Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 3645 - 3657, 2005

ACCOUNTING FOR SERIAL CORRELATION IN COUNT MODELS OF TRAFFIC SAFETY

Sittipan SITTIKARIYA
PhD Candidate
Department of Civil and Environmental Engineering
Pennsylvania State University
104 Transportation Research Building,
University Park, PA 16802 USA
Fax: 001-814-863-7304
Email: sittipan@psu.edu

Venky N. SHANKAR, PhD, PE
Associate Professor
Department of Civil and Environmental Engineering
Pennsylvania State University
226C Sackett Building, University Park,
PA 16802 USA
Fax: 001-814-863-7304
Email: shankarv@engr.psu.edu

Ming-Bang SHYU
PhD Candidate
Department of Civil and Environmental Engineering
Pennsylvania State University
104 Transportation Research Building,
University Park, PA 16802 USA
Fax: 001-814-863-7304
Email: mbshyu@psu.edu

Songrit CHAYANAN, PhD
DKS Associates, Inc.
719 Second Avenue, Suite 1250
Seattle, WA 98104, USA
Email: songrit@u.washington.edu

Abstract: This research explores the impact of inflated variances on statistical significance and identification of variables correlated with median crossover accident frequencies. In roadway and accident data that is comprised of cross-sectional panels, serial-correlation effects across time arise due to shared unobserved effects through random error terms. We use an empirical technique to adjust for downward bias in standard errors in count models of traffic accidents, so that proper identification of variables is ensured. The identified model structures reflect the impact of inflated significance values, and provide specifications that include truly significant geometric, traffic and environmental effects contributing to median crossover accidents. It should be noted that the results from this study are localized and hence limited to inferences from the median crossover context in Washington State in the United States. Further study is required to ensure the transferability of these findings to other contexts.

Key Words: Median crossover accidents, Correlated data, Negative binomial, Zero-inflated count models, Negative multinomial,

1. INTRODUCTION

There are some common modeling problems in estimation of count models. Unobserved heterogeneity is a common occurrence in accident occurrence, leading to the well-known overdispersion problem (Shankar *et al.*, 1995). The negative binomial (NB) model is suitable for overdispersed accident frequencies, as show in several prior works (see for example Poch and Mannering, 1996; Milton and Mannering, 1996).

Another common issue that arises in accident data contexts is the issue of serial correlation. Multiple years of cross-sectional data on highway accident occurrences are often available from public domains, including time series information on traffic volumes, accident counts and roadway geometrics. In addition, weather information is also available from national databases (see for example Shankar *et al.*, 2004.) In the presence of serial correlation, the efficiency of parameter estimates comes into question. Similar to the classical linear regression model, one can expect parameter estimates to be inefficient in the presence of serial correlation in count models (Greene, 2000.) A method to adjust for repeated observation effects on parameter estimates is necessary to adjust for serial correlation effects. One such method relates to the use of the negative multinomial (NM) model (Ulfarsson and Shankar, 2003.) In that model, a joint likelihood based on repeated observations is constructed to modify the traditional negative binomial likelihood. As a result, parameter estimates reflect an adjustment in their standard errors, with much of the adjustment resulting from proximate years. The second type of adjustment for serial correlation involves the development of a random effects or fixed effects specification, where segment-specific correlation is accounted for. In the traffic accident context, the random effects approach has been shown to be reasonable (See for example, Shankar *et al.*, 1998).

The third common issue that arises in traffic accident contexts is one that pertains to partial observability and the resulting problem of excess zero counts. For example, especially in events of potentially high severity such median crossover accidents, the problem of excess zeros is a significant one. A highway segment with historically excess zeros may appear to be a safe location, but there is no guarantee that no median crossover accidents will occur in the future. As interactions between traffic volumes and geometrics, and geometrics and weather conditions exacerbate accident "proneness," one would expect that over time, with increase in traffic volumes and adverse weather conditions, non-zero count probabilities would be expected to rise. Limited history of observation may not adequately capture the "accident proneness" of the highway segment, as interactions increase over time.

It is the combination of the second and third common problems, namely, serial correlation and excess zeros, which we address in this paper. We choose the empirical context of median crossover accident settings on the highway network in Washington State in the USA. Overdispersion, the first problem discussed is also implicitly discussed in this study; although in this particular case, overdispersion arises primarily through the occurrence of excess zeros, and not necessarily through high-valued counts that are best modeled by a negative binomial distribution. The rest of this paper is organized as follows. Model preliminaries are presented first, where a brief discussion of the negative multinomial is presented, along with the preferred model of excess zeros, namely, the zero-inflated Poisson model for median crossover accidents. The empirical setting is then discussed, followed by a discussion of model results, along with adjustments of standard errors for the zero-inflated model, due to serial correlation. Finally, conclusions and recommendations are offered.

