Effects of the Varying Dispersion Parameter of Poisson-gamma models on the estimation of Confidence Intervals of Crash Prediction models

By

Srinivas Reddy Geedipally
Research Assistant
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 862-3446
Fax. (979) 845-6481
Email: srinivas8@tamu.edu

Dominique Lord1
Assistant Professor
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 458-3949
Fax. (979) 845-6481
Email: d-lord@tamu.edu

November 14th, 2007

4,496 words + 7 figures + 4 tables = 7,246 words

1 Corresponding Author
ABSTRACT

The most common probabilistic structure of the models used by transportation safety analysts for modeling motor vehicle crashes are the traditional Poisson and Poisson-gamma (or Negative Binomial) distributions. Since crash data have been shown to exhibit over-dispersion, Poisson-gamma models are usually preferred over Poisson regression models. Up until recently, the dispersion parameter of Poisson-gamma models has been assumed to be fixed, but recent research in highway safety has shown that the parameter can potentially be dependent upon the covariates, especially for flow-only models. Given the fact that the dispersion parameter is a key variable for computing confidence intervals, there is a reason to believe that a varying dispersion parameter could affect the computation of confidence intervals. The primary objective of this paper is to evaluate whether the varying dispersion parameter affects the computation of the confidence intervals for the gamma mean \((m)\) and predicted response \((y)\) on sites that have not been used for estimating the predictive model. To accomplish the objective of the study, predictive models with fixed and varying dispersion parameters were estimated using data collected in California at 537 3-legged rural unsignalized intersections. The study shows that models developed using a varying dispersion parameter usually produce smaller confidence intervals, hence more precise estimates, than models with a fixed dispersion parameter both for the gamma mean and the predicted response. Therefore, it is recommended to develop models with a varying dispersion whenever possible, especially if they are used for screening purposes.
INTRODUCTION

Statistical models have been a very popular method for estimating the safety performance of various transportation elements (1-8). The most common probabilistic structure of the models used by transportation safety analysts for modeling motor vehicle crashes are the traditional Poisson and Poisson-gamma (or Negative Binomial) distributions. Since crash data have been shown to exhibit over-dispersion (see 9), meaning that the variance is greater than the mean, Poisson-gamma models are usually preferred over Poisson regression models. The Poisson-gamma distribution offers a simple way to accommodate the over-dispersion, especially since the final equation has a closed form and the mathematics to manipulate the relationship between the mean and the variance structures is relatively simple (10).

Up until a few years ago, the dispersion parameter of Poisson-gamma models has been assumed to be fixed, but recent research in highway safety has shown that the parameter can potentially be dependent upon the covariates (5, 11-16). This means that each observation has a unique dispersion parameter that varies as function of the covariates, such as traffic flow, segment length or lane width among others. In previous studies, the varying dispersion parameter has been shown to influence empirical Bayes (EB) estimates for disaggregated analyses (5, 14, 16); the dispersion parameter plays an important role in the weight factors assigned to the predicted and observed values of this estimate. Others have reported that the Poisson-gamma models with varying dispersion parameters provide better statistical fit (11, 15, 16). Given the fact that the dispersion parameter is a key variable for computing confidence intervals (17, 18), there is also a reason to believe that a varying dispersion parameter could affect the computation of confidence intervals compared to the fixed dispersion parameter.

The primary objective of this paper is to evaluate whether the varying dispersion parameter affects the computation of the confidence intervals for the gamma mean ($\mu$) and predicted response ($y$) on sites that have not been used for estimating the predictive model. Confidence intervals can be used to examine alternatives where the safety performance is used a screening criterion for selecting highway design alternatives and for identifying hazardous sites. Thus, selecting the appropriate modeling structure of Poisson-gamma models is critical to minimize errors during the selection process. To accomplish the objective of the study, predictive models with fixed and varying dispersion parameters were estimated using data collected in California at 537 3-legged rural unsignalized intersections. Three functional forms describing the variance function were tested (i.e., relationship between the dispersion parameter and the covariates of the model). The study will show that models developed using a varying dispersion parameter greatly influence the confidence intervals of the gamma mean and predicted response. More specifically, models with a varying dispersion parameter usually produce smaller confidence intervals, hence more precise estimates, than with models with a fixed dispersion parameter with an increase the traffic flows for the gamma mean, no matter which functional form is used.
BACKGROUND

