Copula-Based Method for Addressing Endogeneity in Models of Severity of Traffic Crash Injuries

Application to Two-Vehicle Crashes

Tejsingh A. Rana, Sujan Sikder, and Abdul Rawoof Pinjari

An emerging copula-based methodology is used to address endogeneity in models of crash injury severity. Specifically, two important sources of endogeneity are addressed: (*a*) endogeneity due to correlations between the injury severity of the two drivers involved in two-vehicle crashes and (*b*) endogeneity of collision type and injury severity outcomes. Two sets of copula-based joint model systems are formulated and estimated by using data on two-vehicle crashes from the 2007 Generalized Estimates System: a copula-based joint ordered logit–ordered logit model of injury severities of the two drivers involved in two-vehicle crashes and a copula-based joint multinomial logit–ordered logit model of collision type and injury severity outcomes of two-vehicle crashes. The model estimation results and elasticity estimates underscore the importance of accommodating endogeneity in models of crash injury severity. The results shed new light on the determinants of injury severity in two-vehicle accidents.

Automobile crashes cause significant losses to society through fatalities, traffic congestion, medical costs, and property damage. The severity of injuries sustained by individuals involved in these incidents depends on a multitude of factors, including roadway design features, environmental factors (weather, traffic conditions), vehicle characteristics (size, weight, safety features), and driver characteristics (age, gender) and their driving behavior (speed, seat-belt use, etc.). It is important to understand and quantify the influence of each of these factors on crash injury severity so that engineering and policy measures for enhanced safety may be formulated. Transportation safety literature abounds with studies that model the relationship between crash injury severity and the identified factors.

A major limitation of several studies of crash injury severity, however, is the neglect of several sources of endogeneity bias. Even if endogeneity is recognized, the currently used methods are associated with several drawbacks. Thus, this paper uses a new copula-based methodology to address the issue of endogeneity in models of crash injury severity. Specifically, two important sources of endogeneity are addressed in the context of two-vehicle crashes: (*a*) endogeneity due to the correlations between the injury severities of the drivers involved in two-vehicle crashes, and (*b*) endogeneity due to the common unobserved factors affecting the collision type and injury severity outcomes in two-vehicle crashes. To this end, the study formulates and estimates (*a*) a copula-based joint ordered logit–ordered logit (ORL-ORL) model system with which to jointly model the injury severity levels of the drivers involved in two-vehicle crashes, and (*b*) a copula-based joint multinomial logit–ordered logit (MNL-ORL) model system with which to jointly analyze the collision type and injury severity outcomes in two-vehicle crashes.

T. A. Rana and S. Sikder, ENC 1208, and A. R. Pinjari, ENC 2503, Department of Civil and Environmental Engineering, University of South Florida, 4202 East Fowler Avenue, Tampa, FL 33620. Corresponding author: A. R. Pinjari, apinjari@usf.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2147, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 75–87.
DOI: 10.3141/2147-10

ENDOGENEITY IN MODELING OF CRASH INJURY SEVERITY

Econometrically speaking, endogeneity bias occurs because of the presence of nonindependent errors in the model specification (*1*). Such nonindependent errors can occur for several reasons, including the presence of endogenous explanatory variables where the observed explanatory variables in a model are correlated with the unobserved factors in the error terms, and the presence of correlations across error terms of different model equations due to common unobserved factors influencing the outcome variables of interest.

A classic case of the former type of endogeneity in models of crash injury severity corresponds to seat-belt usage. The decision to wear a seat belt is typically used as an exogenous explanatory variable to explain the influence of seat-belt use on crash injury severity without considering the behavioral differences between those who wear seat belts and those who do not. However, it is possible that seat-belt nonusers may be intrinsically unsafe drivers (*2, 3*) and their unsafe driving habits may lead to severe crashes. The impact of such behavioral differences on injury severity outcomes, when not considered, may become confounded with the impact of seat-belt use on injury severity outcomes. This econometric issue arises because the variable of seat-belt use in the model is likely to be correlated to the error term that contains unobserved factors (such as unsafe driving habits) influencing injury severity. In other words, common unobserved factors influence both seat belt use and crash injury severity. Ignoring such endogeneity can lead to biased and inconsistent estimates and distorted policy implications (*1*).

Similar to the endogeneity of seat-belt use, other variables such as crash type (head-on, angle, rear-end, etc.) and vehicle occupancy may be endogenous to the severity of crashes. For example, in the context of crash type, it is well recognized that head-on collisions are likely to result in more severe injuries than are angle, rear-end, or sideswipe collisions. To accommodate this, collision type is typically included as an exogenous explanatory variable. However, several unobserved driver characteristics or roadway features may result in a particular type of collision and may influence the severity of injuries. Because of such influence of common unobserved factors, the collision type variables can potentially be endogenous to injury severity outcomes.

Another important type of endogeneity occurs because of common unobserved factors affecting the injury severity of individuals involved in the same crash. For example, if a driver involved in a two-vehicle crash sustains severe injuries, it is likely that other individuals involved in that crash also sustain severe injuries. This is because several factors (such as speeding, aggressive driving, roadway design, and environment-related factors) that influence the injury severity of one driver also influence the injury severity of another driver. Thus, the injury severity propensities of all the individuals involved in a crash can potentially be endogenous to each other. Despite the likelihood of such endogeneity, most studies of crash injury severity model the injury severity of each crash victim as independent of the injury severity of other victims from the same crash as if each individual were involved in a different crash.

The traffic safety literature is not devoid of studies that recognize and capture the forms of endogeneity discussed here. Table 1 provides a summary of these studies. The first six studies in the table focus on the endogeneity of seat-belt use. Among these, whereas Evans (*2*) addressed this issue by using simple descriptive analysis, Dee (*4*), Derrig et al. (*5*), and Cohen and Einav (*6*) used aggregate-level analyses with data on seat-belt usage rates and fatality rates. Eluru and Bhat performed disaggregate crash victim-level analysis of injury severity to address the issue of seat-belt endogeneity (*3*). They proposed a joint random coefficients modeling approach in which common random terms are used across the seat-belt use and injury severity models to capture the endogeneity due to common unobserved factors affecting seat-belt use and injury severity. To do the same, de Lapparent used the bivariate probit modeling approach (*7*).

Only two studies have endogenously analyzed crash type with injury severity. Kim et al. included crash type as an endogenous variable in a structural equations framework (*8*), and Ye et al. explicitly recognize the endogeneity of two-vehicle collision variables by using a joint random coefficients MNL-ORL model of collision type and injury severity (*9*).

Lee and Abdel-Aty used bivariate probit models to address the endogeneity of vehicle occupancy and passenger characteristics in crash outcome (crash type, injury severity) models (*10*). Such endogeneity of explanatory variables has been found to be an important issue not only in crash injury severity models but also in crash frequency models. For example, Kim and Washington identified that ignoring the endogeneity of left-turn-lane variables may lead to a counterintuitive result that left-turn lanes cause an increased occurrence of angle accidents (*11*).

Another set of studies addressed endogeneity due to the simultaneity of the injury severity outcomes of the individuals involved in a crash (*12–15*). These studies used bivariate ordered (or binary) response modeling methods to jointly model the injury severities of the two individuals under consideration. Other studies recognized the presence of a multilevel hierarchy (such as vehicle level, crash level, and location level) in traffic crash data (*16–19*). These studies provide evidence of significant magnitude of common unobserved factors at each level of the hierarchy.

Each of the methods used in these studies is associated with specific drawbacks. For example, the joint random coefficients modeling approach necessitates the use of simulation-based estimation, which is computationally intensive and saddled with such technical issues as parameter unidentifiability. The bivariate probit method is not easily adaptable to accommodate the endogeneity of polychotomous categorical variables (such as collision type). Another disadvantage is that the bivariate normal distribution used for most bivariate modeling methods is restrictive in the types of correlations (or dependencies) it can accommodate between the variables of interest. The multilevel modeling method, as identified by Lenguerrand et al., is more applicable "when the number of vehicles per crash and the number of occupants per vehicle is high" (*17*). However, when the number of data points in a cluster (or level) is small, such models may be difficult to estimate. In modeling of crash injury severity, since a large proportion of crashes do not involve more than two individuals per vehicle and more than two vehicles per crash, an alternative approach may be preferred. Even if estimable, multilevel models require computation-intensive simulation-based estimation methods.

