Transformation-Based Generalization of Ordered-Response Probit Models with Dummy Endogenous Regressors

Chandra R. Bhat
The University of Texas at Austin
Dept of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172
Phone: 512-471-4535, Fax: 512-475-8744
Email: bhat@mail.utexas.edu

ABSTRACT

In this paper, we propose a Yeonjoo-Johnson (YJ) transformation to accommodate flexible specifications of stochastic terms in multivariate mixed data models in general, and ordered-response models with binary endogenous explanatory variables (EEVs) in particular. The use of such a flexible parametric distribution lends additional robustness to the maximum likelihood (ML) estimator. The resulting bivariate YJ model is as easy to estimate as a bivariate normal model. More generally, the YJ transformation is an efficient way to capture flexible marginal error distributions, which then can be bound together using an implicit copula approach. The proposed approach can be applied to a number of different univariate and multivariate mixed modeling structures. In a demonstration application, in the current paper, the proposed model is applied to investigate the effect of urban living on walking frequency, considering the choice of urban living as being endogenous to walking frequency.

Keywords: Ordered-Response Model, Mixed Data Modeling, Residential Self-Selection, Walking Frequency, Built Environment Effects

1. INTRODUCTION

Econometric consumer choice models may have a variety of dependent outcome variables, including those that are continuous, grouped, binary, ordered-response, unordered-response (or nominal), count, or multiple- discrete-continuous. Increasingly, to recognize the jointness in decision-making regarding multiple dependent outcomes, there has been an emphasis on multivariate modeling approaches (see Bhat and Mondal, 2022 for a recent review of such approaches). In all of these models, typically referred to as multivariate mixed data models, stochasticity is introduced to recognize that there will always be unobserved individual-specific factors that impact the outcome variable. Such stochasticity is commonly assumed to originate in the context of an underlying latent variable for each outcome variable, which is then appropriately mapped to the actual observed outcome (except for the continuous case when the underlying latent variable is also the actual observed variable). For example, to analyze an ordered-response outcome, a latent preference underlying the outcome is assumed, which is then mapped to the observed ordered outcome based on a thresholding mechanism. Or to analyze a nominal outcome, a latent utility valuation associated with each alternative is assumed, which is then mapped tot he observed unordered outcome based on a utility maximizing decision principle. Stochasticity is then included in the latent preference (for the ordered-response case) or the utility valuation (for the nominal case) in the form of a kernel error term with or without random coefficients on exogenous variables. Joint multivariate models can then be developed through accommodating covariance on the kernel error terms (see, for example, Müller and Czado, 2018, Wei et al., 2019, Jiryaie and Khodadadi, 2019, and Kwon et al., 2022) and/or through multivariate parsimonious factoring mechanisms for the random coefficients (see, for example, Bhat, 2015, Ong et al., 2018, Dias et al., 2020, Kang et al., 2021). In the latter case, the jointness across outcomes is typically generated through the use of a limited number of latent constructs (or factors), themselves being underlying stochastic cumulatives of a battery of indicators representing attitudinal or lifestyle characteristics. If these stochastic latent constructs are used as determinants of the underlying latent variables affecting the endogenous outcomes (either with or without interactions with other exogenous variables), it leads to jointness across the endogenous outcomes.

Regardless of how jointness is developed in the models discussed above, almost all multivariate mixed data models are based on a multivariate normal distribution of the underlying latent variables of the outcomes (imposed either directly on the kernel error terms of each outcome and/or through the kernel error terms embedded in the latent constructs or factors). Unfortunately, the multivariate distribution of the error terms of the dependent outcomes (conditional on other observed exogenous variables) is unknown *a priori*. Purely out of convenience, rather than based on any underlying economic theory, it is common to *a priori* assume a multivariate normal distribution for the error terms. But the use of a multivariate normal distribution when a different multivariate distribution characterizes the error terms can, and in general will, lead to biased parameter estimates, especially of the treatment effects of one endogenous outcome on another (see, for example, Han and Lee, 2019, and Bhat and Mondal, 2022).

The most general way to accommodate an unknown multivariate distribution would be, of course, to use a multivariate nonparametric joint distribution, with some assumptions related to smoothness and regularity imposed on the joint distribution (as in Gallant and Nychka, 1987, Vytlacil and Yildiz 2007, or Chesher and Rosen 2013). Such methods, however, quickly get extremely profligate in parameters because of series-based or similar approximations of the density function, especially when moving beyond a univariate distribution to a multivariate distribution (see Chen et al., 2006 and Denzel, 2019). Besides, no clear asymptotic distribution results are available for such models (needed to compute parameter standard errors), and any treatment effects of one endogenous outcome on another are not point-identified (Schwiebert, 2013; Han and Lee, 2019). Similar issues of profligateness and kernel density estimation difficulty can arise when using semi-parametric approaches (such as those of Lewbel, 2000, Dong and Lewbel, 2015, Yildiz, 2013, and Mu and Zhang, 2018). These semi-parametric approaches also provide unbiased estimators only under specific conditions (such as a large support requirement for the "special regressor" in the Dong and Lewbel (2015) approach; see Bontemps and Nauges, 2017).

A convenient alternative to handle skewed or fat-tailed multivariate distributions for the error terms is to use a parametric copula approach. Essentially, rather than a completely non-parametric multivariate distribution, this parametric copula approach breaks down the multivariate distribution into univariate specifications for each marginal distribution that are then tied together using a specific parametric copula dependence structure (see Joe, 1997 and Bhat and Eluru, 2009).. Within this copula approach, the use of non-parametric marginals leads to an infinite-dimensional parameter space and the usual maximum likelihood estimation approach fails to produce any meaningful estimator (Geman and Hwang, 1982). To overcome this, recent studies have considered what is called the sieve maximum likelihood estimation approach (see, for example, Zou et al., 2017, Xu and Lee, 2018; Han and Lee, 2019, Szabo et al., 2020, Huang and Xu, 2022). But this sieve-based approach, while providing clear asymptotic distribution results under usual regularity conditions, continues to be profligate in parameters and is computationally expensive (see Smith et al., 2020). On the other hand, a parametric copula, combined with flexible (but) parametric marginals, has advantages in terms of fewer parameters to estimate and has clear estimation efficiency advantages through the use of a parametric maximum likelihood estimation approach. In fact, earlier studies have shown that, as long as the marginal distributions are specified to be reasonably flexible (allowing for skewness and fat tails), a parametric copula dependence structure is able to very closely mimic any multivariate distribution (Chen, 2006). As further elucidated by Baillien et al. (2022), multivariate skewness and fat tails can be introduced in a multivariate copula either through skewed/fat tail margins, a skewed/fat tail copula, or both. However, as they go on to state, a skewed/fat tail copula over skewed/fat tail marginals substantially increases model complexity and may be considered "overkill". Similarly, Denzer (2019) extensively compare an entire set of different copula dependency structures (including the Clayton copula, the Frank copula, the Gumbel, and the t copula) with the baseline Gaussian copula dependency structure for a binary response model with a binary endogenous explanatory variable (EEV). Based on evaluations of both bias and root mean squared error for the average partial effects (APEs), they find that the Gaussian dependency structure performs very well even when alternate bivariate structures characterize the dependency. In another similar study, Han and Lee (2019), also focusing on a binary response model with a binary EEV, find that a bivariate model with flexible marginals is robust to misspecification of the bivariate dependency structure. Thus, in this paper, we consider the convenient Gaussian copula for the multivariate dependence structure, but build in substantial flexibility in the marginal distribution specification, as discussed below. Doing so ensures robustness, and also enables point identification as well as efficient estimation (Han and Lee, 2019).

1.1. Flexible Parametric Marginal Distributions

Two broad methods are available to generate flexible parametric densities with skew and/or excess kurtosis (fat tails). The first is to use a parametric skew distribution (such as a skew-normal or a skew-t distribution or the broader class of skew-elliptical distributions or even mixtures of skew distributions) and the second is to use a transformation method. In its simplest form, the density in the former approach takes the form of the product of a symmetric density function and a skewing component that typically is the cumulative distribution of the symmetric density (see Lee and McLachlan, 2022). The latter transformation method essentially translates each marginal distribution to a corresponding normal distribution through an implicit multivariate Gaussian copula across the marginals.[1] Of the two, the transformation method affords more flexibility as it is not tied to a specific density function, and can also mimic a whole host of skewed and fat-tailed density functions. As an illustration, Gallaugher et al. (2020) have compared, using Mardia's multivariate skewness and kurtosis metrics (Mardia, 1970) and multiple datasets for cluster analysis, the performance of two types of mixtures of skewed distributions with two transformation approaches (see more on this below). While they undertake their analysis at a multivariate level, their results apply to the univariate marginals that essentially get tied into a multivariate distribution. They conclude that "From the analyses on a variety of datasets…, it appears that no one method consistently outperforms the others and usually the performance is very similar if not identical". Further, as they state, the advantage of the transformation method is that it is much more parsimonious than the skewed approach, because the simple transformation approaches they use perform at least as well as complicated mixtures of skew distributions. The potency of the transformation method relative to the use of the more profligate density mixture approach is also evident in the extensive use of transformation method in the field of data science and analytics as a preprocessing step in classification/regression type models to turn general distributions underlying the outcome variable into a near-normal distribution (see, for example, Jadhav et al., 2023 and Peterson, 2020). Doing so not only improves the specification, but also benefits the power of statistical tests used in data analysis, decreasing the risk of committing Type 1 or Type

[1] Mixtures of univariate skew distributions for each marginal distribution (in the skew approach) or transforming the target density function to a mixture normal distribution (in the transformation approach) provide additional flexibility for each marginal. However, such mixtures entail additional parameters that get particularly difficult to estimate and implement for a high dimensional joint model with many outcome variables.

II errors (Zimmerman, 1998 and Osborne, 2010). Thus, in this paper, we will focus exclusively on the transformation method for generating flexible univariate marginal distributions.

