work on the matter. Then, the data collected and assembled for the study is briefly described, and the two considered models are introduced and their estimation described. Next, the obtained models are discussed and compared, followed by a summary of the findings to conclude the paper.

1.1. Highway work zone safety

There is strong evidence that work zone conditions involve heightened risk of crash, which justifies the ongoing efforts to better understand the safety factors in work zones and to propose effective countermeasures. Early research on this subject compared the crash rates between construction and non-construction conditions (Rouphail, et al., 1988; Ha and Nemeth, 1995; Pal and Sinha, 1996; Khattak et al., 2002). Their research confirmed that crash rates during construction periods tend to be higher than in non-construction periods. More recent research used statistical modeling to examine the link between work zone characteristics and safety measured with crash frequency and severity. Lane closures and construction intensity were found to be significant factors of crash frequency in work zones (Pal and Sinha, 1996; Venugopal and Tarko, 2000) as well as the length of the work zone and the duration of the construction period (Khattak et al., 2002; Venugopal and Tarko, 2000). Different sections of the work zone were found to experience different prevailing crashes types (Garber and Zhao, 2002). Finally, traffic conditions and driver behavior were identified as major contributing factors that increase the frequency of crashes (Harb et al., 2008; Daniel et al., 2000; Wang et al., 1996).

A number of studies investigated the injury severity of work zone crashes (Khattak et al., 2002; Venugopal and Tarko, 2000; Li and Bai, 2008; Khattak and Targa, 2004). Venugopal and Tarko (2000) estimated the relationships between the work zone characteristics and the frequencies of crashes of different severity. The obtained relationships were similar for exposure variables across injury categories, while injury and fatal crashes were more affected by the construction intensity but not by the type of work. Li and Bai (2008) used logistic models to estimate crash severity and developed a Crash Severity Index, in which a set of factors of crash severity in work zones were identified. Khattak and Targa (2004) estimated work zone truck crash severity with an ordered logit model. Lane crossovers and the proximity of construction activities were found to significantly increase crash severity.

The relatively small number of studies on highway work zone crashes could be attributed to both the typically small sample of crashes (most work zones last for a short period) and the limited scope of data about work zones that cannot be easily increased afterwards when the work zone no longer exists. Traditionally, a work zone crash is identified through the police crash reports. Ullman and Scriba (2004) found that in 1992, only 14 states had explicit entries of data indicating work zone presence in their crash reports; 21 states had some work zone information entered, but it was not reported explicitly; and 16 states had no work zone information included in their crash reports. Although this situation has improved since the above study was conducted, missing work zone information still poses a major issue for work zone safety research.

Wang et al. (1996) also reviewed the current state of crash reporting issues along with federal efforts undertaken to improve the situation. They pointed out that work zone crashes are seriously under-represented due to (1) lack of explicit information, (2) non-inclusion of crashes outside the work zones but nonetheless caused by the construction activities, and (3) failure to identify a crash that happened in a work zone. The authors recommended that work zone information and status should be better reported in police crash reports. They also concluded that the existence of a work zone inventory database is essential for gaining a better understanding of work zone safety.

Another issue is the limited data about work zone characteristics. State transportation agencies preserve basic work zone information, but detailed road cross-section data and traffic management plans are not readily available. As a result, researchers either must use the existing limited data or make a great effort to collect detailed work zone characteristics. This situation is responsible for the limited number of variables included in past studies on work zone safety.

1.2. Traffic crash modeling

The methodology of crash modeling has significantly improved since most of the reviewed studies took place. Lord and Mannering (2010) and Savolainen et al. (2011) discussed the state-of-the-art statistical models useful for crash frequency and severity analysis, while Mannering and Bhat (this issue) provide an updated overview and discussion of the current methodology and future directions and opportunities of crash data analysis. A variety of potential issues with a crash frequency analysis were pointed out in Lord and Mannering (2010) along with possible solutions. The most popular method for analyzing work zone crash frequency is negative binomial models, which successfully addressed the over-dispersion issue. Time-varying explanatory variables, omitted variables, crash under-reporting, and varying effects were not well accounted for in the past studies. The first three issues are caused by data deficiencies and are difficult to address without the availability of better data. The varying effects can be dealt with using mixed models by taking a form of random parameters model that relaxes the fixed parameters assumption (Lord and Mannering, 2010). To some extent, the random parameters models also address the undesirable effects of variables omission by attributing the resulted unexplained heterogeneity to the model parameters. The recent methodological improvements, including the random parameters models, applied to better work zone data, may provide an opportunity to learn more about highway work zone safety (Mannering and Bhat, this issue).

