

Article

Trivariate Copula for Modeling Barriers Crash Severity, Accounting for Policy Endogeneity

Mahdi Rezapour * and Khaled Ksaibati

Wyoming Technology Transfer Center, University of Wyoming, 1000 E. University Avenue, Department 3295, Laramie, WY 82071, USA; khaled@uwyo.edu

* Correspondence: rezapour2088@yahoo.com; Tel.: +1-(307)-766-6230; Fax: +1-(307)-766-6784

Abstract: Studies have been implemented in the literature to enhance the safety of traffic barriers by identification of the contributory factors to those crashes. However, almost all those past studies are subject to potential pitfalls of just answering the cause-and-effect question by traditional statistical methods, which fail to account for possible endogeneity. Modeling traffic barrier crash severity with traditional statistical methods might be biased as many parameters such as barrier's types is endogenous to unseen factors including policy makers' decisions in allocating those barriers. Not accounting for the possible endogeneity in the dataset and ignoring correlation between regressors' error terms might result in biased or erroneous coefficients' estimates. That is especially true in the presence of strong correlations across models' error terms. Thus, this study was conducted to model barriers' crash severity by taking into consideration the endogeneity and correlations across the models' error terms. Here, the trivariate copula-based method was implemented to simultaneously model traffic barrier crash severity, shoulder width and barrier's types, while accounting for interrelationships across the models' error terms. The results provide strong evidence of correlations between the unseen factors to the selections of barrier's types, shoulder width installation, and crash injury levels. For instance, we found in the presence of accounting for endogeneity and correlation between unseen factors, concrete traffic barrier type and higher shoulder width installation are negatively correlated with unseen factors contributing to severe barriers' crashes. That is despite the fact that the observed factors of those predictors were found to have a reverse impact on the severity of barriers' crashes.

Keywords: endogeneity; trivariate copula; logit model; barrier crash; traffic safety; subjective policy



Citation: Rezapour, M.; Ksaibati, K. Trivariate Copula for Modeling Barriers Crash Severity, Accounting for Policy Endogeneity. *Future Transp.* **2021**, *1*, 601–614. <https://doi.org/10.3390/futuretransp1030032>

Academic Editor: Dan Graham

Received: 12 August 2021

Accepted: 25 October 2021

Published: 1 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Traffic crashes have been ranked as the 7th causes of death in the U.S. [1]. This is equivalent to \$871 billion dollars based on the crash collisions costs [2]. Run-off-road (ROR) crashes account for significant proportion of those crash collision cost, especially as those crashes result in a significant proportion of severe crashes. Even though traffic barriers are one of the well-known countermeasures, which could be employed to reduce the severity of ROR crashes, the severity of barrier crashes still account for high proportion of severe crashes.

1.1. Traffic Barriers and Shoulder Width

Due to consideration of three models in this study, and those models are related to crash severity and shoulder width, and barrier types the next few paragraphs would discuss some of the related studies.

A study conducted to evaluate the severity of median barrier crash severity [3]. A nested logit model, which considers unobserved impacts among the severity levels was employed. The results indicated that a collision with cable median barriers decrease the probability of sever crashes.

The effects of various geometric characteristics of median barriers were evaluated using the random-parameter model [4]. The results indicated that concrete barriers are

more prone to the severity of barriers' crashes. On the other hand, the impacts of shoulder width on traffic safety have been evaluated extensively in the literature review. Shoulders are set with objectives of providing recovery areas, or emergency stop. Wider shoulder width could encourage drivers to have a higher operating speed, which might result in a higher crash severity [5].

However, none of the past studies took into considering the point that various barrier's types are likely to be installed based on policy makers' discretion on assigning a specific barrier to a specific location. For instance, concrete barriers are more likely to be installed by policy makers at locations with higher traffic, and with no sharp curve. On the other hand, cable barriers, due to not having enough stiffness in crashes, might be installed at locations with less possibility of traffic and crashes. In addition, despite the contribution of the past studies to the safety of traffic barriers, none of them considered the possible endogeneity of the shoulder width on the severity of traffic barriers crashes.

1.2. Bivariate Copula Method

Copula based method could be implemented in the statistical analyses to account for the possible endogeneity. The few next paragraphs would highlight few studies, which employed copula-based method in traffic safety studies.

A joint estimation of injury severity, crash type, vehicle damage and driver error were simultaneously evaluated for intersection crash data [6]. It was found that those factors are highly correlated, and a joint analysis of the variables are needed. In another study, the interaction terms of predictors and various years were considered for temporal instability evaluation. A binary logit copula-based method was used to model the severity of two vehicle crashes by considering endogeneity [7]. In that study, the collision type predictor was considered as endogenous to crash severity outcomes. The significance of copula dependence highlighted the importance of considering the endogeneity of collision type and injury severity.

A bivariate copula-based method in another study was implemented to identify the contributory factors to at-fault and not-at-fault drivers in head-on collisions [8]. A positive dependence was detected between the injury outcome of at-fault and not at fault drivers' models.

The choice of policy makers in setting up warning signs at a more hazardous locations was evaluated by the past study [9]. The results highlighted an important and significant correlation due to the choice of policy makers.

Even though no study, to the best of the knowledge of the authors of this study, has used trivariate copula on transportation problem, a study which implemented with the help of that method on other fields would be outlined in the next few paragraphs.

1.3. Trivariate Copula

A trivariate copulas method was implemented to model sample selection and treatment effects on family health care demand [10]. Three equations were considered in that study: a dichotomous choice equation for insurance status selection, and an equation for the health care use by each spouse. Significant dependences were observed across regressors and the outcome model.