Of course, within the class of transformation methods, there are a whole suite of possible transformations to address skewness and fat tails, including square root transformation, inverse transformation, logarithmic transformation, arcsine transformation, Box-Cox (BC) transformation, and the Yeo-Johnson (YJ) transformation (see Box and Cox, 1964, Yeo and Johnson, 2000, Peterson, 2020, Melnykov et al., 2021). Of these, the BC and YJ transformations have been shown, in analysis with both simulation and real empirical data, to be the best transformations in a wide variety of situations (see Osborne, 2010, Jadhav et al., 2023, Watthanacheewakul, 2021, and Marimuthu et al., 2022). In fact, most other transformations are special cases of the BC transformation for data on the positive half of the real line, depending on the value of the power parameter (for example, the square root transformation is the BC transformation with a power parameter of 0.5; the corresponding power parameters for the inverse and natural logarithm transformation are -1 and 0). Further, the YJ transformation is a simple extension of the restrictive BC transformation (which is applicable only to the positive half of the real line) to the entire real line, which makes it more appropriate for observable dependent outcomes that may take negative values too and to limited-dependent outcome models where the underlying latent variables span the entire real line. Thus, in the current paper, which involves the estimation of a multivariate limited-dependent outcome model (specifically, an ordered-response model with a binary endogenous explanatory variable or EEV), we consider YJ transformations for the latent variables underlying the ordered outcome as well as for the binary EEV.

The YJ transformation for each univariate outcome is then coupled together using an implicit Gaussian copula to provide for a simple and flexible multivariate distribution for the latent variables underlying the multivariate dependent outcomes. Smith et al. (2020) show that using such an approach, which entails an implicit copula using element-wise transformations, considerably simplifies estimation and shows more accurate results than multivariate skew distributions, once again reinforcing the efficacy of the YJ transformation coupled with an implicit Gaussian copula.

1.2. The Current Study in Context

The use of the YJ transformation has only recently been introduced by Bhat and Mondal (BM; 2022) for the flexible specification of stochastic terms in multivariate mixed data models. The application there relates to a joint model of neighborhood living type (a binary variable of high density living versus low density living) and monthly bicycling frequency (an ordinal variable with five categories). The formulation and application in the BM paper is based on an underlying latent construct (environmental consciousness) that is assumed to have a YJ-transformed normal distribution, which affects, and is also correlated with, the assumed normally distributed kernel error terms of the latent variables underlying the neighborhood living type/bicycling frequency outcomes. Their model is akin to a single endogenous non-normally distributed regressor in a joint probit-based binary-ordered response model system with the binary outcome serving as a treatment variable. Unlike BM, in this paper, we specify the error terms of the outcome variables themselves to be YJ-transformed normal variables, resulting in a more flexible (non-probit) binary-ordered response model system (our model can equivalently be viewed as an ordered-response model with an endogenous binary variable). In this regard, the current paper constitutes the first formulation and application of a flexible multivariate non-normal limited-dependent variable model system with element-wise YJ transformation, which can be applied well beyond the specific case of the model system used here for illustration purposes. We then apply our proposed model to an empirical context of residential choice and walking frequency.

2. METHODOLOGY

2.1. YJ Transformation for a Univariate Random Variable

The Yeo and Johnson (2000) or YJ transformation is a widely used approach that enables the transformation of potentially non-normal data to near normality and symmetry. It involves but a single additional transformation parameter, making it an efficient vehicle to achieve parsimony in mixed data modeling. When back-transformed, it has been shown to mimic a host of asymmetric, non-normal, and fat-tailed distributions.

The YJ transformation considers a random variable ε_l and transforms it to an assumed normal random variable G_l (with a mean parameter of μ_l and variance of σ_l^2) on the real line as follows, with an additional parameter $0 < \lambda_l < 2$:

$$G_l \sim N(\mu_l, \sigma_l^2) = t_{\lambda_l}(\varepsilon_l) = \begin{cases} -\dfrac{(-\varepsilon_l + 1)^{2-\lambda_l} - 1}{2 - \lambda_l} & \text{if } \varepsilon_l < 0 \\ \dfrac{(\varepsilon_l + 1)^{\lambda_l} - 1}{\lambda_l} & \text{if } \varepsilon_l > 0 \end{cases} \tag{1}$$

The inverse transformation back from G_l to ε_l is:

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

The notation above of ε_l and G_l is used to make the association easier later on in the modeling system. When 0<λ_l<1, ε_l is skewed to the right with a thick right tail, while if 1<λ_l<2, ε_l is skewed to the left with a thick left tail. The normal distribution is returned for ε_l if λ_l=1. Figure 1 plots ε_l for μ_l=0 and σ_l^2=1, and for different values of λ_l ($0 < \lambda_l < 1$). The plots are restricted to the $0 < \lambda_l \le 1$ range, because the corresponding plots for $1 \le \lambda_l < 2$ are mirror images of the plots shown, except with the skew toward the left. For easy comparison, we have centered the mode for all the resulting plots for ε_l at zero. The plots show the flexibility of the YJ transformation to accommodate different levels of skew and tail thickness, especially given that the skew can be toward the right too (not shown in Figure 1). Also, across different random variables $\varepsilon_1, \varepsilon_2, ..., \varepsilon_L$, the direction and intensity of skew/tail can vary. Then, the different variables $\varepsilon_l \ (l = 1, 2, ..., L)$ can be brought together into a multivariate distribution using an implicit Gaussian copula, as discussed next. For convenience, define $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_L)' \ (L \times 1 \text{ vector})$, $\mathbf{G} = (G_1, G_2, ..., G_L)' \ (L \times 1 \text{ vector})$, and $\mathbf{t}_{\boldsymbol{\lambda}}^{-1}(\mathbf{G}) = \left[t_{\lambda_1}^{-1}(G_1), t_{\lambda_2}^{-1}(G_2), ..., t_{\lambda_L}^{-1}(G_L) \right] \ (L \times 1 \text{ vector})$.

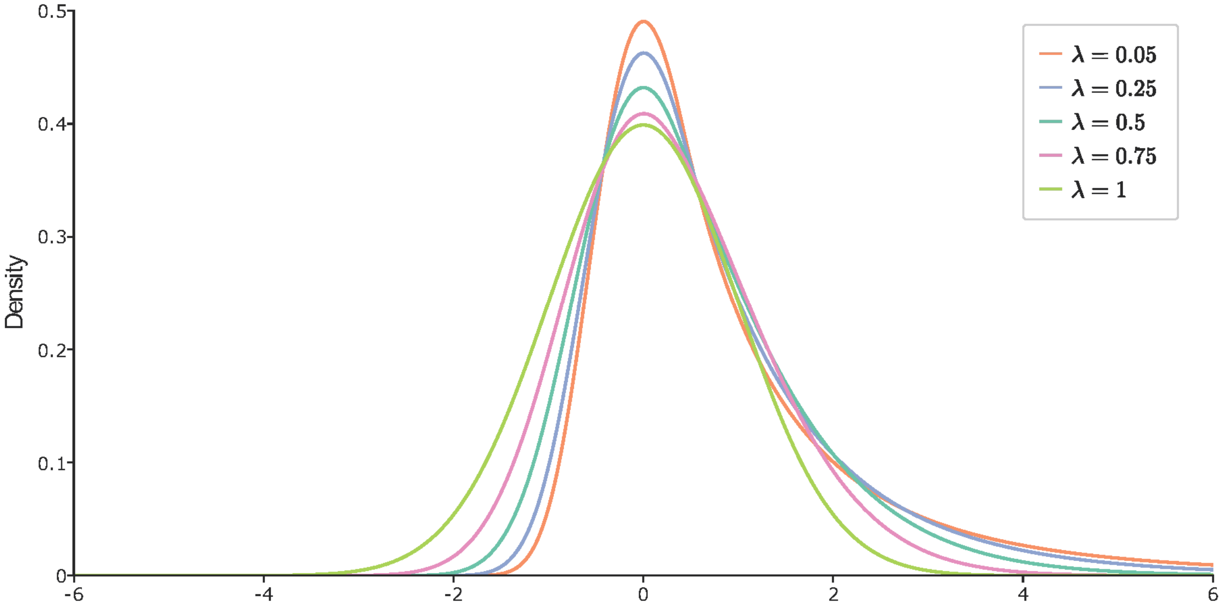

Figure 1: Density of transformed variable for different lambda values

2.2. Construction of a Multivariate Distribution

The most direct way to construct a multivariate distribution with the transformed normal univariate marginals is to assume a multivariate normal distribution for the transformed vector $\mathbf{G}$ (that such an approach is equivalent to an implicit Gaussian copula approach is discussed in Smith et al. (2020), Loaiza-Maya et al. (2022), and Bhat and Mondal (2022). Specifically, we may write the cumulative distribution function for the vector $\boldsymbol{\varepsilon}$ as follows:

$$\begin{aligned} H_{\boldsymbol{\varepsilon}}(\mathbf{z}) &= \text{Prob}(\boldsymbol{\varepsilon} < \mathbf{z}) = \text{Prob}\left[\varepsilon_1 < z_1, \varepsilon_2 < z_2, ..., \varepsilon_L < z_L\right] \\ &= \text{Prob}\left[\left(t_{\lambda_1}^{-1}(G_1)\right) < z_1, \left(t_{\lambda_1}^{-1}(G_2)\right) < z_2, ..., \left(t_{\lambda_1}^{-1}(G_L)\right) < z_L\right] \\ &= \text{Prob}\left[G_1 < t_{\lambda_1}(z_1), G_2 < t_{\lambda_2}(z_2), ..., G_L < t_{\lambda_L}(z_L)\right] \\ &= \text{Prob}(\mathbf{G} < \mathbf{g}),\ \mathbf{g} = (g_1, g_2, ..., g_L)',\ g_l = t_{\lambda_l}(z_l),\ l = 1, 2, ..., L \\ &= F_L\left[\mathbf{g}; \boldsymbol{\mu}, \boldsymbol{\Omega}\right],\ \boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_L)',\ \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1L} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1L} & \sigma_{2L} & \cdots & \sigma_L^2 \end{bmatrix}. \end{aligned} \tag{3}$$

