A New Flexible Skewed Bimodal Distribution with Multivariate Extensions: Theory and Application to Traffic Crash Injury Severity Analysis

Chandra R. Bhat
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
Tel: +1-512-471-4535; Email: bhat@mail.utexas.edu

ABSTRACT
This paper introduces a proper multivariate flexible continuous parametric distribution, a first to our knowledge, that allows for asymmetric bimodality in each univariate dimension. The distribution is developed through a combination of an approach to generate bimodality and a Yeo-Johnson (YJ)-based transformation. A number of properties of the proposed distribution are stated and proved, including a computationally easy way to generate random variates from the proposed multivariate density. An application of the proposed distribution is demonstrated to analyze injury severity using data drawn from the Texas Department of Transportation (TxDOT) crash database of two-vehicle crashes at intersections. The proposed distribution may be applied to a number of different econometric modeling contexts in both a univariate and multivariate context, and in a whole variety of fields to consider bimodal asymmetry in stochastic distributions.

Keywords: YJ transformation, Bimodality, Non-normality, Multivariate skewed distributions, Injury severity analysis.

1. INTRODUCTION

Econometric models recognize the presence of unobserved factors in the choice outcome process, and thus assume one or more forms of stochasticity in the decision mechanism leading up to the observed choice (by choice, we are not confining ourselves to a discrete choice, but more broadly to any outcome type, including continuous, binary, ordered, multinomial unordered, count, grouped, and other limited-dependent outcomes). In the most basic form, the stochasticity is traditionally included as an additive error term in an equation associated with an outcome-generating endogenous variable that is then mapped to the actual observed outcome. In the case of a continuous outcome, the outcome-generating variable corresponds directly to the observed outcome. In the case of binary or ordered or grouped or count or limited-dependent outcomes, a single latent (endogenous) index variable is horizontally partitioned to map to the observed outcome. In the case of multinomial unordered outcomes, most commonly a latent (endogenous) index variable specific to each alternative (also referred to as an alternative-specific utility function) is defined with its own additive error term, and the utility functions across alternatives are then mapped to the observed outcome in the form of a maximum utility mechanism.

In the past two decades or so, there have been several efforts (in the context of the full range of different types of observed choice outcomes listed above) to extend the single additive error structure to recognize unobserved heterogeneity across individuals in the responsiveness to specific exogenous variables (sometimes referred to as response heterogeneity). Such additional stochastic specifications are also identified as "random coefficients on exogenous variables" in the literature. For example, the sensitivity to travel time in a transportation mode choice context may vary across individuals due to such unobserved personality factors as how "chill" an individual is or how time-conscious an individual is (see, for example, Bhat, 2020 for a relatively recent review of such random coefficient studies in mode choice). Or as another example of unobserved variation, the impact of the type of crash on injury severity may vary based on such unknown factors (at least as collected in most crash databases) as the precise evasive maneuvers taken by motorists just before the point of impact and variations in the structural integrity of the vehicle(s) involved in a crash (see Mannering et al., 2016 for a review of such random coefficient studies in the safety analysis). In most such random coefficient studies, the typical distribution assumed for response heterogeneity is a normal distribution, driven by the convenience offered by the distribution for multivariate extensions to allow unobserved correlations across the random coefficients as well as the closure property of the normal distribution under affine transformations (which, when combined with a normally distributed kernel error term collapses back to an overall normal distribution).

There has also been growing recognition in recent years that the normally distributed assumption for the kernel error term or for the random coefficients on exogenous variables may not be symmetric or even unimodal. Even so, almost all such studies retain a symmetric unimodal shape for the kernel error term, but allow more flexible non-symmetric (mixing) and/or multimodal distributions for the random coefficients on exogenous variables (an issue we will get back to later). These flexible multivariate distributions generally take one of three forms (1) a discretevalued random variable vector, (2) a continuous parametric random variable vector, or (3) a combination of the two. Of these, the first corresponds to a non-parametric approach, the second to a parametric approach, and the third, as typically implemented in practice, to a semiparametric approach. The first approach of specifying a non-parametric (discrete) series-based or similar approximation to the density function provides for substantial flexibility, but brings with it parameter profligateness and computational complexity/inference challenges (we forgo full details of the three approaches here, and refer the reader to Bhat et al., 2025). The second approach can be quite restrictive, especially given that the popular implementation of this approach uses unimodal distributions such as the skew-normal distribution. However, it does provide for parameter parsimony and computational ease in estimation. The third approach is almost exclusively specified in the literature as a finite mixture of parametric distributions, which itself can be interpreted as a latent class model (consumers belonging to each of a finite number of segments) with a parametric distribution within each class, or as a direct non-latent class semiparametric assumption about the random coefficients at hand. As the finiteness of the mixture grows, this third approach can mimic literally any multivariate density function, including those with multiple modes as well as the fully nonparametric distribution. From a structural perspective, though, this approach, when it accommodates multimodality, does so only at a population level, not at an individual level. Also, the problem with this approach is that, as the finiteness grows, the approach is saddled with parameter profligateness and computational complexity/inference challenges, similar to the first non-parametric approach. Besides, the popular implementation of this third approach in the econometric literature is in the form of a finite mixture-of-normal distributions. As observed and demonstrated by studies in the statistical density estimation literature (see, for example, Fruhwirth-Schnatter and Pyne, 2010, Lin et al., 2016, Gallaugher et al., 2020, and Dong et al., 2023), the mixture-of-normals approach can, and in general will, provide distorted inferences due to overfit and weak identification caused by needing unnecessarily high number of components when the target multivariate density has substantial skew. But, adopting more general component distributions, such as mixtures of skew-normal and other asymmetric parametric distributions, brings additional inference and computational challenges (see Lee and McLachlan, 2022), and requires large sample sizes to attain favorable asymptotic properties (see Dong and Lewbel, 2015, Mu and Zhang, 2018).

In this study, we return to the second approach of specifying a continuous parametric random variable vector, but now relax the unimodality restriction while also continuing to leverage the parsimony and computational ease benefits. In doing so, and to our knowledge for the first time in the statistical and econometric literature, we propose a new transformation-based parametric multivariate skewed density function with bimodal marginals (with a unimodal margin being a special case of a bimodal margin based on the value of a specific parameter). In this approach, we frame bimodality as a primitive and intrinsic characteristic of the stochastic process (rather than as a mixture of decision groups, each with a strict unimodal stochastic process). In doing so, we shift the econometric lens from cross-sectional preference heterogeneity at a population level to contextual latent state uncertainty at an individual level. Further, we propose and verify an approach to generate random variates from the proposed multivariate distribution, based on identifying a relationship between the proposed density and the reciprocal inverse Gaussian (RIG) distribution. This is important for model estimation purposes. Additionally, we demonstrate the application of our proposed new bimodal distribution to traffic crash injury severity analysis.

Specifically, in the current research, we build off two strands of recent literature. The first strand relates to the use of the Yeo and Johnson (2000) or YJ transformation approach to convert a non-symmetric distribution along each dimension into a symmetric distribution. Earlier studies have shown the efficacy and the parsimony of the YJ transformation approach to mimic a whole gamut of skewed and fat-tailed multivariate distributions for the original untransformed (or equivalently, the reverse YJ transformed) random distribution. In particular, Gallaugher et al. (2020) have shown that the simple transformation approach is comparable to even mixtures of skew-normal distributions (let alone mixtures of normal distributions; see also Jadhav et al., 2023, Watthanacheewakul, 2021, and Marimuthu et al., 2022). Bhat (2024) discusses this YJ transformation in detail, and applies the approach to the formulation and application of a flexible multivariate non-normal limited-dependent variable model system, considering a normal distribution for the symmetric distribution transformation for each dimension, which are then tied together across dimensions in a natural way using a multivariate normal distribution. The second strand corresponds to a general approach to develop bimodal distributions for univariate distributions, based on Gómez-Déniz et al. (2025). We use concepts from this strand and extend it in a specific way to allow a more general bimodal skewed distribution for univariate distributions than has been proposed so far. Then, we bring the univariate distributions together using a multivariate "stitching" mechanism to develop the new transformation-based multivariate asymmetric density function with bimodal margins (which can also collapse to unimodality along selected margins as a special case).[1]

In the next section, we briefly discuss both of the YJ and bimodal strands of research, with an emphasis on the second strand of research given its relatively limited consideration in the literature, after which we discuss the proposed methodology for developing the new proposed density function. Section 3 presents an application of the proposed bimodal model to traffic crash injury severity analysis. Section 4 concludes the paper with a brief summary of the research undertaken, potential other applications of the proposed model, and future research directions.

[1] The only other studies that we are aware of that consider a multivariate bimodal distribution are those by de Waal et al. (2022) and Gómez-Déniz et al. (2021). Wald et al. considers a relatively restrictive triangular distribution, with two triangular distributions joined together using a uniform distribution between the modes of each triangular distribution. Gómez-Déniz et al. resort to the use of a generalized normal distribution using concepts of folded distributions. In this paper, we follow the more general approach of Gómez-Déniz et al. (2025) to propose a distribution that allows for density smoothness throughout the multivariate range as well as accommodates asymmetry and substantial flexibility in the shape of the distribution.

2. METHODOLOGY

We start with the case of a simple univariate random variable, and extend later to the multivariate case.

