MITSUI SUMITOMO INSURANCE WELFARE FOUNDATION
RESEARCH GRANT 2023

Instructions for

REPORT OF RESEARCH RESULTS

(a) Title

Modeling the crash severity at street intersections by the Latent Class Bivariate Generalized Ordered Probit model (LCBGOP)

(b) Primary Researcher

Fu, Chiang, Department of Transportation and Communication Management Science, National Cheng Kung University, Assistant Professor

Co-researcher

Tu, Hsin-Tung, Intelligent Transportation Development Center, National Chung Hsing University, Research Assistant

Hung, Yu-Chen, Department of Transportation and Communication Management Science, National Cheng Kung University, Master Student

Yeh, Rui-Zhe, Department of Transportation and Communication Management Science, National Cheng Kung University, Master Student

(c) Summary

Street intersection crashes often involve two parties (vehicle-vehicle and vehicle-pedestrian). The disregard for the right-of-way of motorcycles and the pedestrian environment in our country has been ignored, making vulnerable users more prone to serious accidents. However, existing improvements have been proven ineffective. Therefore, it is necessary to analyze the factors affecting the injury severity and to make improvements accordingly. The parties involved in crashes can vary considerably. To accurately identify the causality of a two-party crash, it is necessary to assess the damage of both parties simultaneously. While the latent class ordinal model has been used in crash severity studies to capture heterogeneity in crash propensity, most are univariate. They are inappropriate for the context of two-vehicle crashes. We propose a latent class bivariate generalized ordered probit (LCBGOP) model to examine two-party crashes at intersections in the study.

This study collected 32,308 cases of two-party crashes at street intersections in Taipei City from 2018 to 2020. Injury severity is categorized into three levels: property damage only, minor/possible injury, and fatal/evident injury. Here are two classes, low-risk and high-risk, determined as the optimal class number through the latent class method. According to our model, the Ordinary Crash Severity (OCS) group mainly involves two-vehicle crashes colliding with motorcycles; the High Crash Severity (HCS) group comprises vulnerable road users like pedestrians and cyclists, mainly in mixed traffic with high volumes.

Our model-based estimation points out several potential factors, such as drivers (elderly), violations (safety equipment, yielding to vehicles, or hit-and-run), and modes (four-wheeled vehicles, two-wheeled vehicles, or pedestrians). Three elements of traffic engineering, namely people, vehicles, and roads, are some existing risk factors that can influence severity. Through the elasticity effects, the OCS group has a higher magnitude of fatal/evident injury than the HCS does. By variable patterns, the mode of mobility exhibits the highest fatal/evident injury values, underscoring its significant influence. Accordingly, we hope to reduce violations at intersections and prevent large vehicle crashes.

The results show that the party-specific factors contribute to injury severity more than generic factors do, providing invaluable insight into intersection crashes from the perspective of reducing two-party collisions. By integrating the traditional traffic 3E (Engineering, Education, and Enforcement) with Encouragement

into 4E, we develop the corresponding safety measures to reduce the frequency and severity of future crashes. It is recommended that authorities implement the strategies proposed in this study and enhance public awareness of driving. Finally, this study clarifies causal relationships in accidents by analyzing crash severity and fault determination, enabling risk management for insurance.

(d) Aim of Research

This study investigates factors influencing accident occurrence at intersections and offers suggestions for improvement. It delves into the varying severity levels experienced by both parties in the same accident, stemming from distinct driving behaviors, vehicle attributes, traffic conditions, and other risk factors. By examining these differences, we aim to propose effective improvement strategies.

This research aims to tackle the aforementioned unresolved issues by introducing latent class BGOP. The model is unprecedented in the literature. Notably, the proposed approach enables the within-crash correlations of two-party crash severity. Moreover, it categorizes two-party crashes into different risk occurrence patterns based on characteristics, enhancing our understanding of crash variables in the context of two-party crashes. Furthermore, the study utilizes comprehensive and representative samples to estimate the model. This allows for exploring the factors affecting the crash severity of both parties and diving into safety countermeasures to prevent intersection crashes, especially vehicle-pedestrian crashes.

The main objectives of this study are as follows: Establish a novel LCBGOP model to analyze the observable heterogeneity and explore the factors affecting the severity of both drivers. Based on the model estimation results, improve strategies to serve as references for enhancing traffic safety. Clarify causal relationships in accidents by analyzing crash severity and fault determination, effectively enabling risk management for insurance.

(e) Method of Research & Progression

Latent Class Bivariate Generalized Ordered Probit model (LCBGOP)

The framework used for modeling individual driver-level injury severity assumes that two-party drivers can be implicitly sorted into S relatively homogenous (but latent to the analyst) segments (types) based on crash characteristics. Within each segment, the effects of exogenous variables are fixed across two-party drivers in the segment. Let s be the index for segments *(s = 1, 2, …, S)*, q_n be the index representing two parties (i.e. q_1, q_2) involved at the same accident *q (q =1, 2, …, Q)*; k and l are the indices for the two parties' injury severity levels *(k = 1, 2..., K; l =1, 2..., L)*.