$F_L\left[.; \boldsymbol{\mu}, \boldsymbol{\Omega}\right]$ in the equation above is the multivariate normal distribution function of dimension L with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Omega}$. The net result is that we have now transformed the multivariate distribution of $\boldsymbol{\varepsilon}$ to the multivariate normal distribution of $\mathbf{G}$ through the use of the element wise YJ-transformation parameters λ_l. To be noted also is that the elements of the mean vector $\boldsymbol{\mu}$ and the diagonal elements of the covariance matrix $\boldsymbol{\Omega}$ are simply the parameters of the normalized marginal components, so that the only remaining elements in the multivariate dependence structure characterizing the dependence structure of $\boldsymbol{\varepsilon}$ correspond to the off-diagonal covariance elements of $\boldsymbol{\Omega}$. Of course, using the properties of the multivariate normal distribution, we can finally write:

$$H_{\boldsymbol{\varepsilon}}(\mathbf{z}) = F_L\left[\mathbf{g}; \boldsymbol{\mu}, \boldsymbol{\Omega}\right] = \Phi_L\left[\boldsymbol{\omega}^{-1}(\mathbf{g} - \boldsymbol{\mu}), \boldsymbol{\Omega}^*\right] \text{ with } \boldsymbol{\Omega}^* = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}, \tag{4}$$

where $\boldsymbol{\omega}$ is the diagonal matrix containing the square root of the variance elements of $\boldsymbol{\Omega}$. For completeness, we also write the corresponding density function as follows:

$$h_{\boldsymbol{\varepsilon}}(\mathbf{z}) = \frac{\partial H_{\boldsymbol{\varepsilon}}(\mathbf{z})}{\partial \mathbf{z}'} = \frac{\partial \Phi_L\left[\boldsymbol{\omega}^{-1}(\mathbf{g} - \boldsymbol{\mu}), \boldsymbol{\Omega}^*\right]}{\partial \mathbf{g}} \times \left|\frac{\partial \mathbf{g}}{\partial \mathbf{z}'}\right| = \left(\frac{\phi_L\left[\boldsymbol{\omega}^{-1}(\mathbf{g} - \boldsymbol{\mu}), \boldsymbol{\Omega}^*\right]}{\prod_{l=1}^{L} \omega_l}\right) \times \prod_{l=1}^{L} \left(\left|z_l\right| + 1\right)^{\text{sgn}(z_l)(\lambda_l - 1)}, \tag{5}$$

where ω_l refers to the lth diagonal element of $\boldsymbol{\omega}$. $\text{sgn}(z_l)$ take the value of 1 if z_l is positive, the value of -1 if z_l is negative, and the value of 0 if z_l is zero.

2.3. Estimation of Limited-Dependent Variable Models with EEVs

Two primary approaches have been used in the literature to estimate limited dependent variable models (including ordered-response models) with one or more endogenous regressors. The first is the control function (or two stage residual inclusion) method (see Rivers and Vuong 1988, Blundell and Powell 2004, Terza et al., 2008, Petrin and Train, 2010, Wooldridge, 2015, and Terza, 2018), and the second is the maximum likelihood (ML) approach (see Heckman, 1978, Amemiya, 1978, Bhat and Sardesai, 2006, Brownstone and Fang, 2014, Kang and Lee, 2014, Bhat et al., 2014, and Haddad et al., 2023). The first method is, however, applicable only for continuous EEVs, and not for EEVs that have a limited range (including discrete EEVs). The control function approach, as suggested by Terza et al. (2008), may work reasonably okay with limited-dependent EEVs when the endogeneity intensity is small (see Woolridge, 2015). But, as indicated by Wan et al. (2018), Mu and Zhang (2018) and Denzer (2019), such a control function approach requires the limited-dependent EEV to be a strictly increasing function of the first stage error. In other words, the first stage error (obtained as the difference between the limited-dependent EEV and its expected value in the population) is not independent of the exogenous variables in the EEV model, a necessary condition for the control function to be a consistent estimator. Thus, the control function is not, in general, appropriate for models with limited dependent EEVs.

The second ML method is predicated on the correct specification of the multivariate error structure underlying the overall limited dependent variable system (including the main outcome and the limited-dependent EEVs). It is typical to consider a multivariate normal distribution in ML applications, though deviations from such a distribution can lead to inconsistent estimation, especially in the effect of the limited-dependent EEV (sometimes referred to as the treatment effect when this happens to be a discrete EEV). This is where flexible parametric multivariate distributions, particularly the marginal distributions as discussed earlier, can be substantially beneficial, while also being parsimonious in parameters and efficient in estimation. At the same time, because the ML estimator is a single-step estimator, standard errors are immediately available for all parameters. Further, the ML approach provides information on the joint (and marginal) distribution of the error terms.[2,3]

In this paper, we focus on the case of limited dependent variable models with limited-dependent EEVs, and thus use the ML method for estimation, but with flexible parametric error distributions based on the YJ transformation, as discussed next in the specific context of an ordered-response model with a discrete EEV.

[2] Even when the control function approach is valid, as in the case when the EEV is continuous (rather than being limited-dependent), the maximum likelihood (ML) approach we develop here "chips away" at any robustness advantage that the two-stage control function approach may hold (in models with a continuous EEV, the control function approach does not rely on strong parametric assumptions as needed by the single-stage maximum likelihood approach; see Wooldridge, 2015). But, by using flexible parametric assumptions rather than the typical rigid normal parametric assumption for each of the EEV and main outcomes, our ML approach is much more robust and retains the many advantages of the single step ML estimation. In any case, the control function approach is not even valid for the type of models of interest in the current paper.

[3] As discussed at length by Greene (2017) and Wooldridge (2010), there is no difference in the likelihood function development for model systems with or without discrete EEVs. Thus, while the model considered in this paper is for a limited-dependent outcome with a discrete EEV, the procedure we develop is immediately generalizable to a whole host of other multivariate mixed data model types, including sample selection models, endogenous switching models, multivariate mixed data models, and revealed preference-stated preference models, all of which currently typically use a restrictive multivariate error structure (see, for example, Lee, 1983, Hamed and Mannering, 1993, Bhat, 1997, Bhat and Sardesai, 2006, Kim and Brownstone, 2013, and Heydari et al., 2020).

2.4. Ordered Response Model Structure with Discrete EEV

Ordered response model systems are appropriate when analyzing ordinal discrete outcome data. Examples include ratings data or Likert-scale type attitudinal/opinion data. The typical modeling approach considers the observed ordinal outcome as censored (or course) partitioning or thresholding of an underlying continuous random variable that is endowed with a natural ordering (see McKelvey and Zavoina, 1971 and Bhat, 1997). Greene and Hensher, 2010 provide a comprehensive history and treatment of the ordered-response model structure, which has been applied in a variety of fields, including (but not limited to) sociology, biology, marketing, and transportation fields.

In the usual form of the ordered-response model structure with an EEV, we write two equations, one for the EEV and the other for the ordered-response model:

$$\begin{aligned} &y_1^* = \boldsymbol{\beta}'\boldsymbol{x}_1 + \varepsilon_1,\ y_1 = 1 \text{ if } y_1^* \geq 0; y_1 = 0 \text{ if } y_1^* < 0 \\ &y_2^* = \boldsymbol{\gamma}'\boldsymbol{x}_2 + \delta y_1 + \varepsilon_2, y_2 = k \text{ if } \psi_{k-1} \leq y_2^* < \psi_k; k = 1,2,...K;\ \psi_0 = -\infty, \psi_K = +\infty \end{aligned} \tag{6}$$

The latent propensity y_1^* determines the EEV outcome y_1, while the partitioning of the latent propensity y_2^* determines the main ordered-response outcome y_2. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are vectors of exogenous variables (including a constant in each). $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are coefficients to be estimated, as is the treatment effect parameter δ.[4] The elements of the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed independent of the stochastic elements ε_1 and ε_2, but ε_1 and ε_2 are correlated. ψ_k is the upper bound threshold for ordinal level k for the outcome y_2 ($\psi_0 < \psi_1 < \psi_2 ... < \psi_{K-1} < \psi_K$; $\psi_0 = -\infty$, ψ_1=0, $\psi_K = +\infty$). For future use, define $\boldsymbol{\psi} = (\psi_2, \psi_3, ..., \psi_{K-1})'$. For identification, we also maintain the usual exclusion restriction that there is at least one variable ("instrument") that is contained in the vector $\boldsymbol{x}_1$ but does not appear in the vector $\boldsymbol{x}_2$. While there has been some confusion in the literature on whether such an exclusion restriction is necessary in multivariate limited-dependent variable models of the type in Equation (1) (see, for example, Wilde, 2000 and Rhine et al., 2006), Han and Lee (2019) show that such an exclusion restriction will, in general, be necessary and sufficient for identification of the model parameters. While the parameters may be identified even without the exclusion restriction in the specific case of the presence of a common continuous exogenous variable (but not a binary variable) in both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ for a bivariate normal distribution of ε_1 and ε_2, only weak identification is suggested (at the very best) even in this case. And for more flexible marginal distributions for ε_1 and ε_2, as in the current paper, the exclusion condition is necessary. Also, in

[4] The non-constant coefficients in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ themselves can be considered as random coefficient vectors to accommodate unobserved heterogeneity in the sensitivity to the exogenous variables in $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively. Doing so would not substantially change the basic formulation presented here, except that the resulting mixing would need appropriate integration. In this paper, we assume the non-constant coefficients in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to be fixed to maintain focus on the non-normality distributions of the kernel error terms, the main emphasis of the current paper.

the case when flexible marginal distributions for ε_1 and ε_2 are used (either in parametric or non-parametric form), the findings from Han and Lee (2019) hold that, for point identification of the parameters, a restrictive dependence structure has to be specified for the joint distribution of ε_1 and ε_2, even after allowing for an exclusion restriction. This restrictive dependence structure is satisfied by our use of the Gaussian copula with flexible parametric marginals.[5]

Next, write ε_1 and ε_2 in terms of their YJ-transformed normally distrtibuted counterparts G_1 and G_2 as in Equation (1), and then construct a bivariate normal distribution for G_1 and G_2 as a special case of Equation (3):

$$\varepsilon_1 = t_{\lambda_1}^{-1}(G_1),\ G_1 \sim N(0,1); \varepsilon_2 = t_{\lambda_2}^{-1}(G_2),\ G_2 \sim N(0,1), \text{ and}$$
$$\mathbf{G} = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} \sim BVN(\mathbf{0},\mathbf{\Omega}^*), \text{ where } \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \mathbf{\Omega}^* = \left[\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]. \qquad (7)$$

As can be noted from above, the YJ transformation in our model is applied to the error terms. In the context of Equation (6), the transformations are thus applied to the conditional distributions $y_1^* \mid \boldsymbol{x}_1$ and $y_2^* \mid (\boldsymbol{x}_2, y_1)$. The emphasis is on allowing the error distribution to potentially have a non-normal distribution. In contrast, in earlier transformation applications used with continuous observed dependent outcome variables (as discussed in Section 2.1; see also Atkinson et al., 2021), the transformations have been applied directly to the observed outcome variables to account for non-normality and heteroskedasticity in the observed dependent outcome. Similarly, Guerrero and Johnson (1982) applies a Box-Cox transformation to the odds ratio (computed from observed proportion data) for each group of individuals. In our case, however, the underlying continuous random variables y_1^* and y_2^* are not observed, and so we proceed by placing an identical transformation (across individuals) on the conditional (on $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$) underlying latent variables (that is, the error terms in Equation (6)).