2. MODELING PRELIMINARIES

2.1 The Negative Multinomial Model for Serial Correlation in Accident Counts

The negative multinomial (NM) model is suggested as a proper estimator when serial correlations are present in the dataset (Ulfarsson and Shankar, 2003). The derivation of the

NM model begins with the Poisson probability density function, which is an expression for the probability of the frequency equaling a particular count (Guo, 1996):

$$\mathrm{P}(\mathrm{Y}_{\mathrm{it}} = \mathrm{y}_{\mathrm{it}}) = \frac{\mathrm{e}^{-\lambda_{\mathrm{it}}} \lambda_{\mathrm{it}}^{\mathrm{y}_{\mathrm{it}}}}{\mathrm{y}_{\mathrm{it}}!}; \qquad \mathrm{y}_{\mathrm{it}} = 0,1,2,..., \tag{1}$$

where Y_{it} is the observed frequency of median crossover accidents in section i at time t, and λ_{it} is the mean of the number of median crossover accidents. The Poisson model is estimated by defining:

$$\ln \lambda_{it} = x_{it}\beta, \tag{2}$$

where x_{it} is a vector of section- and time-specific explanatory variables; and β is a vector of coefficients to be estimated. To account for the section-specific variation, we proceed as is done for the NB model. A random error term is added to the expression for the mean,

$$\ln \lambda_{it} = x_{it}\beta + \varepsilon_i, \tag{3}$$

where ε_i is a section-specific (not observation-specific as in the NB model) random error term, $\exp(\varepsilon_i)$ is assumed to be independently and identically distributed gamma with mean 1 and variance $\alpha = 1/\theta$. The assumption of mean 1 does not cause loss of generality if (3) includes an intercept term.

The conditional joint density function of all individual counts for a particular section i, given that the individual counts are distributed by (1) and conditioned on ε_i, can now be written as:

$$\mathrm{P}(Y_{i1} = y_{i1},...,Y_{it_i} = y_{it_i} \mid \varepsilon_i) = \prod_{t'=1}^{t_i} \mathrm{P}(Y_{it'} = y_{it'} \mid \varepsilon_i), \tag{4}$$

where t_i denotes the number of time periods observed for section i. This assumes the accident counts in different sections are independent. This is not unreasonable because these sections are generally not next to each other and will therefore only share minimal unobserved effects. The unconditional joint density function for the negative multinomial distribution can now be derived by integrating (4) and by using the assumed distribution of $\exp(\varepsilon_i)$ to give:

$$\mathrm{P}(Y_{i1} = y_{i1},...,Y_{it_i} = y_{it_i}) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta) y_{i1}!...y_{it_i}!} \left(\frac{\theta}{\eta_i + \theta}\right)^{\theta} \left(\frac{\eta_{i1}}{\eta_i + \theta}\right)^{y_{i1}} ... \left(\frac{\eta_{it_i}}{\eta_i + \theta}\right)^{y_{it_i}}, \tag{5}$$

where Γ is the gamma function, $\eta_{it} = e^{x_{it} \cdot \beta}$, $\eta_i = \eta_{i1} + ... + \eta_{it_i}$, and $y_i = y_{i1} + ... + y_{it_i}$. Recall that the variance of $\exp(\varepsilon_i)$ is $\alpha = 1/\theta$. The degenerate case, when each section has only one observation, (i.e. there is no section-specific correlation) yields the negative binomial distribution. The expected value, the variance and covariance for the NM model are:

$$\mathrm{E}(Y_{it}) = \eta_{it}, \qquad \mathrm{Var}(Y_{it}) = \mathrm{E}(Y_{it})[1 + \alpha\, \mathrm{E}(Y_{it})], \qquad \mathrm{Cov}(Y_{it}, Y_{it'}) = \alpha \eta_{it} \eta_{it'}. \tag{6}$$

A likelihood function is written using (5) to give (7):

$$L(\beta,\theta)=\prod_{i=1}^{n}\frac{\Gamma(y_i+\theta)}{\Gamma(\theta)}\left(\frac{\theta}{\eta_i+\theta}\right)^{\theta}\left(\frac{\eta_{i1}}{\eta_i+\theta}\right)^{y_{i1}}\cdots\left(\frac{\eta_{it_i}}{\eta_i+\theta}\right)^{y_{it_i}}, \tag{7}$$

where n is the total number of sections. The estimable coefficients β and $\alpha = 1/\theta$ are estimated by maximizing the likelihood function (7), (Green, 2003).