The crash outcomes are analyzed using an LCBGOP model within each segment. Across segments, the coefficients for the two-party drivers in each BGOP model vary. Under the BGOP model framework, conditional on two drivers q_n=1,2 belonging to segment s, the ordinal injury severity levels pair (y_{q1},y_{q2}) are assumed to be a mapping (or partitioning) of two underlying continuous latent response variables (y^*_{q1},y^*_{q2}) as follows:

$$y^*_{q1|s} = X_q\beta_{1|s} + \varepsilon_{q1|s}, y_{q1|s} = k, if\ \tilde{u}_{q,1,k-1} < y^*_{q1|s} < \tilde{u}_{q,1,k}$$

$$y^*_{q2|s} = X_q\beta_{2|s} + \varepsilon_{q2|s}, y_{q2|s} = l, if\ \tilde{u}_{q,2,l-1} < y^*_{q2|s} < \tilde{u}_{q,2,l}$$

Where, X_q is a variable vector is a variable vector, $\beta_{1|s}$ and $\beta_{2|s}$ are corresponding vectors of unknown parameters specific to segment s. $\varepsilon_{1|s}$ and $\varepsilon_{2|s}$ are segment-specific random disturbance term assumed to be identically and independently normal. To simplify the estimation, the current study assumed the thresholds ($\tilde{u}_{q,1,k}$,$\tilde{u}_{q,2,l}$) associated with two-party drivers *(n=1,2)* and their injury levels (k, l) vary over the two-parties but are fixed across segments. There the ordering conditions expressed as ($-\infty < \tilde{u}_{q,1,1} < \tilde{u}_{q,1,2} < \cdots < \tilde{u}_{q,1,k-1} < +\infty; -\infty < \tilde{u}_{q,2,1} < \tilde{u}_{q,2,2} < \cdots < \tilde{u}_{q,2,l-1} < +\infty$) ∀ *s = 1, 2, …, S.* To maintain the ordering conditions and allow the thresholds to vary across two- party drivers, the non-linear parameterization of the thresholds as mentioned, since the thresholds are invariable across the segments.

Given the above set-up, the joint probability that two-party drivers, q_1 and q_2, suffering certain injury severity injury levels (k, l) and belonging to segment s engaged in the same accident may be formulated as:

$$Pr(\tilde{\mu}_{1,k-1} < y^*_{q,1|s} < \tilde{\mu}_{1,k};\ \tilde{\mu}_{2,l-1} < y^*_{q,2|s} < \tilde{\mu}_{2,l})$$
$$= Pr(\tilde{\mu}_{1,k-1} < \beta'_{1|s}X_q + \varepsilon_{q|s,1} < \tilde{\mu}_{1,k};\ \tilde{\mu}_{2,l-1} < \beta'_{2|s}X_q + \varepsilon_{q|s,2} < \tilde{\mu}_{2,l})$$
$$= Pr(\tilde{\mu}_{1,k-1} - \beta'_{1|s}X_q < \varepsilon_{q|s,1} < \tilde{\mu}_{1,k} - \beta'_{1|s}X_q;\ \tilde{\mu}_{2,l-1} - \beta'_{2|s}X_q < \varepsilon_{q|s,2} < \tilde{\mu}_{2,l} - \beta'_{2|s}X_q)$$
$$= \Phi_2(\tilde{\mu}_{1,k} - \beta'_{1|s}X_q, \mu_{2,l} - \beta'_{2|s}X_q\,; \rho_s) - \Phi_2(\tilde{\mu}_{1,k-1} - \beta'_{1|s}X_q, \tilde{\mu}_{2,l} - \beta'_{2|s}X_q\,; \rho_s)$$
$$- \Phi_2(\tilde{\mu}_{1,k} - \beta'_{1|s}X_q, \tilde{\mu}_{2,l-1} - \beta'_{2|s}X_q\,; \rho_s) + \Phi_2(\tilde{\mu}_{1,k-1} - \beta'_{1|s}X_q, \tilde{\mu}_{2,l-1} - \beta'_{2|s}X_q\,; \rho_s)$$

Where the random errors were assumed as like BOP and BGOP under bivariate normal cumulative distribution Φ_2 (·), and the correlation parameter ρ_s varied across segments.