A fundamental assumption of non-random parameters models is that the parameter values remain fixed for all observations in the sample (Washington et al., 2011). While convenient for estimating the model and predicting the dependent variable value, this assumption is a major limitation for such models. Due to unexplained heterogeneity in a sample, the same variable may have different effects on the dependent variable in different observations. For example, a low impact speed may cause minimal injury to a young driver, but it may turn out to be lethal for an 80-year-old driver. Omission of the age and physical condition of a driver in the model and forcing a fixed value on the parameter associated with the impact speed can lead to model misspecification. Random parameters models are more flexible in that regard, and they are considered more adequate for handling unexplained heterogeneity than fixed parameters models.

The mixed logit model used by Milton et al. (2008) is considered the first application of random parameters models in traffic safety. The motivation for this application was to properly represent strongly heterogeneous vehicles and driver characteristics in models of crash severity (Ulfarsson and Mannering, 2004; Islam and Mannering, 2006) suitable for identification of engineering countermeasures. Using a random parameters model with road and traffic-specific variables included and the heterogeneity of vehicles and drivers accounted for with varying model parameters are believed to yield unbiased estimates of road-specific variables that should lead to adequate engineering countermeasures. Although no comparison between fixed and random parameters models were given, the random parameters model gave much more insights into how various factors affect crash severity. Later, two studies (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009) used random parameters models for estimating crash frequencies. Anastasopoulos and Mannering (2009) assumed normal distribution for all the random parameters and showed that the random parameters model resulted in the ρ_C^2 improving from 0.177 to 0.223 over the fixed parameters model. El-Basyouny and Sayed (2009), on the other hand, explored the prior and posterior distributions with Full Bayes estimation. Poisson-lognormal models were estimated with fixed effects, random effects, and random parameters, respectively, and the results showed that the random parameters model had the best fit. Nonetheless, these applications of random parameters models provided empirical evidence for their potential, both in providing more accurate estimation of model parameters and in obtaining additional insights into the safety effects of predicting variables.

In a previous study by the authors of this paper (Chen and Tarko, 2012), a two-level work zone database with detailed work zone characteristics was assembles to tackle the omitted and time-varying explanatory variables issues, and a random effects model was estimated to account for the temporal correlation. In this effort, a random parameters negative binomial model was estimated using the same database to account for the potential heterogeneity in the data and to get a better understanding of the relations between work zone characteristics and crash frequency.

Estimation of random parameters models delivers the distributions of some of their parameters. Although this model structure seems to better fit reality and reduces the estimation bias, the use of such models to estimate the expected value of the dependent variable requires Monte Carlo or numerical integration techniques, which makes these models less convenient for engineering purposes. Another drawback is the current lack of an estimation tool for large samples. A researcher faces a dilemma in using a random parameters model and a portion of the sample when the sample size is beyond the capability of the estimation tools or when using a traditional model but with an entire sample. This paper is meant to help readers deal with this dilemma by comparing the results obtained from the random parameters model and from the best specified traditional model (random effects). The results are presented and interpreted, followed by conclusions and recommendations.

2. Data

Similar to previous studies by other authors, the work zone data initially available for this study was limited in scope. The Indiana Department of Transportation (INDOT) provided basic data for projects with letting in 2009. This data included the project's location, total award, and type of construction. More detailed cross-section elements and traffic management information were not included. A questionnaire survey was designed and distributed to project engineers in charge of these work zones to obtain additional information. The following information was collected for each distinct construction phase: project starting and ending dates, cross-section geometry during the construction period (number of lanes open, lane shift, etc.), traffic management details (use of barrier, detour sign, etc.), and the presence of police enforcement. The survey forms were mailed to the project engineers for the 189 work zones identified by the authors as appropriate for the purpose of this study. Seventy-eight surveys were returned, and 72 were used in the study after eliminating six responses that had missing important information. The 72 work zones included in the analysis represent a 38% survey return.