Poisson-gamma models in highway safety applications have been shown to have the following probabilistic structure: the number of crashes at the \( i \)-th entity (highway segment, intersections, etc.) and \( t \)-th time period, \( Y_{i,t} \), when conditional on its mean \( \mu_{i,t} \), is assumed to be Poisson distributed and independent over all entities and time periods as:

\[
Y_{i,t} \mid \mu_{i,t} \sim Po(\mu_{i,t}) \quad i = 1, 2, \ldots, I \text{ and } t = 1, 2, \ldots, T
\]  

(1)

The mean of the Poisson is structured as:

\[
\mu_{i,t} = f(X; \beta) \exp(e_{i,t})
\]

(2)

where,

- \( f(\cdot) \) is a function of the covariates (\( X \));
- \( \beta \) is a vector of unknown coefficients; and,
- \( e_{i,t} \) is a model error independent of all the covariates.

It is usually assumed that \( \exp(e_{i,t}) \) is independent and gamma distributed with a mean equal to 1 and a variance \( 1 / \phi \) for all \( i \) and \( t \) (with \( \phi > 0 \)). With this characteristic, it can be shown that \( Y_{i,t} \), conditional on \( f(\cdot) \) and \( \phi \), is distributed as a Poisson-gamma random variable with a mean \( f(\cdot) \) and a variance \( f(\cdot)(1 + f(\cdot) / \phi) \) respectively. (Note: other variance functions exist for Poisson-gamma models, but they are not covered here since they are seldom used in highway safety studies. The reader is referred to Cameron and Trevidi (19) and Maher and Summersgill (20) for a description of alternative variance functions.) The probability density function (PDF) of the Poisson-gamma structure described above is given by the following equation:

\[
f\left(y_{i,t}; \phi, \mu_{i,t}\right) = \left(\frac{y_{i,t} + \phi - 1}{\phi - 1}\right)\left(\frac{\phi}{\mu_{i,t} + \phi}\right)^{\phi} \left(\frac{\mu_{i,t}}{\mu_{i,t} + \phi}\right)^{y_{i,t}}
\]

(3)

Where,

- \( y_{i,t} \) = response variable for observation \( i \) and time period \( t \);
- \( \mu_{i,t} \) = mean response for observation \( i \) and time period \( t \); and,
- \( \phi \) = inverse dispersion parameter of the Poisson-gamma distribution.

The term \( \phi \) is usually defined as the "inverse dispersion parameter" of the Poisson-gamma distribution. (Note: in the statistical and econometric literature, \( \alpha = 1 / \phi \) is usually defined as the dispersion parameter; in some published documents, the variable \( \alpha \) has also been defined as the “over-dispersion parameter.”). This term has traditionally been assumed to be fixed and a unique value applied to the entire dataset in the study. As described above, the inverse dispersion parameter plays an important in safety analyses,
including the computation of the weight factor for the EB method (10, 16) and the estimation of confidence intervals around the gamma mean and the predicted values of models applied to a different dataset than the ones employed in the estimation process (17, 18).

Hauer (11) first raised this issue and reported that the dispersion parameter of Poisson-gamma models should be dependent upon the length of a highway segment. On the other hand, Heydecker and Wu (12) attempted to estimate varying dispersion parameters as a function of sites’ covariates, such as AADT, vertical and horizontal curvatures among others. They asserted that the Poisson-gamma model with a varying dispersion parameter can better represent the nature of the crash dataset than the traditional Poisson-gamma model with a fixed dispersion parameter. The approach proposed by Heydecker and Wu (12) was also used by Lord et al. (13) for modeling the safety performance of freeways as a function of traffic flow characteristics. Miaou and Lord (5) have also noted that the dispersion parameter can be dependent upon the entering flows of crash-flow predictive models, suggesting that the variance function has an unobserved structure. Mitra and Washington (21) confirmed the results of Miaou and Lord (5) and reported that the variance became less structured when the mean function is better specified.

More recently, Miranda-Moreno et al. (14) reported that Poisson-gamma models with a varying dispersion parameter performed better than traditional models for identifying hazardous sites. El-Basyouny and Sayed (15), on the other hand, indicated that this type of model offered better a statistical fit, but did not improve the hazardous site identification process. Lord and Park (16) noted that Poisson-gamma models with a varying dispersion parameter could influence the EB estimates, and the magnitude is dependent upon the functional form. They also reported that the ranking process for identifying hazardous sites can be dependent upon the functional form used for estimating the models.