In median crossover accident contexts, it is important to keep in mind that there are the possibilities of near-crossover cases in the roadway section. For example, as median width increases, the probability that a potential crossover will occur decreases, but may be enhanced with increasing volumes and interactions with adverse weather conditions over time. A near-crossover roadside encroachment usually takes the form of a rollover, fixed object or other similar types of roadside accidents that do not result in the vehicle crossing the median and entering opposing traffic. The segment hence may be characterized as operating in multiple states. For simplicity, we can consider two states, including a zero crossover state and a non-zero crossover state. Depending on how interactions vary over time, the segment may switch to a non-zero state. In such a case, a process that adequately captures the probability of the two states is necessary. The traditional count models, such as the Poisson or negative binomial, only address the non-zero state, thus ignoring the potential shift of excess zeros into non-zero counts. A plausible model presented below to address the excess zero problem is the zero-inflated Poisson or the zero-inflated negative binomial model.

2.2 Zero Inflated Poisson Model of Median Crossover Accidents

Let "Z" represents the zero-crash count state of the median crossover accident site, and "Y^*" denote the crash count state for that roadway section. Neither "Z" nor "Y^*" is observed, but only the observed median crossover count "Y" is, such that $Y=Z*Y^*$. Determining the latent components can then be viewed as a mixing distribution problem, with "Z" being modeled as a dichotomous outcome probability and "Y^*" modeled as a count probability.

In vehicular accident contexts, such distributions have been found to be appropriate (Shankar *et al.*, 1997). In particular these studies have highlighted the importance of roadway design deviations as a motivator for partial observability effects. The effect of such deviations has been found to, at the least, cause partial observability, and in certain design situations, overdispersion as well. In the median crossover accident context, design deviations are a significant issue. To highlight this issue, a brief discussion of median barrier warrants is necessary. Median barrier warrants in Washington State, and for that matter in the entire United States, are based on two simplifying factors, namely, average daily traffic (ADT) and median width. No account of interactions between traffic volumes and geometrics, or weather and geometrics is considered in the decision to install a median barrier. Hence, potential interactions arising due to the presence of deviations from preferred geometric design are ignored, leading to the potential for the excess zero median crossover problem. To formally address the excess zero problem, let Y_i be the annual number of accident counts reported for section i, and let p_i be the probability that section i will exist in the zero-count state over its lifetime. Thus $1-p_i$ is the probability that section i will operate in the non-zero accident state. It is to be noted that in the non-zero count state, the probability of zero accident counts still exists. For our immediate purposes, we assume that this count state follows a Poisson distribution. Given this,

$$Y_i = 0 \text{ with probability } p_i + (1 - p_i)e^{-\lambda_i} \quad , \tag{8}$$

and,

$$\mathrm{Y_i} = \mathrm{k} \text{ with probability } (1 - p_i)(\frac{e^{-\lambda i}\lambda_i^{k_i}}{k!}) , \tag{9}$$

In (8) and (9), the probability of being in the zero-accident state p_i is formulated as a logistic distribution such that $\log \frac{\mathrm{p_i}}{1-\mathrm{p_i}} = \mathbf{G_i}\,\boldsymbol{\gamma}$ and λ_i is defined by $\log(\lambda_i) = \mathbf{H_i}\,\boldsymbol{\beta}$, where $\mathbf{G}_i$ and $\mathbf{H}_i$ are covariate vectors, and γ and β are coefficient vectors. The covariates that affect the mean λ_i of the Poisson state may or may not be the same as the covariates that affect the zero-accident state probability (i.e., $\mathrm{p_i}$). Alternatively, vectors $\mathbf{G}_i$ and $\mathbf{H}_i$ may be related to each other by a single parameter τ. This is essentially a parametric constraint in the sense that the explanatory variables are forced to be the same in both the zero and count states. In such a case, a natural parameterization is $\log(\mathrm{p_i}/1-\mathrm{p_i}) = \tau \cdot \mathbf{H_i}\,\boldsymbol{\beta}$. It is useful in the accident context to begin with unconstrained parameter vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. This allows different effects to be correlated with the zero state and count state respectively. For example, in the median crossover accident context, it may be useful to consider the impact of median widths alone in the specification of the zero-state probability function, while including design, traffic and environmental interactions as well in the count state function.

Equations (8) and (9) combined provide the zero-inflated Poisson (ZIP) model. We refer to a model of unconstrained parametric vectors such as the one discussed above is referred to as the ZIP-Full model. The maximum likelihood estimation using the gradient/line search approach proposed by Green (1996) is performed to estimate the β and γ. A likelihood function is given by (10):

$$\mathrm{L}(\beta, \gamma,) = \prod_{i=1}^{n} \left[(1-\mathrm{p_i}) \frac{\mathrm{e}^{-\lambda_i}\lambda_i^{y_t}}{\mathrm{y_i}!} + \mathrm{Z_i p_i} \right], \tag{10}$$

where $\mathrm{Z_i} = 1$ when $\mathrm{Y_i} =$ is observed to be zero and 0 otherwise. An alternate method to estimate ZIP parameters is to employ the expectations maximization technique. Greene (1996) argues that the method shown in (10) is reliable.