1.4. Why Copula Is Needed for This Case Study

Despite much effort in application of copula-based method in traffic safety studies, a comprehensive application of copula method in a traffic barrier study is missing. This is especially important as many characteristics of barriers such as choice of selection of a specific type of barrier for a specific location might be decided by policy makers based on various criteria such as crash proneness of a road locations, or amount of traffic. For instance, concrete barriers have been installed at the locations with higher likelihood of crashes and traffic such as median barrier's locations. That is due to rigidity of concrete barriers in receiving numerous crashes.

The designs of various aspects of roadway characteristics such as shoulder width is expected to vary based on the associated conditions of the road segments. Those variables are also typically correlated with crashes necessitating the use of copulas for modeling their dependence structures. For instance, shoulder width decision criteria might be impacted by truck traffic for accommodating their needs, or due to a possible reason of a higher traffic so drivers have more time to return to the driveway in case of running off the road.

Despite the importance of implementation of the copula method for safety, especially evaluation of various aspects of roadside designs, not many studies conducted to investigate those factors while accounting for endogeneity. Thus, this study was conducted to study factors to barriers' crash severity after accounting for correlation between error terms of decisions impacting setting up specific barriers' types, or shoulder width at specific locations, and error term of crash severity model itself. Furthermore, by considering the shoulder width and barriers type as independent variables in the main model of the crash severity. The three considered barriers are shown in Figure 1 for reader that are not familiar by included types of barriers.



Figure 1. Examples of various considered barriers.

Finally, it should be noted that simulations have been employed for analyzing the road safety barriers which the paragraph would outlines few of them. The paper addressed numerical simulation of a concrete barriers [11]. The study identified the locations in concrete barriers segments. Safety performance of a precast concrete barriers with numerical method was employed [12]. To gain a precise representation of the state of the damage the stochastic damage-plasticity was adopted. The impact locations, load and boundary constraints were highlighted.

The remainder of this manuscript is organized as follows: the method section comprehensively details proposed methodology implemented in this study. The results section illustrates an implemented approach on real dataset, whereas the discussion section discusses the findings and potential direction for the future studies.

2. Method

The method section would be presented in 4 subsections. First, the general background briefly presents the general framework of copula. The second subsection, Gaussian copula mixture model, would be presented for general, and three univariate variables. The Multivariate model copula talks about how the copula method would be implemented on multivariate model and would briefly outline the process of estimating the results parameters. As endogeneity is a core of copula method, the last section would discuss that concept.

2.1. General Background

The term Copula was derived from a Latin word of copula meaning to connect or join [13]. The concept of this method lies in the question that by knowing an observation X_1 , how much information would be gained to predict an observation X_2 . This is expected due to interrelation or dependence of these two random variables. Copula method helps in the above marginal distributions and interrelationship.

The first step in achieving the above prediction process is to transform the random variables X_i into uniformly distributed random variables U_i . It should be noted that the original values of X_i could also be achieved by using inverse of cumulative density function (cdf) of F , or F^{-1} , $X = F^{-1}(U)$, where F^{-1} is called quantile transformation.

Based on Sklar theorem, for a d -dimensional cdf of F with margins of F_1, \dots, F_d , there would be a copula as:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \tag{1}$$

From above, It should be noted that while x_i can have a range of $[-\infty, \infty]$, the range of F_i is $[0, 1]$ due to the concept of cdf making it as a probability.

One of the importance features of copula-based model is the dependence consideration. Consider a bivariate copula $C(u_1, u_2)$, and a bivariate cdf of $H(x,y)$, where $H(x,y) = C(F(x), G(y))$, we would have:

$$P(U_2 < u_2 | U_1 = u_1) = \frac{P(U_2 < u_2, U_1 = u_1)}{P(U_1 = u_1)} = \frac{\frac{\partial}{\partial u_1} C(F(x), G(y)) \cdot g(y)}{g(y)} = \frac{\partial}{\partial u_1} C(F(x), G(y)) \tag{2}$$

where U_1 and U_2 are two uniform random variables, U_1 is known, and the objective is to find a conditional distribution of the unknown or U_2 .

2.2. Gaussian Copula Mixture Model

It has been argued that Archimedean copulas, such as Frank or Gumble lack the flexibility to accurately model the dependence across large number of variables [14]. In addition, there are associated limitations for extending Archimedean copulas to trivariate or higher orders. One of the main difficulties is related to the fact that the lower-level dependence cannot be preserved due to a reason that the copula cannot provide a unique distribution when the sequence of variables are altered [15].

However, multivariate Gaussian distribution could be implemented without suffering from the aforementioned limitations [16]. That is why we only considered Gaussian copula due to a possible limitation of other copulas in accommodating the trivariate types of copulas.

Every copula could be written based on its variance-covariance matrix. For instance, consider a two-dimensional Gaussian Copula with interrelationship of ρ and correlation matrix ∇ :

$$C_{\rho}^{Gauss}(u_1, u_2) = \Phi_{\nabla}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \tag{3}$$

where Φ^{-1} is the inverse cumulative distribution function, and Φ_{∇} is the multivariate standard normal distribution function. The marginal distributions of u_1 and u_2 are restricted to be Gaussian.

The correlation matrix of ∇ is a resultant of a covariance matrix as:

$$\tilde{\nabla} = \begin{pmatrix} s_1^2 & \rho s_1 s_2 \\ \rho s_1 s_2 & s_2^2 \end{pmatrix} \tag{4}$$

Based on the above formula the representation of Gaussian copula could be written as:

$$F(u_1, u_2 | \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)}\right) ds_1 ds_2 \tag{5}$$

For $\rho = 0$, $\rho = 1$, and $\rho = -1$ the copula would be converted into (a) independent, (b) comonotonicity, and (c) countermonotonicity copulas, respectively [17]. In summary, it can be said that the above equation is equal to multivariate standard normal distribution with a correlation matrix of ρ .