Poisson-gamma models with a varying dispersion parameter use the same PDF shown in equation (3) and estimate the same number of crashes for each observation, similar to those produced from the traditional Poisson-gamma model. However, instead of estimating a fixed dispersion parameter, these models use a varying dispersion parameter that can be estimated using the following expression (12, 22, 23):

$$\alpha_i = \exp(Z_i \times \delta_i)$$

where,

- $\alpha_i$ = dispersion parameter for the site $i$
- $Z_i$ = a vector of secondary covariates (are not necessarily the same as the covariates used in estimating the mean function $\hat{\mu}_i$),
- $\delta_i$ = regression coefficients corresponding to covariates $Z_i$. 

$$\delta_i = \exp(Z_i \times \delta_i)$$

Hauer (11) first raised this issue and reported that the dispersion parameter of Poisson-gamma models should be dependent upon the length of a highway segment. On the other hand, Heydecker and Wu (12) attempted to estimate varying dispersion parameters as a function of sites’ covariates, such as AADT, vertical and horizontal curvatures among others. They asserted that the Poisson-gamma model with a varying dispersion parameter can better represent the nature of the crash dataset than the traditional Poisson-gamma model with a fixed dispersion parameter. The approach proposed by Heydecker and Wu (12) was also used by Lord et al. (13) for modeling the safety performance of freeways as a function of traffic flow characteristics. Miaou and Lord (5) have also noted that the dispersion parameter can be dependent upon the entering flows of crash-flow predictive models, suggesting that the variance function has an unobserved structure. Mitra and Washington (21) confirmed the results of Miaou and Lord (5) and reported that the variance became less structured when the mean function is better specified.

More recently, Miranda-Moreno et al. (14) reported that Poisson-gamma models with a varying dispersion parameter performed better than traditional models for identifying hazardous sites. El-Basyouny and Sayed (15), on the other hand, indicated that this type of model offered better a statistical fit, but did not improve the hazardous site identification process. Lord and Park (16) noted that Poisson-gamma models with a varying dispersion parameter could influence the EB estimates, and the magnitude is dependent upon the functional form. They also reported that the ranking process for identifying hazardous sites can be dependent upon the functional form used for estimating the models.

Poisson-gamma models with a varying dispersion parameter use the same PDF shown in equation (3) and estimate the same number of crashes for each observation, similar to those produced from the traditional Poisson-gamma model. However, instead of estimating a fixed dispersion parameter, these models use a varying dispersion parameter that can be estimated using the following expression (12, 22, 23):

$$\alpha_i = \exp(Z_i \times \delta_i)$$

where,

- $\alpha_i$ = dispersion parameter for the site $i$
- $Z_i$ = a vector of secondary covariates (are not necessarily the same as the covariates used in estimating the mean function $\hat{\mu}_i$),
- $\delta_i$ = regression coefficients corresponding to covariates $Z_i$. 

$$\delta_i = \exp(Z_i \times \delta_i)$$
With equation (4), the model can be used for estimating a different dispersion parameter according to the sites’ attributes (i.e., covariates). If there are no significant secondary covariates for explaining the systematic dispersion structure, the dispersion parameters will only contain a fixed value (i.e., constant term), resulting in a traditional Poisson-gamma regression model.