2.3 Statistical Validation of the ZIP Median Crossover Model

In statistically validating the ZIP model, one has to distinguish the base count model (such as the Poisson model for median crossover counts) from the zero-inflated probability model (such as the ZIP). The reasoning behind this test is that the Poisson model does not adequately capture the entire "zero" density, and that the ZIP structure is more reliable. A statistical test for this has been proposed by Vuong in 1989. The Vuong test is a t-statistic-based test with reasonable power in count-data applications (Green, 1994). The Vuong statistic (V-statistic) is computed as:

$$V = \frac{\overline{m}\sqrt{N}}{S_m} , \tag{11}$$

where $\overline{m}$ is the mean with m = log [f_1(.) /f_2(.)], (with f_1(.) being the density function of the ZIP distribution and f_2(.) is the density function of the parent-Poisson distribution), and S_m and N are the standard deviation and sample size respectively. The advantage of using the Vuong test is that the entire distribution is used for comparison of the means, as opposed to just the excess zero mass. A value greater than 1.96 (the 95 percent confidence level for the t-test) for the V-statistic favors the ZIP-Full model while a value less than -1.96 favors the parent-Poisson (values in between 1.96 and -1.96 mean that the test is indecisive). The intuitive reasoning behind this test is that if the processes are statistically not different, the mean ratio of their densities should equal one. To carry out the test, both the parent and zero-inflated distributions need to be estimated and tested using a t-statistic. Studies (Greene 1994) have shown that a Vuong statistic has reasonable power and hence is quite reliable. Greene indicates that a significant Vuong statistic in favor the ZIP model will also favor the ZIP model over the negative binomial model, which would otherwise be the default model for overdispersion from excess zeros.

2.4 The Relevance of the Negative Multinomial to Median Crossover Accidents

It is important to keep in mind that the estimated parameters from the ZIP-Full model would be efficient if serial correlation did not exist. Such is not the case in median crossover contexts where multiple years of observations are available. In fact, as in the classical linear regression model, one would expect standard errors to be downward biased if serial correlation is not accounted for. To correct the variances in a dual-state distribution (such as the ZIP) with temporal correlation, the negative multinomial model offers an intuitive approach. We showed that the negative multinomial captures serial correlation across years without biasing the parameters of the negative binomial distribution. Empirically, this characteristic is of some use in our current ZIP-Full context. We can benchmark standard errors from a negative multinomial model against a naïve negative binomial model first, when repeated observations are the primary source of serial correlation. The point of importance to note here is that the negative multinomial model adjusts the standard errors upward to account for serial correlation. We call this upward adjustment the "loading factor." In other words, the standard errors of the negative binomial model are "factored up" by "loading factors" derived from the negative multinomial model. As an example, consider the vectors of parameters estimated as "$\boldsymbol{\beta}$" by a negative binomial model. We assume here that $\boldsymbol{\beta}$ has dimensionality k. Consider the negative multinomial model vector of parameters $\tilde{\boldsymbol{\beta}}$ also of dimensionality k, with exactly the same set of regressors as those used for $\boldsymbol{\beta}$. Then, the loading factor for $\boldsymbol{\beta}$ is a vector μ of with dimensionality k, where $\mu_k = \text{s.e.}(\tilde{\beta}_k)/\text{s.e.}(\beta_k)$. The "loading factor" for any given parameter β_k varies depending on the effect of serial correlation in the variance-covariance matrix. The loading factors are then used to adjust upward the standard errors obtained for a naïve ZIP-Full model. The adjusted standard errors are then used to determine parameter significance under serial correlation.