Because of the disadvantages of the currently used methods, this paper uses a recently emerging copula-based approach to address endogeneity in models of crash injury severity for drivers involved in two-vehicle crashes. The copula-based approach, as discussed next, offers several advantages over the currently used methods.

COPULA-BASED METHODOLOGY

Copulas are mathematical constructs used to generate dependency among stochastic variables with known marginal distributions (*20, 21*). Specifically, a copula is a multivariate distribution function defined to link (or tie) several uniformly distributed marginal variables. Following Bhat and Eluru (*21*), if $U_1, U_2, \ldots, U_N$ are N uniformly distributed random variables, then the N-dimensional copula or the N-dimensional joint distribution of these random variables can be defined as

$$C_\theta(u_1, u_2, \ldots, u_N) = \Pr(U_1 < u_1, U_2 < u_2, \ldots, U_N < u_N) \tag{1}$$

where θ is a parameter vector referred to as the dependence parameter vector analogous to (but not the same as) the correlation vector in a multivariate normal distribution. Such a copula function can be applied to prespecified marginal distributions to generate dependency (or correlations) among those marginal distributions. To see this, again following Bhat and Eluru (*21*), consider N univariate random variables $X_1, X_2, \ldots, X_N$, each with continuous marginal distribution functions $F_n(y_n) = \Pr(Y_n < y_n)$, $n = 1, 2, \ldots, N$. Using the integral transform result, the marginal distribution of each random variable X_n can be expressed as

$$F_n(x_n) = \Pr(X_n < x_n) = \Pr(F_n^{-1}(U_n) < x_n) = \Pr(U_n < F_n(x_n)) \tag{2}$$

Then, by Sklar's theorem (*22*), a joint N-dimensional distribution function of the random variables with the marginal distribution functions $F_n(x_n)$ can be generated as

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= \Pr(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n) \\ &= \Pr(U_1 < F_1(x_1), U_2 < F_2(x_2), \ldots, U_n < F_n(x_n)) \end{aligned} \tag{3}$$

TABLE 1 Summary of Endogeneity Literature

No.	Paper	Data Used	Model Structure Used	Type of Endogeneity	Research Method and Findings
1	Evans (*2*)	National Accident Sampling System (NASS) data (1982–91)	Descriptive analysis	Endogeneity of seat-belt use	Explored the relationship between the effectiveness of seat belt and crash severity (measured by the change of velocity from before to after crash, inferred from the extent of vehicle damage). Results: Drivers who do not wear seat belts are more likely to be involved in severe crashes, and without considering this effect seat-belt effectiveness can be overestimated.
2	Dee (*4*)	Seat-belt use data from NHTSA and Centers for Disease Control and Prevention (1985–93)	Linear probability models	Endogeneity of seat-belt use	Panel data models were used to understand the influence of seat-belt laws on fatality rates. Results: unsafe (or accident prone) drivers are less responsive to seat-belt laws (i.e., they tend to continue to not wear seat-belts even after the enactment of seat-belt laws), which causes an attenuation in the benefits of seat-belt laws.
3	Derrig et al. (*5*)	Fatality rates and seat-belt usage data from Fatality Analysis Reporting System (FARS) (1991–96)	Multivariate regression analysis for per-capita fatality rates per vehicle miles traveled	Endogeneity of seat-belt use and state insurance system	Panel data models were used. Instrumental variables were used for the risk-taking incentive of the insurance system by state. Results: increase in seat-belt usage rates in the general population may not lead to reductions in fatality rates, as long as accident prone drivers maintain risky behavior.
4	Cohen and Einav (*6*)	Aggregated data from FARS (1983–97)	Log-linear regression model for per-capita fatality rates in each state in the United States	Endogeneity of seat-belt use	Presence of mandatory seat-belt laws (by state) was used as an instrument variable to control for the endogeneity of seat-belt usage. Ignoring the endogeneity of seat-belt usage rates resulted in a biased assessment of the effect of seat-belt usage rates on the predicted fatality rates.
5	Eluru and Bhat (*3*)	Generalized Estimates System (GES) data (2003)	Joint random coefficients binary logit–ordered logit model	Endogeneity of seat-belt use	Jointly modeled seat-belt usage and injury severity. Results indicate the presence of unobserved factors influencing both seat-belt use and injury severity. The influence of seat-belt use on injury severity was overestimated if the endogeneity of seat-belt usage was not considered.
6	de Lapparent (*7*)	French road accident reports (2003)	Bivariate ordered probit model of seat-belt use and injury severity	Endogeneity of seat-belt use	Jointly modeled seat-belt use and injury severity, separately for the drivers, front passengers, and rear passengers. Results suggest that while seat-belt use is effective in moderating injury severity, drivers may compensate for some of this safety benefit by taking more risks.
7	Ye et al. (*9*)	GES data (2005)	Joint random coefficients multinomial logit–ordered logit model of collision type and injury severity	Endogeneity of collision-type variables in two-vehicle crashes	Jointly modeled collision type and injury severity of two-vehicle crash victims. Results suggest that the unobserved factors contributing to head-on collisions are negatively associated with those contributing to severe injuries, whereas the unobserved factors contributing to rear-end crashes are positively correlated with those contributing to severe injuries.
8	Kim et al. (*8*)	Crash data from Hawaii CODES project (1990)	Structural equations models	Endogeneity of seat-belt use and crash type	Structural equations model with seat-belt use, crash type, and injury severity as endogenous variables. No specific result or discussion was provided on endogeneity.
9	Lee and Abdel-Aty (*10*)	5-year crash records of Interstate-4 freeway in Orlando, Florida (1999–2003)	Bivariate ordered probit models	Endogeneity of passenger characteristics	Jointly modeled passenger characteristics (presence, number, and age of passengers) and crash characteristics (citation, crash type, and injury severity) to capture the endogeneity of passenger characteristics with the crash characteristics (such as injury severity). It was found that drivers display safer driving behavior with the presence of passenger(s), but younger drivers with younger passengers may be more crash prone.

(continued on next page)