The locations of the 72 studied work zones showed that the sample sufficiently represented rural vs. urban and freeway vs. non-freeway conditions. The next step in data preparation was linking the work zone data obtained from the project engineers survey with the Indiana road inventory, which provided the detailed road characteristics and traffic volumes prior to the construction. The continuous geometric characteristics (e.g., right-of-way-width) may change with distance, and they are represented in the sample with values averaged over the road sections with construction activities. The binary variables (e.g., urban—yes, no) are presented with proportions. The lack of traffic volume measurements during the construction periods made it necessary to use the traffic volume for the regular road conditions. To account for the temporal variation in traffic volume, the intensity of construction activities, and the seasonal effects, each construction period was broken down into monthly observations with varying adjusted average daily traffic (ADT), monthly indicators, and construction phases. This created a two-level data structure with the higher level corresponding to a work zone and the lower level corresponding to a month.

Finally, crashes were assigned to the two-level work zone database. As Ullman and Scriba (2004) and Wang et al. (1996) discussed, assigning crashes based on information (explicit or inexplicit) in police crash reports may lead to under-representation of work zone crashes. With the geo-coded work zone database and starting and ending dates available, crashes were assigned to a work zone if the data indicated that (1) the crash happened on a road segment located within a work zone and (2) the crash happened during the construction period. More specifically, the crashes were assigned to corresponding monthly observations. Although this approach in identifying work zone crashes is still subject to the accuracy of the crash locations in the crash report, it addresses the lack of explicit information in the crash report; and it allows identifying crashes that occurred on road segments upstream of work zones and affected by these work zones.

3. Methodology

For crash frequency analysis, Poisson and negative binomial models, along with their variants, have been the prevailing methods (Washington, et al., 2011; Lord and Mannering, 2010) used in past studies. For data displaying over-dispersion (variance greater than mean), Poisson distribution is not appropriate. In negative binomial distribution, a Gamma-distributed dispersion term α is added to relax the equal mean and variance assumption. The mean and variance of the negative binomial models are expressed in Eqs. (1) and (2).

$$E(y_i) = \lambda_i = \mathrm{EXP}(\beta X_i + \varepsilon_i) \tag{1}$$

$$\mathrm{VAR}(y_i) = E(y_i)(1 + \alpha E(y_i)) = E(y_i) + \alpha E(y_i)^2 \tag{2}$$

where y_i is the number of work zone crashes per month for observation i, λ_i is the mean number of crashes for observation i, β is a vector of estimable parameters, X_i is a vector of covariates (explanatory variables) for observation i, ε_i is a gamma distributed dispersion term, and α is a dispersion parameter (for additional details see Washington et al., 2011). To account for heterogeneity, all the previously mentioned random parameters model applications specified the parameters to vary at the observation level as shown in Eq. (3).

$$\beta_i = \beta + \omega_i \tag{3}$$

where subscript i denotes each observation, and ω_i is a randomly distributed error term following a priori distribution. However, due to the structure of the data used in this study, it is suspected that, with the monthly indicator variable and adjusted ADT for the monthly observations, heterogeneity among the monthly observations for the same work zone will be relatively insignificant; on the other hand, at the work zone level, even though several cross-sectional and traffic management variables were collected in the survey and included in the model, there could still be significant heterogeneity unaccounted for by the covariates. Thus, in this study, the parameters were specified to vary at the work zone level, as did Wu et al. (2013). The formulation for this model structure is shown in Eq. 4.

$$\beta_{ij} = \beta + \omega_j \tag{4}$$

where subscript i denotes each monthly observation, while subscript j denotes each work zone, and ω_j is a randomly distributed error term (assumed to be normally distributed with mean 0 and the variance estimated by the model) varying at the work zone level. This error structure reflects the assumption that the unaccounted heterogeneity is mainly due to the differences between the work zones. With this, the Poisson parameter is now conditioned on the error term ω_j, as shown in Eq. 5.

$$\lambda_{ij}|\omega_j = \mathrm{EXP}(\beta X_{ij}) \tag{5}$$

The sample log-likelihood for the negative binomial model with the corresponding probabilities $P(n_{ij}|\omega_j)$ is shown in Eq. (6).

$$LL = \textstyle\sum_{\forall_i} ln \displaystyle\int_{\omega_j} g(\omega_j) P(n_{ij}|\omega_j)\, d\omega_j \tag{6}$$

The g(.) is the probability density form of the random distribution term assumed to be normal. The estimation of random parameters count models was developed by Greene (2007). Maximum likelihood estimation was found to be computationally cumbersome for such applications, and a simulation technique called Halton draws (Halton, 1960) has been the most popular estimation technique in estimating random parameters models (Bhat, 2003; Train, 2009). Their studies also showed empirically that Halton draws converge quicker than random draws, with 200 draws usually sufficient to provide accurate estimation without being too time consuming (Bhat, 2003; Train, 2009). If only the intercept of the model is specified as a random parameter, then the model reduces to a random effects model, where the expected value of the dependent variable is shown in Eq. (7).

$$\lambda_{ij} = \mathrm{EXP}(\beta x_{ij} + \varepsilon_{ij} + \eta_j) \tag{7}$$

where subscript j indicates each work zone, and η_j is the random effects.