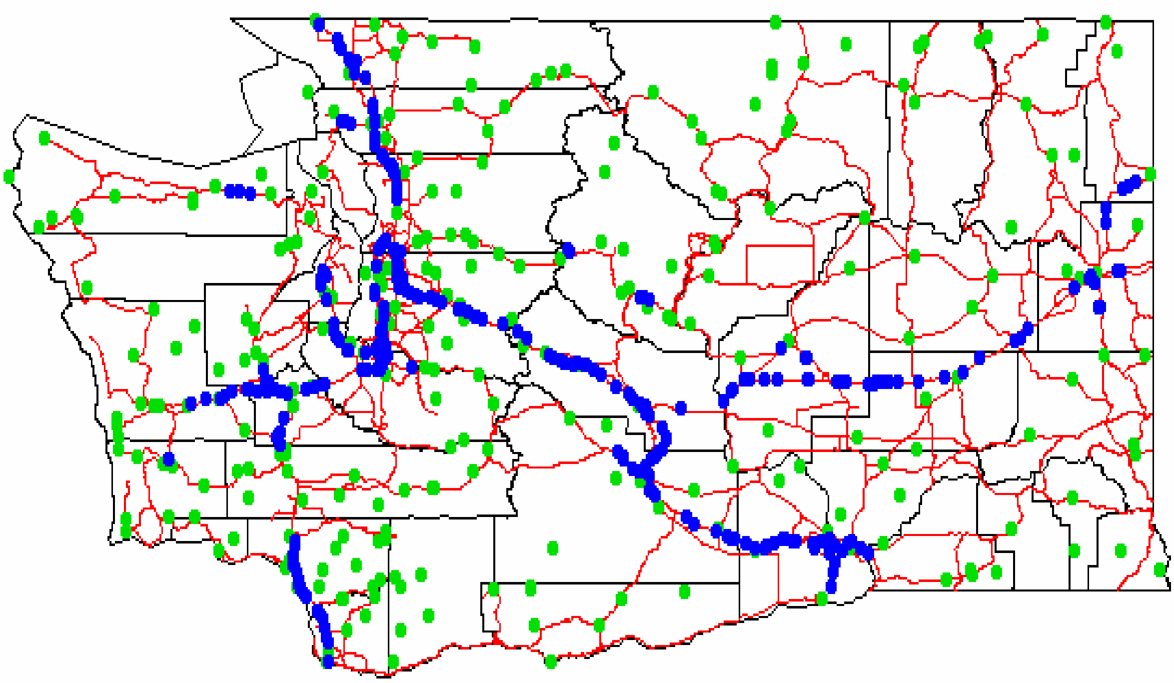
To illustrate the development of "loading factors" for standard errors empirically, we present findings from a negative binomial/negative multinomial analysis of median crossover accidents in Washington State. Prior to presenting these findings, we discuss the empirical setting used for the development of the negative binomial and negative multinomial models.

3. EMPIRICAL SETTING OF MEDIAN CROSSOVER ACCIDENTS IN WASHINGTON STATE

In this discussion, information regarding the data, framework, and criteria for model analysis and evaluations is presented. The Washington State highway system contains over 7,000 centerline miles of state highways. The annual records in the database developed by the Washington State Department of Transportation (WSDOT) contain almost every roadway and roadside incident that occurs during the year. The accident records from this database are the primary source of data for the current median crossover accidents study.

The panel data in this research consists of five years (1990–1994 inclusive) of annual accident counts for 275 roadway sections in Washington State. The panel is balanced, with all sections having a full five-year history. This panel represents all sections (longer than 2,624 ft) without median barriers on divided state highways. The reasons why only sections longer than 2,624 ft are selected are that about 95% of shorter sections on divided highways have barriers, and that the shorter sections are more affected by access controls and intersections (Ulfarsson and Shankar, 2003). In addition to median crossover accident information, other components of data extracted from the database included roadway geometrics, median characteristics and traffic volumes. The database did not provide any weather information required for the study.

The GIS program ArcView 3.2 was used to match the roadway sections to their weather attributes stored in the historical weather database provided by the Western Regional Climate Center. The mapping criteria involved linking the non-median barrier roadway sections to the nearest corresponding weather stations. Each weather station provided climate data including daily, monthly and annual measurements of temperature, precipitation and snowfall including snow depth, with the records dating back to 1948. The selected roadway sections and all weather stations in Washington State are illustrated in figure 1. Table 1 provides descriptive statistics of key variables in the median crossover accident dataset.



*Blue (darker color) = The Selected Roadway Sections, **Green (lighter color) = The Weather Stations

Figure 1. Selected Roadway Sections and Weather Stations in Washington State

Table 1. Descriptive Statistics of Key Median Crossover Accident Related Variables

Variable	Mean	Std. Dev.	Min	Max
Number of crossover accidents in section	0.24	0.65	0	7
AADT	37,355	36,975	3,350	172,560
AADT per lane	7,445	5,830	835	28,690
Single truck percentage	4.20	1.22	1.90	10
Double truck percentage	7.76	4.62	0.55	17.80
Truck-train percentage	2.21	1.60	0	7
Percentage of AADT in the peak hour	10.72	7.314	0	19.40
Speed limit	60	5.5	35	65
Maximum median shoulder width in feet	5.31	2.49	0	18
Minimum median shoulder width in feet	4.48	1.6833	0	10
Percent medians narrower than 40 feet	32.36	46.80		
Percent medians between 40 feet and 50 feet in width	11.64	32.08		
Percent medians between 50 feet and 60 feet in width	5.81	23.42		
Percent medians wider than 60 feet in width	50.18	50.02		
Percent medians that are paved	4.36	20.44		
Length of the roadway section in miles	2.43	2.69	0.50	19.30
The number of interchanges in section	0.85	0.84	0	4
The number of horizontal curves in section	2.75	2.86	0	29
The number of horizontal curves per mile	1.44	0.96	0	5
Maximum horizontal central angle in degrees	30.29	23.88	0	111.49
Minimum radius of horizontal curve in feet	4267.24	4875.08	0	38,400
Average annual snow depth in inches	19.44	45.66	0	652*
Average annual precipitation in inches	29.86	21.98	4.53	131.74
Number of grade changes	3.865	4.089	0	28
Average roadway width	57.42	15.47	24	121