Another important point of note is our normalization of the transformed variables G_1 and G_2 to be standard univariate normally distributed. Of course, this does not imply that the mean and standard deviation of the original (untransformed) error terms ε_1 and ε_2 will be normalized to a mean of zero and standard deviation of 1. The mean and standard deviation of these error terms will be determined by the YJ parameters λ_1 and λ_2. In effect, λ_1 and λ_2 are estimated so as to characterize the distribution of the error terms (including the mean, the standard deviation, the skew, and tail thickness) that provide the best data fit to the observed discrete outcomes. This approach of normalizing the transformed variables G_1 and G_2 to standard normal distributions and allowing λ_1 and λ_2 to dictate the distribution of the untransformed error terms ε_1 and ε_2 is

[5] These same identification issues also apply to joint models (such as a binary selection equation and an ordered-response equation) that do not have treatment parameters, since the likelihood function take identical forms with or without the treatment parameters in such cases.

compatible with the location and scale invariance of the latent variables y_1^* and y_2^*. Besides doing so also nests the bivariate probit model (with the usual standard normalization on the error terms) as a special case of the proposed model when $\lambda_1 = \lambda_2 = 1$.[6]

2.5. Model Estimation

From Equation (4), we may write the cumulative distribution of ε_1 and ε_2 as follows:

$$H_{\varepsilon_1,\varepsilon_2}(z_1,z_2) = H_{\varepsilon_1,\varepsilon_2}(\mathbf{z}) = F_2\left[\mathbf{g};\mathbf{0},\mathbf{\Omega}^*\right] = \Phi_2\left[\mathbf{g},\mathbf{\Omega}^*\right] = \Phi_2(g_1,g_2,\rho),\ g_1 = t_{\lambda_1}(z_1);\ g_2 = t_{\lambda_2}(z_2). \quad (8)$$

where $\Phi_2(.,.,\rho)$ refers to the standardized bivariate normal distribution with correlation ρ. Next, define $\theta = -\boldsymbol{\beta}'\boldsymbol{x}_1, \varphi_{k,1} = \psi_k - \boldsymbol{\gamma}'\boldsymbol{x}_2 - \delta$, and $\varphi_{k,0} = \psi_k - \boldsymbol{\gamma}'\boldsymbol{x}_2$. The joint probabilities for the two regimes of $y_1 = 0$ and $y_1 = 1$ are:

$$\begin{aligned} P(y_1 = 0, y_2 = k) &= P(y_1^* < 0,\ \psi_{k-1} < y_2^* < \psi_k) = P(\varepsilon_1 < \theta, \varphi_{k-1,0} < \varepsilon_2 < \varphi_{k,0}) \\ &= H_{\varepsilon_1,\varepsilon_2}(\theta,\varphi_{k,0}) - H_{\varepsilon_1,\varepsilon_2}(\theta,\varphi_{k-1,0}) \\ &= \Phi_2\left(t_{\lambda_1}(\theta), t_{\lambda_2}(\varphi_{k,0}),\rho\right) - \Phi_2\left(t_{\lambda_1}(\theta), t_{\lambda_2}(\varphi_{k-1,0}),\rho\right), \text{ and} \end{aligned} \quad (9)$$

$$\begin{aligned} P(y_1 = 1, y_2 = k) &= P(y_1^* > 0,\ \psi_{k-1} < y_2^* < \psi_k) = P(\varepsilon_1 > \theta, \varphi_{k-1,1} < \varepsilon_2 < \varphi_{k,1}) \\ &= \left[H_{\varepsilon_2}(\varphi_{k,1}) - H_{\varepsilon_1,\varepsilon_2}(\theta,\varphi_{k,1})\right] - \left[H_{\varepsilon_2}(\varphi_{k-1,1}) - H_{\varepsilon_1,\varepsilon_2}(\theta,\varphi_{k-1,1})\right] \\ &= \left[\Phi\left(t_{\lambda_2}(\varphi_{k,1})\right) - \Phi_2\left(t_{\lambda_1}(\theta), t_{\lambda_2}(\varphi_{k,1}),\rho\right)\right] - \left[\Phi\left(t_{\lambda_2}(\varphi_{k-1,1})\right) - \Phi_2\left(t_{\lambda_1}(\theta), t_{\lambda_2}(\varphi_{k-1,1}),\rho\right)\right] \\ &= \Phi_2\left(-t_{\lambda_1}(\theta), t_{\lambda_2}(\varphi_{k,1}),-\rho\right) - \Phi_2\left(-t_{\lambda_1}(\theta), t_{\lambda_2}(\varphi_{k-1,1}),-\rho\right) \end{aligned} \quad (10)$$

Now introduce the index q for individuals. Define a set of dummy variables M_{qk} (k=1,2,…,K): $M_{qk} = 1$ if $y_2 = k$, and $M_{qk} = 0$ otherwise. Assuming independence across individuals, the likelihood function for estimation of the parameters in the model system is:

$$L(\boldsymbol{\beta},\boldsymbol{\gamma},\delta,\boldsymbol{\psi},\rho) = \prod_{q=1}^{Q}\left(\left[P_q(y_{q1} = 0, y_{q2} = k)\right]^{(1-y_{q1})M_{qk}} \times \left[P_q(y_{q1} = 1, y_{q2} = k)\right]^{y_{q1}M_{qk}}\right). \quad (11)$$

3. APPLICATION TO URBAN RESIDENCE CHOICE AND WALKING FREQUENCY

3.1. Background

Walking is a travel mode that requires no motorization, is a natural human-learned skill from childhood for most individuals, entails zero monetary cost, can ease traffic congestion on our roadways, and contributes to a reduction of the transportation sector's carbon footprint (see Xia et al., 2013; Longo et al., 2015). At the individual level, walking, as a contributor to physical activity,

[6] There is no known closed form for the moments of the untransformed error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$ as a function of the YJ parameter vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ and the correlation ρ. But these moments can be estimated through numerical integration based on the probability density function in Equation (5).

can provide substantial health benefits, including both mental and physical. From a mental wellness standpoint, walking has known to be a calming technique that releases stress, relieves anxiety, and reduces depression (Pearce et al., 2022; Vetrovsky et al., 2017; Makela and Aktas (2023). The very act of being outdoors and feeling the associated social vibrancy has been noted to buoy spirits and generate a positive mindset (Pearson and Craig, 2014; Jimenez et al., 2021). From a physical wellness standpoint, walking, as part of a broader regimen of regular moderate-intensity physical activity, can help reduce the risk of type 2 diabetes, chronic heart disease, hypertension, cardio-vascular diseases, various forms of cancer, and also other chronic diseases and disorders (Paluch et al., 2022, 2023; World Health Organization (WHO) guidelines on physical activity and sedentary behaviour: at a glance, 2020[7]). As importantly, the benefits of walking can be accrued either through utilitarian walking (that is walking with a specific activity at the destination end) or recreational walking (such as walking without a specific destination, such as around the neighborhood for exercise or walking the dog). During the height of the pandemic, while there was fear of increased exposure to the COVID virus when outdoors (see Lima et al., 2020, Sallis et al., 2020), there was some engagement in "recreational walking" as individuals walked around their homes as a means to take a break from being bottled up within their homes (Przybylowski et al., 2021; de Haas et al., 2020). Further, after the peak of the pandemic was behind us, and vaccinations became widely available, walking became quite prevalent as the fear of contracting COVID from fellow walkers decreased even further. This was particularly the case in specific demographic groups, such as those with high education and income levels (Kyan and Takakura, 2022; Hwang et al., 2023).

An important related issue beyond demographic determinants of walking that has attracted substantial interest in the land use-transportation literature in the past is the differential walking tendency based on residential built environment. Of keen interest in particular over the years has been to disentangle associative effects from causal effects in the residential built environment (BE) influence on active travel behavior (for example, see Van Wee, 2009, Van Acker *et al.*, 2014, Bhat et al., 2016). This is particularly important when using cross-sectional data that entails the joint observation, at a specific point in time, of the residential location of households and activity-travel choices. From an econometric perspective, the "single-point-in-time" observation co-mingles the "true" BE effect of residential location with an associative effect due to the non-random assignment of individuals to residential locations. Of course, if this non-random location assignment can be magically explained entirely through a combination of observed non-travel (including sociodemographic) characteristics and BE attributes, then a traditional travel model for the activity-travel dimension with BE attributes as explanatory variables would be adequate to tease out "true" causal BE effects. However, the more likely situation is the presence of unobserved antecedent personality, attitude, and lifestyle characteristics of individuals/households that simultaneously impact residential location choice and activity-travel behavior. The question then

[7] WHO guidelines on physical activity and sedentary behaviour: at a glance. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO., https://iris.who.int/bitstream/handle/10665/337001/9789240014886-eng.pdf, accessed January 5, 2024.

is whether the co-movement between walking intensity and residential density represents a true "causal" effect of density on walking intensity or simply an associative residential self-selection effect. Bhat and Guo (2007) discuss this residential self-selection issue at length, emphasizing its importance for policy-making. More recently, Guan et al., 2020 provided an exhaustive review of the earlier self-selection literature investigating BE effects on a whole host of travel-related decisions, including walking. While the fraction of the contribution due to associative and "true casual" residential BE effects has varied across studies based on the geographic, cultural, data type, analytic methodology, time-period, and travel outcome of interest, the importance of considering self-selection effects in assessing BE effects has not ever been in question. This underscores the importance of explicitly modeling the jointness in (that is, correlation in unobserved factors impacting) residential living choice and walking behavior.