TABLE 1 *(continued)* **Summary of Endogeneity Literature**

No.	Paper	Data Used	Model Structure Used	Type of Endogeneity	Research Method and Findings
10	Kim and Washington (*11*)	Intersections data set of 38 counties within the state of Georgia	Joint negative binomial model for angle crashes and logit model for left-turn lanes	Endogeneity of left-turn lane presence in angle crash occurrence models	Installation of left-turn lanes at intersection appears to contribute to crashes when endogeneity is not considered. Recognizing endogeneity results in a negative effect of left-turn lanes on occurrence of angle crashes, which is intuitive and concurrent with engineering judgment.
11	Hutchinson (*12*)	British road accidents data (1969–72)	Bivariate normal distributed model of injury severities of two drivers in a two-vehicle crash	Simultaneity (common unobserved factors influencing injuries of different individuals)	Developed a bivariate ordered probit type model of injury severity of the two drivers involved in a two-vehicle crash. Results indicate that unobserved factors common to both drivers (assumed to be the relative speed in this case) play an important role in determining injury severity levels.
12	Hutchinson (*13*)	British road accidents data (1969–72)	Bivariate normal distributed model of driver's and front passenger's injury severity	Simultaneity (common unobserved factors influencing injuries of different individuals)	Research method same as above. Significant positive correlation was found between the injury severities of the two occupants in a vehicle. The correlation was interpreted as largely due to the speed of the crash.
13	Ouyang et al. (*14*)	Washington State accident records database (1990–96)	Simultaneous binary logit model	Simultaneity (common unobserved factors influencing injuries of different individuals)	Jointly modeled the most severe injury in each vehicle for car–truck collisions. Results show significant positive correlation between the two injury severity propensities and that considering jointness ensures more efficient and less-biased estimates.
14	Yamamoto and Shankar (*15*)	Washington State accident records database (1990–96)	Bivariate ordered-response probit model	Simultaneity (common unobserved factors influencing injuries of different individuals)	Jointly modeled the injury severity of the driver and that of the most severely injured passenger in a single-vehicle accident. It was found that the error term of the driver's injury severity propensity is positively correlated with that of the most severely injured passenger.
15	Jones and Jørgensen (*16*)	Norwegian road accident police records (1985–96)	Multilevel logistic regression	Common unobserved factors at accident level and accident location level	Three-level (crash victim, crash, and crash location levels) regression models were estimated to disentangle the unobserved factors at the individual, crash, and crash location levels. Results signify the presence of intraunit correlations in the data set at both the crash and crash location (municipality, in this case) levels.
16	Lenguerrand et al. (*17*)	French road crash data (1996–2000)	Multilevel logistic, generalized estimating equations, and simple logistic regression models were compared	Common unobserved factors at vehicle level and accident level	Three-level (crash victim, vehicle, and crash levels) regression models were estimated using multilevel and generalized estimating equations approaches. Results indicate nonnegligible correlations at crash level, at the same time indicating a need for large data sets and great care to estimate multilevel models.
17	Kim et al. (*18*)	Crash data for 91 two-lane rural intersections in the state of Georgia (1996–97)	Multilevel binomial logistic models of crash type	Common unobserved factors at the intersection level	Two-level (crash-level and intersection-level) binary logistic models of crash type (angle, rear-end, sideswipe, etc.). Results indicate a significant presence of intersection-level unobserved factors affecting the crash type outcomes.
18	Helai et al. (*19*)	Database of crashes at urban intersections in Singapore	Multilevel logistic regression	Correlation between the individuals involved in same crash	Two-level (individual level and crash level) binary logistic models were estimated. The results show that 28.9% of unexplained variation in severity level results due to the between-crash variance (i.e., crash-level unobserved factors). Significant correlation was found between the accident severities of the individuals involved in the same crash.

which is nothing but a copula function (as in Equation 1) as follows:

$$F(x_1, x_2, \ldots, x_n) = C_\theta\big(u_1 = F_1(x_1), u_2 = F_2(x_2), \ldots, u_n = F_n(x_N)\big) \tag{4}$$

There are several advantages to using copulas. First, copulas can capture more general forms of dependency (asymmetric dependence, asymptotic dependence, etc.) than the simple, symmetric, and asymptotically independent forms of dependency exhibited in bivariate normal distributions of the bivariate probit models. Second, the dependency form is independent of the marginal distributions of the stochastic variables. Thus the stochastic variables of interest (for example, injury severity propensity and collision type-propensity) need not necessarily follow the same marginal distribution. Third, a variety of copula functions can be used to explore different forms of dependency other than the usual bivariate normal distributions used in earlier studies. Fourth, several copulas offer closed-form probability expressions and obviate the need for computationally intensive simulation-based model estimation.

In modeling of crash injury severity, as discussed next, copulas can be used to join the injury severity models of individuals involved in an accident into a simultaneous equations framework. Further, copulas can be applied to incorporate endogeneity of collision type (or other) variables.

Independent Injury Severity Model

Let q_{dj} be an index to represent the two drivers d ($d = 1, 2$) involved in a two-vehicle collision q ($q = 1, 2, \ldots, Q$) of type j($j = 1, 2, \ldots, J$) and let k_{dj} ($k_{dj} = 1, 2, 3, \ldots, K$) be an index to represent injury severity of the drivers. Here, j takes the values of head-on ($j = 1$), angle ($j = 2$), rear-end ($j = 3$), sideswipe ($j = 4$), and other ($j = 5$), where the index k_{dj} takes the values of no injury ($k_{dj} = 1$), possible injury ($k_{dj} = 2$), nonincapacitating injury ($k_{dj} = 3$), incapacitating injury ($k_{dj} = 4$), and fatal injury ($k_{dj} = 5$). Let $y_{q_{dj}}$ denote the observed injury severity sustained by the drivers involved in a two-vehicle collision, let $y^*_{q_{dj}}$ denote the latent (unobserved) injury severity propensity of those drivers, and let $\psi_{k_{dj}}$ be the thresholds used to map the observed injury severity levels to the latent injury severity propensities. With this notation, define $y^*_{q_{dj}}$ as

$$y^*_{q_{dj}} = \beta'_j x_{q_{dj}} + \xi_{q_{dj}} \tag{5}$$

where β'_j is a vector of coefficients of the observed factors $x_{q_{dj}}$ affecting the driver's injury severity if she or he was involved in an accident of type j and $\xi_{q_{dj}}$ is a random component that captures the unobserved factors affecting the injury severity. The latent injury severity propensity $y^*_{q_{dj}}$ of each driver d is mapped to his or her injury severity level $y_{q_{dj}}$ by the $\psi_{k_{dj}}$ thresholds:

$$y_{q_{dj}} = k_{dj} \qquad \text{if } \psi_{k_{dj}-1} < y^*_{q_{dj}} < \psi_{k_{dj}}$$

or

$$y_{q_{dj}} = k_{dj} \qquad \text{if } \psi_{k_{dj}-1} < \beta'_j x_{q_{dj}} + \xi_{q_{dj}} < \psi_{k_{dj}}$$

or

$$y_{q_{dj}} = k_{dj} \qquad \text{if } \left(\psi_{k_{dj}-1} - \beta'_j x_{q_{dj}}\right) < \xi_{q_{dj}} < \left(\psi_{k_{dj}} - \beta'_j x_{q_{dj}}\right) \tag{6}$$

Given this mapping, and assuming that the $\xi_{q_{dj}}$ terms are Gumbel distributed, the probability of a driver d involved in a two-vehicle accident q of type j sustaining an injury severity of level k_{dj} ($y_{q_{dj}} = k_{dj}$) is given by the familiar ordered logit formula:

$$P\left(y_{q_{dj}} = k_{dj}\right) = F_{\xi_d j}\left(\psi_{k_{dj}} - \beta'_j x_{q_{dj}}\right) - F_{\xi_d j}\left(\psi_{k_{dj}-1} - \beta'_j x_{q_{dj}}\right) \tag{7}$$

where $F_{\xi dj}(.)$ is the cumulative distribution function of the error term $\xi_{q_{dj}}$. This probability expression represents an independent injury severity model (for each driver and accident type j) that does not capture any form of endogeneity.

Joint Injury Severity Model for Two Drivers Involved in Two-Vehicle Crashes

The equation system for simultaneously modeling the injury severity of two drivers q_{1j} and q_{2j} involved in a two-vehicle accident q of type j can be written as

$$\begin{aligned} y_{q_{dj}} &= k_{1j} \qquad \text{if } \left(\psi_{k_{1j}-1} - \beta'_j x_{q_{1j}}\right) < \xi_{q_{1j}} < \left(\psi_{k_{1j}} - \beta'_j x_{q_{1j}}\right) \\ y_{q_{dj}} &= k_{2j} \qquad \text{if } \left(\psi_{k_{2j}-1} - \beta'_j x_{q_{2j}}\right) < \xi_{q_{2j}} < \left(\psi_{k_{2j}} - \beta'_j x_{q_{2j}}\right) \end{aligned} \tag{8}$$