* Weather station data for mountainous section

3.1. Development of Loading Factors for Standard Error Adjustment

Table 2 presents a comparative analysis of standard errors for significant parameters in the negative binomial and negative multinomial models of median crossover accidents. It is to be noted here that only significant effects appearing in the negative multinomial model are used to develop the loading factors. It may be that variables significant in the negative binomial may not be significant in the negative multinomial due to inflation in the standard error. As is noted in table 2, the loading factor varies by parameter, from a minimum adjustment of 1.1889 for the precipitation-horizontal curves interaction variable to 2.3759 for the length variable for medians narrower than or equal to 40 feet. The standard error adjustment for the "alpha" parameter (overdispersion effect) is ignored in our discussion, since the alpha parameter does not appear in the ZIP model of median crossover accidents. The ZIP model is discussed next, considering the loading factors developed in table 2.

Table 2. Load Factors Developed from NB and NM Models for Adjusting Standard Errors in Zero-Inflated Poisson Model of Median Crossover Accidents

Variable	**Negative Binomial Model**		**Negative Multinomial Model**		**Load Factor***
	Estimated Coefficient	**Standard. Error**	**Estimated Coefficient**	**Standard Error**	
Constant	-1.8990	0.1326	-1.9990	0.1987	1.4980
Per-lane AADT indicator (1 if per-lane AADT <= 5000 vehicles, 0 otherwise)	-0.9455	0.1725	-0.9476	0.2110	1.2233
Length of the roadway section on median widths less than or equal to 40 feet	0.3177	0.0433	0.3560	0.1029	2.3759
Length of the roadway section on median widths greater than or equal to 41 feet and less than or equal to 60 feet	0.4324	0.0619	0.4448	0.0951	1.5357
Length of the roadway section on median widths greater than 60 feet	0.1122	0.0259	0.1198	0.0389	1.5010
The number of interchanges in the section	0.2398	0.0815	0.2692	0.1185	1.4538
Interaction 1 between average monthly precipitation indicator and the number of horizontal curves per mile indicator (1 if average monthly precipitation <= 1.5 inches and the number of horizontal curves per mile <= 0.5, 0 otherwise)	-1.0116	0.34954	-1.0294	0.4788	1.3698
Interaction 2 between average monthly precipitation indicator and the number of horizontal curves per mile indicator (1 if average monthly precipitation > 4.0 inches and the number of horizontal curves per mile > 0.5, 0 otherwise)	-0.5329	0.2166	-0.4624	0.2575	1.1889
α	0.8295	0.2037	-	-	N.A.
θ^{**}	-	-	2.3318	3.1036	N.A.
Restricted log-likelihood (All parameters = 0, α = 0)	-1,462.3480		-1,462.3480		
Log-likelihood at convergence	-745.6914		-723.1721		
Adjusted ρ^2	0.4839		0.4993		
Number of observations	1,375***				

* The load factor was the proportion of the S.E. of NM to the S.E. of NB

** $\theta = 1/\alpha$

*** 5 years of data for 275 separate non-median barrier sections

4. AN EMPIRICALLY ADJUSTED SERIAL CORRELATION ZERO INFLATED POISSON MODEL OF MEDIAN CROSSOVER ACCIDENTS

A ZIP-Full model of median crossover accidents was estimated and loading factors from table 2 were used to adjust for serial correlation in the five-year median crossover dataset. Median crossover accident counts were the dependent variable in this estimation. The ZIP-Full model was estimated and validated by the Vuong statistic. The result revealed a Vuong statistic of 3.1618 suggesting that the ZIP-Full as the favorable model compared to the basic Poisson model.

In estimating the significance of key variables in the ZIP-Full model, standard errors estimated using the likelihood function shown in equation 10 were adjusted through multiplication with the "loading factors" developed in table 2. By so doing, we empirically account for serial correlation present in the dataset. Table 3 shows the empirically adjusted serial correlation ZIP-Full model of median crossover accidents.

