報告番号	甲　第　11897　号

主　論　文　の　要　旨

論文題目　Analysis on Crash Types Frequency Models Considering Correlation
(相関を考慮した交通事故種別頻度モデルに関する研究)

氏　名　MOTHAFER　Ghasak Ibrahim
Mohamed Amen

論　文　内　容　の　要　旨

Despite the evident unobserved heterogeneity correlation among the crash types that frequently occur on freeway segments, inadequate research has been devoted in safety analysis to accommodate such correlation. Furthermore, ignoring such correlation could possibly lead to an enormous misleading conclusions and judgments since the former affects the model parameter efficiency. The correlation components in dynamic states alter with the length of the observation time which makes even more difficult to trace. Modeling the unobserved heterogeneity can improve the predictions of the count outcomes of interest as well. Thus, these improvements can be achieved via introducing the multivariate count model concept. Recent advancements in multivariate count econometric models have allowed researchers to investigate the correlations using simulation based techniques which are not so tractable in a sense of time consuming and efficiency. Our main objective in this research is to find more flexible model and easy to be used by analysts, then apply the obtained knowledge to model the traffic crash types counts that frequently occur on freeway segments. We will investigate the correlations and covariances among the rear end, sideswipe, fixed object and other crash types on freeway sections using three-year crash data for 274 multilane freeway segments in the State of Washington, U.S.A.

To comprehend correlations among different types of accidents and explanatory variables, while taking full benefit of the available crash count record, a multivariate

Poisson gamma mixture (MVPGM) count model has been implemented. The model consents a restricted correlation pattern allowing for positive correlation among accident types. The model parameters are estimated using a maximum likelihood approach. Based on the empirical results presented in this study error correlations across accidents types are significantly presented. These significant error correlations occur due to common unobserved heterogeneity affecting the specific accident frequency type. The proposed model shows significant unobserved correlations among different types of accidents frequencies. It also provides a better representativeness for the variance and covariance structure of each accident type Furthermore, the results reveal that rear end accident type is more likely to be affected by geometric and traffic characteristics of freeway. It is also found that considering the types of accidents is efficient and similar to modeling total number of accidents.

For the second part of this study we tried to contribute to methodology enhancement of the multivariate count data modeling by introducing a simple and practical formula. The formulation started from modifying the standard ordered response model to adopt the count outcomes nature. This modification is accomplished by introducing a non-linear asymmetric interdependence structure among the error terms using the copula-based model. To avoid using the simulation maximum-likelihood to solve the joint probability of multi-integrations among the count outcomes dimensions in the joint probability function, we proposed to utilize the composite marginal likelihood (CML) approach. It is proven that this approach with the copula formulation works efficiently and easy to be implemented for the discrete data. The proposed model allows the positive and negative dependency among the count outcomes as well as a variety of dependent structures including radially asymmetric or tail dependency without a need for a simulation mechanism.

We apply these techniques to study the interdependence structure for the same crash count dataset. The developed second model is applicable for parameter estimates using the maximum likelihood approach. The empirical results show a significant presence of the unobserved heterogeneity dependency across these types of crashes. The results also show that considering the unobserved heterogeneity are highly recommended to enhance the covariance and the variance structure estimation when they are compared to the observed ones. Another finding is that the characteristics of the horizontal curves on the designated freeway segment increase the likelihood of these types of crashes occurring, when compared to the characteristics of vertical curves.

Later, we shifted our scope to the serial correlation problem using the same

crash-count data set that we used before but this time considering the time of observation. The unobserved heterogeneity now is in dynamic status, thus, time invariant heterogeneity arising through multiple years of observation (between 2005 and 2007) for each segment is viewed as a common unobserved effect at the segment level, and typically treated with panel models involving fixed or random effects. Random effects model unobserved heterogeneity through the error term, typically following a gamma or normal distributions. We exploited the fact that gamma heterogeneity in a multi-period Poisson count modeling framework is equivalent to a negative binomial distribution for a dependent variable which is the summation of crashes across years. The Poisson panel model is the random effects Poisson gamma (REPG). In the REPG model, the dependent variable is an annual count of crashes of a specific type. The multi-year crash sum model is a negative binomial (NB) model that is based on three consecutive years of crash data (2005-2007). In the multi-year crash sum model, the dependent variable is the sum of crashes of a specific type for the three-year period. Four categories (in addition to total crashes) of crash types are considered in this study including rear end, sideswipe, fixed objects and all-other types. The empirical results show that the three-year crash sum model is a computationally simpler alternative to a panel model for modeling time invariant heterogeneity while imposing fewer data requirements such as annual measurements.

Within panel cash-count context, as our final target in this thesis, we utilized all the knowledge we gain through all the developed previous models to construct an econometric framework to model the multivariate panel crash count by type data. The point of emphasis is that modeling multivariate count panel data has more superior econometric benefits, which is clarified in producing more efficient parameter estimates compared to the ones arising from the multivariate cross-sectional models. Therefore, we considered the intertemporal (serial) correlations of a given crash type among the years of observations. Moreover, we have considered the inter-type correlations formulated by jointing the probabilities among different crash types. Both of these correlation components added a higher intricacy to seek a conceivable inference. We developed two flexible models to overcome this problem: Multivariate Panel Poisson Gamma Copula (MVPPGC) and Multivariate Panel Copula-Copula (MVPCC) model. These two models are in no need for a simulation mechanism, which is a common issue to model the multivariate count outcomes. The source of flexibility of these models is demonstrated through allowing a non-linear asymmetric shape of these correlation components generated among the unobserved heterogeneity of each crash type and across the years of observations. The empirical results suggest that Frank copula

statistically outperforms other copula types to fit the serial correlation among the years of observations of each crash type. Moreover, MVPCC model offers a better prediction of the crash-type count, since it more accurately represents the variance-covariance structure.