Confidence intervals are important to assess the uncertainty associated point estimates of predictive models. Wood (17) proposed a method for estimating the confidence intervals for the mean response ($\mu$), for the gamma mean ($m$), and the predicted response ($y$) at a new site having similar characteristics as the sites used in the original dataset from which the model was developed. For Poisson-gamma models, the 95% confidence intervals for $m$ and $y$ are described in Table 1. In this table, $\eta$ is the logarithm of the estimated mean response $\hat{\mu}$, while $\phi$ is the inverse dispersion parameter of the Poisson-gamma model, as described above. In this paper, the confidence intervals were estimated for the fixed dispersion parameter as well as for the varying dispersion parameter. The method proposed by Wood (17) can be used for a varying dispersion parameter. This has been confirmed by the author via personal communication.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\left[ \frac{\hat{\mu}}{e^{1.96 \sqrt{\text{Var}(\hat{\eta})}}}, \hat{\mu} e^{1.96 \sqrt{\text{Var}(\hat{\eta})}} \right]$</td>
</tr>
<tr>
<td>$m$</td>
<td>$\left[ \max \left{ 0, \hat{\mu} - 1.96 \sqrt{\frac{\hat{\mu}^2 \text{Var}(\hat{\eta}) + \hat{\mu} \text{Var}(\hat{\eta}) + \hat{\mu}^2}{\phi}} \right}, \hat{\mu} + 1.96 \sqrt{\frac{\hat{\mu}^2 \text{Var}(\hat{\eta}) + \hat{\mu} \text{Var}(\hat{\eta}) + \hat{\mu}^2}{\phi}} \right]$</td>
</tr>
<tr>
<td>$y$</td>
<td>$\left[ 0, \left[ \hat{\mu} + 1.96 \sqrt{\frac{\hat{\mu}^2 \text{Var}(\hat{\eta}) + \hat{\mu} \text{Var}(\hat{\eta}) + \hat{\mu}^2}{\phi}} \right] \right]$</td>
</tr>
</tbody>
</table>

Note:

$\text{Var}(\hat{\eta}) = x_0 (XWX')^{-1} x'_0$

$\lfloor x \rfloor$ denotes the largest integer less or equal than $x$

For computing any other % CI for $y$, $\sqrt{19}$ can be substituted with $\sqrt{\frac{1.0 - \alpha}{\alpha}}$, where $\alpha = 1 - \text{percentile in } \%$; For instance, for estimating the upper boundary 90% CI, the upper value becomes $\sqrt{\frac{1.0 - 0.1}{0.1}} = \sqrt{9} = 3$. 

Table 1: 95% Confidence Intervals for $\mu, m, y$. (17)
METHODOLOGY

In this analysis, the statistical models were estimated using the following functional form

\[ \mu_i = \beta_0 F_{1i}^{\beta_1} F_{2i}^{\beta_2} \]

for the mean function, where \( \mu \) corresponds to the mean of observation \( i \); \( F_{1i} \) and \( F_{2i} \) are the entering flows from the major and minor approaches, respectively; and, \( \beta s \) are the regression coefficients associated with the covariates of the model.

Although traffic-flow only models could suffer from omitted variables bias, they are still the most popular type of models developed and used by transportation safety analysts \((10, 24)\). They are often preferred over models that include several covariates because they can be easily re-calibrated when they are developed in one jurisdiction and applied to another \((25, 26)\). In fact, this type of model will be the kind of model used for estimating the safety performance of rural and urban highways as well as for intersections in the forthcoming Highway Safety Manual (HSM) \((27)\).

The functional form described in the previous paragraph is not the most adequate for describing the relationship between crashes and exposure (the mean function). As reported by Miaou and Lord \((5)\), this form does not appropriately fit the data near the boundary conditions. Nonetheless, this form is still relevant for this study, as it is considered an established functional form in the highway safety literature. The most adequate functional form proposed by Miaou and Lord \((5)\), a model with two distinct mean functions, cannot be estimated via a generalized linear modeling (GLM) framework, as it was done in this study.

The modeling procedure was accomplished using the following 4-stage process:

1. In the initial step, the model was estimated using a fixed dispersion parameter in SAS \((28)\). The negative binomial function in SAS was used in this step. It should be noted that the number of years for each site was used as an offset.

2. The confidence interval for gamma mean \( (m) \) and the prediction interval \( (y) \) for a new observation with the same traffic flow characteristics was then calculated using the values in Table 1 \((17)\). In the equation, \( \eta \) is the logarithm of the estimated mean \( (\mu) \) while \( \phi \) is the inverse dispersion parameter estimated during the fitting process. \( Var(\eta) \) is calculated using \( X'I^{-1}X \) where \( I^{-1} \) represents the variance-covariance matrix and \( X \) is the matrix containing the observed values. The variance-covariance matrix was provided by SAS \((28)\).