The results show that ***before accounting for serial correlation***, all the variables included in the ZIP-Full model were highly significant and thus have a high level of confidence (exceeding 95%). However, ***after accounting for serial correlation***, the level of significance for all variables decreased. The length variable for medians wider than 60 feet had the lowest t-statistic of 1.3154.

It is also noted that the parameters in non-zero accident state always negatively correlate with the same parameters in zero accident state (e.g. opposite signs). For example, the length variable for medians narrower than 40 feet positively correlates with median crossover accidents, while it negatively correlates with the probability of a section being in a zero count state in its lifetime.

In the non-zero count state, factors positively correlating with median crossover accidents included the interaction variables between length and three categories of median widths,, as well as the number of interchanges in the section. The three categories of the median width interacted with the length of the section were 1) less than or equal to 40-foot median width, 2) between 40 to 60-foot median width, and 3) greater than 60-foot median width. Different ranges of the median width were experimented but the result showed that these three categories, when interrelated with the section length, had the greatest impact on the crossover accident likelihood. The magnitudes of the coefficients of the length variables suggest that all for a given length, the likelihood of median crossover accidents increases the greatest on sections with median widths between 40 and 60 feet wide in comparison to the other width categories. On the other hand, sections with median width wider than 60 feet have the least contribution on the likelihood of median crossover frequency. In fact, in the presence of serial correlation, their impact is statistically marginal.

Three factors negatively correlate with median crossover accident frequencies. These included the indicator variable if average annual daily traffic was less than 5,000, the interaction variable between the number of horizontal curves less than or equal to 0.5 per mile in the section and the average monthly precipitation being less than or equal 1.5 inches, and the interaction variable between the number of horizontal curves greater than 0.5 per mile in the section and the average monthly precipitation being greater than 4 inches.

Table 3. Zero-Inflated Poisson Model of Median Crossover Accidents with Serial Correlation Adjusted Standard Errors

Variables	Zero-Inflated Poisson Model			Load Factor	Adjusted Standard Error	Adjusted T-Statistic
	Estimated Coefficient	Standard. Error	T-statistic			
Non-zero accident Poisson probability state						
Constant	-1.0012	0.1583	-6.3250	1.4980	0.2371	-4.2224
Per-lane AADT indicator (1 if per-lane AADT <= 5000 vehicles, 0 otherwise)	-1.0077	0.1754	-5.7440	1.2232	0.2146	-4.6960
Length of section where medians are less than 40 feet wide	0.1499	0.0263	5.7020	2.3759	0.0625	2.4000
Length of section where medians are between 40 feet and 60 feet wide	0.3954	0.0604	6.5500	1.5357	0.0927	4.2652
Length of section where medians are wider than 60 feet	0.0495	0.0251	1.9740	1.5010	0.0377	1.3154
Number of interchanges in section	0.1834	0.0711	2.5800	1.4538	0.1034	1.7747
Interaction 1 between average monthly precipitation indicator and the number of horizontal curves per mile indicator (1 if average monthly precipitation <= 1.5 inches and the number of horizontal curves per mile <= 0.5, 0 otherwise)	-0.8648	0.3235	-2.6740	1.3698	0.4431	-1.9517
Interaction 2 between average monthly precipitation indicator and the number of horizontal curves per mile indicator (1 if average monthly precipitation > 4.0 inches and the number of horizontal curves per mile > 0.5, 0 otherwise)	-0.4366	0.2185	-1.9990	1.1889	0.2598	-1.6810
Zero accident probability state as logistic function						
Constant	1.1049	0.2950	3.7460	1.4980	0.4419	2.5004
Length of section where medians are less than 40 feet wide	-1.4137	0.4414	-3.2020	2.3759	1.0488	-1.3479
Length of section where medians are between 40 feet and 60 feet wide	-0.3691	0.1663	-2.2200	1.5357	0.2554	-1.4455
Length of section where medians are wider than 60 feet	-0.5098	0.1692	-3.0130	1.5010	0.2539	-2.0076
Restricted log-likelihood (constant only)	-889.7221*					
Log-likelihood at convergence	-718.6493					
Vuong statistic	3.1618					

* This restricted log-likelihood was obtained from the restricted log-likelihood computed by the Poisson model.

The weather effect appearing in the form of the interaction variables played a significant role in median crossover likelihood. In a section where the average number of horizontal curves was less than or equal to 0.5 per mile, the median crossover counts were expected to decrease if the average monthly precipitation was less than or equal to 1.5 inches. Crossover accident frequency also decreases when average monthly precipitation exceeds 4 inches on sections with greater than 0.5 horizontal curves per mile. Both effects point to the range of interactions between precipitation and horizontal curves on median crossover frequencies.