4. Estimation

To check to what extent the monthly observations structure and the use of the random parameters model improve the model fitting and the results in general, two models were estimated for comparison: (1) a random parameters model with monthly observations and (2) a random effects model with monthly observations (with only the intercept specified as random to capture the shared unobserved correlation at the work zone level). The models were estimated with LIMDEP (Greene, 2007). The obtained models include the combination of variables that delivers the best model fit to the data. The overall model fitting statistics are shown in Table 1, and the parameter estimates are shown in Table 2. Although the log-likelihood values indicate that the random parameters model better fits to the data, the Akaike Information Criterion (AIC) statistic points out that the random effects model has a slightly better predictive power because there are fewer parameters included in this model.

As expected, the two models presented in Table 2 differ from each other. The random effects model has fewer variables included. Nevertheless, the differences between the random parameters model and the random effects model are relatively small. By specification, the work zone-related error term in the random effects model is varying across work zones, while the random parameters model has six variables (including the intercept) with the parameters significantly varying across work zones. It seems that most of the parameter values in the random effects model have slightly higher values than the corresponding expected parameter values in the random parameters model. Also, in the random effects model, the standard deviation for the intercept was of much higher magnitude (0.514) compared to the random parameters model (only 0.062, but still significant). It is suspected that the standard deviation for the intercept in the random effects model was grasping the heterogeneities among the work zones, whereas in the random parameters model, the heterogeneity was (supposedly) to a good extent accounted for by the standard deviations of the explanatory variables, and only a small portion of heterogeneity at the work zone level was left to be captured by the standard deviation of the intercept. This may also explain why the log-likelihood functions of both the random parameters and random effects models are very close.

Marginal effects for both the random parameters model and the random effects model are presented in Table 3. The marginal effects measure the changes in the conditional mean of the dependent variable in response to unit changes in the explanatory variables. Differences could be observed between the two models, but most of them are of relatively small magnitude.

5. Discussion

The variables included in the models were divided into four groups: (1) exposure variables, (2) roadway characteristics prior to construction, (3) work zone features, and (4) temporal variables. The random parameters model is considered superior and it is given more focus in the following discussion.

5.1. Exposure variables

The model parameter estimate for the work zone length of 0.797 indicates that the crash frequency tends to increase with the length of the work zone at a decreasing rate. The crash density in long work zones is expected to be lower than in short work zones with the same traffic volume. The total traffic volume used in the model is the accumulated ADT over the

Table 1
Model fitting statistics.

Models	Random parameters	Random effects
Number of observations	547	547
LL(0) (log-likelihood with nothing)	−2811.234	−2811.234
LL(C) (log-likelihood with constant only)	−1242.688	−1242.688
LL(β) (log-likelihood with covariates)	−923.248	−927.879
Rho(0) square	67.16%	66.99%
Rho(C) square	25.71%	25.33%
AIC	1894.496	1893.757

Table 2
Parameter estimation of work zone crash frequency models.