3. In the third step, two models were estimated using the dispersion parameter that varies from site to site according to the entering flows. As documented in Miaou and Lord \((5)\) and confirmed by Mitra and Washington \((21)\), flow-only models will have a structured variance. Thus, structure of the variance can be captured using the same covariates as the ones used for mean function. Although the mean function remained the same, different functional forms were used to model the dispersion parameter:
Model 1: Fixed dispersion parameter $\alpha$
Model 2: $\alpha_i = \frac{1}{\phi_i} = \beta_i F_i^{\beta_4} F_i^{\beta_5}$
Model 3: $\alpha_i = \frac{1}{\phi_i} = \beta_i (F_i + F_i)^{\beta_i}$

The models were estimated using the PROC NLMIXED in SAS (28). As described above, the variance-covariance matrix was used for estimating the confidence intervals.

4. Similar to step 2, the confidence interval for the gamma mean ($m$) and the prediction interval ($y$) were computed for each variance function described in step 3.

DESCRIPTION OF DATA

The statistical models were developed using data collected in California at 537 3-legged unsignalized intersections between 1992 and 2001. The data were obtained from the Federal Highway Administration’s (FHWA) Highway Safety Information System (HSIS), maintained by the University of North Carolina. The data included information on crash severity, entering flows as well as other variables not included in the analysis. The descriptive statistics are presented in Table 2. This table shows that entering flows for the major approach varied between 2,430 and 78,300 vehicles per day. For the minor approach, the entering flows varied between 10 and 23,110 vehicles per day.

Table 2. Summary Statistics for Input Data

<table>
<thead>
<tr>
<th></th>
<th>Maximum</th>
<th>Minimum</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major Flow</td>
<td>78,300</td>
<td>2,430</td>
<td>16,454.12</td>
<td>13,445.57</td>
</tr>
<tr>
<td>Minor Flow</td>
<td>23,111</td>
<td>10</td>
<td>441.568</td>
<td>1,322.93</td>
</tr>
<tr>
<td>Crashes/site/year</td>
<td>13.9</td>
<td>0</td>
<td>1.10</td>
<td>1.73</td>
</tr>
</tbody>
</table>

MODELING RESULTS

This section describes the modeling results for the three functional forms or models described above. Table 3 shows the output for the four models. As expected, the coefficients show that the mean predicted number of crashes follows a non-linear relationship with the entering traffic flows. This table indicates that the coefficients do not change significantly between Models 2 and 3. However, the differences between the model with a fixed dispersion parameter (Model 1) and the other two models are more noticeable. This disparity can be attributed by the differences in the values observed in the variance-covariance matrix (see Table 4).
Table 3. Estimates of model coefficients (and standard errors)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
</tr>
<tr>
<td>Mean Function</td>
<td></td>
</tr>
<tr>
<td>Intercept (ln(β₀))</td>
<td>-10.7190 (0.5113)³</td>
</tr>
<tr>
<td>β₁</td>
<td>0.9703 (0.0530)</td>
</tr>
<tr>
<td>β₂</td>
<td>0.2550 (0.0265)</td>
</tr>
<tr>
<td>α</td>
<td>0.7337 (0.0560)</td>
</tr>
<tr>
<td>Intercept (ln(β₄))</td>
<td>-</td>
</tr>
<tr>
<td>β₄</td>
<td>-</td>
</tr>
<tr>
<td>β₅</td>
<td>-</td>
</tr>
<tr>
<td>β₆</td>
<td>-</td>
</tr>
<tr>
<td>Variance Function</td>
<td></td>
</tr>
<tr>
<td>(dispersion parameter)</td>
<td></td>
</tr>
<tr>
<td>Deviance</td>
<td>602.46</td>
</tr>
<tr>
<td>AIC</td>
<td>3250.6</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-1621.3</td>
</tr>
</tbody>
</table>

¹ Standard Error

Table 4 below shows the output of the variance-covariance matrix calculated using a fixed dispersion parameter and the different functional forms describing the varying variance function. This table shows that, similar to the mean function, the values between the fixed and varying dispersion parameters are different.