From this equation system, the joint probability that one driver sustains injuries of severity level k_{1j} and another driver sustains injuries of severity of level k_{2j} is

$$\begin{aligned} &P\left(y_{q_{1j}} = k_{1j}, y_{q_{2j}} = k_{2j}\right) \\ &\quad = P\Big(\Big[\left(\psi_{k_{1j}-1} - \beta'_j x_{q_{1j}}\right) < \xi_{q_{1j}} < \left(\psi_{k_{1j}} - \beta'_j x_{q_{1j}}\right)\Big], \\ &\qquad \Big[\left(\psi_{k_{2j}-1} - \beta'_j x_{q_{2j}}\right) < \xi_{q_{2j}} < \left(\psi_{k_{2j}} - \beta'_j x_{q_{2j}}\right)\Big]\Big) \\ &\quad = P\Big[\xi_{q_{1j}} < \left(\psi_{k_{1j}} - \beta'_j x_{q_{1j}}\right), \xi_{q_{2j}} < \left(\psi_{k_{2j}} - \beta'_j x_{q_{2j}}\right)\Big] \\ &\qquad - P\Big[\xi_{q_{1j}} < \left(\psi_{k_{1j}} - \beta'_j x_{q_{1j}}\right), \xi_{q_{2j}} < \left(\psi_{k_{2j}-1} - \beta'_j x_{q_{2j}}\right)\Big] \\ &\qquad - P\Big[\xi_{q_{1j}} < \left(\psi_{k_{1j}-1} - \beta'_j x_{q_{1j}}\right), \xi_{q_{2j}} < \left(\psi_{k_{2j}} - \beta'_j x_{q_{2j}}\right)\Big] \\ &\qquad + P\Big[\xi_{q_{1j}} < \left(\psi_{k_{1j}-1} - \beta'_j x_{q_{1j}}\right), \xi_{q_{2j}} < \left(\psi_{k_{2j}-1} - \beta'_j x_{q_{2j}}\right)\Big] \end{aligned} \tag{9}$$

The form of the preceding probability expression depends on the specification of the dependency form between the random terms $\xi_{q_{1j}}$ and $\xi_{q_{2j}}$. Specifically, one may use copula functions to write the joint probability expression as

$$\begin{aligned} P\left(y_{q_{1j}} = k_{1j}, y_{q_{2j}} = k_{2j}\right) &= C_\theta\left(u^j_{qk_1}, u^j_{qk_2}\right) - C_\theta\left(u^j_{qk_1}, u^j_{qk_2-1}\right) \\ &\quad - C_\theta\left(u^j_{qk_1-1}, u^j_{qk_2}\right) + C_\theta\left(u^j_{qk_1-1}, u^j_{qk_2-1}\right) \end{aligned} \tag{10}$$

where $C_\theta(.,.)$ is a copula function defining the dependency form between $\xi_{q_{1j}}$ and $\xi_{q_{2j}}$, and

$$\begin{aligned} u^j_{qk_1} &= F_{\xi_1 j}\left(\psi_{k_1 j} - \beta'_j x_{q_{1j}}\right), u^j_{qk_1-1} = F_{\xi_1 j}\left(\psi_{k_1 j-1} - \beta'_j x_{q_{1j}}\right) \\ u^j_{qk_2} &= F_{\xi_2 j}\left(\psi_{k_2 j} - \beta'_j x_{q_{2j}}\right), u^j_{qk_2-1} = F_{\xi_2 j}\left(\psi_{k_2 j-1} - \beta'_j x_{q_{2j}}\right) \end{aligned}$$

Using the joint probability expression of Equation 10 for the injury severity of the two drivers involved in a two-vehicle accident q, the likelihood function of injury severity outcomes for all the Q two-vehicle accidents is

$$L = \prod_{q=1}^{Q} \left\{ \prod_{k_{1j},k_{2j}=1}^{K} \left[P\left(y_{q_{1j}} = k_{1j}, y_{q_{2j}} = k_{2j}\right)\right]^{\delta_{qk_{1j}}\delta_{qk_{2j}}} \right\}^{w_q} \quad (11)$$

where $\delta_{qk_{1j}}$ and $\delta_{qk_{2j}}$ are dummy variables taking the value 1 if Driver 1 and Driver 2 involved in accident q of type j sustain injuries of levels k_{1j} and k_{2j}, respectively, and 0 otherwise. w_q is the weight for accident q used to represent an unbiased sample of two-vehicle crashes. Given this likelihood function, a maximum likelihood approach is used to estimate the model parameters.

Joint Model of Collision Type and Injury Severity

In the joint model of collision type and injury severity, the injury severity model component takes the ordered logit specification as in Equations 6 and 7, and the collision-type model component takes the familiar discrete choice formulation. Consider the following equation that represents the propensity of an accident type j:

$$u_{qj}^{*} = \alpha'_j z_{qj} + \epsilon_{qj} \quad (12)$$

where

- u_{qj}^{*} = propensity that qth accident is of type $j (j = 1, 2, \ldots, J)$;
- z_{qk} = column vector of roadway design and environment, vehicle, and other attributes (including a constant) affecting the propensity;
- α'_j = corresponding coefficient vector; and
- ϵ_{qj} = error term capturing the effects of unobserved factors on the propensity associated with accident type j.

With this propensity specification, the accident type outcome of an accident q is assumed to be of type j if it is associated with the maximum propensity among all J accident types, that is, if

$$u_{qj}^{*} > \max_{l=1,2,\ldots,J,\ l\neq J} u_{ql}^{*} \quad (13)$$

Next, following Spissu et al. (*23*), the preceding polychotomous outcome model is recast into a series of binary outcome model formulations, one for each collision type. Let R_{qj} be a binary variable that takes a value of 1 if accident q is of type j and 0 otherwise. Subsequently, substituting $\alpha'_j z_{qj} + \epsilon_{qj}$ for u_{qj}^{*} (from Equation 12), rewrite Equation 13 as

$$R_{qj} = 1 \qquad \text{if } \alpha'_j z_{qj} > v_{qj},\ \left(j = 1,2,\ldots,J\right) \quad (14)$$

where

$$v_{qj} = \left\{ \max_{l=1,2,\ldots,J,\ l\neq j} u_{ql}^{*} \right\} - \epsilon_{qj} \quad (15)$$

Equation 14 represents a series of binary outcome model formulations, one for each collision type j, which is equivalent to the multinomial discrete choice model of collision type. An assumption that the ϵ_{qj} terms are assumed to be independent (across j) and identical Gumbel distributed results in logistic distribution for the v_{qj} terms, and consequently the collision-type probability expressions resemble the multinomial logit probabilities.

Now, the joint probability that an individual gets involved in a collision of type j and sustains injuries of severity level k_{dj} is given by

$$\begin{aligned} P\left(R_{qj} = 1, y_{q_{dj}} = k_{dj}\right) &= P\left\{\left(\alpha'_j z_{qj} > v_{qj}\right), \left(\left(\psi_{k_{dj}-1} - \beta'_j x_{q_{dj}}\right) < \xi_{q_{dj}} \right.\right.\\ &\qquad \left.\left. < \left(\psi_{k_{dj}} - \beta'_j x_{q_{dj}}\right)\right)\right\} \\ &= P\left(\left(\alpha'_j z_{qj} > v_{qj}\right), \left(\xi_{q_{dj}} < \psi_{k_{dj}} - \beta'_j x_{q_{dj}}\right)\right) \\ &\quad - P\left(\left(\alpha'_j z_{qj} > v_{qj}\right), \left(\xi_{q_{dj}} < \psi_{k_{dj}-1} - \beta'_j x_{q_{dj}}\right)\right) \end{aligned} \quad (16)$$

This joint probability expression depends on the dependence structure between the random variables v_{qj} and $\xi_{q_{dj}}$. As indicated earlier, copula-based methods are used in this paper to capture these dependences. First the marginal distributions of v_{qj} and $\xi_{q_{dj}}$, $F_{vj}(.)$ and $F_{\xi_{dj}}(.)$, are transformed into uniform distributions by using their inverse cumulative distribution functions. Subsequently, copula functions are applied to couple the marginal inverse cumulative distribution functions into a joint distribution $F_{vj,\ \xi_d j}(.,.)$. Thus, for a driver involved in an accident q, the joint probability that the collision type outcome is j and the injury severity outcome is k_{dj} can be expressed by using the following copula functions:

$$P\left(R_{qj} = 1, y_{q_{dj}} = k_{dj}\right) = \left[C_\theta\left(u_{qk_d}^{j}, u_q^{j}\right) - C_\theta\left(u_{qk_d-1}^{j}, u_q^{j}\right)\right] \quad (17)$$

where $C_{\theta_j}(.,.)$ is the copula corresponding to $F_{vj,\xi_d j}(u_{qk_d}^{j}, u_q^{j})$ and $F_{vj,\xi_d j}(u_{qk_d-1}^{j}, u_q^{j})$ with $u_{qk_d}^{j} = F_{\xi_d j}\ (\psi_{k_{dj}} - \beta'_j x_{q_{dj}})$, $u_{qk_d-1}^{j} = F_{\xi_d j}\ (\psi_{k_{dj}-1} - \beta'_j x_{q_{dj}})$, and $u_q^{j} = F_{vj}\ (\alpha'_j z_{qj})$. This copula function captures the dependency between v_{qj} $(= \left\{\max_{l=1,2,\ldots,J,l\neq j} u_{ql}^{*}\right\} - \epsilon_{qj})$ and $\xi_{q_{dj}}$ terms.