Variable description	Random parameters		Random effects	
	Estimate	*t* Value	Estimate	*t* Value
Constant	−5.454	−31.179	−5.441	−31.965
Standard deviation of parameter distribution	*0.062*	*3.077*	*0.514*	*20.789*
Exposure variables				
Logarithm of work zone length (in miles)	0.797	23.601	0.881	26.974
Logarithm of total traffic volume (in thousands)	0.853	29.239	0.814	29.200
Roadway characteristics prior to construction				
Left shoulder width (in feet)	−0.037	−4.622	−0.059	−7.144
Right-of-way width (in feet)	−0.004	−12.113	−0.006	−11.200
Urban land development fraction	1.011	10.952	1.409	12.993
Standard deviation of parameter distribution	*0.293*	*8.058*	–	–
Park lane fraction	−1.675	−3.495	−2.191	−4.484
Work zone features				
Indicator for multi-lane freeway work zones (no system interchange)	0.634	7.005	0.942	7.481
Indicator for multi-lane freeway work zones (with system interchange)	0.406	4.381	0.364	3.531
Indicator for freeway work zones restricted to one lane (per direction)	Not significant		0.268	2.301
Indicator for detour sign	−0.240	−4.546	−0.295	−5.386
Indicator for lane shift	0.237	5.051	0.257	5.456
Indicator for lane split	0.261	2.947	0.307	3.411
Indicator for low construction intensity (cost $<$ \$10,000/(day*mi))	0.419	6.798	0.592	8.775
Standard deviation of parameter distribution	*0.488*	*17.753*	–	–
Indicator for high construction intensity (cost $>$ \$35,000/(day*mi))	0.431	4.370	0.622	6.318
Standard deviation of parameter distribution	*0.087*	*2.432*	/	/
Temporal variables				
Indicator for May–July	0.125	3.115	0.180	4.946
Standard deviation of parameter distribution	*0.207*	*5.885*	–	–
Indicator for November–December, in Northern Indiana	0.097	1.360	Not significant	
Standard deviation of parameter distribution	*0.402*	*4.657*	–	–
Indicator for November–December, in Central Indiana	0.240	2.592	0.241	2.629
Indicator for November–December, in Southern Indiana	0.251	2.171	0.229	1.906
Dispersion parameter for negative binomial distribution				
Dispersion parameter, α	0.02553	2.653	0.036362	3.359

Table 3
Marginal effects.

Variable description	Random parameters model	Random effects model
Exposure variables		
Logarithm of work zone length (in miles)	1.170	1.270
Logarithm of total traffic volume (in thousands)	1.251	1.173
Roadway characteristics prior to construction		
Left shoulder width (in feet)	−0.054	−0.085
Right-of-way width (in feet)	−0.006	−0.008
Urban land development fraction	1.483	2.031
Park lane fraction	−2.458	−3.159
Work zone features		
Indicator for multi-lane freeway work zones (no system interchange)	0.930	1.358
Indicator for multi-lane freeway work zones (with system interchange)	0.596	0.524
Indicator for freeway work zones restricted to one lane (per direction)	–	0.386
Indicator for detour sign	−0.351	−0.425
Indicator for lane shift	0.348	0.370
Indicator for lane split	0.382	0.443
Indicator for low construction intensity (cost $<$ \$10,000/(day*mi))	0.615	0.853
Indicator for high construction intensity (cost $>$ \$35,000/(day*mi))	0.633	0.897
Temporal variables		
Indicator for May–July	0.184	0.260
Indicator for Nov–Dec, in Northern Indiana	0.142	–
Indicator for Nov–Dec, in Central Indiana	0.351	0.348
Indicator for Nov–Dec, in Southern Indiana	0.368	0.330

days of work. The use of the total volume instead of keeping the average ADT and the number of days of work separate was justified by the close values of the corresponding two parameters. The total traffic volume exhibits a similar relationship with safety as does the work zone length; the corresponding parameter is 0.853.

5.2. Roadway characteristics prior to construction

Work zones on roads with a wide right-of-way and a wide left shoulder tended to experience less crashes. The results obtained were similar to Chen and Tarko (2012), but the estimated parameters were of smaller magnitude (-0.037 and -0.004, respectively).

The model parameter of urban land development fraction (the proportion of the total work zone length in an urban area) turned out to vary significantly between locations (with mean 1.011 and standard deviation 0.293, see Table 2). This result indicates that urban work zones tend to have more crashes on average than rural work zones. The distribution of this parameter, depicted in Fig. 1, shows that this difference between urban and rural locations varies considerably, but urban work zones seem to have more frequent crashes than their rural counterparts in practically all cases. All the distribution figures used the template developed by Wittwer (2004).

Work zones that had a high fraction of parking lanes (the proportion of the total work zone length with parking lanes prior to the construction) were found to have significantly less crashes. The parking lane presence in regular non-work-zone conditions serves as a surrogate for a low functional class road with a low speed limit. Another possible explanation of this result could be that a wider roadway allows a less restricted traveled way during the construction period.

5.3. Work zone features

Freeway work zones with multiple lanes open to traffic during the construction period were found to experience a significantly higher frequency of crashes than freeway single-lane and all non-freeway work zones. Marginal effects in Table 3 show that multilane freeway work zones with system interchanges had 0.93 more crashes per month on average; while for those freeway work zones without system interchanges had 0.596 more crashes per month on average relative to single-lane freeway work zones and non-freeway work zones.