2.1. YJ Transformation

In canonical form, most econometric models include an additive error term that is, conditional on observed exogenous variables, assumed normally distributed. To generalize this to accommodate asymmetry and skewness, models with continuous dependent outcomes transform the dependent variable directly using a strictly monotonic transformation such as $Z = t_\lambda(G)$, where G is closer to a symmetric distribution with non-fat tails (λ is a transformation parameter vector to be estimated) (see, for example, Atkinson et al., 2021 and Riani et al., 2023). In models where the dependent outcome is not continuous, Bhat (2024) proposes the use of the reverse transformation on a normally distributed variable to generate asymmetry and skewness as follows: $Z = t_\lambda^{-1}(G)$. This inverse transformation exists because the original transformation $Z = t_\lambda(G)$ is strictly monotonic. In the current paper, as in Bhat (2024), we consider the YJ-transformation for transforming Z into a normal distribution for $G \sim N(\mu, \sigma^2)$ as follows (0< λ <2):

$$G \sim N(\mu,\sigma^2) = t_\lambda(Z) = \begin{cases} -\dfrac{(-Z+1)^{2-\lambda} - 1}{2-\lambda} & \text{if } Z < 0 \\[2ex] \dfrac{(Z+1)^{\lambda} - 1}{\lambda} & \text{if } Z > 0 \end{cases} \tag{1}$$

Next, we apply the transformation in Equation (1) in reverse to generate asymmetry and skew in Z:

$$Z = t_\lambda^{-1}(G) = \begin{cases} 1 - \left[1-(2-\lambda)G\right]^{\left(\frac{1}{2-\lambda}\right)} & \text{if } G < 0 \\[2ex] \left[1+G\lambda\right]^{\left(\frac{1}{\lambda}\right)} - 1 & \text{if } G > 0 \end{cases} \tag{2}$$

To be noted is that $t_\lambda(0) = 0$ and $t_\lambda^{-1}(0) = 0.$ When 0< λ <1, as explained and plotted in Bhat (2024), Z gets skewed to the right with a long right tail. If 1< λ <2, Z gets skewed toward the left with a long left tail. λ =1 returns the original distribution of η for Z. The cumulative distribution function (CDF) and probability density function (PDF) of Z may be derived conveniently from those of the standardized normal distribution CDF $(\Phi[.])$ and PDF $(\phi[.])$ as follows after defining $g(z) = \left[\dfrac{t_\lambda(z) - \mu}{\sigma}\right]$. For convenience, we will write $g(z)$ simply as g when obvious.

$$\begin{aligned} H_Z(z) &= \text{Prob}(Z < z) = \text{Prob}(t_\lambda^{-1}(G) < z) = \text{Prob}(G < t_\lambda(z)) = \Phi\left[\sigma^{-1}(t_\lambda(z) - \mu)\right] = \Phi\left[g\right] \\ h_Z(z) &= \frac{\partial H_Z(z)}{\partial z} = \frac{\partial \Phi\left[\sigma^{-1}(t_\lambda(z)-\mu)\right]}{\partial t_\lambda(z)} \times \left|\frac{\partial t_\lambda(z)}{\partial z}\right| = \left(\frac{\phi[g]}{\sigma}\right) \times \left(|z|+1\right)^{\text{sgn}(z)(\lambda-1)}. \end{aligned} \tag{3}$$

2.2. Motivation for an Asymmetric and Skewed Bimodal Distribution

Most econometric models assume a unimodal symmetric distribution for the error terms embedded in their models. Over the years, some studies have considered alternative asymmetric unimodal error term distributions, including the skew-normal, or a skew-t distribution, or the broader class of skew-elliptical distributions (see Azzalini and Capitanio, 2013 and Teimouri and Nadarajah, 2016 for detailed reviews of such skew distributions and applications). The density of such distributions takes the form of a symmetric density function multiplied by a skewing component that typically is the cumulative distribution of the symmetric density (see Lee and McLachlan, 2022 for a detailed discussion). However, transformation approaches (such as the YJ transformation discussed earlier) are more parsimonious and flexible enough to accommodate a variety of different skew distributions, as illustrated in Bhat et al. (2025).

The YJ-based transformation approach, while proving to be a powerful approach to accommodate asymmetric distributions, does not, at least as it has been applied in the statistical and econometric literature so far, consider multimodality in the distribution. On the other hand, there are many reasons why such multimodality may appear in the stochasticity characterizing data. In particular, the focus in the current paper is on bimodality, which has been documented to be prevalent in many situations. For instance, in the transportation field, Mei et al. (2004) observe that the combination of normal vehicles and overloaded trucks on highways causes a distinct bimodal pattern in vehicle loadings on highways, while Jintanakul et al. (2009) indicate the bimodality of travel time distributions on freeways due to a mix of vehicle types and driving patterns. Taylor and Somenahalli (2014) and Ji et al. (2015) similarly show that urban link travel time distributions are much better characterized as bimodal distributions, given the influence of signal controls, vehicle-pedestrian interactions, and cross-street traffic. In this urban link case, two vehicles traveling next to each other can have very different traveling times if one vehicle just about makes it through a green light, while the succeeding vehicle has to stop. Das et al. (2014) similarly observes the bimodal nature of ocean shipping transit time, associating this bimodal nature to a sequence of (typically unreported in data sets) independently managed activities that characterize the movement from the initial origin point of the manufacturer to the final destination close to the customer end. In the traffic safety literature, Xiong and Mannering (2013) make a theoretical case for the bimodality of injury severity, given that a whole host of unobserved factors in most crash data sets (such as those associated with aggressiveness in driving, and actions to reduce injury severity as the crash is unfolding, vehicle responsiveness to driver actions as they relate to weather and pavement characteristics, and vehicle crash protection equipment and performance) can combine together to have quite different crash severities for the same set of observed crash characteristics. Similarly, multiple studies (see, for example, Kleber et al., 2012 and Rauf et al., 2019) have analyzed the temporal pattern of deaths of individuals after a severe injury sustained in a traumatic incident (mostly traffic crashes, but also including other types of incidents such as a high fall), and suggested the presence of a continuous fatality rate with two modes – one that is immediate (within 60 minutes of the traumatic incident) caused by central nervous system injuries and major vascular collapse, and another within a day after admission to a hospital caused by brain injury and/or multiple organ failure.

Almost all of the studies above (and many other studies in the transportation literature) adopt a finite mixture of normal distributions within the interpretation of a latent class model (consumers belonging to each of a finite number of segments), considering a unimodal distribution within each consumer segment. Interestingly, such studies almost exclusively consider multimodality to be at play only in responsiveness heterogeneity (that is, on the random coefficients on observed exogenous variables), and completely ignore possible multimodality more fundamentally in overall preference heterogeneity (that is, the studies assume a unimodal – typically a normal distribution or a logistic distribution or an extreme value distribution – for the kernel error term). Certainly an argument can be made that, given it is the kernel error term that absorbs a combination of different unobserved factors at play (unlike response unobserved heterogeneity that is tied more specifically to a single observed factor), allowing multimodality in the kernel error term is at least as important to test (if not much more so) than in the random coefficient distributions. In this paper, we focus on a bimodal distribution for the kernel error term, while considering the bimodality in the kernel error term to be a result of unobserved factor effects within a specific individual or within a specific context (rather than at the population level due to mixing as in latent class approaches). For example, in the Xiong and Mannering injury severity framework, say that aggressive driving is an unobserved factor. Then, it is wholly possible that the same motorist may drive aggressively at one point in time (say due to time pressure), and not so aggressively at another point in time, leading to bimodality in injury severity within the context of "observationally-equivalent" instances of two crashes. In this regard, as also stated by Vila et al. (2021), mixture-free bimodal distributions can play an important role, especially because they are not saddled with identifiability/computational problems of mixture distributions during estimation. Besides, if there is strong reason to maintain the clustering/classification interpretation of the mixture approach, one can extend current mixture approaches to include mixtures of bimodal distributions for each segment, as discussed in the conclusions section of this paper.

Finally, in most analysis contexts, there is no a priori reason to believe that any bimodality should be of the same intensity at the two modal points or that the density function should be close to being symmetric or have similar variance in the vicinity of each modal point, highlighting the importance of asymmetric and skewed bimodal distributions. For example, when not driving aggressively, injury severity may be more clustered around the low injury severity level with a rather small amount of diffusion, while when driving aggressively, injury severity may be more loaded toward the high injury severity level though the variance in the injury severity sustained may be high (more diffuse injury severity) because of last minute pre-crash maneuvers taken (or not taken) by specific motorists involved in the crash. In most crash analysis, without a clear marker for aggressive driving, this would indeed lead to an asymmetric and skewed injury severity kernel error term, as we demonstrate in our empirical analysis.

2.3. The Proposed Approach to Construct an Asymmetric and Skewed Bimodal Distribution

The foundation for the proposed approach is a theorem proposed and proved by Slobin (1927), which has also been invoked recently by Gómez-Déniz et al. (2025) to derive two probability density functions (one being a symmetric density function, and another being an asymmetric function based on a skew-normal unimodal distribution). In this paper, we derive another probability density function based on the YJ transformation that allows for more general forms of bimodality skew (as illustrated in Bhat et al., 2025, the YJ transformation approach is quite general, and can represent a variety of unimodal distributions very closely, including the skew-normal, the skew-t, the power log-normal, and the extreme value). Besides, the skew-normal and skew-t approaches manifest a singularity in Fisher's information matrix for the skew parameter, and can become somewhat difficult to estimate (see Arellano-Valle and Azzalini, 2006. Bhat and Sidharthan, 2012, and Gómez-Déniz et al., 2021), while our transformation approach does not exhibit any such problems. Our approach also provides immediately for a straightforward extension to a multivariate distribution with bimodal asymmetry along each dimension (and varying asymmetries and shapes across dimensions). We present our new proposed bimodal distribution in a set of theorems.

Theorem 1: The following function represents a proper probability density function (pdf) for a random variable Z spanning the real line and $\alpha \geq 0,\ 0 < \lambda < 2,\ \sigma > 0$:

$$f_Z(z) = \sigma^{-1}\phi\left[g - \frac{\alpha}{g}\right] \times \left(|z|+1\right)^{\mathrm{sgn}(z)(\lambda-1)},\ g = \left[\frac{t_\lambda(z)-\mu}{\sigma}\right]. \tag{4}$$

Proof: The density function above is clearly positive and continuous over the entire range of Z, given that the standard normal density function $\phi(.)$ is always positive and continuous, and so is $\sigma^{-1}\left(|z|+1\right)^{\mathrm{sgn}(z)(\lambda-1)}$. The cumulative distribution function (CDF) of Z may be written as:

$$F_Z(z) = \int_{t=-\infty}^{z} f_Z(t)dt = \int_{t=-\infty}^{z} \sigma^{-1}\phi\left[g(t) - \frac{\alpha}{g(t)}\right] \times \left(|t|+1\right)^{\mathrm{sgn}(t)(\lambda-1)} dt\,. \tag{5}$$

Let $v = g(t)$. Then $dv = \dfrac{dg(t)}{dt} = \sigma^{-1}\left(|t|+1\right)^{\mathrm{sgn}(t)(\lambda-1)}$, and

$$F_Z(z) = \int_{v=-\infty}^{g(z)} \phi\left(v - \frac{\alpha}{v}\right) dv\,. \tag{6}$$

To complete the proof, it needs to be shown that

$$F_Z(\infty) = \int_{v=-\infty}^{g(\infty)} \phi\left(v - \frac{\alpha}{v}\right) dv = \int_{v=-\infty}^{\infty} \phi\left(v - \frac{\alpha}{v}\right) dv = 1. \tag{7}$$

This can be achieved by following a generalization of the proof given in Slobin (1927) for improper definite integrals.[2] In particular, note that $\int\limits_{v=-\infty}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv = 2\int\limits_{v=0}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv$ because the standard normal density function $\phi(s)$ is an even function; that is, $\phi(s)=\phi(-s)$. Now, consider the one-sided integral. $\int\limits_{v=0}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv$, and let $\tilde{h}=\frac{\alpha}{v}$. Then,

$$\int\limits_{v=0}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv = \begin{cases} = \int\limits_{\tilde{h}=\infty}^{0} \phi\left(\frac{\alpha}{\tilde{h}}-\tilde{h}\right)d\left(\frac{\alpha}{\tilde{h}}\right) \\ = \int\limits_{\tilde{h}=\infty}^{0} \phi\left(\frac{\alpha}{\tilde{h}}-\tilde{h}\right)d\left(\frac{\alpha}{\tilde{h}}-\tilde{h}\right) + \int\limits_{\tilde{h}=\infty}^{0} \phi\left(\frac{\alpha}{\tilde{h}}-\tilde{h}\right)d\tilde{h}. \\ = \int\limits_{u=-\infty}^{\infty} \phi(u)\,d(u) - \int\limits_{\tilde{h}=0}^{\infty} \phi\left(\tilde{h}-\frac{\alpha}{\tilde{h}}\right)d\tilde{h}, \text{ with } u=\left(\frac{\alpha}{\tilde{h}}-\tilde{h}\right) \end{cases} \tag{8}$$