It is worth discussion of the simulation process of Gauss copula, where all margins are univariate Gaussian, which could be summarized as follows:

- Cholesky decomposition of the covariance matrix, and set A for the lower triangular Metrix
- Consider the iteration for the following steps
 - Generate a vector of z from a random normal distribution where $z = z_1, \dots, z_d$
 - Now set $x = A^*z$
 - Return U as $(\Phi(x_1), \dots, \Phi(x_1))$

Now consider three-dimensional joint of $F \sim (X_1, X_2, X_3)$, we know:

$$f(x_1, x_2, x_3) = f_1(x_1)f(x_2|x_1)f(x_3|x_1, x_2) \tag{6}$$

where f_n is the probability density function (pdf).

From the above equation, $f(x_2|x_1)$ and $f(x_3|x_1, x_2)$ are parameters that need to be estimated. Based on Sklar’s theorem, $f(x_3|x_1, x_2)$ could be written as:

$$f(x_3|x_1, x_2) = \frac{f(x_2, x_3|x_1)}{f(x_2|x_1)} = \frac{c_{2,3|1}(F(x_2|x_1), F(x_3|x_1))f((x_2|x_1)f(x_3|x_1))}{f(x_2|x_1)} = c_{2,3|1}(F(x_2|x_1), F(x_3|x_1))c_{1,3}(F_1(x_1)|F_3(x_3))f_3(x_3) \tag{7}$$

On the other hand, from a binary dimension, $f(x_2|x_1)$ could be written as:

$$f(x_2|x_1) = \frac{c(x_1, x_2)}{f_1(x_1)} = c_{1,2}(F_1(x_1), F_2(x_2))f_2(x_2) \tag{8}$$

Now based on Equations (6)–(8) we would have:

$$f(x_1, x_2, x_3) = f_1(x_1)c_{1,2}(F_1(x_1), F_2(x_2))f_2(x_2)c_{2,3|1}(F(x_2|x_1), F(x_3|x_1))c_{1,3}(F_1(x_1)|F_3(x_3))f_3(x_3) \tag{9}$$

It should be noted that $f_1(x_1)$, $f_2(x_2)$, and $f_3(x_3)$ are univariate pdf’s.

Now the $C_{1,2}, C_{2,3|1}, C_{1,3}$ could be written as the pair copulas of $c_{1,2}, c_{2,3|1}, c_{1,3}$ [14,18]. In addition, thetas, dependence, are incorporated in various copulas. For instance, θ_{12} would be incorporated in c_{12} .

2.3. Multivariate Model Copula

For a three multivariate model, the copula would be implemented on error terms instead of the univariate variable itself. It should be noted that the Gaussian copula resulted from multivariate normal distribution. For instance, consider a trivariate model as follows:

$$\begin{aligned} y_1^* &= \gamma z + \varepsilon_1 \\ y_2^* &= \beta x + \varepsilon_2 \\ y_3^* &= \varphi g + \alpha y_2 + \varepsilon_3 \end{aligned} \tag{10}$$

where y_1^* and y_2^* , and y_3^* are three latent equations γ , β , and φ are coefficients related to three modeling components. From the above equations, the resultant distributions would be fully known if the joint distribution of ε_1 and ε_2 , and ε_3 are known [19]. For trivariate model, the joint distribution of the three models could be written as:

$$F(\varepsilon_1, \varepsilon_2, \varepsilon_3) = \Phi(\varepsilon_1, \varepsilon_2, \varepsilon_3, \theta_{12}, \theta_{23}, \theta_{13}) = \Phi \left[\Phi^{-1}(\Phi(\varepsilon_1)), \Phi^{-1}(\Phi(\varepsilon_2)), \Phi^{-1}(\Phi(\varepsilon_3)) \right] \tag{11}$$

The current study considered trivariate copulas for three considered equations. In this study, the logit link function was implemented for the marginal distribution of the three models. Here, the difference between simple logistic regression model and copula based logistic model is that copula model accounts for dependence between the residuals of the marginal models while simple logistic regression does not. In other words, the multivariate models’ copula could be seen as a joint distribution of residuals, while the errors follow parameter θ_i [20]. The correlations across the error terms might capture the presence of unobservable factors impacting various decision-makers’ processes.

Parameter estimates of logistic regression could be computed as follows. First, we know the likelihood function of logistic regression, which could be written for binomial mass function as:

$$l(\beta|Y) = \prod_{i=1}^n \binom{n_i}{y_i} \pi_i^{y_i} \frac{1 - \pi_i^{n_i}}{1 - \pi_i^{y_i}} \tag{12}$$

For n_i trials, there are $\binom{n_i}{y_i}$ ways to arrange y_i successes of those n_i trials [21]. Furthermore, the probability of y_i success is $\pi_i^{y_i}$, and the probability of $n_i - y_i$ failure is $1 - \pi_i^{n_i - y_i}$.