The presence of a detour sign was found to significantly reduce crashes by an average of 0.351 crashes per month as indicated by the marginal effects in Table 3. It is believed that a detour provision with detour signs guiding drivers around a work zone significantly reduced the traffic volume inside the work zone and subsequently reduced the frequency of crashes inside the work zone. Diverting traffic not only alleviates traffic congestion in highway work zones, but it also brings safety benefits in these work zones. The possible negative effect of detours on road network safety around the work zone was not investigated in this study.

Lane shifts and lane splits were found to increase the crash frequency by 0.348 and 0.382 crashes per month, respectively (see Table 3). This result is intuitive since these geometry departures from regular conditions make driving in the work zone more challenging. This result also should prompt transportation agencies to revisit the current practice of designing traffic lane shifts and splits.

Both high and low construction intensity indicators were found to significantly increase crash frequency. However in this random parameters model, both were found to have significant standard deviations. For low construction intensity, the mean parameter estimation was 0.419, while the standard deviation for the distribution was 0.488. This suggested that, although an overall low construction intensity increases crash frequency (by 0.615 crashes per month on average as shown in Table 3), the effects differed across work zones; and according to the normal distribution assumption, about 20% of these low intensity work zones even had less crashes, as shown in Fig. 2. It might be expected that low-intensity work zones experience fewer crashes due to less frequent distractions of drivers from the work crew. Indeed, it seems that in some cases the safety was higher in low-intensity work zones (negative range of the coefficient in Fig. 2). Nevertheless, the majority of low-intensity work zones experienced a higher frequency of crashes than medium-intensity work zones. This result may be interpreted as the effect of driver relaxation and overconfidence due to the lack of construction activities. On the other hand, all high-intensity work zones exhibited a higher crash frequency than medium-intensity work zones (Fig. 3). The mean value of the parameter was estimated as 0.431 with a standard deviation of 0.087, which gave an average increase of 0.633 crashes per month (see Table 3).

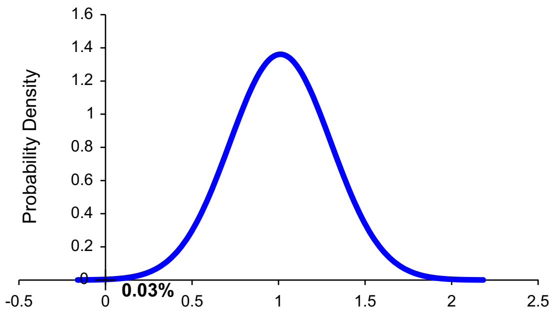

Fig. 1. Distribution of parameter estimation for urban fraction.

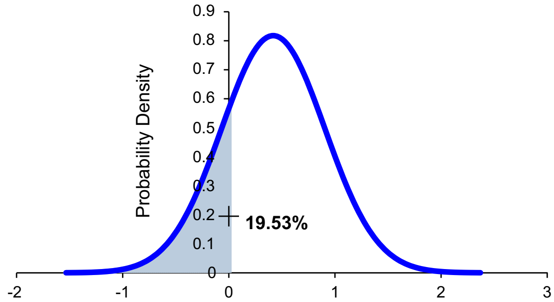

Fig. 2. Distribution of parameter estimation for low intensity.

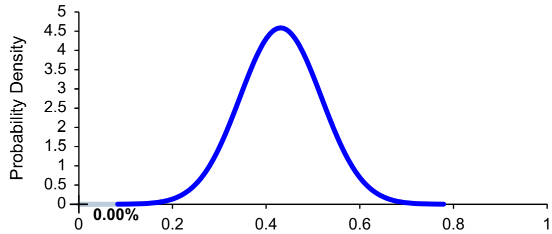

Fig. 3. Distribution of parameter estimation for high intensity.

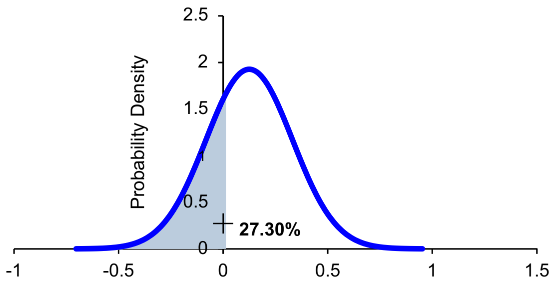

Fig. 4. Distribution of parameter estimation for summer indicator.