The importance of considering self-selection effects when examining residential BE effects on walking intensity takes on particular significance in a COVID-changed landscape, where the residential choice process of individuals has seen substantial upheaval. In particular, evidence indicates a substantial move of residences away from urban areas during the pandemic as individuals attempted to avoid contact with others (after all, about 90% of COVID infections worldwide were associated with urban highly walkable areas; see Lima et al., 2021 and Choi and Denice, 2022). Besides, with hybrid work locations (working from the office a few days and working remotely on other days) becoming a more accepted practice relative to pre-COVID times (see Asmussen et al., 2024), other residential factors (such as a desire for more space within homes, better school quality, and lower reported crime neighborhoods) have altered the trade-offs among attributes in residential decisions, contributing further to an exodus from urban to suburban/rural areas (Robbennolt et al., 2023). In this regard, there is a distinct possibility that those who relocated to the suburbs post the onset of the pandemic may have, at least in part, done so specifically to have more of a sense of mental comfort and personal safety to walk, away from the hustle-bustle of crowded urban areas that heighten the threat of contracting COVID (and other similar viruses in the future) and the general stress of walking in crowded locales. Indeed, Paydar and Fard (2021) allude to Alfonso (2005)'s five levels of needs that influence walking decisions, and discuss how the pandemic has altered these five levels, especially by way of elevating the importance of the basic lower-level needs of feasibility (related to personal limitations such as health conditions) and safety (considerations related to virus spread) when making walking decisions. Given that many datasets that capture walking behavior do not collect information on such basic lower-level walking needs, such considerations then can lead to a correlation in unobserved factors between residential choice and walking intensity. In fact, this correlation can be in ways that are even different from a pre-pandemic world. Unlike the usual finding in the pre-pandemic scientific literature that unobserved attitude/lifestyle factors (such as green consciousness) appear to lead to a positive correlation in the error terms of urban living and walking intensity (thus, overestimating the positive effect of the urban built environment if such correlation is ignored), the self-sorting into residential neighborhoods in a post-pandemic world may be such that those with a higher propensity to walk, due to unobserved basic-level walking needs of safety/comfort, may be more resident in non-dense neighborhoods, possibly leading to a negative correlation in the error terms of urban living and walking intensity (thus, underestimating the "true" positive effect of the urban built environment itself). If this latter situation were the case, this would suggest that investing in urban-like walking infrastructure facilities, while being conscious of accommodating basic-level walking needs, would be a good policy instrument to promote walking across the board of residential neighborhoods. Of course, considerations of safety/comfort associated with COVID concerns are also likely to recede with time, and thus it is imperative anyway to account for any transient associative (i.e., correlation) effects when evaluating what are likely to be the "true" residential BE effects in a COVID-disrupted transportation and walking environment.

In the current paper, we contribute to the extant land use-transportation relationship by examining residential BE effects on walking intensity in a post-COVID world. Further, we focus on not-so-young adults (50+ age group) in our analysis, given that some earlier descriptive research (see, for example, Rantanen et al., 2021 in Finland, Shaer and Hagshenas, 2021 in Iran, Yamada et al. 2020 in Japan, and Choe et al. in 2022 in Hong Kong) has indicated the particularly large impact of COVID-19 on the active-travel (including walking) patterns of the older adults. However, unlike these earlier descriptive before-after COVID studies undertaken during the height of the pandemic or just after, and in non-US contexts, we examine the walking behavior of older US adults employing a survey collected in late (October-November) 2022 (when the peak of the pandemic was decidedly in the rear view mirror). The emphasis is on current walking behavior, and we explore the effects of sociodemographic variables, as well as the endogenous effect of residential BE infrastructure, on walking intensity. Such an analysis is of value not just for transportation and social reasons (including reducing loneliness; see, for example, van de Berg et al., 2016), but also for health considerations. In particular, walking has been established to be particularly beneficial to older age adults in terms of reducing the risk of cardio-vascular disease, depression and dementia (Won et al., 2019, Roe et al, 2020, Paluch et al.,2022, 2023). A recent study (Garcia et al., 2023) further indicates that even as low as 10 minutes per day of brisk walking activity prevents premature deaths and lowers the risk of heart disease, stroke, and cancer.

The BE attributes of the residential location of individuals in our study are captured using the single attribute of residential density, given the well-established strong association between density and other BE elements. This is clearly reflected in the long and strong precedent for doing so, as evidenced in the literature (see, for example, Bhat and Singh, 2000, Ewing and Cervero, 2010, Cao and Fan, 2012, Kim and Brownstone, 2013, and Wang et al., 2021). As in most of these earlier studies, we use a binary classification of residential density as living in an urban area (associated with pedestrian-friendly BE) or a non-urban area(associated with a car-centric infrastructure environment) to capture BE effects (this urban/non-urban classification was based on the metropolitan statistical area of residence). The ordered-response outcome in our empirical analysis corresponds to an ordinal scale of weekly walking frequency (in days per week of walking for at least 10 minutes). Residence in an urban area (capturing walk-friendly BE attributes) is included as an *endogenous* outcome in the ordered-response walk frequency equation, to capture residential self-selection effects. The correlation in the error terms of (unobserved factors affecting) the two equations may be positive or negative (as discussed earlier). After all, we do not know what these unobserved factors may be. The important thing is to recognize the potential endogeneity of urban residence to capture the "true" causal effect of urban BE on walking intensity.

3.2. Data and Sample Description

The walking survey of older adults in the US population used in the current paper was undertaken through the Foresight 50+ Consumer Omnibus panel survey, which is a mixed mode survey (online and telephone) funded and operated by the National Opinion Research Center (NORC) at the University of Chicago. The survey constitutes a probability-based panel designed to be representative of the US household population age 50 or older. The survey questions and survey instrument in the Foresight 50+ survey vary by month. The specific walking survey instrument used in the current paper was funded by the American Association of Retired Persons (AARP) and undertaken in July 2022.[8] A sample of 1691 individuals aged 50 years or older was obtained with deliberate oversampling of individuals of non-White origin. The unweighted sample, therefore, includes 44.6% of White, non-Hispanic individuals (the corresponding census statistic is 71% for individuals aged 50 and over), 21.8% of individuals of Black, non-Hispanic origin (the corresponding census statistic is 11%), 9.6% of Asian, non-Hispanic origin (the corresponding census statistic is 6%), and 22.7% of individuals of Hispanic origin (the corresponding census statistic is 12%), and 1.3% of other (including mixed and other non-Hispanic races) race categories (the corresponding census statistic is 1%). The average age in the unweighted sample is 63.7 years (the corresponding census statistic is 65 years). Men were slightly over-represented (52% in the unweighted sample versus 48% in the census). The sample was weighted to the current population survey (CPS) benchmarks based on gender, age, education, and race/ethnicity to develop a representative (weighted) sample of the US adult population 50 years and older.

The skew of the unweighted sample implies that descriptive statistics for the endogenous outcomes of urban residence and walking frequency from the unweighted sample cannot be directly generalized to the U.S. 50+ adult population. However, in the aggregate, the unweighted and weighted samples do not show too much divergence in the endogenous outcomes. The unweighted sample indicates 1461 (86.4%) respondents residing in urban areas compared to 1386 (81.9%) in the weighted sample. Similarly, by way of walking frequency (number of days of walk per week for at least 10 minutes at a time), the unweighted and weighted percentages were as follows: (a) Do not walk any day (19.2 unweighted versus 22.1 weighted), (b) Walk 1-2 days per week (17.0 unweighted versus 17.7 for weighted), (c) Walk 3-4 days per week (27.9 unweighted versus 25.0 weighted), (d) Walk 5-6 days per week (21.1 unweighted versus 18.8 weighted), and (e) Walk 7 days per week (14.8 unweighted versus 16.4 weighted).

[8] Additional details of the survey administration and methodology are available at https://www.aarp.org/pri/topics/health/prevention-wellness/walking-attitudes-habits-adults-50-older/. Accessed November 13, 2023.

In estimation, we use the unweighted sample because the focus of the current paper is on estimating individual-level relationships (how changes in exogenous demographics and the urban variable affect the outcomes). In such analyses, the critical consideration is whether the sampling is dependent (sampling strategy is endogenous to modeled outcomes) or independent (sampling strategy is exogenous to modeled outcomes). Weighting is needed for consistent estimation in the former case, but not the latter case. The sample in our case corresponds to the latter. In this exogenous sampling situation, the unweighted approach provides more precise parameter estimates and is to be preferred . Accordingly, in our model estimations, we use the unweighted approach (see also Wooldridge, 1995 and Solon et al., 2015). Besides, the sample exhibits good variation in the range of sociodemographic variables, allowing us to test a variety of functional forms for these variable effects. For example, age is collected as a continuous variable in the sample with good variation from 50 years to 92 years, which allows us to test linear, non-linear, as well bracket-specific dummy variable specifications for age. In combination, the exogenous sampling approach and the good variation in demographic variables enable us to estimate individual-level relationships that should be applicable to the larger population.

3.3. Model Results

A number of different specifications in terms of variables and the functional forms of variables were attempted in arriving at the final model specification. All variables, except age, are in either bracketed categories (income), or are naturally discrete. Examples of the latter include gender, race/ethnicity, education level, housing tenure, dwelling type, employment status (paid employee, self-employed, unemployed/laid off temporarily, unemployed/looking for job, unemployed/retired, and unemployed/ physically disabled), household structure (single adult, single parent, married or unmarried couple, nuclear family, joint family of related individuals with more than two adults aged 18 or more and at least one individual of 17 years of age or younger, household with more than two adults aged 18 or more and no individual 17 years or younger), and residence region of the U.S.[9] The influence of the bracketed and discrete exogenous variables were tested as dummy variables in the most disaggregate form possible, and progressively combined for parsimony based on statistical tests. For the continuous age variable, alternative functional forms (such as a continuous linear form, a continuous logarithm form, a piece-wise linear form, and a set of dummy variables for different ranges) were tested. At the end the effect of age was best represented in dummy variable form. Further, we examined interaction effects across variables, including between urban and gender, urban and race/ethnicity, and urban and household structure in the walk frequency equation (after accounting for the jointness through correlation effects between the urban residence and walk frequency endogenous outcomes), but none of these came out to be statistically significant.

[9] The region of the U.S. was classified into four regions; Northeast, South, West, and Midwest; based on the Census Division of residence as follows: (a) New England and Mid-Atlantic = Northeast, (b) South Atlantic, West South Central and East South Central = South, (c) Pacific and Mountain=West, (d) East North Central and West North Central = Midwest.