With the joint probability expression of Equation 17 for the collision type and injury severity outcomes, the likelihood function for all the Q two-vehicle accidents is

$$L = \prod_{q=1}^{Q} \left[\prod_{j=1}^{J} \prod_{k_{dj}=1}^{K} \left\{ P\left(R_{qj} = 1, y_{q_{dj}} = k_{dj}\right)\right\}^{R_{qj}\delta_{qk_{dj}}} \right]^{w_q} \quad (18)$$

where $\delta_{qk_{dj}}$ takes the value 1 if driver d involved in accident q sustained an injury of level k_{dj} and 0 otherwise (R_{qj} and w_q are as defined earlier). The logarithm of the preceding likelihood function is used in a maximum likelihood estimation routine (coded in GAUSS) to estimate the model parameters.

EMPIRICAL ANALYSIS

The data used in this study were obtained from the 2007 NHTSA Generalized Estimates System (GES). The 2007 GES includes information regarding 61,282 crashes (of which 35.7% were single-vehicle crashes and 55.9% were two-vehicle crashes) involving 152,727 individuals and 107,202 vehicles. Information pertaining to two-vehicle crashes (28,775 crashes) was extracted from this database. Further, crashes involving commercial vehicles or large trucks were discarded, and data with missing information were removed. Subsequently, 5,027 two-vehicle crash records (with 10,054 driver records) were randomly sampled for model estimation and analysis. Accident-level weights were developed such that the weighted distribution of injury severity in this sample was the same as that in the full sample of two-vehicle drivers.

Among the 10,054 driver records in the sample, close to 69.9% of the drivers experienced no injury, 14.3% driver records indicate possible injury, 9.1% indicate nonincapacitating injury, 6.1% indicate incapacitating injury, and 0.6% indicate fatal injury. Among the 5,027 two-vehicle crashes, 6.2% were head-on collisions, 46.4% were angle collisions, 38.5% were rear-end collisions, 7.6% were sideswipe crashes, and 1.2% were other types of crashes.

Joint Injury Severity Model for Two Drivers Involved in Two-Vehicle Crashes

This section presents the results of the bivariate copula-based joint ORL-ORL model (of the injury severity of the two drivers involved in two-vehicle crashes) as well as a simple independent ORL model (that does not consider correlations between the injury severities of the two drivers). For both models, Table 2 presents the parameter estimates (*t*-statistics in parenthesis) and marginal effects of each explanatory variable for the fatal injury category.

The first set of variables in the table corresponds to driver characteristics. Among these, the female dummy variable coefficients and marginal effects indicate that females are more susceptible to higher injury severities than males. Age-related variable effects indicate that drivers younger than 65 are less prone to higher-severity injuries than are drivers 65 or older. Effects of the alcohol or drug influence variable indicate that drivers under the influence of alcohol or drugs are likely to experience higher injury severities than those who do not drive under the influence. Further, the coefficients and the marginal effects of the alcohol or drug use by the driver of the partner vehicle indicates that in two-vehicle accidents, even if a driver is not under the influence, he or she is likely to experience higher injury severities

TABLE 2 Injury Severity Models for Two Drivers Involved in Two-Vehicle Crash

		Parameter Estimates (*t*-stats.)		Marginal Effects	
Variable	Descriptive Statistics (%)	Independent Model (ORL)	Copula-Based Joint Model (ORL–ORL)	Independent Model (ORL)	Copula-Based Joint Model (ORL–ORL)
Driver characteristics					
Gender—female	46.60	0.55(8.72)	0.53(9.05)	54.26	53.01
Age (>64 years is base)					
<25 years	27.00	−0.58(−5.22)	−0.59(−5.74)	−51.12	−52.43
25–64 years	63.90	−0.27(−2.75)	−0.27(−2.88)	−28.71	−28.06
Alcohol or drug use					
Driver	1.90	0.82(3.65)	0.83(3.82)	124.52	125.61
Driver of partner vehicle	1.90	0.66(3.10)	0.63(3.07)	91.35	86.42
Use of seat belts	97.40	−1.65(−9.39)	−1.47(−8.60)	−389.12	−312.89
Roadway characteristics					
Surface condition (dry–wet is base): snow, ice	3.30	−0.25(−1.32)	−0.28(−1.35)	−22.11	−24.57
Profile (level is base): grade	18.20	−0.16(−1.95)	−0.16(−1.83)	−15.18	−15.12
Speed limit (<26 mph is base)					
Medium (26–65 mph)	85.30	0.60(6.19)	0.51(5.25)	49.28	43.22
High (>65 mph)	2.00	0.41(1.52)	0.32(1.12)	50.00	37.27
Environmental factors					
Lighting condition (daylight is base): dark	5.30	0.23(1.75)	0.18(1.21)	25.74	19.56
Land use, population >100,000	40.40	0.31(4.51)	0.29(3.97)	28.59	27.42
Crash characteristics					
Manner of collision (rear-end and other collision is base)					
Head-on	6.20	1.87(14.55)	1.73(12.06)	487.48	417.18
Angle	46.40	0.58(8.92)	0.51(7.20)	61.09	53.24
Sideswipe collision	7.60	−0.90(−6.09)	−0.91(−5.73)	−62.01	−62.65
Vehicle role (striking other vehicle is base)					
Struck by other vehicle	44.40	0.54(8.46)	0.50(9.64)	55.42	51.55
Struck by and strikes other vehicle	2.30	1.67(8.04)	1.40(6.77)	409.38	296.19
Vehicle characteristics					
Body type (sedan is base)					
Pickup truck	15.80	−0.48(−4.9)	−0.49(−5.34)	−40.32	−41.26
Utility vehicle	18.90	−0.19(−2.4)	−0.19(−2.51)	−18.10	−17.63
Body type of the partner vehicle: Nonsedan (pickup, utility vehicle, minivan)	42.30	0.18(3.1)	0.20(3.41)	18.79	19.70
Age of vehicle: >10 years	26.10	0.24(3.4)	0.22(3.52)	25.79	23.46
Threshold parameters					
Threshold 1	—	1.20(5.26)	1.23(5.53)	—	—
Threshold 2	—	2.49(10.78)	2.51(11.14)	—	—
Threshold 3	—	3.87(16.00)	3.87(16.30)	—	—
Threshold 4	—	6.86(15.32)	6.76(15.65)	—	—
Copula dependency parameter (θ)	—	—	1.31(11.04)	—	—
Log likelihood at convergence	—	−4,923.45	−4,747.31	—	—

if the other driver is under the influence of alcohol or drugs. This result indicates that those who drive under the influence pose a risk of higher injury severity (and fatality) not only to themselves but also to other individuals involved in the accident. Although an intuitive finding, few previous studies documented this result. Finally, among the driver characteristics, the seat-belt effects indicate that use of safety belts helps protect occupants from severe injuries.

Among the roadway characteristics, roadway surface condition related effects indicate that accidents occurring during snow or ice on the road tend to be less severe than those occurring during dry or wet conditions. This may be because drivers exercise higher caution during such roadway conditions than during normal (dry or wet) conditions. Similarly, drivers appear to be more cautious on steeper roads than on level roads. The speed limit variable effects indicate that driver injury tends to be most severe for crashes on roads having medium (26 to 65 mph) and high speed limits (>65 mph) when compared to roads having low speed limits. Since vehicle speeds are higher on roads with medium and high speed limits, the injury severity on such roads is higher than that on roads with low speed limits. Since speed limit variables can be viewed as surrogates for (and are highly correlated with) road type classification (freeways, arterials, local streets, etc.), road type classification variables were not included in the model.