It should be noted that Y is a linear combination of various coefficients. Now maximum likelihood could be implemented to find β s which maximize the above equation. The values could be identified when the first derivative of above is set to be zero, and the second derivative of above is less than zero. Now based on log odds formula we have:

$$\frac{\pi_i}{1 - \pi_i} = e^{\sum_{k=0}^K x_{ik}\beta_k} \tag{13}$$

In addition, for π_i we have:

$$\pi_i = \frac{e^{\sum_{k=0}^K x_{ik}\beta_k}}{1 + e^{\sum_{k=0}^K x_{ik}\beta_k}} \tag{14}$$

Now substituting Equation (14) in Equation (12) we would have:

$$\prod_{i=1}^N e^{\sum_{k=0}^K x_{ik}\beta_k y_i} \left(1 - \frac{e^{\sum_{k=0}^K x_{ik}\beta_k y_i}}{1 + e^{\sum_{k=0}^K x_{ik}\beta_k y_i}}\right)^{n_i} \tag{15}$$

After some algebra, the above equation would turn into the Kernel of likelihood function to be maximized. After taking the log of the kernel equation we would have:

$$l(\beta) = \sum_{i=1}^N y_i \left(\sum_{k=0}^K x_{ik}\beta_k\right) - n_i \cdot \log(1 + e^{\sum_{k=0}^K x_{ik}\beta_k}) \tag{16}$$

Now equating the derivative of above to zero for β_k , and solving based on Newton-Raphson method would solve for betas values without a need for second derivative of the log likelihood function.

In summary, Maximum likelihood is implemented to make the observed data most probable, where θ is a density function which would be maximized:

$$L(\theta|x) = f(x|\theta) \tag{17}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} L(x|\theta) \tag{18}$$

The variance-covariance matrix would be estimated by the Hessian matrix as follows:

$$\operatorname{var} - \operatorname{covar}(\beta) = (-E[H(\beta)])^{-1} = -E\left[\frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta^2}\right] \tag{19}$$

The above likelihood estimates with some minor modifications. using penalized additive method, could be implemented for copula method as follows [22,23]:

$$l(\delta) = \sum_{i=1}^n \log\{c(F_1(y_{1i}|\mu_{1i}, s_{1i}), F_2(y_{2i}|\mu_{2i}, s_{2i}), F_3(y_{3i}|\mu_{3i}, s_{3i}); \theta_i)\} + \sum_{i=1}^n \sum_{m=1}^n \log\{f_m(y_{mi}|\mu_{3i}, s_{3i})\} \tag{20}$$

where μ and s are the location and scale parameter of logistic models, m are marginal distribution parameters, θ is dependence coefficient, ζ is degree of freedom. The marginal distributions of Y s could be specified by cdf's and densities as F and f , respectively. In

addition, δ consists of coefficient vectors of β s. Furthermore, the copula density c could be written as:

$$c = \frac{\partial^2 C(F_1(y_{1i}), F_2(y_{2i}))}{\partial F_1(y_{1i}) \partial F_2(y_{2i})} \tag{21}$$

Now consider a conditional distribution of (Y_1, Y_2) based on our case of the binary logistic regression for marginal distribution for two scenarios would be based on Table 1 [24] as:

Table 1. Conditional distributions of binary and trivariate copulas.

	$Y_2 = 0$	$Y_2 = 1$	$Y_2 = 0, Y_3 = 0$
$Y_1 = 0$	$C(\pi_1, \pi_2)$	$\pi_1 - C(\pi_1, \pi_2)$	—
$Y_1 = 1$	—	—	$C(\pi_1, \pi_2, \pi_3)$

2.4. Endogeneity

The role of copulas is to account for dependence across residuals of the marginal models [24]. The endogeneity resulted from correlation across error terms reflected from unobserved factors impacting two endogenous models, which could not be measured, and thus would be set as error terms.

In this study, the endogeneity was considered as we are interested in estimating the impacts of various predictors on the severity of traffic barrier crashes in the presence of unobserved confounders. This problems in economics studies arise when important covariates have been omitted, and as a result become part of error terms [20]. After setting the treatment by considering the endogeneity in the model, if the correlation across the treatment and the outcome error terms were found to be not important, the treatment would be turned to be exogenous in the main model, or crash severity. Various techniques could be implemented to account for endogeneity effect. For instance, latent instrumental variable (LIV) could be implemented by utilizing a discrete latent variable to account for dependencies between regressors and error terms.

The trivariate copula-based model in this study controls for unobserved confounders by three equations: the main equation describing the binary outcome in terms of the two treatments, where the other two treatments model the contributory factors to the treatments. On the other hand, the error terms describe the omitted coefficients being not considered in the model.

On the other hand, endogeneity resulted from incorporating the dependent predictor as an independent predictor in the main equation. In this study, we considered barrier types and shoulder width, which were dependent variables in other models, as independent variables in the crash severity model. To have a vision regarding the process Figure 2 is provided, which highlight while the standard copula account for correlation across the error terms, we further extend the models to account for endogeneity by considering the barrier types and should width as explanatory variables in the main model of crash severity.

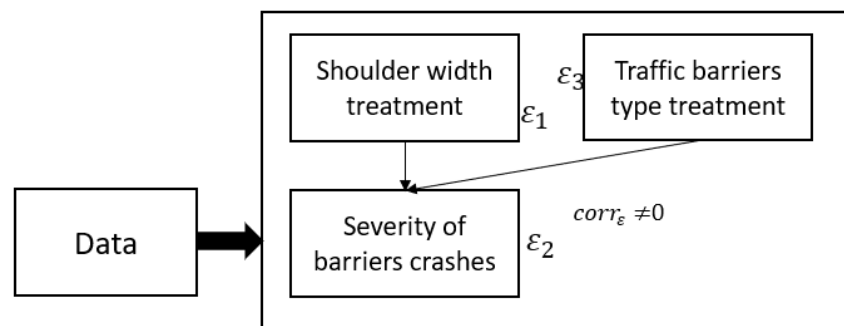


Figure 2. Depiction of consideration of correlation across error terms and endogeneity of the models.