Table 4. Variance-Covariance Matrix Output

<table>
<thead>
<tr>
<th>Parameters</th>
<th>β₀</th>
<th>β₁</th>
<th>β₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₀</td>
<td>0.26147</td>
<td>-0.02609</td>
<td>-0.002141</td>
</tr>
<tr>
<td>β₁</td>
<td>-0.02609</td>
<td>0.002812</td>
<td>-0.000146</td>
</tr>
<tr>
<td>β₂</td>
<td>-0.002141</td>
<td>-0.000146</td>
<td>0.0006999</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₀</td>
<td>0.2698</td>
<td>-0.02671</td>
<td>-0.00166</td>
</tr>
<tr>
<td>β₁</td>
<td>-0.02671</td>
<td>0.002858</td>
<td>-0.0002</td>
</tr>
<tr>
<td>β₂</td>
<td>-0.00166</td>
<td>-0.0002</td>
<td>0.000676</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₀</td>
<td>0.2692</td>
<td>-0.02653</td>
<td>-0.00206</td>
</tr>
<tr>
<td>β₁</td>
<td>-0.02653</td>
<td>0.002812</td>
<td>-0.00014</td>
</tr>
<tr>
<td>β₂</td>
<td>-0.00206</td>
<td>-0.00014</td>
<td>0.000668</td>
</tr>
</tbody>
</table>
Figure 1 illustrates the frequency distribution of the varying dispersion parameter for Models 2 and 3. Recall that Model 1 provided a dispersion parameter equal to 0.7337. Figure 1a depicts the frequency of the varying dispersion parameter values for the functional form $\alpha_i = \beta_3 F_{1i}^{\beta_{4i}} F_{2i}^{\beta_{5i}}$ (each observation has its own dispersion parameter estimate). This figure shows that the highest frequency (i.e., the mode) occurred between the range of 0.50-0.70 and the average dispersion value was found to be 0.83, which is slightly higher than the value found for the fixed dispersion parameter. Figure 1b illustrates the frequency of the varying dispersion parameter values for the functional form $\alpha_i = \beta_3 \left( F_{1i} + F_{2i} \right)^{\beta_5}$. The highest frequency occurred between 0.70 and 0.90 with an average dispersion value equal to 0.80.
Figure 2 shows the plot of the predicted dispersion parameters between Models 2 and 3. In this figure, the predicted dispersion parameters for each observation are compared with each other. This figure illustrates that Model 3 seems to underestimate the dispersion parameter compared to Model 2 for values above 1.

![Figure 2. Relationship between Models 2 and 3 as a function of the dispersion parameter.](image)

Figure 3 below illustrates the relationship between the dispersion parameter and entering flows for Models 2 and 3. This figure shows that low entering flows for both approaches are associated with larger dispersion values compared to that of higher flows for every model. The difference in the predicted dispersion parameters between minor entering flows of $F_2=100$ veh/day and $F_2=3,000$ veh/day is about 30% when the entire range of entering flows for the major approaches for Model 2. On the other hand, for Model 3, the difference in predicted values between the entering flows for the minor approaches is almost zero for major flows above 30,000. Figure 3 also illustrates that the dispersion parameter is much more sensitive for intersections with low entering flows. This has also been reported by Lord and Park (16) using a different dataset.
a) Model 2: $\alpha = \beta_3 F_1^{\beta_4} F_2^{\beta_5}$

**Figure 3. Relationship between the dispersion parameter and entering flows $F_1$ and $F_2$**

**CONFIDENCE INTERVALS**

This section describes the results of the confidence interval analysis between the model with a fixed dispersion parameter and the models with a varying dispersion parameter. The equations presented in Table 1 were used to find the 95% confidence interval for the gamma mean and predicted response for the Models 2 and 3. For this exercise, three values describing the minor entering flows were used: $F_2 = 3,000, 1,000, \text{ and } 100 \text{ veh/day}$.

Figures 4 and 5 show the 95% confidence intervals for the gamma mean and predicted response for Model 2, respectively. Figure 4 illustrates that the difference in the width (i.e., the distance between the mean response and the upper confidence interval boundary) between the models with a fixed and a varying dispersion (w.r.t. fixed) for
gamma mean \textquote{m} varied between (i.e., the value between the fixed dispersion parameter less the value produced from the varying dispersion parameter):

- -4\% and 32\% for F2=3000
- -10\% and 28\% for F2=1000
- -23\% and 20\% for F2=100

Figure 5 shows that the difference in the width between the models with a fixed and varying dispersion (w.r.t. fixed) for predicted response \textquote{y} varied between:

- -0.6\% and 28\% for F2=3000
- -0.6\% and 25\% for F2=1000
- -40\% and 16\% for F2=100