5.4. *Temporal variables*

Both the summer months (May through July) and the winter months (November through December) were found to experience an increase in the crash frequency. The distribution of the parameter estimates for these months obtained in the random parameters model provided more insight. Although the summer indicator had a positive mean estimate of 0.125 (0.184 more crashes per month on average), the considerably larger standard deviation of 0.239 indicated a relatively large portion of work zones with the negative effect of summer (27.3% in Fig. 4). As summer temperatures are not considered a major cause of crashes in Indiana, the authors suspect that such variations in crash frequency could be caused by varying population of drivers during the summer season (e.g., tourism and recreation activities).

The strong variability of the safety effect of the winter months (more precisely, late fall and early winter) was identified. The parameter was originally estimated to be 0.149 with a large standard deviation of 0.154. To further investigate the unknown heterogeneity associated with the winter month conditions, Indiana was divided into three regions based on the INDOT districts. The LaPorte and Fort Wayne districts were categorized as the Northern region, the Crawfordsville and Greenfield districts as the Central region, and the Vincennes and Seymour districts as the Southern region. This treatment was dictated by the hypothesis that varying weather conditions (geographically and temporally) may lead to different safety levels at different locations. The Northern region tends to experience particularly unstable and varying weather condition during fall and winter because of the proximity to Michigan Lake and its well-known "lake effect."

Separate winter month parameters were estimated for the three regions (see Table 2). The winter month parameters for the Central and Southern regions turned out to be fixed and close to each other (0.240 and 0.251, respectively). The corresponding effects on crash frequency are equivalent to 0.351 and 0.368 more crashes per month as indicated by marginal effects (see Table 3). For the Northern region, the winter month parameter turned out to be random with a mean close to zero (0.097) with a large standard deviation of 0.402 (see Fig. 5). The parameter distribution indicates that about

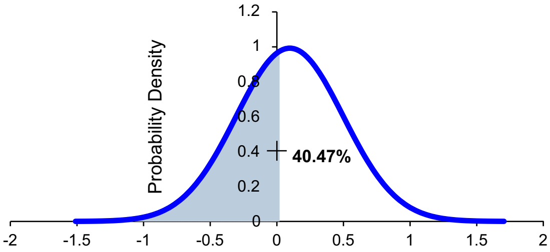

Fig. 5. Distribution of parameter estimation for winter indicator (northern region).

40% of the work zones in the Northern region had fewer crashes in November and December than during other months and 60% had more crashes.

6. Conclusions

Due to the data limitations faced by past studies and the recent opportunities presented by vastly improved methodologies to investigate this topic in a new way, the authors conducted this current study to achieve a better understanding of highway work zone safety issues.

Data improvements were sought first in this study; and it is clear from the modeling output that the additional work zone information retrieved from the project engineers survey significantly improved the model fitting and the understanding of the underlying causalities for work zone crashes, especially issues related to work zone design and traffic management. Use of detour signs, lane shifts, and lane splits were all found to significantly affect the crash frequency, all of which had not been identified by any previous modeling study. The study by Wang et al. (1996) recommended that a work zone inventory database would be most helpful in improving the understanding of work zone safety issues, and the present study confirmed their conclusion with strong evidence. The other data improvement in this study was the creation of a monthly data structure, which allowed accounting for the time varying effects. Both the random parameters and random effects models, while also taking care of the shared unobserved correlation issues associated with repeated observations, showed significant improvements over the model estimated with data representing the entire construction period (model not shown in this paper). This finding suggests that work zone crash frequency varies over time and this variability provides the opportunity for better temporal allocation of safety treatment resources such as speed enforcement.

Improved modeling methodologies were also implemented. In this analysis, both the random parameters and random effects models were shown with the monthly observation data. It is important to point out that the model with random effects (this improvement is equivalent to estimating the intercept as a random parameter) had both reasonably estimated fixed parameters and overall model fitting comparable to the random parameters model. The AIC measure, which favors parsimonious models, indicated that the random effects model was slightly better in predicting crash frequencies. If the purpose of the model is to conveniently and accurately predict crash frequency in work zones, then the random effects model is a good choice in the considered case. This is particularly true for large samples when an estimation method for a random parameters model may be troublesome.

On the other hand, if the main purpose of the model is to estimate certain safety effects represented by the model variables, then particular caution should be given to the choice of a model. Although the differences in most of the estimated parameters were found to be relatively small in the studied case some of the differences were considerable. From the results of this analysis, it is recommended that, whenever the methodology is available, random parameters model or more recent model techniques (see Mannering and Bhat, this issue) be used as they are theoretically more advanced, provide more insights, and have the most accurate parameter estimation.