Finally, from the last equation and rearranging, we get

$$\begin{aligned} \int\limits_{v=-\infty}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv &= \int\limits_{v=0}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv + \int\limits_{v=0}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv \\ &= \left[\int\limits_{u=-\infty}^{+\infty} \phi(u)\,du - \int\limits_{\tilde{h}=0}^{\infty} \phi\left(\tilde{h}-\frac{\alpha}{\tilde{h}}\right)d\tilde{h}\right] + \int\limits_{v=0}^{\infty} \phi\left(v-\frac{\alpha}{v}\right)dv = \int\limits_{u=-\infty}^{\infty} \phi(u)\,du = 1. \end{aligned} \tag{9}$$

Henceforth, we will write $Z \sim BYJN(\mu,\, \sigma^2,\, \lambda,\, \alpha)$ when the random variable follows the pdf in Equation (4) to denote that Z represents a bimodal YJ transformed-to-normal (BYJN) random variable. Note also that our proposed BYJN distribution with λ=1 nests the symmetric bimodal distribution proposed by Gómez-Déniz et al. (2025) as a special case.

Next, we establish the following theorem that helps with generating random realizations from the density function of *Z*.

Theorem 2: Let $Z \sim BYJN(\mu,\, \sigma^2,\, \lambda,\, \alpha)$. Then, the random variable $W=\frac{\alpha^2}{S}$, with $S=G^2$ and $G=\left[\frac{t_\lambda(Z)-\mu}{\sigma}\right]$, follows an inverse Gaussian distribution (scaled by one half) with mean parameter $\tilde{\mu}=\alpha$ and shape parameter $\tilde{\lambda}=\alpha^2$; $W \sim IG(\alpha,\alpha^2), W \in \{0,\infty\}$

[2] Later, we will provide an expression for the CDF, from which too the above result may be obtained.

Proof: Equation (4) may be rewritten as follows:

$f_Z(z) = \phi\left[g - \frac{\alpha}{g}\right] \times \left|\frac{dg}{dz}\right|$, from which we get using the usual change of variable technique

$$
\begin{aligned}
f_G(g) &= f_Z\left[z = t_\lambda^{-1}(\sigma g(z) + \mu)\right] \times \left|\frac{dz}{dg(z)}\right| \\
&= \phi\left[\frac{t_\lambda(t_\lambda^{-1}(\sigma g + \mu) - \mu)}{\sigma} - \frac{\alpha}{\frac{t_\lambda(t_\lambda^{-1}(\sigma g + \mu) - \mu)}{\sigma}}\right] \times \frac{dg(z)}{dz} \times \frac{dz}{dg(z)} \qquad (10) \\
&= \phi\left[g - \frac{\alpha}{g}\right]
\end{aligned}
$$

Again using successive change of variables $S = G^2$ and $W = \frac{\alpha^2}{S}$, we get:

$$f_S(s) = \frac{1}{2}\phi\left[\sqrt{s} - \frac{\alpha}{\sqrt{s}}\right] s^{-1/2}, \text{ and}$$

$$f_W(w) = \frac{\alpha}{2}\phi\left[\sqrt{w} - \frac{\alpha}{\sqrt{w}}\right] w^{-3/2}. \qquad (11)$$

The inverse Gaussian (IG) distribution is given by:

$$f_{\tilde{W}}(\tilde{w}) = \sqrt{\tilde{\lambda}}\,\phi\left[\frac{\sqrt{\tilde{\lambda}}(\tilde{w} - \tilde{\mu})}{\tilde{\mu}\sqrt{\tilde{w}}}\right] \tilde{w}^{-3/2}. \qquad (12)$$

Putting $\tilde{\mu} = \alpha$ and $\tilde{\lambda} = \alpha^2$, the expression collapses to the same expression as Equation (11) scaled by one half. Equivalently, given that the random variable $(1/W)$ is reciprocal inverse Gaussian (RIG) distributed, the random variable $S = G^2$ is also distributed RIG (with the RIG having a mean parameter α and a scale parameter α^2) scaled by a factor of $\alpha^2 / 2$.

Theorem 2 provides an easy way to generate realizations from the distribution of $Z \sim BYJN(\mu, \sigma^2, \lambda, \alpha)$, given the ease of drawing variates from the inverse Gaussian.[3] Further,

[3] Our approach is easier than that proposed by Gómez-Déniz et al. (2025). Their approach requires draws from the Generalized inverse Gaussian (GIG) distribution that is more challenging and time-consuming. Our approach only requires drawing from the inverse Gaussian. This generation procedure is based on the result that $\frac{\sqrt{\tilde{\lambda}}(\tilde{w} - \tilde{\mu})}{\tilde{\mu}\sqrt{\tilde{w}}}$ is normally distributed when $\tilde{w}$ is inverse Gaussian distributed as in Equation (12). Equivalently, $\frac{\tilde{\lambda}(\tilde{w} - \tilde{\mu})^2}{\tilde{\mu}^2 \tilde{w}}$ is chi-

our procedure also makes it simple to extend to the multivariate case, given the way we propose the multivariate extension. The procedure for the univariate case is as follows:

(1) Draw a variate ϑ from the standardized normal distribution, and square the value $\varphi = \vartheta^2$.

(2) Compute $v = \alpha + (\varphi / 2) - 0.5\sqrt{4\alpha\varphi + \varphi^2}$ (this is the lower root).

(3) Generate another random variate u sampled from a standardized uniform distribution.

(4) If $u \le \dfrac{\alpha}{\alpha + v}$, $w = v$; otherwise $w = \alpha^2 / v$. w is a variate from the inverse Gaussian.

(5) Compute $s = \alpha^2 / w$.

(6) If $u < 0.5$, $g = -\sqrt{s}$; otherwise $g = +\sqrt{s}$.

(7) $z = t_\lambda^{-1}(\mu + \sigma g)$.

Theorem 3: The random variable $Z \sim BYJN(\mu,\ \sigma^2,\ \lambda,\ \alpha)$ is bimodal for $\alpha > 0$ and unimodal for α =0. There is no closed-form expression for the two modes when $\alpha > 0$ and $\lambda \ne 1$, but they are obtained as solutions to the following quadratic equation:

$$g^2 - \tau(z)g - \alpha = 0;\ g = \frac{t_\lambda(z) - \mu}{\sigma},\ \tau(z) = \frac{\sigma(\lambda - 1)}{\left(1 + |z|\right)^{1 + \mathrm{sgn}(z)(\lambda - 1)}}. \tag{13}$$

Proof: This is obtained by the straightforward differentiation of the density function of *Z* in Equation (4) and setting to zero (the second differential of the density function at the resulting real roots of the Equation (13) above can also be shown to be negative). Of note is that when λ =1, the result is a symmetric bimodal distribution for *Z*. In this special case, $\tau(z)$ =0 and $t_\lambda(z) = z$, and the two modes are solutions to the equation $g^2 - \alpha = 0$, and are obtained as $z = \mu \pm \sigma\sqrt{\alpha}$, which is the case of Gómez-Déniz et al.'s symmetric distribution. Also, in the special case when α =0, the single mode of *g* is at $g = \tau(z)$. Equivalently, the single mode for *Z* in this special case is obtained as the solution to the following equation: $t_\lambda(z) = \sigma\tau(z) + \mu$.

The trough (that is, anti-mode) level between the two modes in the case of $\alpha > 0$ is at $G = 0$ or $z = t_\lambda^{-1}(\mu)$. This should be obvious from the expression in Equation (10) for the density function

---

squared distributed (see Shuster, 1968). In the context of our analysis with $\tilde{\mu} = \alpha$ and $\tilde{\lambda} = \alpha^2$, $\sqrt{w} - \dfrac{\alpha}{\sqrt{w}}$ is normally distributed and $\left(\sqrt{w} - \dfrac{\alpha}{\sqrt{w}}\right)^2 = \left(w + \dfrac{\alpha^2}{w} - 2\alpha\right)$ is chi-squared distributed. The second through fourth steps in the generation process below are based on back-solving for a realization of *w* by putting $\varphi = \left(\sqrt{w} - \dfrac{\alpha}{\sqrt{w}}\right)^2$, computing the two possible roots for the resulting realization of *w*, and then assigning the final realization for *w* by the relationship that the product of the two possible roots is α^2 (see Michael et al., 1976).

of G. This is also the median point of the distribution, with $\text{Prob}(Z < t_\lambda^{-1}(\mu)) = 0.5$ (see next theorem).

Theorem 4: The cumulative distribution function (CDF) of the random variable $Z \sim BYJN(\mu, \sigma^2, \lambda, \alpha)$ is as follows (with $\Phi(.)$ being the standard normal CDF):

$$F_Z(z) = \text{Prob}(Z < z) = \text{Prob}(G < g) = \text{Prob}(S < g^2),\; g = \left[\frac{t_\lambda(z) - \mu}{\sigma}\right]$$

$$= \begin{cases} 0.5 \times \left[\Phi\left(g - \dfrac{\alpha}{g}\right) + \exp(2\alpha)\left(\Phi\left(g + \dfrac{\alpha}{g}\right)\right)\right] \text{ if } g < 0 \\ 0.5 \text{ if } g = 0 \\ 0.5 + 0.5 \times \left[\Phi\left(g - \dfrac{\alpha}{g}\right) - \exp(2\alpha)\left(1 - \Phi\left(g + \dfrac{\alpha}{g}\right)\right)\right] \text{ if } g > 0 \end{cases} \quad (14)$$

Proof: The above result is immediate from the expressions for the cumulative distribution function of the RIG distribution. Essentially, the bimodal distribution includes two RIG distributions, one RIG to the right of $z = t_\lambda^{-1}(\mu)$ and another mirrored about the y-axis (but not symmetrically so except when λ=1) RIG distribution to the left of $z = t_\lambda^{-1}(\mu)$. As expected, $F_Z(\infty) = 1$. Importantly, the ability to write the univariate cumulative distribution function in literally a closed form (given the univariate cumulative normal distribution function is easily computed) implies that estimating a univariate model with the proposed distribution is as easy as estimating a model that considers the error term to be a normal distribution. This is unlike the case of a finite mixture-of-normals that can be unstable in estimation and be computationally intensive.