As a final thought, it should be noted that the impact of median widths on the probability of a median crossover accident occurring in a section's lifetime (zero versus non-zero state) follows a trend that is different from their impacts on positive crossover frequencies (the Poisson state.) In the zero state, for a given length, median widths in excess of 60 feet have the highest odds ratio of being in a non-zero accident state, while the count state effects suggests they will have the least number of positive crossover counts.

5. CONCLUSIONS AND RECOMMENDATIONS

This paper presents an empirical technique to adjust for serial correlation effects on standard errors in median crossover accident models. Similar to the classical linear regression model, standard errors in count models are downward biased as observed in empirical analyses on median crossover accidents in this study. To develop a technique for adjusting standard errors, we used the negative multinomial specification to account for correlation across time. Using the ratio of standard errors from the NM model to the naive NB model, we developed loading factors which represented the level of inflation in standard errors required to account for serial correlation. We then developed a ZIP-Full model of median crossover accident counts to account for the presence of excess zeros in the dataset. This excess zero is unique to median crossover accidents, especially due to the fact that only five years of observation are available. Using the loading factors developed on the basis of the NM model, we adjusted the standard errors of the ZIP-Full model to identify key factors affecting median crossover accident frequencies. It is our belief that the empirically standard errors in the ZIP-Full model are closer to efficient estimates that would be theoretically derived. An obvious recommendation from this study is to compare theoretically derived standard errors of a ZIP-Full model with empirical findings from this study. Such a study would shed significant light on the impact of serial correlation in a variety of count contexts.

To summarize our empirical findings, we conclude that the ZIP-Full model is the most appropriate model among zero-altered probability processes for predicting the number of median crossover accidents. The main effects such as average daily traffic, and the number of interchanges were investigated to have statistically significant impact on the frequency of median crossover accidents. The interaction variables of length of the roadway section and the median width as well as the interaction between monthly precipitation and the number of horizontal curves per mile are also presented as important contributors to median crossover likelihoods.

REFERENCES

Chayanan, S., Shankar, V.N., Sittikariya, S., Ulfarsson, G.F., Shyu, M.B., and Juvva, N.K. (2004) Median crossover accident analyses and the effectiveness of median barriers. Final

Research Report, WA-RD 580.1, Washington State Department of Transportation, Washington.

ESRI (2002) **ArcView GIS Version 3.2**, Environmental Systems Research Institute, Inc., California.
Greene, W.H. (1994) Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working papers EC-94-10, Stern School of Business, New York University, New York.

Greene, W.H. (2004) **LIMDEP, Version 8.0, User's Manual.** Econometric Software Inc., New York.

Greene, W.H. (2003) **Econometric Analysis**. Fifth Edition, Prentice Hall Inc., New Jersey.

Guo, G. (1996) Negative multinomial regression models for clustered event counts, **Sociological Methodology, Vol. 26**, 113–132.

Milton, J.C. and Mannering, F.L., (1996) The relationship between highway geometrics, traffic related elements and motor vehicle accident. Final research report, WA-RD 403.1, Washington State Department of Transportation, Washington.

Poch, M. and Mannering, F.L. (1996) Negative binomial analysis of intersection-accident frequencies, **Journal of Transportation Engineering, Vol. 122, No. 3**, 105-113

Shankar, V.N., Albin, R.B., Milton, J.C. and Mannering, F.L. (1998) Evaluating median cross-over likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model, **Transportation Research Record, Vol. 1635**, 44–48.

Shankar, V.N., Chayanan, S., Sittikariya, S., Shyu, M.B., Juvva, N.K. and Milton J.C. (2004) The marginal impacts of design, traffic, weather, and related interactions on roadside crashes, **Transportation Research Record**, **Vol. 1897**, 156-163.

Shankar, V.N., Mannering, F.L. and Barfield, W. (1995) Effect of roadway geometrics and environmental conditions on rural accident frequencies, **Accident Analysis and Prevention, Vol. 27, No.3**, 371-389.

Shankar, V.N., Milton, J.C. and Mannering, F.L. (1997) Modeling statewide accident frequencies as zero-altered probability processes: an empirical inquiry, **Accident Analysis and Prevention, Vol. 29, No.6**, 829–837.

Ulfarsson, G. F. and Shankar V.N. (2003) An accident count model based on multi-year cross-sectional roadway data with serial correlation, **Transportation Research Record, Vol. 1840**, 193-197.

Voung, Q. (1989) Likelihood ratio tests for model selection and non-nested hypotheses, **Econometrica, Vol. 57, No. 2**, 307-333.

Western Regional Climate Center (2002) Historical Climate Information, Desert Research Institute, Retrieved from http://www.wrcc.dri.edu/summary/climsmwa.html.