3. Study Continuations

Finally, the study contribution could be summarized as follows:

1. It is expected that the choice of various traffic safety barriers to be confounded by various policy-making decisions, so here to account for that unseen factor, we accounted for the correlation across the error terms by copula method.
2. It is also important to account for the endogeneity of various models. For instance, while copula model could account for the error terms correlations across the models, considering barrier types and should width, those factors are also expected to be important predictors for the severity of barriers crashes. So, we accounted for the endogeneity by incorporating those factors as explanatory variables in the main model of barriers crash severity.
3. Accounting for the aforementioned points is especially important while studying traffic barriers crashes due to their confounding effects of policy makers in assigning various barriers types to specific locations.
4. This one of the earliest study implemeneted the copula technique to account for policy makers subjectivity in making an especific policy.

4. Data

The dataset used in this study was obtained from Wyoming department of transportation (WYDOT) through the critical analysis reporting environment (CARE). The data was filtered to include only crashes occurred due to hitting barriers as their first harmful events. Due to variation of factors to crashes in multivehicle crashes, also only single vehicle hitting a barrier considered in this study. Due to low number of cable barriers and significant variation in the designs of temporary barriers, those barriers were excluded from this study. The data was also filtered to include only observation in the state's interstate system.

The traffic barrier geometric characteristics were collected from 4176 miles of roadway in Wyoming, including various characteristics such as barrier heights, shoulder width, type of barriers, and offset. Offset is defined as the outer edge of shoulder width to the side of a barrier. The Wyoming roads map is presented in Figure 3 to provide the summary of the road sketch in the state.



Figure 3. Wyoming road maps.

Originally, there were 3 datasets including: crash, traffic, and barriers geometric characteristic. The crash dataset was filtered to include only those crashes occurred between 2007 and 2016. More recent data was not incorporated due to lack of availability at the time of analysis. The traffic, and crash dataset were aggregated to the barriers' geometric characteristics. The aggregation was implemented by matching highway system, milepost, and direction of travel. The averages of various explanatory variables were used for aggregating the crash to the barrier dataset while the total of various crash severity was considered for aggregation. For instance, the average of having no improper action in the crashes is 0.131 (see Table 2), that means that the majority of the drivers hitting barriers had some form of improper driving as this binary predictor is closer to the reference category of improper driving as zero.

Table 2. Summary statistics of important trivariate copula-based model.

Attributes	Average	Variance	Min	Max
Crash severity injury (vs. PDO *)	0.26	0.194	0	1
Shoulder width, ≤12 * (vs others)	0.02	0.022	0	1
Barrier types, box/W beam * (vs concrete barriers)	0.05	0.046	0	1
Posted speed limit, continuous	71.89	50.921	15	75
AADT, continuous	797	165,023	128	2227
Barrier length, continuous, ft	678	1,393,700	19.7	10,280
Barrier height, continuous, inches	30	8.466	1.200	56.39
Barrier offset, continuous, inches	5.72	10.649	0	28
Negotiating a curve * (vs. others)	0.08	0.072	0	1
Normal driver emotional condition * (vs. others)	0.25	0.124	0	1
Improper drivers' actions * (vs others)	0.13 0.131	0.078	0	1
Wyoming residence * (vs. others)	0.87	0.168	0	1
Alcohol was involved (vs. others) *	0.04	0.025	0	1

* reference category.

Due to similarities of box beam and W-beam, compared to concrete barriers, those two types were combined and set as reference as 0. The average of this predictor highlight that the majority of barriers were non-concrete barriers, mean = 0.048.

The model consists of three equations: barrier crash severity's equation for identification of contributory factors to barriers severity, and two variables being endogenous to the traffic barrier crash severity: shoulder width and barrier types. A barrier is considered as hazardous if during the analysis period, it has experienced at least a fatal or severe crash. Shoulder width is converted into two categories based on the cutting point of 12 feet; the federal highway administration (FHWA) recommends a minimum shoulder width of 10 feet for the interstate system. However, a shoulder width of 12 feet is recommended when truck traffic exceeds 250 design hourly volume (DHV) (the design hourly volume for one direction) [25]. Barrier type is set to be binary, reference category as 0, when the barriers were box beam or W-beam, and 1 when the barriers were concrete.

5. Results

The results of the trivariate copula would be presented in 4 subsections. The first two subsections would present the results of two endogenous binary variables of shoulder width and barrier types. Then it would detail the results of the main model of barrier crash severity. At the end, it would go over the results of copula dependence of the three considered models.

5.1. Endogenous Factor of Shoulder Width

The results of this part of the model highlight a policy makers' decisions for setting a wider shoulder width of 12 feet in the state. The results highlight that it is more likely to have a wider shoulder width when truck traffic and barrier length increases, $\hat{\beta}_{truck\ traffic} = 0.001$, $\hat{\beta}_{barrier\ length} = 0.0004$. For locations with higher truck traffic, it is intuitive to accommodate the truck drivers with wider shoulder width for various purpose such as pulling over to the shoulder width for mechanical issues, recovery, or resting. The impact of barrier length also could be dependent on other unseen factors that cause policy makers to make a decision about shoulder width.

5.2. Endogenous Factor of Barrier Types

For this model, a binary response was created by considering W-beam and box-beam as reference category, compared to concrete barriers. The aggregation across two barriers for a reference category was due to the similarity across those two barriers. As expected, barrier heights, $\hat{\beta} = 0.27$ and offset, $\hat{\beta} = -0.395$ are two variables that varied across these two predictors based on the policy design of these two barriers. The results are presented in Table 3.