The mean, variance, and other higher moments are not computable in a closed form, but may be estimated through numerical integration based on the probability density function in Equation (4), or through generating draws based on Theorem 2 and estimating the desired quantities. Figure 1 provides the plots of the proposed BYJN distribution maintaining $\mu = 0$ and $\sigma^2 = 1$, but allowing for different values of λ and α (in our empirical context with the BYJN distribution being applied to the kernel error term of an ordered-response model, we maintain $\mu = 0$ and $\sigma^2 = 1$ for identification purposes). The top panel shows three graphs for $\alpha = 0.5$ with different values of the YJ parameter λ. When $\lambda = 0.5$ (more generally when $\lambda < 1$), the bimodal distribution has a sharp peak to the left of the zero point with a more diffuse peak to the right. The exact reverse is the case when $\lambda = 1.5$ (more generally when $\lambda > 1$), with the case of $\lambda = 1$ (the middle graph) being the symmetric bimodal case of Gómez-Déniz et al. (2025). The bottom panel shows three graphs for $\alpha = 2.0$ with the same values of the YJ parameter λ as the top panel. This parameter α may be viewed as a diffusion factor. As α increases, the spread of the distribution increases (especially around the zero point). In the case when $\lambda = 1$, as discussed earlier, the modes are located at $\pm\sqrt{\alpha}$ and the density function take the identical magnitude value of $\phi(0)$=0.3989 (this may be observed by plugging in $g=\left[\frac{t_\lambda(z)-\mu}{\sigma}\right]=z$ when $\mu=0$, $\sigma^2=1$, and $\lambda=1$ in Equation (4) with $z=\pm\sqrt{\alpha}$). In this specific case, α simply pulls the two modes more apart but the density remains the same at the modal points regardless of the value of α. However, when $\lambda\neq 1$, the α parameter also affects the density value at the modal points. Specifically, when $\lambda<1$, the distribution to the left of the zero point gets tighter and the modal density becomes higher, while the distribution to the right gets more diffuse and the modal density becomes lower. When $\lambda>1$, the reverse holds. Clearly, the combination of λ and α when $\lambda\neq 1$ provides substantial flexibility to the bimodal distribution. Of course, when α =0, the result is a unimodal distribution with skew based on the YJ parameter, as discussed in Bhat (2024).

2.4. Extension to the Multivariate Case

The univariate skewed bimodal distribution proposed in the previous section provides a straightforward and very effective way to extend to multiple dimensions, though we have not seen this approach to extend univariate bimodal distributions in the extant literature. Importantly, the procedure to generate draws, as discussed for the univariate case, can also be extended for the multivariate case after careful modification. The key to our extension of the univariate asymmetric bimodal distribution to the multivariate case is the introduction of asymmetry using the YJ transformation. By doing so, and as should be obvious from Equation (10), $g-\frac{\alpha}{g}$ is normally distributed. This immediately allows an extension to the multivariate case as discussed next.

Consider a vector $\mathbf{Z}=(Z_1,Z_2,...,Z_L)'$ $(L\times 1 \text{ vector})$. The nature of the bimodal distribution can vary across the random elements Z_l through different values of μ_l, σ_l^2, λ_l, and α_l. The procedure begins with the multivariate YJ distribution with a single mode along each dimension. Using the same approach as for the univariate case, we can extend it to up to two modes along each dimension. Define the random variable $G_l=\frac{t_{\lambda_l}(Z_l)-\mu_l}{\sigma_l}$, and let $g_l=\frac{t_{\lambda_l}(z_l)-\mu_l}{\sigma_l}$ for a specific value of $Z_l=z_l$. Let $\mathbf{G}=(G_1,G_2,...,G_L)'$, $\mathbf{z}=(z_1,z_2,...,z_L)'$, and $\mathbf{g}=(g_1,g_2,...,g_L)'$. Similarly, collect other component-wise elements into vectors $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\alpha}$. For compactness, we will also write $\mathbf{t}_{\boldsymbol{\lambda}}(\mathbf{z})=\left(t_{\lambda_1}(z_1),t_{\lambda_2}(z_2),...,t_{\lambda_L}(z_L)\right)$, and $\mathbf{g}=\frac{\mathbf{t}_{\boldsymbol{\lambda}}(\mathbf{z})-\boldsymbol{\mu}}{\boldsymbol{\sigma}}$. $\frac{\boldsymbol{\alpha}}{\mathbf{g}}$, as written in the next few expressions, will refer to the vector of element -by-element division along each dimension.

Theorem 5: The following function represents a proper multivariate probability density function (pdf) for a random variable vector $\mathbf{Z}$ spanning the multi-dimensional real line; $0 < \boldsymbol{\lambda} < 2$ and $\alpha \geq 0$ (the last two notations taken to mean that each element of the vector follows the scalar limits):

$$h_Z(\mathbf{z}) = \left(\frac{\phi_L\left[\mathbf{g} - \dfrac{\boldsymbol{\alpha}}{\mathbf{g}}, \boldsymbol{\Omega}^*\right]}{\prod_{l=1}^{L} \sigma_l} \right) \times \prod_{l=1}^{L} \left[\left(|z_l| + 1 \right)^{\mathrm{sgn}(z_l)(\lambda_l - 1)} \right], \ \boldsymbol{\Omega}^* \text{ representing a correlation matrix.} \qquad (15)$$

Proof: The proof follows exactly the one for the univariate case, because the multivariate normal distribution is also a symmetric distribution, with each element of $\mathbf{g} - \dfrac{\boldsymbol{\alpha}}{\mathbf{g}}$ being univariate normally distributed. From the properties of the multivariate normal distribution, it is immediate that the marginal along each dimension is BYJN distributed; $Z_l \sim BYJN(\mu_l, \sigma_l^2, \lambda_l, \alpha_l)$. Essentially, the elements Z_l along each dimension are being brought together using an implicit Gaussian copula to generate a correlation across the elements (see Bhat et al., 2025). This observation also provides a pathway to generate draws from the multivariate distribution of Equation (15). The procedure we propose is as follows, which does not need the inverse CDF of Z_l (this inverse CDF has no closed form, and numeric computation can be expensive):

(1) Follow steps (1) through (6) for the univariate case to generate independent draws (across dimensions) for G_l (that is, obtain realizations g_l).

(2) Compute $\tilde{g}_l = g_l - \dfrac{\alpha_l}{g_l}$ for each dimension *l*. These realizations, as per Equation (15), are normally distributed and correlated across dimensions through the matrix $\boldsymbol{\Omega}^* = \mathbf{L}\mathbf{L}'$, where $\mathbf{L}$ corresponds to the lower diagonal Cholesky matrix. Obtain $\vec{\mathbf{g}} = \mathbf{L}\tilde{\mathbf{g}}$, with elements $\vec{g}_l$, which now need to be back-transformed to obtain $\breve{g}_l$ through the equation $\vec{g}_l = \breve{g}_l - \dfrac{\alpha_l}{\breve{g}_l}$.

To do so, compute the two roots of $\breve{g}_l$ for each dimension as follows:

$$\breve{g}_l^{low} = \frac{\vec{g}_l - \sqrt{\vec{g}_l^2 + 4\alpha}}{2}, \text{ and } \breve{g}_l^{high} = \frac{\vec{g}_l + \sqrt{\vec{g}_l^2 + 4\alpha}}{2}.$$

Note also that $\breve{g}_l^{low} \times \breve{g}_l^{high} = -\alpha$, or $\breve{g}_l^{low} = -\left(\dfrac{\alpha}{\breve{g}_l^{high}} \right)$.

(3) From $f_{G_l}(g_l) = \phi\left[g_l - \dfrac{\alpha_l}{g_l} \right]$ (see Equation (10)),

$$\frac{f_{G_l}\left(\breve{g}_l^{low}\right)}{f_{G_l}\left(\breve{g}_l^{high}\right)}=\frac{\phi\left[\breve{g}_l^{low}-\left(\dfrac{\alpha_l}{\breve{g}_l^{low}}\right)\right]}{\phi\left[\breve{g}_l^{high}-\left(\dfrac{\alpha_l}{\breve{g}_l^{high}}\right)\right]}=\frac{\phi\left[-\left(\dfrac{\alpha_l}{\breve{g}_l^{high}}\right)+\breve{g}_l^{high}\right]}{\phi\left[\breve{g}_l^{high}-\left(\dfrac{\alpha_l}{\breve{g}_l^{high}}\right)\right]}=1,$$

$$\frac{\partial\overleftrightarrow{g}_l / \partial\breve{g}_l\big|_{\breve{g}_l=\breve{g}_l^{low}}}{\partial\overleftrightarrow{g}_l / \partial\breve{g}_l\big|_{\breve{g}_l=\breve{g}_l^{high}}}=\frac{\left(1+\dfrac{\alpha_l}{\left(\breve{g}_l^{low}\right)^2}\right)}{\left(1+\dfrac{\left(\breve{g}_l^{low}\right)^2}{\alpha_l}\right)} \text{ and compute } \tilde{u}_l=\left[1+\frac{\left(1+\dfrac{\alpha_l}{\left(\breve{g}_l^{low}\right)^2}\right)}{\left(1+\dfrac{\left(\breve{g}_l^{low}\right)^2}{\alpha_l}\right)}\right]^{-1}.$$

(4) Generate another random variate $\overleftrightarrow{u}$ sampled from a standardized uniform distribution.

(5) Based on application of the result of Michael et al. (1976), select the root from between $\breve{g}_l^{low}$ and $\breve{g}_l^{high}$ as follows: if $\overleftrightarrow{u} \le \tilde{u}$, assign $\underset{\frown}{g}_l=\breve{g}_l^{low}$; otherwise, $\underset{\frown}{g}_l=\breve{g}_l^{high}$.

(6) Finally, generate variates $z_l=t_{\lambda_l}^{-1}(\mu_l+\sigma_l\underset{\frown}{g}_l)$.

There is no simple closed-form expression for the CDF or the moments of this flexible multivariate distribution, but the easy and powerful generation process just discussed allows the computation of all of these quantities numerically. The generation approach has been verified through numerically integrating the density function in Equation (15) with given parameters between different bounds, and estimating the same value by generating random variates and computing the fraction of random variates between the integration bounds. Of course, all the univariate properties discussed earlier remain in effect for each marginal.