Crashes occurring in the dark tend to be more severe than those occurring during daylight. Land use variable effects indicate that crashes occurring in areas of larger population (>100,000) tend to be more severe. The reason for this result is not clear, as one would expect areas with larger populations to be associated with lower traffic speeds (because of high traffic volumes and congestion) and hence lower injury severity. The manner of collision variable effects indicate that head-on and angle collisions lead to more severe injuries, and sideswipe collisions lead to less severe injuries than rear-end and other types of collisions. [Ye et al. made a similar finding (*9*).] The marginal effects indicate that head-on collisions show the largest propensity to cause fatalities. Effects of the vehicle role variable suggest a higher injury severity level if the vehicle is struck or is struck and strikes another vehicle, relative to striking another vehicle.

Among the body type variables, as expected, drivers in pickup trucks and utility vehicles involved in two-vehicle accidents appear to be less prone to greater injuries than those in sedans. Further, if the partner (or other) vehicle involved in the accident is a nonsedan (i.e., a pickup truck, a utility vehicle), driver injury severity tends to be higher. Drivers of older vehicles indicate propensity for higher injury severity than those of new vehicles, perhaps because of the improved safety features in new vehicles.

The next set of parameters corresponds to the thresholds. The copula dependency parameter θ (for the copula-based joint model) represents the level of association (or correlation) between the injury severity propensities of the two drivers in two-vehicle accidents. This model explored different types of copula functions (including Gaussian, Frank, Gumbel, Joe, and Clayton copula functions) to model the association between the two propensities for injury severity. The Gumbel copula provided the best model fit. According to the properties of the Gumbel copula, a θ value greater than 1 indicates a nonzero correlation. The reported *t*-statistic of the parameter (against a null value of 1) shows that the θ parameter value of 1.31 is statistically different from 1, indicating significant positive correlations between the injury severities of the two drivers involved in an accident. [Similar results are shown by Hutchinson (*12*) and Ouyang et al. (*14*).]

The positive correlation indicates that the unobserved factors that increase the injury severity of one driver involved in a two-vehicle accident also increase the injury severity of the other driver involved in that accident. Such correlations may arise because of the presence of several common (to the equations for injury severity propensity of the two drivers) unobserved, but influential, factors that affect the injury severity of both drivers involved in two-vehicle accidents. Such factors include vehicle speeds, aggressive or risky driving behavior (such as speeding), and other roadway-related features (such as presence or absence of guardrails and traffic conditions) that are not usually well recorded in crash reports.

The log likelihood ratio statistic between the independent and the copula-based joint models is $-2 * (-4747 - -4923) = 352$, which is much higher than that the critical chi-square value for a degree of freedom of 1 (for one additional copula parameter in the joint model) at any level of significance. This indicates the superior statistical fit of the joint model over the independent model and indicates that driver injury severity should be modeled in a joint fashion for two-vehicle crashes. The independent injury severity model treats the two drivers involved in a two-vehicle accident as if from two separate accidents. Such an assumption may result in distorted estimates of the influence of various roadway, environmental, vehicle, and driver characteristics on injury severity. This can be observed by comparing the model estimates of the independent and copula-based joint models in Table 2. Further, the differences in the marginal effects obtained from the two models are nonnegligible for certain variables. Specifically, the joint model shows smaller marginal effects than the independent model. For example, the magnitude of the marginal effect of the variable seat-belt use from the joint model is considerably lower than that from the independent model. This may be because nonusers of safety belts may be intrinsically unsafe drivers (*3*), and because of this, both drivers in a two-vehicle accident are likely to experience injuries of high severity (especially fatal injuries). The joint model captures such common unobserved factors that affect the injury severity of both the drivers in the copula dependency parameter and isolates the seat-belt effect from any such confounding effects. Thus, the model indirectly controls for the endogeneity of the seat-belt variable. The independent model simply ignores such unobserved factors that become confounded into the effect of seat-belt use and lead to spuriously inflated estimates of the influence of seat-belt use. Similar inferences can be made in the context of the differences in marginal effects of the head-on collision variable and the vehicle role (i.e., struck by and strikes other vehicle) variable. Thus, an advantage of jointly modeling the injury severities of the individuals involved in a crash is that different sources of endogeneity can be controlled for.

Results of Joint Model of Collision Type and Injury Severity

This section presents the results of the copula-based joint MNL-ORL model of collision type and injury severity. Estimates of the collision-type model component are shown in Table 3. These estimates indicate the influence of various roadway, crash, and driver characteristics on the collision-type outcome of a two-vehicle crash (given the occurrence of a two-vehicle crash). For example, a two-vehicle crash is likely to be a head-on collision when people drive under the influence of alcohol or drugs, in snow or ice road conditions, on road segments that either are curved or have no median, and in the dark. Similarly, a two-vehicle crash is likely to be an angle collision when it occurs on straight road segments (rather than curved segments), roads with no median, roads with multiple lanes, in the dark, in areas with large populations (>100,000), at intersections, and at roadway sections with stop signs and yield signs. Two-vehicle crashes that occur in snow or ice conditions, on steep road segments, at intersections, and at roadway sections with yield signs and traffic signals are likely to be rear-end collisions. Two-vehicle crashes that occur during snow or ice conditions, on roads with multiple lanes, in the dark, and in areas with large populations (>100,000) are likely to be sideswipes.

Estimates of the injury severity component are also shown in Table 3. Two copula dependency parameters (θ_j) were estimated, one each for head-on and angle crashes. (The dependency parameters for the other three types of crashes were not statistically significant.) For both these types of crashes, the copula functions corresponding to Frank copulas provided the best model fit. For Frank copula functions, a dependency parameter significantly different from zero indicates a significant dependency (or correlation) between the marginal variables of interest (*21*). In this context, the implication is that there is significant positive correlation between the v_{qj} ($= \{\max_{l=1,2,\ldots,J, l\neq j} u^*_{ql}\} - \epsilon_{qj}$) and $\xi_{q_{dj}}$ terms for head-on and angle collisions (see the section on the joint model of collision type and injury severity). This implies that the implied correlation between collision-type propensity error term (ϵ_j) and injury severity propensity error term ($\xi_{q_{dj}}$) is negative and statistically significant. To interpret this result, it is important to note that the negative sign of the correlation term does not necessarily imply that injury severity in head-on or angle collisions is likely to be lower. Rather, the negative correlation indicates that the unobserved factors that contribute to the likelihood of head-on and angle collisions (within two-vehicle crashes) are negatively correlated with the unobserved factors that contribute to higher injury severity. [Ye et al. offer a similar result (*9*).] Further investigation is needed for understanding of the unobserved factors that lead to such negative correlations. Nonetheless, the statistically significant dependency parameters highlight the endogeneity of collision-type outcomes with injury severity outcomes. Ignoring such endogeneity may lead to poor model fit and biased estimates of variable coefficients. In the current context, the log-likelihood value deteriorated from −15,439 (for the joint MNL-ORL model) to −15,459 when collision type was not considered endogenous to the model system. This log-likelihood difference is equivalent to a log-likelihood ratio of 40, which is greater than the 95% critical chi-square value for two degrees of freedom (for the two copula parameters), indicating the statistical superiority of the joint model system.

As shown in the table, several of the injury severity model estimates (both thresholds and coefficients) vary across collision types. This allows the analyst to examine the differential impact of various factors on injury severity by collision type. For example, the threshold values for the head-on collisions are all smaller than the thresholds for other types of collisions. This implies an intuitive result that the probability of higher injury severity categories (fatality, etc.) tends to be higher for head-on collisions than for other types of collisions. Furthermore, the snow or ice conditions variable has been found to be insignificant for head-on collisions. This implies that the severity of injuries resulting from head-on collisions does not differ by surface conditions, perhaps because head-on collisions tend to result in severe accidents anyway, because of the high relative speed and impact of the crash. Similarly, the severity of injuries resulting from head-on collisions does not differ by the age of vehicle. Almost all studies in the literature ignore such differential effects and assume that all factors have the same impact on injury severity irrespective of the type of collision.