Table 3. Results of trivariate copula-based method.

Parameters	Estimates	Standard Error	p-Value
1st copula: shoulder width, treatment			
Intercept	-2.88	1.479	0.051
Posted speed limit	-0.03	0.019	0.090
Truck traffic	0.001	0.0005	0.043
Barrier length	0.0004	6.42×10^{-5}	<0.005
2nd copula: Barrier types, treatment			
Intercept	-5.65	1.779	0.001
Barrier height	0.33	0.051	<0.005
Barrier offset	-0.40	0.074	<0.005
Negotiating a curve	11.82	4.465	0.008
Posted speed limit	-0.09	0.016	<0.005
Negotiating a curve \times Posted speed limit	-0.16	0.067	0.017
3rd copula: Crash severity, outcome			
intercept	0.43	0.697	0.54
Posted speed limit	-0.02	0.010	0.04
No improper action	-0.72	0.312	0.02
Driver emotional condition	0.58	0.209	0.01
Residency	-0.54	0.177	0.00
Alcohol involvement	1.14	0.434	0.01
Barrier types	1.45	0.520	0.01
Shoulder width	2.33	0.985	0.02
$\theta_{12} = 0.586(0.241, 0.783), \theta_{13} = -0.844(-0.977, -0.233), \theta_{23} = -0.403(-0.708, -0.0947)$			

For instance, concrete barriers have a higher height than other barriers, which could be seen by direction of this coefficients, $\hat{\beta} = 0.27$. Or it is more likely to assign a higher offset for w-beam compared with concrete barriers. The interaction terms across important predictors were considered as well: when there is a possibility of negotiating a curve at a location with a higher posted speed limit, it is less likely to have a concrete barrier. This is likely due to the nature of concrete barrier that could not be installed at curves segments.

5.3. Injury Severity Model (Main Model)

In addition to incorporating important variables such as posted speed limit or improper restrain, the two other variables discussed above were used as endogenous variables in this model. The results indicated that an increase in posted speed limit, $\hat{\beta} = -0.02$, results in a decrease in the severity of barriers' crashes. Even though the impact seems to be counterintuitive, the negative impact of posted speed limit on the severity of barrier crashes might resulted from the fact that higher posted speed limit is at locations with less critical geometric characteristics resulting in a reduction of the severity of barriers' crash severity.

It was found that having no improper driving action, $\hat{\beta} = -0.72$, would result in a reduction of the severity of barrier crashes compared with the time that drivers showed some signs of improper driving action such as speeding or driving under influence (DUI). Numerous studies have been conducted on various aspects of the impacts of speeding or DUI on the severity of crashes.

Moving to the variable of drivers' emotional condition. For this variable, various emotional drivers' conditions such as being angry, sad, or irritable at the time of crash were compared with the time when drivers were observed to be under normal condition. As expected, being under non normal emotional condition, $\hat{\beta} = 0.58$, increases the likelihood of severe barriers crashes. The impact highlights the importance of emotional condition on the severity of crashes. This impact could be attributed to the fact that drivers under normal conditions, in terms of emotional conditions, assess physical and emotional traffic hazards accurately resulting in a decrease in crash risk [26]. The results of residency indicated that residence drivers of Wyoming, $\hat{\beta} = -0.54$ are more prone to the severity of barriers crashes. The impact might result from the confidence that resident drivers have while driving on Wyoming roadways. This might be due to distinguished fear to possible hazard on the roads for confident resident drivers, believing that they are not at risk [27].

Another factor found to be important is alcohol involvement, $\hat{\beta} = 1.14$. The impact is intuitive, and many studies have been conducted on the importance of the effect of DUI on the severity of crashes. There is worldwide evidence of higher risks of driver under the influence in crashes and crash severities. Alcohol usage significantly impact the driving skills [28].

Even though the impacts of the other two models were evaluated through the correlation across the error terms, the factors of barrier types and shoulder width considered in the third model as exogenous variables. Barrier type variable highlights the impact of the concrete barriers $\hat{\beta} = 1.45$ increase the severity of crashes compared with other barriers. The impacts of the barriers are intuitive and related to the capacity of energy absorption of W-beam barriers compared with concrete barriers, lacking flexibility and being rigid.

Moving to the factor of shoulder width $\hat{\beta} = 2.33$. The results indicated that a wider shoulder width of 12 feet increase the severity of barrier crashes. The impact was attributed to the fact that vehicles going off the roadway need to be stopped by barriers as soon as possible, and before going over a wider shoulder width [29].

5.4. Copula Dependence Parameters

The copula dependence parameters, θ , highlights the presence of unobservable factors. If the unobservable factors impacting the outcomes were uncorrelated, having all the models in the framework simultaneously would be unnecessary. The copula could account for dependence between the two treatments and the outcomes. The large and significant dependence parameters across the models highlight that there are strong negative association between the unstructured terms of the model parameters (see the bottom of Table 3).

Ignoring the dependencies might result in erroneous parameters' estimates. Accounting for the correlated models is important as it could be hypothesized that concrete barriers are installed at the locations with higher crash frequency or higher crash severity such as median barriers. On the other hand, wider shoulder width is installed at the locations where wider shoulder width was necessitated by policy makers.

Even though all the studies in the literature, along with this study highlight the impact of concrete barriers on the severity of crashes, it is worth investigating the relationship between unseen factors across these two models. The results of copula dependency factor indicated that unseen factors that install concrete barrier at specific locations, at the same time they decrease the severity of barrier crashes. This result is contrary to the impact of seen factors of barrier types on barrier crash severity. This could be related to the decision being made by the WYDOT on installation of concrete barriers at right locations, which consequently reduce the severity of barriers crashes.