3. APPLICATION

3.1. Background

We now demonstrate the application of our proposed asymmetric bimodal distribution in the context of traffic crash injury severity analysis. In the crash injury severity field, two model structures have been widely used – the ordered-response formulation and the unordered-response formulation. There has been, and continues to be, a debate regarding which one of these two should be the basis for modeling injury severity. While the workhorse formulation within the ordered-response formulation is the ordered probit or ordered logit, more advanced versions of these formulations that incorporate varying thresholds and varying exogenous variable effects have been considered (see Eluru et al., 2008, Yasmin and Eluru, 2013, Balan and Paleti, 2018, and Huang et al., 2025). Similarly, while the workhorse formulation within the unordered-response formulation is the multinomial logit, more advanced versions such as the mixed logit model and the multinomial probit have also been considered (see Chen et al., 2023, Wu et al., 2023, and Haddad et al., 2024). Several studies have investigated these different formulations, with a detailed discussion of the advantages and limitations of each approach (see, for example, Eluru, 2013, Abay, 2013, and Nasri et al., 2022). While there is general consensus that formulations such as the generalized-ordered response framework that straddle the space between the ordered and unordered formulations may be most appropriate, the results also suggest that the conclusions from the two model frameworks tend to be rather similar. In this paper, as a first use case of the proposed bimodal distribution, we consider the simple ordered-response framework, leaving investigations with more advanced versions of the ordered-response framework and the unordered framework to future research.

In the canonical form of an ordered-response model structure (see McKelvey and Zavoina, 1975; Bhat, 1997; and Greene and Hensher, 2010), consider the following relationship that maps the latent injury severity risk index y^* to the observed injury severity level y.

$$y^* = \boldsymbol{\gamma}'\mathbf{x} + \varepsilon, y = k \text{ if } \psi_{k-1} \le y^* < \psi_k; k = 1,2,...K; \ \psi_0 = -\infty, \psi_1 = 0, \psi_K = +\infty \tag{16}$$

In the above equation, $\mathbf{x}$ represents an exogenous variable vector (excluding a constant), and $\boldsymbol{\gamma}$ is a corresponding coefficient vector to be estimated. Assume that the elements of the vector $\mathbf{x}$ are independent of the error term ε. ψ_k represents the upper bound threshold for ordinal level k for the outcome y $(\psi_0 < \psi_1 < \psi_2 ... < \psi_{K-1} < \psi_K; \ \psi_0 = -\infty, \psi_K = +\infty)$. For later use, let $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, ..., \psi_{K-1})'$. In this paper, we assume that the coefficient vector $\boldsymbol{\gamma}$ is fixed, and focus on a bimodal asymmetric distribution specification for the kernel error term ε. The correspondence between the discussion in Section 2.3 and Equation (16) is straightforward by equating the error term ε to the random variable Z with location and scale normalization as is needed in any ordered-response model. Specifically, we assume $\varepsilon \sim BYJN(0,1,\lambda,\alpha)$. We refer to the model in (16) with this error distribution as the ordered-response probit BYJN model. Defining $\varphi_k = \psi_k - \boldsymbol{\gamma}'\mathbf{x}$, the required probability corresponding to Equation (16) for maximum likelihood estimation is:

$$\text{Prob}(y = k) = \text{Prob}(\varphi_{k-1} < \varepsilon < \varphi_k) = F_\varepsilon\left[\varphi_k\right] - F_\varepsilon\left[\varphi_{k-1}\right], \tag{17}$$

where $F_\varepsilon\left[\varphi_k\right]$ is computed based on Equation (14).

To ensure that the conditions $0 < \lambda < 2$ strictly holds during estimation, we use a parameterization as follows:

$$\lambda = \frac{2}{1 + \exp(-\lambda^*)}. \tag{18}$$

Also, to ensure that $\alpha \ge 0$, we parameterize $\alpha = \exp(\alpha^*)$. In the analysis, we first run the model in parameterized form with λ^* and α^*. After this estimation, we run a final iteration with the implied unparametrized values of λ and α to obtain the standard errors of all parameters.

3.2. Data and Sample Used in Analysis

The sample used in estimation is drawn from the Texas Department of Transportation (TxDOT) crash database of two-vehicle crashes at intersections recorded between January 1, 2018, and December 31, 2019. While five levels of injury severity for the most severely injured person are reported; no injury, possible injury, non-incapacitating injury, incapacitating injury, and fatal injury; we consolidate the last two levels into a single severe injury category due to the very few number of observations in the fatal crash category. We use a two-year period to obtain a reasonable number of crashes in the highest level "severe" injury severity category. The final dataset used in analysis comprises 2,757 crashes, with the following split in the injury severity categories: (1) Non-injury – 1314 (47.7%), (2) Possible injury – 691 (25.0%), (3) Non-incapacitating injury – 664 (24.1%), and (4) Severe injury – 88 (3.2%).

The data used is the same as that employed for part of the analysis in Haddad et al. (2024). As discussed there, an array of data sources were used to compile a comprehensive set of explanatory variables influencing crash severity for each crash, including (a) crash characteristics (intersection control type and number of intersection approach legs at crash location, crash time and day of week, and crash weather/lighting conditions) from TxDOT's CRIS database, (b) road network features (such as functional classification of the approach roadways at intersection, and AADT/number of lanes/posted speed limits of approach roadways from the TxDOT roadway network inventory database, (c) CBG-level land-use distribution splits in the Census Block Group (CBG) of crash location from the City of Austin's Open Data Portal, (d) motorized vehicle ownership data, aggregated to the CBG-level using GIS tools, as obtained from the U.S. Environment Protection Agency (EPA) Smart Location Database (or SLD; see Chapman et al., 2021, and Ramsey and Bell, 2014) (e) sociodemographic and commute mode split data, again aggregated to the CBG level, as extracted from the American Community Survey (ACS) 2021 five-year estimates.

3.3. Exogenous Variable Specification and Model Results

The exogenous variable specification for inclusion in the final model was based on considering a number of different combinations of variables and different functional forms for variables, along with insights from earlier research and parsimony considerations. In terms of functional form, for variables in grouped form (such as age of most severely injured individual and time of crash) and those naturally discrete (such as gender, race, lighting conditions, time-of-day of crash, and weather conditions at the time of crash), dummy variables were created in the most disaggregate form and then, to achieve an efficient specification, progressively combined based on statistical tests. For variables in continuous form (most BE and CBG variables), various functional forms were tested, including a continuous linear form, a continuous logarithm form, a piece-wise linear form, and a set of dummy variables for different ranges, and the best specification among the many possibilities was selected based on data fit.

Four different models are considered and estimated in our analysis: (1) the simple ordered probit (ORP) model with a standard unimodal and symmetric non-skewed distribution for the normal distribution (λ=1, α=0), (2) the unimodal skewed ORP (or the SORP) model in which a single mode is retained, but the error distribution is allowed to be skewed (λ free for estimation, α=0), (3) the bimodal symmetric ORP (or BSORP) model in which two modes are allowed, but the distribution is symmetric $(\lambda=1, \alpha \text{ free for estimation})$, and (4) the proposed BYJN ORP (or the BYJNORP) model in which bimodality and asymmetry are allowed (both λ and α left free for estimation). In all models, for identification, $\mu=0$ and $\sigma^2=1$. The corresponding shapes of the error distributions, as estimated in the models, are shown in Figure 2 (the implied/estimated values of λ and α are reported underneath each figure, and are also reported later in Table 1). The SORP model shows a substantial rightward skew, indicative of a long but thin tail at high injury propensity values. The BSORP model indicates the double modality, but imposes the condition of long and thin tails at both very low and very high injury propensity values. The BYJNORP model, on the other hand, indicates a sharp spike at the low injury propensity range and a much more tempered spike at high injury severity values, with a long and thin tail only at high injury propensity values. That is, our results indicate that, after controlling for observed determinants of injury severity, unobserved factors tend to substantially increase the possibility of no or possible injury severity and moderately increase the possibility of non-incapacitating-to-severe injury severity, a trend that none of the other models are able to replicate.

The estimation results for the four models are presented in Table 1. The first four numeric columns of the table include all exogenous variables that were statistically significant at the 90% confidence level in at least one of the four models. We do so to highlight the potential pitfalls in terms of exogenous variable effects if a more restricted normal error distribution is used rather than the BYJN error distribution. Doing so also allows us to use the more powerful likelihood ratio test to compare models rather than use non-nested likelihood ratio tests. The last column of the table presents the final specification retained for our proposed BYJNORP model. Interestingly, weather/lighting conditions and time-of-day of crash did not turn up statistically significant even at the 80% confidence level in any of the models. In the rest of this section, we briefly discuss the results, though the intent here is not on substantive explanations but more on the comparison of the results across models to demonstrate the value of the proposed model. We should point out here that all the models took about the same time (less than 10 seconds) to estimate as the ordered-response model, demonstrating the ease with which the proposed bimodal distribution may be implemented relative to the more computationally expensive discrete mixtures-of-normals approach to generate bimodality.

In all the ordered-response models, among characteristics of the most severely injured person, the injury risk propensity is elevated if the most severely injured person in the crash is an individual under the age of 13 years or over the age of 60 years or is a woman. These results are quite expected, given the smaller body frames/muscle masses of young children and women (Bose et al., 2011), and the lower bone densities as one ages (Kabli et al., 2020). Individuals of Black origin involved in a crash (as the most severely injured individual), per the results of all the non-BYJN models, are likely to be associated with a higher injury severity propensity than individuals of other races. This is consistent with the findings from earlier studies (see, for example, Adanu and Jones, 2017 and Haddad et al., 2024) that attribute this result to more limited access to advanced safety features prevalent in newer and safer vehicles (Hanks et al., 2018) and disparities in healthcare access/response times (Hanchate et al., 2019). However, the race of the most severely injured person in a crash does not turn up statistically significant in the proposed BYJNORP model. Overall, because the BYJNORP models (the last two columns) accommodate the sharp peak at the lower end of the injury severity spectrum (thus capturing the high fraction of injury severity at the no/possible injury levels), it is able to better distinguish between low and high injury severity propensities based on the characteristics of the most severely injured person, accentuating the differences in injury severity propensity based on age and gender but tempering the difference based on race (of course, one should be careful in comparing the magnitudes of any variable coefficient across different models, because the thresholds are at different points, as discussed later; but a qualitative observation may be made regarding the differences in variable coefficient estimates across the models). A similar trend is observable for the characteristics of the at-fault driver. In all models, if the at-fault driver is over the age of 60 (relative to younger individuals) or is a woman (rather than a man), the result is a lower propensity of injury severity propensity, which may be attributable to the more cautious driving among older individuals and men (see, for example, Song et al., 2021). But these effects again are rather accentuated (and estimated with higher degree of precision) in the BYJNORP models. In contrast, while the non-BYJNORP models all indicate a rather non-intuitive lower injury propensity if the at-fault driver is under the influence of alcohol, the BYJNORP model does not evidence such a non-intuitive result.