Table 4 presents the marginal effects (for the fatal injury category) from the joint model (which considers collision type as endogenous) as well as an independent model system (which does not consider collision type as endogenous). Several important observations can be made from this table. First, the marginal effects of both models show differential impacts of several variables by collision type. This result reiterates the need to examine the differential impact of various factors by collision type. Second, for both model systems, the marginal effects of several variables (gender, age, alcohol or drug use, seat-belt use, and environmental and crash characteristics) are smaller in magnitude for head-on collisions than for all other types of collisions. This result suggests that injury severity resulting from head-on collisions tends to be less moderated by various factors. This is especially the case with seat-belt use; seat belts are less protective in head-on collisions than in other types of collisions. Since protective measures are less effective in the event of a head-on collision, it is important to reduce the likelihood of a head-on collision by using control measures such as installation of median barriers. Third, there are nonnegligible differences in the marginal effects between the independent and joint models for certain variables. For example, in the context of the medium speed limit variable, the independent model system shows rather small magnitudes of marginal effect compared with that in the joint model system. The seat-belt-use variable shows higher marginal effects in the independent model system compared with that in the joint model system.

CONCLUSIONS

This paper used an emerging copula-based methodology to address endogeneity in models of crash injury severity. Two important sources of endogeneity were addressed in the context of two-vehicle collisions: endogeneity due to correlations between the injury severities of the two drivers involved in a two-vehicle crash and endogeneity of collision type due to the common unobserved factors affecting the collision type and injury severity outcomes. Two sets of joint model systems were formulated and estimated by using data on two-vehicle crashes from the 2007 GES: a copula-based joint ORL-ORL model of injury severity of the two drivers involved in two-vehicle crashes and a copula-based joint MNL-ORL model of collision type and injury severity outcomes of two-vehicle crashes.

Model estimation results from use of the two joint model systems show a statistically significant presence of the two types of endogeneity. Both model systems provide intuitive results on the impact of various roadway, environmental, vehicle, and driver characteristics on injury severity of the drivers involved in two-vehicle accidents. Furthermore, the joint model systems perform better than the independent counterparts that do not accommodate the corresponding endogeneity for model fit and policy implications. These results underscore the importance of accommodating endogeneity in modeling of crash injury severity, as well as the potential of the copula-based methods in traffic crash modeling and analysis.

Several important findings surfaced from this analysis. First, drivers under the influence of alcohol or drugs pose a risk of high (fatal) injury not only to themselves but also to other individuals involved in the accident. Such drivers are more likely to be involved in head-on collisions. Thus, strict enforcement policies need to be implemented

TABLE 3 Joint MNL-ORL Model

Variable	Collision Type (MNL) Model Component					Injury Severity (ORL) Model Component				
	Head-On	Angle	Rear-End	Sideswipe	Others	Head-On	Angle	Rear-End	Sideswipe	Others
Copula dependency type						Frank	Frank	None	None	None
Copula dependency parameter (θ)	—	—	—	—	—	3.03(5.1)	1.29(3.1)	—	—	—
Constant	−1.59(−13.1)	−0.23(−3.5)	—	−1.07(−11.8)	−3.04(−14.1)	—	—	—	—	—
Threshold (1)	—	—	—	—	—	−1.49(−5.8)	0.37(1.6)	0.53(2.3)	2.05(7.9)	1.52(6.4)
Threshold (2)	—	—	—	—	—	−0.64(−2.5)	1.52(6.4)	1.82(7.5)	3.19(10.1)	2.38(6.5)
Threshold (3)	—	—	—	—	—	0.37(1.6)	2.92(11.6)	2.92(11.6)	4.44(9.9)	4.44(9.9)
Threshold (4)	—	—	—	—	—	2.38(6.5)	6.01(10.1)	6.76(6.4)	6.76(6.4)	6.76(6.4)
Driver characteristics										
Gender: female	—	—	—	—	—	0.49(8.5)	0.49(8.5)	0.49(8.5)	0.49(8.5)	0.49(8.5)
Age (> 64 years is base)										
<25 years	—	—	—	—	—	−0.51(−4.9)	−0.51(−4.9)	−0.51(−4.9)	−0.51(−4.9)	−0.51(−4.9)
25–64 years	—	—	—	—	—	−0.24(−2.6)	−0.24(−2.6)	−0.24(−2.6)	−0.24(−2.6)	−0.24(−2.6)
Alcohol or drug use										
Driver	0.75(2.8)	—	—	—	—	0.75(3.6)	0.75(3.6)	0.75(3.6)	0.75(3.6)	0.75(3.6)
Driver of partner vehicle	—	—	—	—	—	0.62(3.2)	0.62(3.2)	0.62(3.2)	0.62(3.2)	0.62(3.2)
Use of seat belts	—	—	—	—	—	−1.50(−8.9)	−1.50(−8.9)	−1.50(−8.9)	−1.50(−8.9)	−1.50(−8.9)
Roadway characteristics										
Surface condition (dry–wet is base): snow, ice	0.77(2.2)	—	0.58(2.9)	0.60(2.1)	—	—	−0.49(−1.7)	−0.28(−1.1)	—	—
Profile (level is base): grade	—	—	—	—	—	−0.15(−2.1)	−0.15(−2.1)	−0.15(−2.1)	−0.15(−2.1)	−0.15(−2.1)
Alignment (straight is base): curve	0.35(2.2)	−0.93(−8.9)	—	—	—	—	—	—	—	—
Speed limit (<26 mph is base)	—	—	—	—	—	0.94(3.9)	0.54(4.9)	0.44(3.3)	0.44(3.3)	0.44(3.3)
Medium (26–65 mph)	—	—	—	—	—	0.36(1.5)	0.36(1.5)	0.36(1.5)	0.36(1.5)	0.36(1.5)
High (>65 mph)	1.20(7.6)	0.56(10.8)	—	—	—	—	—	—	—	—

Absence of median	—	0.15(3.2)	—	0.24(3.4)	—	—	—	—	—	—
Number of lanes—three or more										
Environmental factors										
Lighting condition										
Daylight	—	—	—	—	—	—	—	—	—	—
Dark	0.54(4.2)	0.12(2.1)	—	0.19(2.2)	0.49(2.0)	0.24(1.4)	0.24(1.4)	—	0.24(1.4)	0.24(1.4)
Land use: population >100,000	—	0.09(1.9)	—	0.10(1.3)	0.65(2.5)	0.27(4.3)	0.27(4.3)	0.27(4.3)	0.27(4.3)	0.27(4.3)
Crash characteristics										
Vehicle role (striking is base)										
Struck	—	—	—	—	—	0.50(8.6)	0.50(8.6)	0.50(8.6)	0.50(8.6)	0.50(8.6)
Struck by and strikes	—	—	—	—	—	1.63(8.2)	1.63(8.2)	1.63(8.2)	1.63(8.2)	1.63(8.2)
Accident at intersection	—	1.08(15.9)	0.58(8.2)	—	—	—	—	—	—	—
Traffic control device										
Stop sign	—	0.98(12.6)	—	—	—	—	—	—	—	—
Stop sign not at intersection	—	1.39(5.3)	—	—	—	—	—	—	—	—
Yield sign	—	1.06(2.9)	2.12(6.6)	—	—	—	—	—	—	—
Traffic control signal	—	—	0.31(5.2)	—	—	—	—	—	—	—
Vehicle characteristics										
Body type (sedan is base)										
Pickup truck	—	—	—	—	—	−0.84(−2.5)	−0.57(−4.0)	−0.33(−2.7)	—	—
Utility vehicle	—	—	—	—	—	—	−0.20(−1.7)	−0.32(−3.1)	—	—
Body type of the partner vehicle										
Nonsedan (pickup/utility vehicle)	—	—	—	—	—	0.26(3.3)	0.26(3.3)	—	—	—
Age of vehicle: >10 years	—	—	—	—	—	—	0.24(3.5)	0.24(3.5)	—	—