Moving to the relationship between error terms of shoulder width model and barriers' crash severity, $\theta_{13} = -0.844$. The results indicated that while the main effect of wider shoulder width increase the likelihood of severe barriers' crashes, there is a negative correlation across wider shoulder width and barriers crash severity error terms. In other words, the unseen factors might be related to policy makers that increase the width of shoulder at the same time decrease the severity of barriers' crashes.

Moving to the last copula dependence between shoulder width and barrier types, $\theta_{12} = 0.586$. The results indicated that unseen factors that install concrete barrier, they also increase the shoulder width to their maximum values. This highlights the designs of concrete barriers associated with higher shoulder width. It should be emphasized that the unseen factors should have been part of the models' predictors but due that fact that they have been omitted, they have become part of the model's error terms

6. Conclusions

Hitting a fixed objects crashes have been considered as one of the leading causes of roads' fatalities around the world. Traffic barriers have been installed with an objective of mitigating the severity of those crashes. However, traffic barriers still account for a significant proportion of severe crashes. Extensive effort has been made in identification of contributory factors to those types of crashes with the hope that the severity of barriers' crashes would be reduced. However, despite many efforts being made regarding identification, and studying the factors to barriers crashes, none of the past studies considered the error terms' correlations across endogenous variables and barrier crash severity.

Not addressing the methods' shortcoming, has been justified by arguing that those unseen factors cannot be measured, or as those unobserved factors would be stored in error terms, no significant issues would be raised. However, there is a considerable chance that the error terms of those endogenous variables, and outcome models are correlated. Not accounting for those unseen factors through considering error terms' correlations might result in bias or erroneous modeling results.

That is especially important for traffic barrier analysis as it is expected that there are unseen factors that impact the decision of various treatments, and thus those unseen factors are expected to end up in error terms. It is expected that a significant portion of error term resulted from the policy makers in the state in designing various aspects of roadway safety features such as allocating specific barrier types or shoulder width are based on traffic or other safety criteria.

Copulas are general tools for describing dependence structure by considering correlations across separate incorporated equations. In this study, a unified likelihood function framework was used to join different marginal models. The method accounts for pairwise dependence across models' residuals, given the explanatory variables. Three joint models were considered and estimated using copula-based method.

Two treatment models along with an outcome model were considered in this study. Here the selection of various barriers or various shoulder width settings, as treatments, are biased toward the safety of roadways or various policy makers' objectives. This selection bias which has been often ignored, cannot be measured by traditional statistical method and termed as endogenous selection bias. Ignoring the selection bias could result in erroneous results, especially when high correlations exist.

While Archimedean copulas have been extensively employed on bivariate problems, there are some limitations when it comes to their applications for higher dimensions copulas, e.g., trivariate copula. Thus, in this study Gaussian copula were implemented to accommodate trivariate copula methods.

The Gaussian copula was selected over Archimedean copulas due to the fact that the Archimedean copulas suffer from inflexible structure in higher dimensions. The Gaussian copula implemented in this study to conveniently express the endogeneity in the form of correlation matrix of a multivariate Gaussian distribution [30].

While similar to the previous studies, the results highlight, for instance, concrete barriers, or wider shoulder width increase the odds of severe barrier crashes, the results of copula dependencies indicated that there are negative, large, and significant correlation across the error terms of models. This justifies the policy makers' decision in installation of concrete barriers at specific locations or setting up a wider shoulder width in reduction of the barriers' crash severity.

In other words, the WYDOT decisions on barrier type selection or decision on the width of shoulder width are in line with the safety of Wyoming interstate system. Accounting for correlations of error terms is important in this study especially because there were strong and significant correlations between regressors, and outcome models' error terms.

More studies are needed to take into account endogeneity issue while analyzing the safety aspects of various treatments. That is because it is expected those choices to be influenced by the safety of those specific locations or limitation of installation of different types of safety measures for a unique location. Our results provide the WYDOT with the assurance that despite the highlighted contributory factors of concrete barriers on higher severity of barriers, their choice of installation of concrete barriers negatively correlate with severity of barriers crashes. It would be interesting to have more available inputs regarding the choices for the selection of various barrier types or shoulder width. In that way, this study would have been able to confirm the findings with the real justification for making a decision to set up various barrier types or shoulder width. Future studies are encouraged to consider and evaluate those factors in their analyses.

Author Contributions: Conceptualization, M.R.; Formal analysis, M.R.; Investigation, K.K. All authors have read and agreed to the published version of the manuscript.

Funding: Funding is supported by the WYDOT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: There is no conflict of interest across the authors.