Among the intersection-level and CBG-of-intersection location variables, the control type at the intersection, functional class of approach roads, and number of lanes of each approach all did not have any influence at any reasonable level of significance in our analysis. The average annual daily traffic has a positive influence on injury severity propensity, though fades in statistical significance especially in the BYJNORP model (and is dropped in the final BYJNORP specification). A similar result is obtained for the effects of intersection density in the CBG of intersection location and the proportion of individuals residing in the CBG of intersection location who commute by driving (both of which may be viewed as exposure measures, but are statistically insignificant in the BYJNORP model). Notably, however, at the CBG level of intersection location, the fraction of industrial and agricultural land use, and the proportion of low-income households, appear as statistically significant factors. In particular, crashes in CBGs with a higher fraction of industrial and agricultural land use tend to result in higher propensity of injury severity, presumably due to higher speed limits, large farm vehicles, or more conflict points. Crashes in lower-income CBGs also increase injury severity propensity, potentially due to well-established infrastructure deficiencies in low-income areas relative to high-income areas (Haddad et al., 2023).

Finally, the thresholds, while not having any substantive behavioral interpretation, map the latent propensities underlying injury propensity to the observed injury severity level. But the substantially higher range of the threshold values is apparent in the BYJNORP model, because it accommodates the sharp spike at the lower propensity level and a right skew at the upper end, allowing a better spread of the thresholds while still fitting the observed injury severity levels.

3.4. Data Fit Measures

We compare our proposed BYJNORP model with the other three restrictive versions, and then also the BYJNORP with the final specification version of the same model, using nested likelihood ratio tests. For completeness, we also compute the adjusted likelihood ratio index of each model with respect to the log-likelihood at zero (log-likelihood $L(0)$ for the model with equal shares model) and the log-likelihood at sample shares (that is, the log-likelihood $L(c)$ for the model with only thresholds) as follows:

$$\bar{\rho}_0^2 = 1 - \frac{L(\hat{\boldsymbol{\theta}}) - M_0}{L(0)} \text{ and } \bar{\rho}_c^2 = 1 - \frac{L(\hat{\boldsymbol{\theta}}) - M_c}{L(c)}. \tag{19}$$

In the above equation, $L(\hat{\boldsymbol{\theta}})$ represents the log-likelihood function at convergence, M_0 represents the total number of parameters in the model, and M_c is the number of parameters excluding the thresholds in the model.

The results are provided in Table 2. The $\bar{\rho}^2$ values for our proposed model are better than that for all the other models. The likelihood ratio (LR) tests (when the proposed model is compared to the three nested and restricted versions) yield values that are higher than the critical chi-squared table values at any reasonable significance level (at the respective degrees of freedom). Further, the final specification for the BYJNORP model in the final column of the table cannot be rejected in comparison with the BYJNORP model with all variables that turned out to be statistically significant in one of the other models. This is also evidenced in the final model of Table 2 having the best $\bar{\rho}^2$ values of all the models.

To confirm that the superior data fit of the proposed BYJNORP model is not simply an artifact of overfitting, we undertake further data fit tests using market segment prediction tests (see Ben-Akiva and Lerman,1985, page 208) in which we compare the implied predictive log-likelihood and the average probability of correct prediction from the proposed model with other restrictive models in each of many market segments. To conserve on space, Table 3 presents these data fit statistics for six market segments based on selected variables. For the predictive log-likelihoods, we use informal predictive likelihood ratio (IPLR) tests (please see the third numeric column of Table 3) to compare data fits. In this regard, we present the statistics only for the first four models presented in Table 3, because these can be compared with the IPLR tests. The fit statistics for the final refined BYJNORP model (final model in Table 2) are almost identical to those of the fourth model in Table 3). In each market segment, the predictive log-likelihood is better for our proposed model, and the predictive informal likelihood ratio tests reject all the restricted models in favor of our proposed model The average probability of correct prediction is also higher for our proposed model. These observations provide additional support and validation that the joint model indeed offers an improved robust data fit that is not simply an artifact of overfitting.

4. CONCLUSIONS

In this paper, we have proposed a bimodal flexible continuous parametric distribution that allows for asymmetry through a combination of an approach to generate bimodality and using a YJ-transformation. A number of properties of the proposed distribution are stated and proved. We also extend the approach to develop a multivariate version of the proposed bimodal distribution. A convenient way to generate random variates from this multivariate density is proposed.

We demonstrate an application of the proposed distribution to model injury severity using an ordered-response formulation and data drawn from the Texas Department of Transportation (TxDOT) crash database of two-vehicle crashes at intersections recorded between January 1, 2018, and December 31, 2019. Supplemental data from many other sources are added to include information on road network features and land-use/demographic distribution of the Census Block Group (CBG) of the crash location. The results indicate the benefits of employing the proposed distribution (rather than the typically used normal distribution or skewed univariate versions or a symmetric bimodal distribution) for better characterizing the effects of variables on injury severity propensity, avoiding the unintuitive effects estimated by the other models. In terms of data fit too, the model employing the new proposed error distribution performs vastly superior to the other more restricted versions, while maintaining an estimation time that is about the same order as that of the simple ordered-response model used commonly in injury severity modeling.

The proposed distribution may be applied to a number of different econometric modeling contexts in both a univariate and multivariate context, and in a whole variety of fields to consider bimodal asymmetry in stochastic distributions. In this paper, we have applied the distribution to a relatively simple ordered-response formulation. Future research can extend the use of the distribution to more advanced ordered-response and unordered-response formulations, including for kernel error distributions and random coefficients on exogenous variables. Also, finite discrete mixtures of the proposed asymmetric bimodal distributions may be considered for clustering/classification, though the estimation stability and computational intensity of such an approach is sure to be even more of a challenge than the current use of finite discrete mixtures of unimodal distributions.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Department of Transportation through the Center for Understanding Future Travel Behavior and Demand (TBD) (Grant No. 69A3552344815 and No. 69A3552348320). The author is grateful to Lisa Macias for help in formatting this document.

REFERENCES

Abay, K.A. (2013). Examining pedestrian-injury severity using alternative disaggregate models. *Research in Transportation Economics*, 43(1), 123–136.

Adanu, E.K., and Jones, S. (2017). Effects of human-centered factors on crash injury severities. *Journal of Advanced Transportation*, 2017, e1208170.

Arellano-Valle, R.B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33(3), 561–574.

Azzalini, A. and Capitanio, A. (2013). *The Skew-Normal and Related Families*. Cambridge University Press, Cambridge (Institute of Mathematical Statistics Monographs).
Atkinson, A.C., Riani, M., and Corbellini, A. (2021). The Box–Cox transformation: Review and extensions. *Statistical Science*, 36(2) 239–255.
Balan, L., and Paleti, R. (2018). Modified mixed generalized ordered response model to handle misclassification in injury severity data. *Transportation Research Record*, 2672(30), 53–63.
Ben-Akiva, M.E., and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.
Bhat, C.R. (1997). Work travel mode choice and number of nonwork commute stops. *Transportation Research Part B*, 31(1), 41–54.
Bhat, C.R. (2020). Consumer choice modeling: The promises and the cautions. *Mapping the Travel Behavior Genome*, Chapter 5, 63–80, edited by K.G. Goulias and A.W. Davis, Elsevier.
Bhat, C.R. (2024). Transformation-based flexible error structures for choice modeling. *Journal of Choice Modelling*, 53, 100522.
Bhat, C.R., and Sidharthan, R. (2012). A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B*, 46, 817–833.
Bhat, C.R., Mondal, A., and Pinjari, A.R. (2025). A flexible non-normal random coefficient multinomial probit model: Application to investigating commuters' mode choice behaviors in a developing economy context. *Transportation Research Part B*, 195, 103186.
Bose, D., Segui-Gomez, S., Maria, and Crandall, J.R. (2011). Vulnerability of female drivers involved in motor vehicle crashes: An analysis of US population at risk. *American Journal of Public Health*, 101(12), 2368–2373.
Chapman, J., Fox, E.H., Bachman, W., and Frank, L.D. (2021). Smart Location Database: Version 3.0 Technical Documentation and User Guide. U.S. Environment Protection Agency. https://www.epa.gov/sites/default/files/2021-06/documents/epa_sld_3.0_technicaldocumentationuserguide_may2021.pdf
Chen, Z., Wen, H., Zhu, Q., and Zhao, S. (2023). Severity analysis of multi-truck crashes on mountain freeways using a mixed logit model. *Sustainability*, 15(8), 6499.
Das, L., Kalkanci, B., and Caplice, C. (2014). Impact of bimodal and lognormal distributions in ocean transportation transit time on logistics costs. Transportation Research Record, 2409(1), 63–73.
deWaal, D., Harris, T. deWaal, A. and Mazarura, J. (2022). Modelling bimodal data using a multivariate triangular-linked distribution. *Mathematics*, 10(14), 2370.
Dong, Y., and Lewbel, A. (2015). A simple estimator for binary choice models with endogenous regressors. *Econometric Reviews*, 34(1-2), 82–105.
Dong, A., Melnykov, V., Wang, Y., and Zhu, X. (2023). Conditional mixture modelling for heavy-tailed and skewed data. *Stat*, 12(1), e608.
Eluru, N. (2013). Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. *Accident Analysis & Prevention*, 55, 1–11.
Eluru, N., Bhat, C.R., and Hensher, D.A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention*, 40(3), 1033–1054.
Fruhwirth-Schnatter, S., and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-*t* distributions. *Biostatistics*, 11(2), 317–336.

Gallaugher, M.P.B, McNicholas, P.D., Melnykov, V., and Zhu, X. (2020). Skewed distributions or transformations? Modelling skewness for a cluster analysis. https://doi.org/10.48550/arXiv.2011.09152, accessed June 11, 2025.

Gómez-Déniz, E., Calderín-Ojeda, E., and Sarabia, J.M. (2025). Bimodal and multimodal extensions of the normal and skew normal distributions. *REVSTAT-Statistical Journal*, 23(2), 253–271.

Gómez-Déniz, E., Sarabia, J.M., and Calderín-Ojeda, E. (2021). Bimodal normal distribution: Extensions and applications. *Journal of Computational and Applied Mathematics*, 388, 113292.

Greene, W.H., and Hensher, D.A. (2010). *Modeling Ordered Choices: A Primer*. Cambridge University Press.

Haddad, A.J., Mondal, A., Bhat, C.R., Zhang, A., Liao, M.C., Macias, L.J., Kyung Lee, M., and Watkins, S.C. (2023). Pedestrian crash frequency: Unpacking the effects of contributing factors and racial disparities. *Accident Analysis & Prevention*, 182, 106954.

Haddad, A., Mondal, A., Eluru, N., and Bhat, C.R. (2024). A novel integrated approach to modeling and predicting crash frequency by crash event state. *Analytic Methods in Accident Research*, 41, 100319.

Hanchate, A.D., Paasche-Orlow, M.K., Baker, W.E., Lin, M.-Y., Banerjee, S., and Feldman, J., (2019). Association of race/ethnicity with emergency department destination of emergency medical services transport. *JAMA Network Open*, 2(9), e1910816.