The final model specification is presented in Table 1. The parameters for the urban or non-urban residence model component represent the elements of the $\boldsymbol{\beta}$ vector (that is, the effect of exogenous variables on the propensity to live in an urban region), while those for the walk frequency component refer to the elements of the $\boldsymbol{\gamma}$ vector and the urban "treatment" effect parameter δ (that is, the effect of exogenous variables and "urban" residence on the propensity underlying walking frequency). In Table 1, categories that do not appear in the column for urban living propensity or walk frequency propensity constitute the base categories. A '-' in the table implies that the corresponding coefficient was not statistically significant at even the 10% level of significance.

The results indicate that individuals from higher income households, with higher than a bachelor's degree, and older individuals (80 years or older) are more likely than their peers to be urban residents. In the context of a post-COVID landscape, these are quite interesting and suggest some shifts in residential choice patterns from before. There appears to be more of a move away from urban areas (to non-urban areas) among individuals with lower incomes and with low formal education degrees, in their quest to get better quality housing farther away, facilitated also by increased teleworking potential for erstwhile "blue collar" jobs (see Asmussen et al., 2023). Besides, recent studies have suggested that there has been a tempering in the market potential differential for employability between urban and non-urban areas, which would once again spur moves of low income individuals and those with low formal education degrees toward non-urban areas (Blumenberg and Wander, 2022). The higher tendency of older individuals to be located in urban areas may be tied to the reluctance to move away from current residences relative to younger individuals. Older individuals typically are averse to changes in life rhythms because stability provides a form of mental self-esteem boost (a sense of control and reduced stress/anxiety) for them at a time when their physical self-esteem may be on the decline (Duque et al., 2021).

While some effects on residential location may have shifted in the aftermath of the COVID pandemic, others do not appear to have. In particular, individuals who identify as white non-Hispanic and those who live in an owned single-unit house (that is, do not live in rented apartment dwellings) are unlikely urban area residents. This is tied to non-white individuals having limited opportunities of residential mobility over decades, containing them to urban areas with a high supply of rental apartment dwellings. This may be traced to the long history of racial discrimination in the housing sector in the United States, offering more opportunities to white families to secure government-insured mortgages, while non-white families were not afforded the same level of benefits (Faber, 2020; Bhat et al., 2022). Also, individuals in the south and midwest of the country are less likely to be located in urban areas, as these parts of the country have large swaths of land with more spatial dispersion of residences.

Table 1: Main outcome results (coefficients represent effects on underlying latent propensity)

Variables	Urban Neighborhood Living		Walk Frequency (days per week)	
	Coeff.	t-stat	Coeff.	t-stat
Annual Household Income				
$75,000-$149,999	0.375	3.704		
$150,000 and higher	0.884	4.771		
$100,000 and higher	-	-	0.114	1.983
Higher than Bachelor’s Degree	0.307	3.117	-	-
Woman	-	-	-0.124	-2.362
Age				
75 years or older	-	-	-0.231	-2.750
80 years or older	0.475	2.312	-	-
White Non-Hispanic	-0.736	-8.523	-	-
House Owned	-0.353	-2.953	-	-
Apartment Dwelling	0.523	3.390	-	-
Unemployed and Disabled	-	-	-0.397	-3.789
Unemployed and Retired	-	-	-0.124	-2.258
Geographic Residence Region				
South and Midwest	-0.358	-3.942	-	-
West	-	-	0.181	3.090
Correlation between the Normally Transformed Underlying Propensities between Urban Residence and Walking Frequency Error Terms	-0.257 (t-statistic -2.687)			
“True” Causal Effect of Urban Neighborhood Living	NA	NA	0.467	2.540
YJ Parameters (t-statistic computed with respect to the value of 1.000)	1.000*	-	0.526	2.843
Constant and Thresholds				
Constant	1.683	12.731	0.446	2.802
Threshold between walk 1/2 days per week and 3/4 days per week	-	-	0.440	1.934
Threshold between walk 3/4 days per week and 5/6 days per week	-	-	1.168	5.197
Threshold between walk 5/6 days per week and 7 days per week	-	-	2.068	9.053

* YJ parameter for the urban residence error term fixed to one because it was not statistically significantly different from one at any reasonable level of significance.

In terms of walking frequency, individuals from high income households (100K per year or over) are more likely to walk over multiple days of the week relative to individuals from low income households (less than 100K per year), while women tend to have a lower walking propensity compared to men. Both of these results may be tied to time availability, with both low income earners and women well known to be time poor because of work-related and, in the case of women, also home-related responsibilities (see Bernardo et al., 2015, Cerrato and Cifre, 2018, and Mondal and Bhat, 2021). The result that older individuals (75 years and over), and those unemployed due to disabilities or who have retired, have a depressed walking propensity than their peers may be attributed to actual or perceived mobility limitations. From a geographic standpoint, residents from the west of the country are, in general, more predisposed to walk frequently.

As discussed earlier, we control for possible association in the residential choice-walking frequency choices when investigating the "true" causal effect of residential choice on walking frequency. Our results suggest that rural areas are better characterized (than urban areas) by unobserved basic-level walking needs of safety/comfort factors, which also promote walking frequency. This is evidenced by the negative correlation in the error terms between the urban area living propensity and the walk frequency equations. After accommodating for this association, the results show a clear positive "true" causal effect of urban living on walking frequency. If the association effects due to unobserved effects were ignored, the "true" urban effect gets underestimated. In fact, if the association is ignored, the urban effect becomes statistically insignificant (coefficient is 0.0740 with a t-statistic of 0.926). Overall, the results indicate that, in the aftermath of the worst of the pandemic, built environment factors associated with urban residence continue to have an important positive causal influence on walking activity engagement of "not-so-young" adults over the course of the week, consistent with earlier studies (see, for example, Lotfata et al., 2022). At the same time, the negative association between urban living and walk frequency suggest that walking environments need to be designed to feel less crowded (such as through the design of wide sidewalks and a well-distributed network of pedestrian walkways throughout the urban landscape to reduce pedestrian movement clustering), so that basic-level walking needs related to safe health environments are fulfilled.

The YJ parameters indicated that the urban equation error is not statistically significantly different from one at any reasonable level of significance, implying that we cannot reject that the urban error term is indeed normally distributed. So, we constrained the YJ parameter to the value of one. However, the YJ parameter for the walking frequency error term is significantly lower than one, indicating a highly rightward skew; that is, a small number of individuals have a high walk frequency. As we will note in the next section, ignoring this rightward skew has substantial implications for the estimated causal effect of urban form on walk frequency.

Finally, the constants and thresholds toward the end of the table do not have any substantive behavioral interpretation, though they serve the important purpose of mapping the latent propensities underlying the urban residence and walk frequency latent propensities to the corresponding categorical outcomes.

3.4. Data Fit Measures

We compare our proposed bivariate YJ model with three other restrictive versions: (i) a model that imposes error normality for both the urban residence and walk frequency outcomes as well as ignores endogeneity of urban residence (independent normal model), (ii) a model that considers the endogeneity in urban residence, but considers bivariate normality (bivariate normal model), (iii) a model that ignores the endogeneity in urban residence, but considers YJ-transformations for the walk frequency outcome (independent YJ model; as in our proposed model, in this model too we could not reject the hypothesis of normality of the error term in the urban residence outcome), and. All these models are nested within our proposed "bivariate YJ model", making it easy to compare performances using the likelihood ratio test. The relevant likelihood-based data fit measures are provided in the top panel of Table 2. Likelihood ratio (LR) tests (when the proposed model is compared to the three restrictive versions) yield values that are higher than the critical chi-squared table values at any reasonable significance level (at the respective degrees of freedom).

In addition to the likelihood-based data fit measures, we also compute more intuitive non-likelihood based data fit metrics. At a disaggregate level, we compute the average (across individuals) probability of correct prediction (for the joint outcome of urban/non-urban living and walk frequency). At an aggregate level, we compare the predicted and actual (observed) shares in each of the ten combinations of residence type and walk frequency, and compute a weighted average percentage error (WAPE) across all the ten possible combinations. The middle panel of Table 2 presents these non-likelihood data fit measures. The results clearly show that the predictions from our proposed model are closer to the observed values at both the disaggregate and aggregate levels. Indeed, the aggregate match of the predicted and actual values for our proposed bivariate YJ model is remarkable, especially given that the proposed model includes just one additional parameter (corresponding to the error term shape for the walk frequency model) relative to the bivariate normal model. Of course, we will also note that the WAPE for the other models are also not too bad, but the superior fit of the proposed model for each and every combination is clear. Interesting too is that, while the bivariate normal model does have a statistically better log-likelihood at convergence at the 95% confidence level (and also a better average probability of correct prediction) relative to the independent normal model, the difference between these two models from a data fit perspective at the aggregate level is not too substantial (in fact, the independent normal model does marginally better; this can happen because the emphasis of the maximum likelihood estimation procedure is on maximizing the probability of the chosen alternative for each individual, not on aggregate predictions). In contrast, the proposed bivariate YJ model has a clear superior fit at both the disaggregate and aggregate levels.