The analyst does not observe the segment to which the two parties were within. So, the analyst specifies this segment assignment to be a function of a column vector of observed crash factors ω_q. To also recognize the presence of unobserved factors that may change this assignment, the analyst develops an expression for the probability of two parties q_1 and q_2, belonging to segment s. While many parametric expressions may be used for this probability expression (the only requirement is that the probabilities sum to one across the segments for two parties in the same crash), the most commonly used form corresponds to the multinomial logit structure (see Bhat, 1997; Greene and Hensher, 2003; Eluru et al., 2012):

$$P_{qs} = \frac{exp(\eta_s \omega_q)}{\sum_s exp(\eta_s \varpi_q)}$$

Where η_s is a row vector of parameters to be estimated. Then, the unconditional probability of two parties leading up to injury severity levels (k, l) can be written as

$$P_q(q_1, q_2) = \sum_{s=1}^{S}(P_q(q_1, q_2)|s) * (P_{qs})$$

The log-likelihood function for the entire dataset can be written as

$$L = \sum_{i=1}^{N} log\ [\sum_{s=1}^{S}(P_q(q_1, q_2)|s)\ * \left(P_{qs}\right)]$$

The parameters to be estimated in the LCBGOP model are the segment parameters $(\beta_{1|s}, \beta_{2|s}, \rho_s, \gamma'_{q,k}, \delta'_{q,l}, \alpha_{qk}, \theta_{ql})$, the class probability parameters (η_s) or each s, and the appropriate number of segments S. For identification reasons, we need to restrict one of the η_s vectors to zero.

(f) Results of Research

The study employs an LCBGOP model to analyze two-party crashes at intersections. This model incorporates threshold values and within-crash correlations that vary across both classes. Notably, the within-crash correlation is found to vary across classes and is related to the exogenous covariates. The research has dived into the interrelationship of those crash variables that are party-specific or generic among the functions of LCBGOP. Our results confirm that party-specific factors (e.g., large vehicles, intoxicated, and wearing safety equipment) significantly impact each party's injury severity in both classes than generic factors (e.g., visibility, road surface, and intersection configuration).

Furthermore, our findings of two-party crashes also confirm that these risk factors are highly intertwined. Specifically, crash types frequently involved specific violations, resulting in different injuries for each party involved. For instance, angular (CSA) and T-bone (CST) intersection collisions during high crash severity occurrences are commonly associated with violations such as "turning without following the right-of-way" (VLOT) and "not yielding to the right-of-way vehicles" (VLOY). In such crashes, both parties are likely to sustain injuries classified as level "K/A/B" if the second party is a motorcyclist (VHM) being inattentive to the vehicles ahead (VLOI).

Last but not least, the study also identifies two injury severity occurrence groups: OCS and HCS. In the OCS group, crashes are likely PDOs and level "C" injuries, whereas those belonging to the HCS group are likely injuries for "C" and "K/A/B" levels. Major crash features serve as the basis for this classification, such as violations, mode of mobility, and roadway conditions. Within the OCS group, two-vehicle collisions are likely to be a negative correlation of the injury for two parties. Typically, these incidents occur when the first-party driver/rider crossing intersections fail to give way to the vehicle with the right-of-way. Moreover, elderly individuals are often the second-party victims. Collisions within the HCS group are likely to involve small vehicles colliding with either bike riders or pedestrians because the first-party driver/rider crossing an intersection may not yield to pedestrians and vehicles with right of way or watch the road ahead. The correlation within HCS might be somewhat ambiguous, which can be illustrated by certain variables.

(g) Future Areas to Take Note of, and Going Forward

As for future research directions, the current estimation sets the threshold parameters to be invariant across the two classes to obtain stable estimates. It may be a price for a sophisticated model specification. The adoption of a more efficient algorithm, such as the expectation-maximization (EM) algorithm (Sfeir et al., 2021; Sfeir et al., 2022), could be considered to address this issue.

Additionally, it is noteworthy that the current classification of injury severity levels combines fatal and evident injuries into fatal/evident injury ("K/A/B") to maintain sufficient samples at the highest injury level. However, this classification fails to reflect the unique characteristics of fatal crashes. Redesigning the severity levels is desirable to study the relevant factors in fatal crashes. Also, as the sample is confined to two parties (exactly two people), many crashes involving more than two occupants and self-collisions are excluded, which may have influenced the current estimate. Perhaps a more complete study should include these cases in the context of multivariate analysis.

(h) Means of Official Announcement of Research Results

Proposed as a dissertation for master's degree in Graduate program of transportation science, Department of Transportation and Communication Management Science, National Cheng Kung University.

Part of the research results were published at the 2024 Conference of the Chinese Institute of Transportation: Fu, C., & Tu, H. T. Investigating the two-party crash severity at street intersections by the Latent Class Parameterized Correlation Bivariate Generalized Ordered Probit.

Part of this study was submitted to the journal Accident Analysis & Prevention: Fu, C., & Tu, H. T. (2024). Investigating vehicle-vehicle and vehicle-pedestrian crash severity at street intersections with the latent class parameterized correlation bivariate generalized ordered probit. Accident Analysis & Prevention, 207, 107745.