Hanks, A., Solomon, D., and Weller, C. (2018). Systematic inequality: How America's structural racism helped create the black-white wealth gap. Center for American Progress. https://www.americanprogress.org/article/systematic-inequality/

Huang, P., Xiong, Y., Tang, S., Wang, S., Zeng, Q., and Lee, J-J (2025). Driver injury severity analysis of work zone crashes: A Bayesian hierarchical generalized ordered probit approach. *Transportation Safety and Environment*, 7(1), tdaf001. https://doi.org/10.1093/tse/tdaf001

Jadhav, A., Dhaulakhandi, D., Kumar, S., Malviya, L., and Mewada, A. (2023). Data transformation: A preprocessing stage in machine learning regression problems. In *Artificial Intelligence Techniques in Power Systems Operations and Analysis*, Singh, N., Tamrakar, S., Mewada, A., and Gupta, S.K. (Eds.). Auerbach Publications.

Ji, Y., Jiang, S., Du, Y., and Zhang, H.M. (2015). Estimation of bimodal urban link travel time distribution and its applications in traffic analysis. *Mathematical Problems in Engineering*, 615468. http://dx.doi.org/10.1155/2015/615468

Jintanakul, K., Chu, L., and Jayakrishnan, R. (2009). Bayesian mixture model for estimating freeway travel time distributions from small probe samples from multiple days. *Transportation Research Record*, 2136, 37–44.

Kabli, A., Bhowmik, T., and Eluru, N. (2020). A multivariate approach for modeling driver injury severity by body region. *Analytic Methods in Accident Research*, 28, 100129.

Kleber, C., Giesecke, M.T., Tsokos, M., Haas, N.P., Schaser, K.D., Stefan, P., and Buschmann, C.T. (2012). Overall distribution of trauma-related deaths in Berlin 2010: Advancement or stagnation of German trauma management? *World Journal of Surgery*, 36(9), 2125–2130.

Lee, S.X., and McLachlan, G.L. (2022). An overview of skew distributions in model-based clustering. *Journal of Multivariate Analysis*, 188, 104853.

Lin, T. I., McLachlan, G. J., and Lee, S. X. (2016). Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 143, 398–413.

Mannering, F.L., Shankar, V., and Bhat, C.R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, 11, 1–16.

Marimuthu, S., Mani, T., Sudarsanam, T.D., George, S., Jeyaseelan, L. (2022). Preferring Box-Cox transformation, instead of log transformation to convert skewed distribution of outcomes to normal in medical research. *Clinical Epidemiology and Global Health*, 15, 101043.

McKelvey, R.D., and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4, 103–120.

Mei, G., Qin, Q., and Lin, D.-J. (2004). Bimodal renewal processes models of highway vehicle loads. *Reliability Engineering and System Safety*, 83, 333–339

Michael, J.R., Schucany, W.R., and Haas, R.W. (1976). Generating random variates using transformations with multiple roots. *The American Statistician*, 30(2), 88–90.

Mu, B., and Zhang, Z. (2018). Identification and estimation of heteroscedastic binary choice models with endogenous dummy regressors. *Econometrics Journal*, 21(2), 218–246.

Nasri, M., Aghabayk, K., Esmaili, A., and Shiwakoti, N. (2022). Using ordered and unordered logistic regressions to investigate risk factors associated with pedestrian crash injury severity in Victoria, Australia. *Journal of Safety Research*, 81, 78–90.

Ramsey, K., and Bell, A. (2014). Smart Location Database. Version 2.0 User Guide. U.S. Environment Protection Agency. https://www.epa.gov/sites/default/files/2014-03/documents/sld_userguide.pdf

Rauf, R., von Matthey, F., Croenlein, M., Zyskowski, M., van Griensven, M., Biberthaler, P. et al. (2019). Changes in the temporal distribution of in hospital mortality in severely injured patients—An analysis of the Trauma Register DGU. PLOS ONE, 14(2), e0212095. https://doi.org/10.1371/journal.pone.0212095

Riani, M., Atkinson, A.C., and Corbellini, A. (2023). Automatic robust Box–Cox and extended Yeo–Johnson transformations in regression. *Statistical Methods & Applications*, 32, 75–102, https://doi.org/10.1007/s10260-022-00640-7

Shuster, J. (1968). On the inverse Gaussian distribution function. *Journal of the American Statistical Association*, 63(324), 1514–1516.

Slobin, H. (1927). A theorem on improper definite integrals. *The American Mathematical Monthly*, 34(5), 265–266.

Song, X., Yin, Y., Cao, H., Zhao, S., Li, M., and Yi, B. (2021). The mediating effect of driver characteristics on risky driving behaviors moderated by gender, and the classification model of driver's driving risk. *Accident Analysis & Prevention*, 153, 106038.

Taylor, M.A.P. and Somenahalli, S. (2014). Urban arterial road travel time variability modeling using Burr regression. In *Proceeding of the Transportation Research Board 93rd Annual Meeting*, Paper No. 14-5654, Washington D.C., January.

Teimouri, M. and Nadarajah, S. (2016). On simulating Balakrishnan skew-normal variates. *Computational Statistics and Data Analysis*, 45(11), 3310–3322.

Vila, R., Saulo, H., and Roldan, J. (2021). On some properties of the bimodal normal distribution and its bivariate version. *Chilean Journal of Statistics*, 12, 125–144.

Watthanacheewakul, L. (2021). Transformations for left skewed data. *Proceedings of the World Congress on Engineering 2021 (WCE 2021)*, July 7-9, 2021, London, U.K.

Wu, Q., Song, D., Wang, C., Chen, F., Cheng, J., Easa, S.M., Yang, Y., and Yang, W. (2023). Analysis of injury severity of drivers involved in different types of two-vehicle crashes using random-parameters logit models with heterogeneity in means and variances. *Journal of Advanced Transportation*, 3399631. https://doi.org/10.1155/2023/3399631

Xiong, Y., and Mannering, F. (2013). The heteroscedastic effects of guardian supervision on adolescent driver-injury severities: A finite mixture-random parameters approach. *Transportation Research Part B*, 49, 39–54.
Yasmin, S. and Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis & Prevention*, 59, 506–521.
Yeo, I.K., and Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954–959.

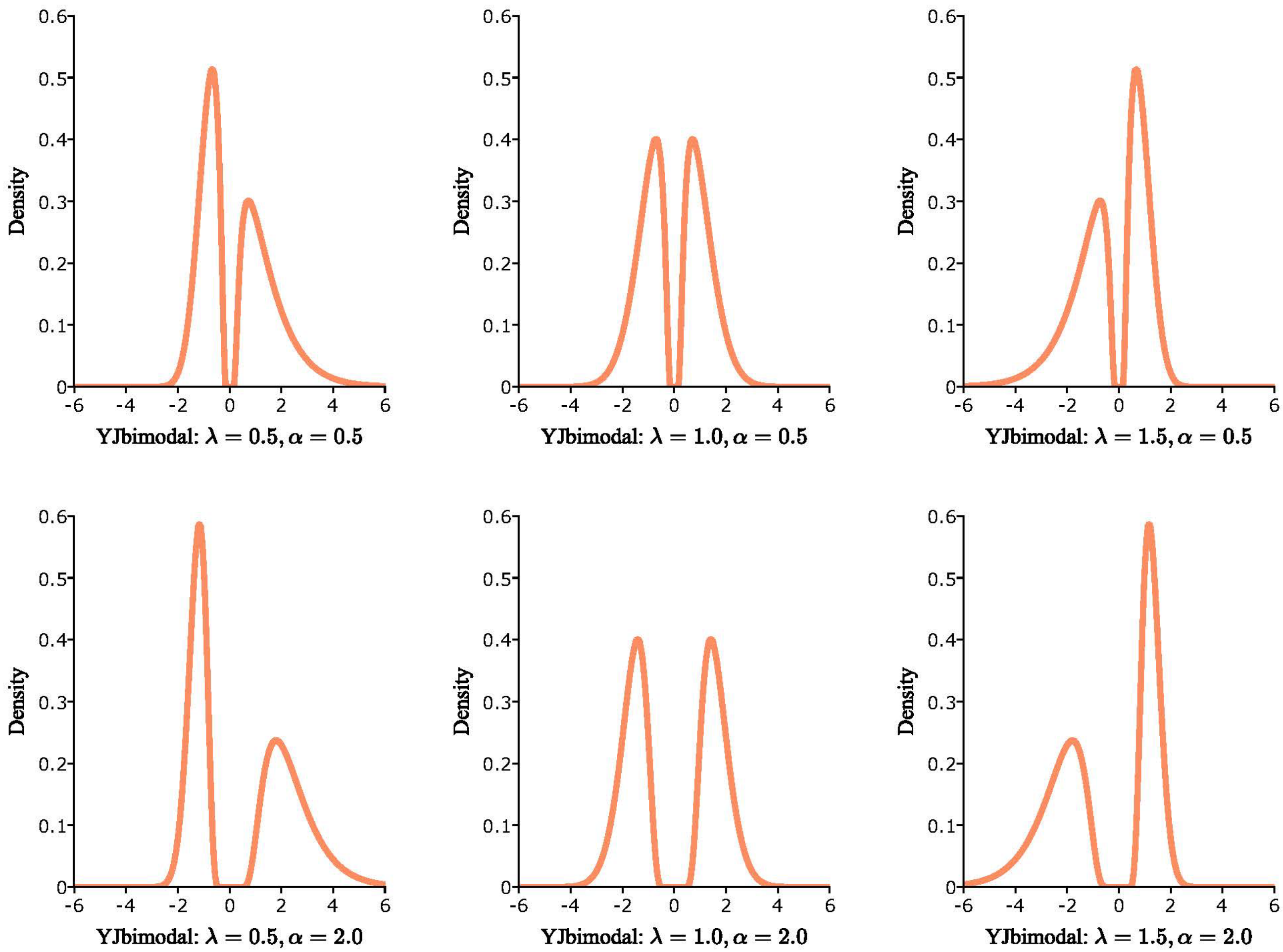

Figure 1. Plots of the proposed BYJN distribution maintaining μ=0 and σ^2=1, but allowing for different values of λ and α