In conclusion, efforts were made in this study both to improve the quality and scope of data and to implement more sophisticated modeling techniques for highway work zone safety analysis. Not only better estimation is promised by the random parameters/random effects negative binomial models, but several work zone design/management features also were identified in this study as significantly affecting crash frequency, which have never been available for analyzing work zone crashes in the past. Significant time varying factors were found as well that could only be identified using disaggregated data, which can provide new insights to guide better allocation of safety countermeasure resources.

This study's findings pointed out the following directions for future studies concerning highway work zone safety: (1) work zone inventory datasets should be maintained by all state transportation agencies for identifying safety issues and potential improvements; and (2) disaggregated data and advanced statistical models should be used to account for time varying effects and to achieve the best estimates possible.

Disclaimer

This research was funded by the Indiana Department of Transportation through the Joint Transportation Research Program (JTRP). The contents of this paper reflect the views of the authors, who are responsible for the facts and the

accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Indiana Department of Transportation or the Federal Highway Administration at the time of publication. This report does not constitute a standard, specification, or regulation.

References

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41 (1), 153–159.
Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (9), 837–855.
Chen, E., Tarko, A., 2012. Analysis of Work Zone Crash Frequency with Focus on Police Enforcement. Compendium of papers, Transportation Research Board 91st Annual Meeting, Paper No. 12-2416, Washington DC.
Daniel, J., Dixon, K., Jared, D., 2000. Analysis of fatal crashes in Georgia work zones. Transportation Research Record 1715, 18–23.
El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 41 (5), 1118–1123.
Garber, N.J., Zhao, M., 2002. Distribution and characteristics of crashes at different work zone Locations in Virginia. Transportation Research Record 1794, 19–25.
Greene,W., 2007. Limdep, Version 9.0. Econometric Software Inc., Plainview, NY.
Ha, T., Nemeth, Z., 1995. Detailed study of accident experience in construction and maintenance zones. Transportation Research Record 1509, 38–45.
Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2 (1), 84–90.
Harb, R., Radwan, E., Yan, X., Pande, A., Abdel-Aty, M., 2008. Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression. Journal of Transportation Engineering 134 (5), 203–214.
Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: some additional evidence. Journal of Safey Research 37 (3), 267–276.
Khattak, A., Targa, F., 2004. Injury severity and total harm in truck-involved work zone crashes. Transportation Research Record 1877, 106–116.
Khattak, A., Khattak, A., Council, F., 2002. Effects of work zone presence on injury and non-injury crashes. Accident Analysis and Prevention 34 (1), 19–29.
Li, Y., Bai, Y., 2008. Development of crash-severity-index models for the measurement of work zone risk levels. Accident Analysis and Prevention 40 (5), 1724–1731.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44 (5), 291–305.
Mannering, F., Bhat, C.R. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods Accident Research, http://dx.doi.org/10.1016/j.amar.2013.09.001, this issue.
Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), 260–266.
Pal, R., Sinha, K., 1996. Analysis of crash rates at interstate work zones in Indiana. Transportation Research Record 1529, 19–29.
Rouphail, N., Yang, Z., Fazio, J., 1988. Comparative study of short and long term urban freeway work zones. Transportation Research Record 1163, 4–14.
Savolainen, P., Mannering, F., Lord, D., Quddus, M., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43 (5), 1666–1676.
Train, K., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press, Cambridge, UK.
Ulfarsson, G., Mannering, F., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis and Prevention 36 (2), 135–147.
Ullman, G., Scriba, T., 2004. Revisiting the influence of crash report forms on work zone crash data. Transportation Research Record 1897, 180–182.
Venugopal, S., Tarko, A., 2000. Safety models for rural freeway work zones. Transportation Research Record 1715, 1–9.
Wang, J., Hughes, W., Council, F., Paniati, J., 1996. Investigation of highway work zone crashes: what we know and what we don't know. Transportation Research Record 1529, 54–62.
Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis, second ed Chapman and Hall/CRC, Boca Raton, FL.
Wittwer, J., 2004. Graphing a Normal Distribution in Excel. [Online] Available at: Vertex42.com [Accessed April 4 2013].
Wu, Z., Sharma, A., Mannering, F., Wang, S., 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accident Analysis and Prevention 54, 90–98.