TABLE 4 Marginal Effects for Fatal Injury Severity

	Marginal Effects from Independent Model System					Marginal Effects from Joint Model System				
Variable	Head-On	Angle	Rear-End	Sideswipe	Others	Head-On	Angle	Rear-End	Sideswipe	Others
Driver characteristics										
Gender: female	52.36	52.88	53.69	53.16	60.62	45.82	48.13	48.94	48.51	54.52
Age (>64 years is base)										
<25 years	–50.80	–51.54	–50.40	–50.79	–48.57	–43.75	–46.12	–45.23	–45.55	–43.78
25–64 years	–27.53	–28.42	–29.06	–28.92	–29.51	–22.78	–24.71	–25.23	–25.09	–25.55
Alcohol or drug use										
Driver	109.36	119.32	119.51	120.06	112.00	93.16	108.71	108.97	109.36	102.66
Driver of partner vehicle	89.22	95.57	95.47	95.22	93.58	73.38	83.96	84.01	83.77	82.41
Use of seat belts	–327.04	–379.94	–390.00	–387.73	–406.32	–249.96	–323.42	–330.99	–329.71	–342.19
Roadway characteristics										
Surface condition (dry–wet is base): snow, ice	—	–38.44	–33.03	—	—	—	–39.44	–24.52	—	—
Profile (level is base): grade	–16.15	–16.44	–16.61	–16.48	–16.74	–13.40	–14.26	–14.39	–14.29	–14.49
Speed limit (<26 mph is base)										
Medium (26–65 mph)	28.51	59.04	33.41	33.23	33.46	69.11	46.04	37.42	37.22	37.57
High (>65 mph)	34.53	36.00	35.86	35.79	35.24	38.49	42.65	42.48	42.37	41.63
Environmental factors										
Lighting condition: dark	33.26	35.21	—	35.10	33.52	24.26	26.84	—	26.78	25.87
Land use: population >100,000	27.81	28.17	28.48	28.16	27.34	23.35	24.73	24.95	24.70	24.08
Crash characteristics										
Vehicle role (striking is base)										
Struck by other vehicle	68.60	56.26	54.56	58.14	63.12	56.48	50.37	49.14	51.79	55.84
Struck by and strikes other vehicle	392.26	436.94	460.55	410.93	375.64	299.22	387.40	406.13	367.14	336.82
Vehicle characteristics										
Body type (sedan is base)										
Pickup truck	–72.26	–44.10	–38.38	—	—	–60.11	–46.12	–29.16	—	—
Utility vehicle	—	–14.57	–35.28	—	—	—	–18.72	–28.59	—	—
Body type of the partner vehicle										
Nonsedan (pickup, utility)	28.02	29.22	—	—	—	24.26	26.35	—	—	—
Age of vehicle: >10 years	—	28.37	29.02	—	—	—	24.34	24.82	—	—

to reduce driving under the influence of alcohol or drugs. Second, the injury severity resulting from head-on collisions tends to be less moderated by other factors (such as seat-belt use) than for injuries from other collisions. This result underscores the importance of control measures that reduce the likelihood of head-on collisions.

REFERENCES

1. Washington, S., M. Karlaftis, and F. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman and Hall, Boca Raton, Fla., 2003.
2. Evans, L. Safety-Belt Effectiveness: The Influence of Crash Severity and Selective Recruitment. *Accident Analysis and Prevention,* Vol. 28, No. 4, 1996, pp. 423–433.
3. Eluru, N., and C. R. Bhat. A Joint Econometric Analysis of Seat Belt Use and Crash Related Injury Severity. *Accident Analysis and Prevention,* Vol. 39, No. 5, 2007, pp. 1037–1049.
4. Dee, T. S. Reconsidering the Effects of Seat Belts Laws and their Enforcement Status. *Accident Analysis and Prevention,* Vol. 30, No. 1, 1997, pp. 1–10.
5. Derrig, R. A., M. Segui-Gomez, A. Abtahi, and L. L. Liu. The Effect of Population Safety Belt Usage Rates on Motor Vehicle-Related Fatalities. *Accident Analysis and Prevention,* Vol. 34, No. 1, 2000, pp. 101–110.
6. Cohen, A., and L. Einav. The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities. *Review of Economics and Statistics,* Vol. 85, No. 4, 2003, pp. 828–843.
7. de Lapparent, M. Willingness to Use Safety Belt and Levels of Injury in Car Accidents. *Accident Analysis and Prevention,* Vol. 40, No. 3, 2008, pp. 1023–1032.
8. Kim, K., L. Nitz, J. Richardson, and L. Li. Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention,* Vol. 27, No. 4, 1995, pp. 469–481.
9. Ye, X., R. M. Pendyala, F. S. Al-Rukaibi, and K. Konduri. A Joint Model of Accident Type and Severity for Two-Vehicle Crashes. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
10. Lee, C., and M. Abdel-Aty. Presence of Passengers: Does It Increase or Reduce Driver's Crash Potential? *Accident Analysis and Prevention,* Vol. 40, No. 5, 2008, pp. 1703–1712.
11. Kim, D. G., and S. Washington. The Significance of Endogeneity Problems in Crash Models: An Examination of Left-Turn Lanes in Intersection Crash Models. *Accident Analysis and Prevention,* Vol. 38, No. 6, 2006, pp. 125–134.
12. Hutchinson, T. P. A Bivariate Normal Model for Intra-Accident Correlations of Driver Injury with Application to the Effect of Mass Ratio. *Accident Analysis and Prevention,* Vol. 15, No. 3, 1983, pp. 215–224.
13. Hutchinson, T. P. Statistical Modeling of Injury Severity, with Special Reference to Driver and Front Seat Passenger in Single-Vehicle Crashes. *Accident Analysis and Prevention,* Vol. 18, No. 2, 1986, pp. 157–167.
14. Ouyang, Y., T. Yamamoto, and V. Shankar. Modeling the Simultaneity in Injury Causation in Multivehicle Collisions. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 143–152.
15. Yamamoto, T., and V. N. Shankar. Bivariate Ordered-Response Probit Model of Driver's and Passenger's Injury Severities in Collision with Fixed Objects. *Accident Analysis and Prevention,* Vol. 36, No. 5, 2004, pp. 869–876.
16. Jones, A. P., and S. H. Jorgensen. The Use of Multilevel Models for the Prediction of Road Accident Outcomes. *Accident Analysis and Prevention,* Vol. 35, No. 1, 2003, pp. 59–69.
17. Lenguerrand, E., J. L. Martin, and B. Laumon. Modeling the Hierarchical Structure of Road Crash Data: Application to Severity Analysis. *Accident Analysis and Prevention,* Vol. 38, No. 1, 2006, pp. 43–53.
18. Kim, D. G., Y. Lee, S. Washington, and K. Choi. Modeling Crash Outcome Probabilities at Rural Intersections: Application of Hierarchical Binomial Logistic Models. *Accident Analysis and Prevention,* Vol. 39, No. 1, 2007, pp. 1094–1100.
19. Helai, H., C. H. Chor, and M. M. Haque. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accident Analysis and Prevention,* Vol. 40, No. 1, 2008, pp. 45–54.
20. Nelson, R. B. *An Introduction to Copulas,* 2nd ed. Springer-Verlag, New York, 2006.
21. Bhat, C. R., and N. Eluru. A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling. *Transportation Research Part B,* Vol. 43, No. 7, 2009, pp. 749–765.
22. Sklar, A. Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika,* Vol. 9, 1973, pp. 449–460.
23. Spissu, E., A. R. Pinjari, R. M. Pendyala, and C. R. Bhat. A Copula-Based Joint Multinomial Discrete-Continuous Model of Vehicle Type Choice and Miles of Travel. *Transportation,* Vol. 36, No. 4, 2009, pp. 403–422.

The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.