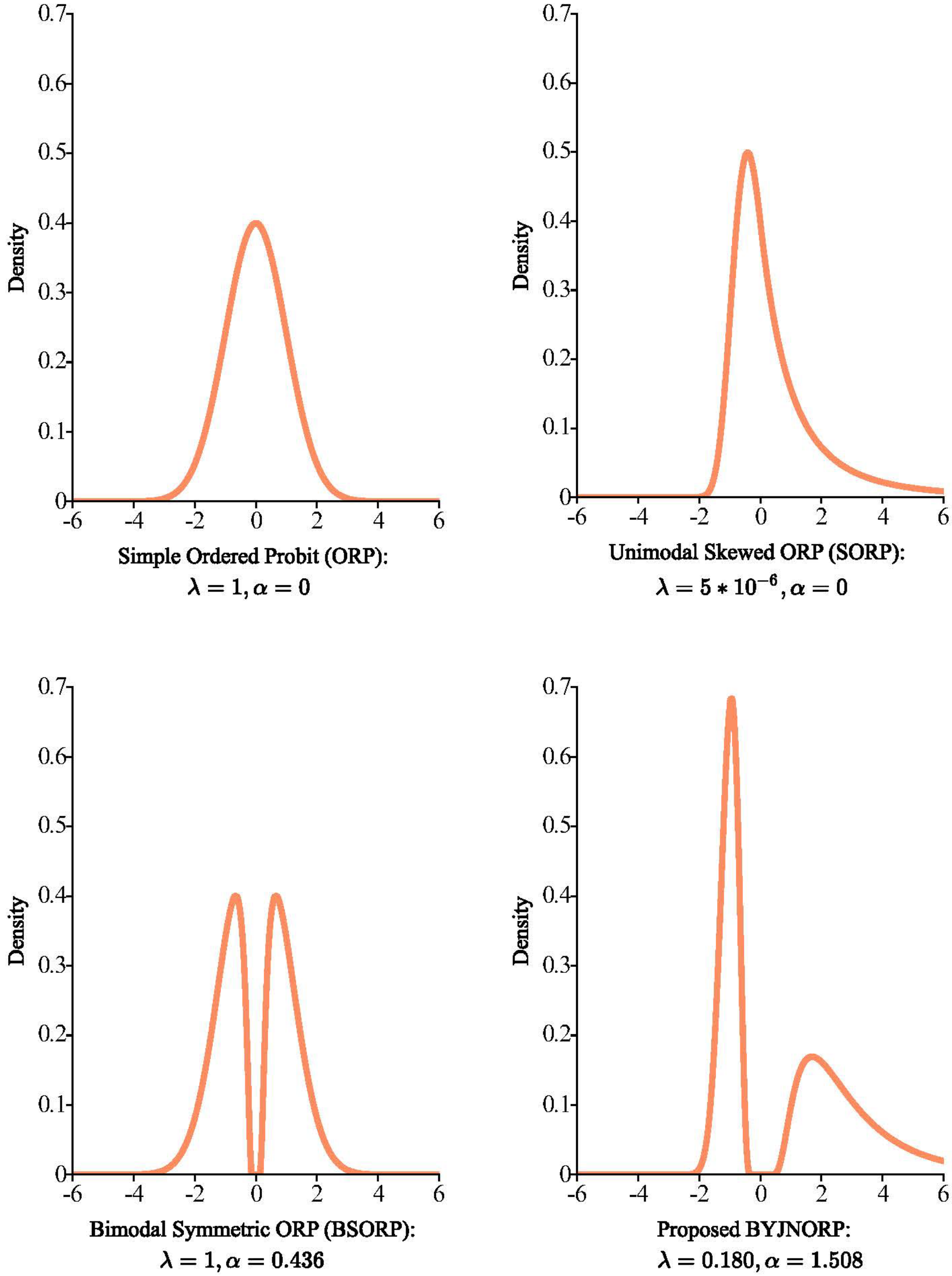

Figure 2. Distribution of kernel error terms in the different models

Table 1. Model results (coefficients represent effects on underlying latent injury severity propensity)

Variables	Ordered Probit (ORP)		Unimodal Skewed Ordered Probit (SORP)		Bimodal Symmetric Ordered Probit (BSORP)		Proposed BYJN Ordered Probit (BYJNORP)		Final Specification for BYJNORP	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Characteristics of most severely injured person										
Individual < 13 years	0.822	9.85	1.057	9.79	0.994	8.41	1.824	5.76	1.851	6.14
Individual > 60 years	0.391	4.07	0.412	3.47	0.594	7.52	1.109	5.71	1.128	6.68
Female	0.483	8.56	0.810	10.92	0.605	6.63	1.010	8.83	1.015	10.07
Black or African American	0.198	3.34	0.197	3.41	0.114	1.73	0.001	0.01	0.000	-
Characteristics of at-fault driver										
Individual > 60 years	-0.018	-0.17	-0.194	-1.36	-0.323	-3.44	-0.846	-3.91	-0.852	-4.88
Female	-0.228	-4.13	-0.520	-7.66	-0.493	-5.31	-0.886	-8.16	-0.896	-10.01
Under the influence of alcohol	-0.731	-2.90	-1.203	-2.31	-0.126	-2.00	-0.023	-0.22	0.000	-
Crash intersection location variables										
Major approach AADT (in 100,000)	0.066	1.60	0.096	2.28	0.020	0.70	0.022	0.36	0.000	-
Census block group of intersection location										
# intersections per mi^2 (in 100s)	-0.088	-2.14	-0.070	-1.61	-0.036	-1.36	-0.020	-0.40	0.000	-
Fraction of industrial and agricultural land-use	0.401	3.06	0.260	2.08	0.333	2.92	0.494	2.47	0.561	3.65
Proportion of low income households	1.740	2.22	1.295	1.60	1.559	2.55	3.390	3.75	3.400	3.89
Proportion of individuals commuting by car	0.272	1.79	0.262	1.33	0.198	1.76	0.170	0.76	0.000	-
λ (YJ) parameter (t-statistic computed w.r.t 1.000)	1.000	-	5×10^{-6}	*	1.000	-	0.183	13.81	0.180	13.90
α (bimodality) parameter	0.000	-	0.000	-	0.436	4.41	1.493	3.73	1.508	2.75
Thresholds										
Between no injury/possible injury	0.560	3.52	0.592	2.82	0.050	0.411	0.208	1.07	0.084	0.52
Between possible/non-incapacitating injury	1.257	7.86	1.542	7.23	1.292	9.54	2.881	5.75	2.771	5.11
Between non-incapacitating/severe injury	2.550	15.12	6.068	16.62	2.582	17.05	6.914	7.29	6.841	6.56

A '-' entry in the t-statistics column indicates that the corresponding coefficient is fixed due to identification considerations or because of a restriction imposed by the model or, in the case of the last column of the table, because of restricting the coefficient to zero due to statistical insignificance.

*In parameterized form, the estimated value of λ^* is -12.869 with a standard error of 0.769. This yields an implied value of $\lambda = 2/(1+\exp(\lambda^*)) \approx 5 \times 10^{-6}$ and standard error of λ equal to 4×10^{-6}. Then, with respect to a value of 1, the implied t-statistic for λ is very large of the order of $1/(4 \times 10^{-6})$ or 2.5×10^5.

Table 2. Data fit measures

Metric	Ordered Probit (ORP)	Unimodal Skewed Ordered Probit (SORP)	Bimodal Symmetric Ordered Probit (BSORP)	Proposed BYJN Ordered Probit (BYJNORP)	Final Specification for BYJNORP
Log-likelihood at convergence	-3075.35	-3034.05	-3074.69	-2984.78	-2985.39
Number of non-constant/non-threshold parameters	15	16	16	17	12
Log-likelihood at zero (equal shares)	-3822.01				
Log-likelihood at constants/thresholds only	-3178.35				
Rho-Bar Squared Value (w.r.t. zero) - $\bar{\rho}_0^2$	0.1906	0.2012	0.1905	0.2138	0.2150
Rho-Bar Squared Value (w.r.t. constants/thresholds) - $\bar{\rho}_c^2$	0.0277	0.0404	0.0276	0.0556	0.0569
LR test: Proposed BYJNORP vs. ORP	LR = 181.1> $\chi^2_{(2,0.05)} = 5.90$				
LR test: Proposed BYJNORP vs. SORP	LR = 98.5 > $\chi^2_{(1,0.05)} = 3.84$				
LR test: Proposed BYJNORP vs. BSORP	LR = 179.8> $\chi^2_{(1,0.05)} = 3.84$				
LR test: Proposed vs. Final BYJNORP	LR = 1.22< $\chi^2_{(5,0.05)} = 11.07$				

Table 3. Measures of fit on various market segments of the estimation sample

Market Segment	Most severely injured person is less than 13 years of age				Most severely injured person is more than 60 years of age				Most severely injured person is female			
Measures of Fit	**ORP**	**SORP**	**BSORP**	**BYJN ORP**	**ORP**	**SORP**	**BSORP**	**BYJN ORP**	**ORP**	**SORP**	**BSORP**	**BYJN ORP**
Number of observations	78				230				1492			
Pred. log-likelihood	-96.8	-90.6	-88.7	-76.0	-284.9	-268.7	-276.5	-252.0	-1740.0	-1706.8	-1689.4	-1657.2
Informal Pred. LR Tests												
BYJNORP vs. ORP	41.6 > $\chi^2_{(2,0.05)}$=5.90				65.8 > $\chi^2_{(2,0.05)}$=5.90				165.6 > $\chi^2_{(2,0.05)}$=5.90			
BYJNORP vs. SORP	29.2 > $\chi^2_{(1,0.05)}$=3.84				33.4 > $\chi^2_{(1,0.05)}$=3.84				99.2 > $\chi^2_{(1,0.05)}$=3.84			
BYJNORP vs. BSORP	25.4 > $\chi^2_{(1,0.05)}$=3.84				49.0 > $\chi^2_{(1,0.05)}$=3.84				64.4 > $\chi^2_{(1,0.05)}$=3.84			
Average Probability of Correct Prediction	0.3028	0.3192	0.3614	0.4268	0.3139	0.3313	0.3449	0.3738	0.3410	0.3440	0.3578	0.3693
Market Segment	**Most severely injured person is of Black race**				**At-fault driver is over 60 years of age**				**At-fault driver is female**			
Measures of Fit	**ORP**	**SORP**	**BSORP**	**BYJN ORP**	**ORP**	**SORP**	**BSORP**	**BYJN ORP**	**ORP**	**SORP**	**BSORP**	**BYJN ORP**
Number of observations	399				214				1313			
Pred. log-likelihood	-486.9	-481.0	-476.7	-461.0	-263.8	-261.3	-264.0	-254.4	-1542.7	-1535.8	-1555.5	-1505.4
Informal Pred. LR Tests												
BYJNORP vs. ORP	51.8 > $\chi^2_{(2,0.05)}$=5.90				18.8 > $\chi^2_{(2,0.05)}$=5.90				74.6 > $\chi^2_{(2,0.05)}$=5.90			
BYJNORP vs. SORP	40.0 > $\chi^2_{(1,0.05)}$=3.84				13.8 > $\chi^2_{(1,0.05)}$=3.84				60.8 > $\chi^2_{(1,0.05)}$=3.84			
BYJNORP vs. BSORP	31.4 > $\chi^2_{(1,0.05)}$=3.84				19.2 > $\chi^2_{(1,0.05)}$=3.84				100.2 > $\chi^2_{(1,0.05)}$=3.84			
Average Probability of Correct Prediction	0.3254	0.3306	0.3405	0.3566	0.3145	0.3225	0.3325	0.3429	0.3437	0.3446	0.3454	0.3573