Table 2: Data fit and ATE measures

Measure Type	Metric		Independent Normal Model	Bivariate Normal Model	Independent YJ Model	Proposed Bivariate YJ Model
Likelihood Based Measures	Log-likelihood at convergence		-3215.54	-3213.46	-3211.29	-3207.49
	Number of non-constant/non-threshold parameters		15	16	16	17
	Log-likelihood at constants and thresholds only		-3353.17			
	Log-likelihood at zero (equal shares)		-3893.67			
	LR test: Proposed Bivariate YJ vs Independent Normal		LR = 16.1> $\chi^2_{(2,0.05)} = 5.99$			
	LR test: Proposed Bivariate YJ vs Bivariate Normal		LR =11.94 > $\chi^2_{(1,0.05)} = 3.84$			
	LR test: Proposed Bivariate YJ vs Independent YJ		LR = 7.6> $\chi^2_{(1,0.05)} = 3.84$			
Non-likelihood Based Measures	Average Probability of Correct Prediction (APCP)		0.1729	0.1731	0.1741	0.1748
	Actual versus Predicted Number of Individuals		Predicted Number			
	Combination Category	Actual Number				
	Non-urban residence and do not walk at all	57	48.48	48.34	53.04	56.16
	Non-urban residence/walk 1-2 days a week	36	41.11	41.51	40.89	37.41
	Non-urban residence/walk 3-4 days a week	56	64.35	64.93	61.44	55.63
	Non-urban residence/walk 5-6 days a week	45	45.90	46.04	44.39	44.61
	Non-urban residence/walk 7 days a week	36	30.26	30.04	30.33	36.82
	Urban residence and do not walk at all	268	276.62	276.76	272.84	269.62
	Urban residence/walk 1-2 days a week	252	248.64	248.74	246.90	251.05
	Urban residence/walk 3-4 days a week	416	407.94	407.43	410.00	415.94
	Urban residence/walk 5-6 days a week	311	307.77	306.99	311.63	310.63
	Urban residence/walk 7 days a week	214	219.93	220.22	219.54	213.13
	Weighted Average Percentage Error (WAPE)		3.42%	3.60%	2.53%	0.46%
Estimated Urban ATE Effect	ATE		0.062	0.812	0.163	1.004
	%ATE		1.8%	29.9%	5.0%	39.2%

3.5. Estimated Causal Effect of Urban Area Residence

The results from the previous section illustrate the better data fit of our proposed model, but does not provide an indication of the extent of mis-estimation (if at all) of the substantive "true" causal effect of urban residence on walk frequency. For this, we examine the impact of urban area residence using an average treatment effect or ATE (see Heckman and Vytlacil, 2000; Heckman and Vytlacil, 2001; and Bhat and Eluru, 2008).[10] In the current empirical context, the average treatment effect (ATE) refers to the expected walk frequency shift for a randomly picked household (from the entire pool of households independent of currently observed residential neighborhood) if it were to reside in an urban neighborhood relative to a non-urban neighborhood. It represents the "true" causal effect of urban residence on walk frequency. For ease in the computation of this ATE effect, we assign a cardinal value c_k to each ordinal walk frequency category *k* as follows: (a) Do not walk any day (k=1, $c_1 = 0$), (b) Walk 1-2 days per week (k=2, $c_2 = 1.5$), (c) Walk 3-4 days per week (k=3, $c_3 = 3.5$), (d) Walk 5-6 days per week (k=4, $c_4 = 5.5$), and (e) Walk 7 days per week (k=5, $c_5 = 7$). Then, the ATE metric is computed for our proposed model as follows (using the same notations as earlier):

$$\begin{aligned} ATE &= \frac{1}{Q}\sum_{q=1}^{Q}\left[\sum_{k=1}^{K} c_k \times \left[\Phi\left(t_{\lambda_2}(\varphi_{k,1})\right) - \Phi\left(t_{\lambda_2}(\varphi_{k-1,1})\right)\right] - \sum_{k=1}^{K} c_k \times \left[\Phi\left(t_{\lambda_2}(\varphi_{k,0})\right) - \Phi\left(t_{\lambda_2}(\varphi_{k-1,0})\right)\right]\right] \\ &= \frac{1}{Q}\sum_{q=1}^{Q}\left[\sum_{k=1}^{K} c_k \times \left\{\left[\Phi\left(t_{\lambda_2}(\varphi_{k,1})\right) - \Phi\left(t_{\lambda_2}(\varphi_{k-1,1})\right)\right] - \left[\Phi\left(t_{\lambda_2}(\varphi_{k,0})\right) - \Phi\left(t_{\lambda_2}(\varphi_{k-1,0})\right)\right]\right\}\right] \end{aligned} \qquad (12)$$

For the more restrictive bivariate normal model, after estimation, we compute the ATE effect without the YJ transformations appearing in the equation above. For the independent YJ model, the ATE takes the same form as Equation (12) after estimation, while, for the independent normal model, the ATE takes the same form as for the bivariate normal model except that the coefficients correspond to the case when the error correlation is ignored. In addition, we also compute the % ATE effect, considering the walk frequency predicted by our proposed model for a random household if it were located in an urban location as the base. To be noted here is that we do not observe the walk frequency for a random household if were in an urban location; this is a counterfactual scenario that can be obtained only through prediction. This counterfactual prediction is obtained from each model and used as the basis to compute the %ATE for the model (this counterfactual walk frequency prediction, in days of walk per week, for a random household if located in a non-urban area is 3.37 for the independent normal model; 2.72 for the bivariate normal model; 3.28 for the independent YJ model, and 2.55 for the bivariate YJ model).

The bottom panel of Table 2 presents these ATE effects for each of the four models. The first numeric value (0.062) indicates that a random household would walk an additional 0.062 days per week if in an urban setting rather than a non-urban setting, corresponding to a %ATE effect of

[10] The ATE effect, as discussed in this section to determine the urban neighborhood residence effect on walk frequency, can also be used to estimate the magnitude effect of any other sociodemographic, employment, or geographic residence region variable on walk frequency. However, we confine attention here to the urban residence effect.

1.8% (that is, the independent normal model predicts that a random household would have 1.8% additional walk days if in an urban environment relative to a non-urban environment). The difference in ATE effects across the many models is rather clear. For the independent models, the ATE effect is substantially underestimated because of the negative correlation in the error terms of the urban residence equation and the walk frequency equation (that is, the fact that individuals with a high walk frequency propensity tend to locate themselves in non-urban areas gets comingled with the true "causal" effect, lowering the ATE substantially). The bivariate models disentangle the correlation effect from the "true" causal effect, which shows up as a much higher ATE effect. Even so, the bivariate normal model does underestimate the "true" causal effect (an ATE of 0.812 or about 30% ATE in the bivariate normal relative to an ATE of 1.004 or about 40% ATE in the bivariate YJ model). This underestimation of the "true" causal ATE effect of urban residence on walk frequency in the bivariate normal model is due to the asymmetry and rightward skew of the error term of the walk frequency propensity (as correctly recognized by the bivariate YJ model but ignored by the bivariate normal model). Of course, the extent of differences in the ATEs between a restrictive distribution model and the proposed more general distribution model will be specific to each empirical context considered, which implies that it is best to estimate the proposed model proposed rather than *a priori* settling for a restrictive structure that may provide inaccurate results.

Finally, in terms of the residential self-selection effect, one may estimate this from the realization that any model that ignores the correlation between the urban residence and walk frequency equations comingles the associative effect of urban residence and the "true causal" effect of urban residence. Thus, considering the two YJ-based models, the ATE for the independent YJ model comingles the "spurious" self-selection and "true" causal urban residence effects, while the ATE for the bivariate model estimates the "true" causal effect. From the ATEs, one may then estimate the total of the two effects to be 1.004+(1.004-0.163)=1.845. Then, the self-selection effect amounts to 45.6% and the "true" causal effect amounts to 54.4%.

4. CONCLUSIONS

In this paper, we have shown the promise of the YJ transformation to accommodate flexible specifications of stochastic terms in multivariate mixed data models in general, and ordered-response models with discrete EEVs in particular. To our knowledge, this is the first such formulation and application in the econometric literature. The use of such a flexible parametric distribution leads to added robustness of the maximum likelihood (ML) estimator. The resulting bivariate YJ model is as easy to estimate as a bivariate normal model. More generally, the YJ transformation is an efficient way to capture flexible marginal error distributions, which then can be bound together using an implicit copula approach. The proposed approach can be applied to a number of different univariate and multivariate mixed modeling structures, including sample selection models, endogenous switching models, multivariate mixed data models, and revealed preference-stated preference models. It can also be employed for empirical analysis in a variety of travel behavior, traffic safety, urban planning, education, public health, geography, and environmental economics fields, among other fields.

The proposed model is applied in the current paper to investigate the effect of urban living on walking frequency, considering the choice of urban living as being endogenous to walking frequency. While this endogeneity has been attributed in the past to unobserved individual factors such as green consciousness, our results, in the context of a post-pandemic world, suggest that lower level needs of walking feasibility and safety related to health and virus spread considerations may be the more dominant individual factors affecting walking frequency today and also leading to a generic predisposition of those with a high walking propensity to locate themselves in non-crowded non-urban areas. This is consistent with the findings from Payd and Fard who note that high-density cities have been more vulnerable to the spread of COVID and so “more basic walking needs may play a more important role in the daily walking patterns of inhabitants”. This is particularly so for the “not-so-young”, because of the increased vulnerability of older individuals to COVID. The net result is a need to understand how best to design walking infrastructure that not only adheres to the traditional notions of pedestrian friendliness (such as narrow streets, continuity in walkways, frequent crossing points, and low speed limits), but also provides a sense of health safety from contagion spread. Doing so proactively would be beneficial in preparation for future pandemics. The takeaway also is that, with shifting residential choices, improved walking infrastructure (as proxied by the “urban” variable in our analysis) can elevate walking frequency across the board in all geographic areas, pointing to the importance of supporting policies to invest in walk-friendly environments in a post-pandemic world.

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

Alfonzo, M.A. (2005). To walk or not to walk? The hierarchy of walking needs. *Environment and Behavior*, 37(6), 808-836.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. *Econometrica*, 46(5), 1193-1205.

Asmussen, K.E., Mondal, A., and Bhat, C.R. (2023). The interplay between teleworking choice and commute distance. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Asmussen, K.E., Mondal, A., Batur, I., Dirks, A., Pendyala, R.M., and Bhat, C.R. (2024). An investigation of individual-level telework arrangements in the COVID-era. *Transportation Research Part A*, 179, 103888.

Atkinson, A.C., Riani, M., and Corbellini, A. (2021). The Box–Cox transformation: Review and extensions. Statistical Science, 36(2) 239-255. https://doi.org/10.1214/20-STS778

Baillien, J., Gijbels, I., and Verhasselt, A. (2022). Estimation in copula models with two-piece skewed margins using the inference for margins method. *Econometrics and Statistics*, https://doi.org/10.1016/j.ecosta.2022.05.002.

Bernardo, C., Paleti, R., Hoklas, M., and Bhat, C.R. (2015). An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. *Transportation Research Part A*, 76, 71-91.

Bhat, A.C., Almeida, D.M., Fenelon, A., Santos-Lozada, A.R. (2022). A longitudinal analysis of the relationship between housing insecurity and physical health among midlife and aging adults in the United States. *SSM - Population Health*, 18, 101128. https://doi.org/10.1016/j.ssmph.2022.101128

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. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B*, 79, 50-77.

Bhat, C.R., and Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B*, 43(7), 749-765.

Bhat, C.R., and Guo, J.Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B*, 41(5), 506-526.

Bhat, C.R., and Mondal, A. (2022). A new flexible generalized heterogeneous data model (GHDM) with an application to examine the effect of high density neighborhood living on bicycling frequency. *Transportation Research Part B*, 164, 244-266.