Both these figures show that models with a varying dispersion parameter produce smaller confidence intervals for both the gamma mean and the predicted response. The upper boundary for the fixed dispersion parameter is almost always above the one estimated using the varying dispersion parameter model.
Figure 4. 95% Confidence Intervals for the Gamma Mean ($m$) (Model 2)
[Note: The dotted lines represent the upper and lower confidence limits while the continuous lines show the mean values.]
Figure 5. 95% Confidence Intervals for the Predicted Values ($y$) (Model 2)
The 95% confidence intervals for the gamma mean and predicted response for Model 3 are illustrated in Figures 6 and 7, respectively. Figure 6 shows that the difference in the width between the models with a fixed and a varying dispersion (w.r.t fixed) for gamma mean ‘m’ varied between:

-9% and 21% for F2=3000
-16% and 21% for F2=1000
-21% and 20% for F2=100

Figure 7 shows that the difference in the width between the models with a fixed and varying dispersion (w.r.t fixed) for predicted response ‘y’ varied between:

-21% and 20% for F2=3000
-30% and 19% for F2=1000
-40% and 17% for F2=100

Similar to the previous results, Figures 6 and 7 illustrates that models following the functional form for Model 3 with a fixed dispersion parameter produces larger confidence intervals than models with a varying dispersion parameter.
Figure 6. 95% Confidence Intervals for the Gamma Mean ($m$) (Model 3)
Figure 7. 95% Confidence Intervals for the Predicted Values ($y$) (Model 3)
SUMMARY AND CONCLUSIONS

The main objective of this study was to examine the effects of varying dispersion on the width of confidence intervals for the gamma mean of Poisson-gamma models and the predicted response. To accomplish the objective of this study, several flow-only predictive models were developed using 357 3-legged rural unsignalized intersections located in California. Three functional forms or model types were used to describe the relationship between the dispersion parameter and the covariates of the model. For this study, the following results were obtained from the analysis:

1. The dispersion parameter varies from one site to another and depends on the flows entering the intersection. The same conclusion was observed for the two models with a varying dispersion parameter. This support the findings documented in previous studies (5, 21).
2. Since the variance-covariance matrix depends on the dispersion parameter, var(η) is also dependent on the dispersion parameter. Thus, the mean accident rate calculated with the fixed dispersion and varying dispersion parameter yielded different results. The models with a varying dispersion parameter provided better statistical fit.
3. The confidence intervals for both the gamma mean and predicted response had larger values with the fixed dispersion than with the varying dispersion; in extreme cases, the difference could be as large as 40%.
4. Therefore, it is recommended to develop and use models with a varying dispersion (particularly for flow-only models), since they usually provide better statistical fit and more precise values. This is important if they are used for screening different highway design alternatives based on safety (e.g., 29) or identifying hazardous sites (14).

Some recommendations for further work on this topic are as follows:

1. This study was carried out using 3-legged unsignalized intersections. It is suggested to conduct further studies for other types of at-grade intersections and highway segments and compare the results with this study.
2. The independent variables considered only included entering flows at major and minor approaches of intersections. The authors recommend that further research considers other explanatory variables. As reported by Mitra and Washington (21), the varying dispersion parameter may not be needed when the functional form describing the mean function contains several covariates.
3. Although the functional form used in this study is the most commonly used for intersection models, further research should be conducted to examine the effects of a varying dispersion parameter using different functional forms, as suggested by Miaou and Lord (5) and Lord and Park (16).
ACKNOWLEDGEMENTS

The authors would like to thank Dr. Peter Y. Park from iTRANS Consulting Inc. for providing useful suggestions to improve this manuscript.

REFERENCES


11. Hauer, E (2001) Overdispersion in modelling accidents on road sections and in


Flow-V/C Ratio for Rural and Urban Freeway Segments. *Accident Analysis &


Regression Techniques in Developing Accident Prediction Models, Presented at
Transportation Research Board 85th Annual Meeting, Washington, D.C.

dispersion parameters of Poisson-gamma models on empirical Bayes estimates.
Research Paper, Zachry Department of Civil Engineering, Texas A&M University,
College Station, TX


Statistic and Confidence Intervals of Crash Prediction Models Subjected to low
Sample Mean Values. *Transportation Research Record 1950*, pp. 35-43.


Fitting Predictive Accident Models. *Accident Analysis & Prevention*, Vol. 28, No. 3,
pp.281-296.

459-468.