References

1. National Highway Traffic Safety Administration. *Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2012–2014*; US Department of Transportation: Washington, DC, USA, 2016.
2. Blincoe, L.; Miller, T.R.; Zaloshnja, E.; Lawrence, B.A. *The Economic and Societal Impact of Motor Vehicle Crashes*; National Highway Traffic Safety Administration: Washington, DC, USA, 2010.
3. Hu, W.; Donnell, E.T. Median barrier crash severity: Some new insights. *Accid. Anal. Prev.* **2010**, *42*, 1697–1704. [[CrossRef](#)] [[PubMed](#)]
4. Rezapour, M.; Wulff, S.S.; Ksaibati, K. Examination of the severity of two-lane highway traffic barrier crashes using the mixed logit model. *J. Saf. Res.* **2019**, *70*, 223–232. [[CrossRef](#)]
5. Stamatiadis, N.; National Research Council. *Impact of Shoulder Width and Median Width on Safety*; The National Academies Press: Washington, DC, USA, 2009; Volume 633.
6. Wang, K.; Bhowmik, T.; Yasmin, S.; Zhao, S.; Eluru, N.; Jackson, E. Multivariate copula temporal modeling of intersection crash consequence metrics: A joint estimation of injury severity, crash type, vehicle damage and driver error. *Accid. Anal. Prev.* **2019**, *125*, 188–197. [[CrossRef](#)] [[PubMed](#)]
7. Rana, T.A.; Sikder, S.; Pinjari, A.R. Copula-Based Method for Addressing Endogeneity in Models of Severity of Traffic Crash Injuries: Application to Two-Vehicle Crashes. *Transp. Res. Rec. J. Transp. Res. Board* **2010**, *2147*, 75–87. [[CrossRef](#)]
8. Wali, B.; Khattak, A.J.; Xu, J. Contributory fault and level of personal injury to drivers involved in head-on collisions: Application of copula-based bivariate ordinal models. *Accid. Anal. Prev.* **2018**, *110*, 101–114. [[CrossRef](#)]
9. Rezapour, M.; Ksaibati, K. Evaluation of traffic warning signs on truck safety considering endogeneity, a copula-based method. *J. Transp. Saf. Secur.* **2020**, *12*, 1–14. [[CrossRef](#)]
10. Zimmer, D.M.; Trivedi, P.K. Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *J. Bus. Econ. Stat.* **2006**, *24*, 63–76. [[CrossRef](#)]
11. Pachocki, L.; Bruski, D. Modeling, simulation, and validation of a TB41 crash test of the H2/W5/B concrete vehicle restraint system. *Arch. Civ. Mech. Eng.* **2020**, *20*, 1–23. [[CrossRef](#)]
12. Li, Z.; Gao, X.; Tang, Z. Safety Performance of a Precast Concrete Barrier: Numerical Study. *Comput. Model. Eng. Sci.* **2020**, *123*, 1105–1129. [[CrossRef](#)]
13. Sklar, M. Fonctions de répartition à dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **1959**, *8*, 229–231.
14. Brechmann, E.; Schepsmeier, U. Cvine: Modeling dependence with c- and d-vine copulas in R. *J. Stat. Softw.* **2013**, *52*, 1–27. [[CrossRef](#)]
15. Kao, S.-C.; Govindaraju, R. Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resour. Res.* **2008**, *44*, W02415. [[CrossRef](#)]

16. Renard, B.; Lang, M. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Adv. Water Resour.* **2007**, *30*, 897–912. [[CrossRef](#)]
17. Schmidt, T. Coping with copulas. In *Copulas-From Theory to Application in Finance*; Riskbooks: London, UK, 2006; pp. 3–34.
18. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2007**, *44*, 182–198. [[CrossRef](#)]
19. Winkelmann, R. Copula bivariate probit models: With an application to medical expenditures. *Health Econ.* **2011**, *21*, 1444–1455. [[CrossRef](#)] [[PubMed](#)]
20. Marra, G.; Radice, R. A joint regression modeling framework for analyzing bivariate binary data in R. *Depend. Model.* **2017**, *5*, 268–294. [[CrossRef](#)]
21. Czepiel, S.A. Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. pp. 1825252548.1564645290–1331571101.1559893676. Available online: Czep.net/stat/mlelr.pdf (accessed on 27 September 2021).
22. Vatter, T.; Nagler, T. Generalized Additive Models for Pair-Copula Constructions. *J. Comput. Graph. Stat.* **2018**, *27*, 715–727. [[CrossRef](#)]
23. Marra, G.; Radice, R. Bivariate copula additive models for location, scale and shape. *Comput. Stat. Data Anal.* **2017**, *112*, 99–113. [[CrossRef](#)]
24. Genest, C.; Nikoloulopoulos, A.K.; Rivest, L.-P.; Fortin, M. Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. *Braz. J. Probab. Stat.* **2013**, *27*, 265–284. [[CrossRef](#)]
25. Stein, W.J.; Neuman, T.R. Mitigation Strategies for Design Exceptions. Available online: <https://trid.trb.org/view.aspx?id=883633> (accessed on 27 September 2021).
26. Legree, P.J.; Heffner, T.S.; Psotka, J.; Martin, D.E.; Medsker, G.J. Traffic crash involvement: Experiential driving knowledge and stressful contextual antecedents. *J. Appl. Psychol.* **2003**, *88*, 15–26. [[CrossRef](#)] [[PubMed](#)]
27. Job, R. The application of learning theory to driving confidence: The effect of age and the impact of random breath testing. *Accid. Anal. Prev.* **1990**, *22*, 97–107. [[CrossRef](#)]
28. Domingues, S.C.A.; Mendonça, J.B.; Laranjeira, R.; Nakamura-Palacios, E.M. Drinking and driving: A decrease in executive frontal functions in young drivers with high blood alcohol concentration. *Alcohol* **2009**, *43*, 657–664. [[CrossRef](#)] [[PubMed](#)]
29. Rezapour, M.; Wulff, S.S.; Molan, A.M.; Ksaibati, K. Application of Bayesian ordinal logistic model for identification of factors to traffic barrier crashes: Considering roadway classification. *Transp. Lett.* **2020**, *13*, 308–314. [[CrossRef](#)]
30. Masarotto, G.; Varin, C. Gaussian copula marginal regression. *Electron. J. Stat.* **2012**, *6*, 1517–1549. [[CrossRef](#)]