Bhat, C.R., and Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B*, 40(9), 709-730.

Bhat, C.R., and Singh, S.K. (2000). A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A*, 34(1), 1-22.

Bhat, C.R., Astroza, S., Sidharthan, R., Jobair Bin Alam, M., and Khushefati, W.H. (2014). A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. *Transportation Research Part B*, 68, 31-51.

Bhat, C.R., Astroza, S., Bhat, A.C., and Nagel, K. (2016). Incorporating a multiple discrete-continuous outcome in the generalized heterogeneous data model: Application to residential self-selection effects analysis in an activity time-use behavior model. *Transportation Research Part B*, 91, 52-76.

Blumenberg, E., and Wander, M. (2023). Housing affordability and commute distance. *Urban Geography*, 44(7), 1454-1473.

Blundell, R., and Powell, J. (2004). Endogeneity in semiparametric binary response models. *The Review of Economic Studies*, 71(3), 655-679.

Bontemps, C., and Nauges, C. (2017). Endogenous variables in binary choice models: Some insights for practitioners. Working paper # 17-855, Toulouse School of Economics, https://publications.ut-capitole.fr/id/eprint/25726/1/wp_tse_855.pdf, accessed January 7, 2024.

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211-252.

Brownstone, D., and Fang, H. (2014). A vehicle ownership and utilization choice model with endogenous residential density. *Journal of Transport and Land Use*, 7(2), 135-151.

Cao, X., and Fan, Y. (2012). Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B*, 39(3), 459-470.

Cerrato, J., and Cifre, E. (2018). Gender inequality in household chores and work-family conflict. *Frontiers in Psychology*, 9, 1330.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475), 1228-1240.

Chesher, A., and Rosen, A. (2013). What do instrumental variable models deliver with discrete dependent variables? *American Economic Review*, 103(3), 557-62.

Choe, E.Y., Du. Y., and Sun, G. (2022). Decline in older adults' daily mobility during the COVID-19 pandemic: The role of individual and built environment factors. *BMC Public Health*, 22, 2317.

Choi, K., and Denice, P. (2022). Inclusion, walkability will be key to rebuilding cities after the COVID-19 pandemic. *The Conversation*, https://theconversation.com/inclusion-walkability-will-be-key-to-rebuilding-cities-after-the-covid-19-pandemic-174313, accessed January 7, 2024.

de Haas, M., Faber, R., and Hamersma, M. (2020). How COVID-19 and the Dutch 'intelligent lockdown' change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands. *Transportation Research Interdisciplinary Perspectives*, 6, 100150.

Denzer, M. (2019). Estimating causal effects in binary response models with binary endogenous explanatory variables: A comparison of possible estimators. Discussion paper number 1916, Gutenberg School of Management and Economics, Johannes Gutenberg University Mainz, Germany.

Dias, F.F., Lavieri, P.S., Sharda, S., Khoeini, S., Bhat, C.R., Pendyala, R.M., Pinjari, A.R., Ramadurai, G., and Srinivasan, K.K. (2020). A comparison of online and in-person activity engagement: The case of shopping and eating meals. *Transportation Research Part C*, 114, 643-656.

Dong, Y., and Lewbel, A. (2015). A simple estimator for binary choice models with endogenous regressors. *Econometric Reviews*, 34(1-2), 82-105.

Duque, M. (2021). Performing healthy ageing through images: From broadcasting to silence. *Global Media and China*, 6(3), 303-324.

Ewing, R., and Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association*, 76(3), 265-294.

Faber, J.W. (2020). We built this: Consequences of new deal era intervention in America's racial geography. *American Sociological Review*, 85(5), 739-775.

Gallant, A.R., and Nychka, D.W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2), 363-390.

Gallaugher, M.P.B, McNicholas, P.D., Melnykov, V., Zhu, X. (2020). Skewed distributions or transformations? Modelling skewness for a cluster analysis. https://doi.org/10.48550/arXiv.2011.09152, accessed January 7, 2024

Garcia, L., Pearce, M., Abbas, A., Mok, A., Strain, T., Ali, S., Crippa, A., Dempsey, P.C., Golubic, R., Kelly, P., Laird, Y., McNamara, E., Moore, S., de Sa, T.H., Smith, A.D., Wijndaele, K., Woodcock, J., and Brage, S. (2023). Non-occupational physical activity and risk of cardiovascular disease, cancer and mortality outcomes: A dose-response meta-analysis of large prospective studies. *British Journal of Sports Medicine*, 57(15), 979-989.

Guerrero, V.M. and Johnson, R.A. (1982). Use of the Box-Cox transformation with binary response models. *Biometrika*, 69(2), 309-314.

Greene, W.H. (2017). *Econometric Analysis* (8th edition). Pearson Education.

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

Guan, X., Wang, D., and Cao, X.J. (2020). The role of residential self-selection in land use-travel research: A review of recent findings. *Transport Reviews*, 40(3), 267-287.

Haddad, A.J., Mondal, A., and Bhat, C.R. (2023). Eat-in or eat-out? A joint model to analyze the new landscape of dinner meal preferences. *Transportation Research Part C*, 147, 104016.

Hamed, M.M., and Mannering, F.L. (1993). Modeling travelers' postwork activity involvement: toward a new methodology. *Transportation Science*, 27(4), 381-394.

Han, S., and Lee, S. (2019). Estimation in a generalization of bivariate probit models with dummy endogenous regressors. *Journal of Applied Econometrics*, 34(6), 994-1015.

Heckman, J.J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46(4), 931-959.

Heckman, J.J., and Vytlacil, E.J. (2000). The relationship between treatment parameters within a latent variable framework. *Economics Letters*, 66(1), 33-39.

Heckman, J.J., and Vytlacil, E.J. (2001). Policy-relevant treatment effects. *The American Economic Review*, 91, 107-111

Heydari, S., Miranda-Moreno, L., and Hickford, A.J. (2020). On the causal effect of proximity to school on pedestrian safety at signalized intersections: A heterogeneous endogenous econometric model. *Analytic Methods in Accident Research*, 26, 100115.

Hogan, C.L., Mata, J., and Carstensen, L.L. (2013). Exercise holds immediate benefits for affect and cognition in older and younger adults. *Psychology and Aging*, 28(2), 587-594.

Huang, X., and Xu, J. (2022). Nonparametric sieve maximum likelihood estimation of semi-competing risks data. *Mathematics*, 10(13), 2248.

Hwang, H., Haddad, A., Batur, I., Saxena, S., Pendyala, R.M., and Bhat, C.R. (2023). An analysis of walking frequency before and after the pandemic. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

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, https://doi.org/10.1201/9781003301820

Jimenez, M.P., DeVille, N.V., Elliott, E.G., Schiff, J.E., Wilt, G.E., Hart, J.E., and James, P. (2021). Associations between nature exposure and health: A review of the evidence. *International Journal of Environmental Research and Public Health*, 18(9), 4790.

Jiryaie, F., and Khodadadi, A. (2019). Simultaneous optimization of multiple responses that involve correlated continuous and ordinal responses according to the Gaussian copula models. *Journal of Statistical Theory and Applications*, 18(3), 212-221.

Joe, H., (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

Kang, C., and Lee, M.-J. (2014). Estimation of binary response models with endogenous regressors. *Pacific Economic Review*, 19(4), 502-530.

Kang, S., Mondal, A., Bhat, A.C., and Bhat, C.R. (2021). Pooled versus private ride-hailing: A joint revealed and stated preference analysis recognizing psycho-social factors. *Transportation Research Part C*, 124, 102906.

Kim, J. and Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, 40, 196-206.

Kwon, S., Ha, I.D., Shih, J.H., and Emura, T. (2022). Flexible parametric copula modeling approaches for clustered survival data. *Pharmaceutical Statistics*, 21(1), 69-88.

Kyan, A., and Takakura, M. (2022). Socio-economic inequalities in physical activity among Japanese adults during the COVID-19 pandemic. *Public Health*, 207, 7-13.

Lee, L.F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(2), 507-512.

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

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1), 145-177.

Lima, F.T., Brown, N.C., Duarte, J.P. (2021). Understanding the impact of walkability, population density, and population size on COVID-19 spread: A pilot study of the early contagion in the United States. *Entropy*, 23(11), 1512.

Longo, A., Hutchinson, W. G., Hunter, R. F., Tully, M. A., and Kee, F. (2015). Demand response to improved walking infrastructure: A study into the economics of walking and health behaviour change. *Social Science & Medicine*, 143, 107-116.

Lotfata, A., Gemci, A., and Ferah, B. (2022). The changing context of walking behavior: Coping with the COVID-19 pandemic in urban neighborhoods. *Archnet-IJAR: International Journal of Architectural Research*, 16(3), 495-516.

Makela, M. and Aktas, B.M. (2023). Learning with the natural environment: How walking with nature can actively shape creativity and contribute to holistic learning. *International Journal of Art & Design Education*, DOI: 10.1111/jade.12447

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530.

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.

Melnykov, Y., Zhu, X., and Melnykov, V. (2021). Transformation mixture modeling for skewed data groups with heavy tails and scatter. *Computational Statistics*, 36, 61-78.

Mondal, A., and Bhat, C.R. (2021). A new closed form multiple discrete-continuous extreme value (MDCEV) choice model with multiple linear constraints. *Transportation Research Part B*, 147, 42-66.

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

Müller, D., and Czado, C. (2018). Representing sparse Gaussian DAGs as sparse R-vines allowing for non-Gaussian dependence. *Journal of Computational and Graphical Statistics*, 27(2), 334-344.

Ong, V.M.H., Nott, D.J., and Smith, M.S. (2018). Gaussian variational approximation with a factor covariance structure. *Journal of Computational and Graphical Statistics*, 27(3), 465-478.

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, 15, Article 12.

Paluch, A.E., Bajpai, S., Bassett, D.R., Carnethon, M.R., Ekelund, U., Evenson, K.R., Galuska, D. A., Jefferis, B.J., Kraus, W.E., Lee, I.M., Matthews, C.E., Omura, J.D., Patel, A.V., Pieper, C.F., Rees-Punia, E., Dallmeier, D., Klenk, J., Whincup, P.H., Dooley, E.E., Pettee Gabriel, K., … Steps for Health Collaborative (2022). Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *The Lancet: Public Health*, 7(3), e219-e228.

Paluch, A.E., Bajpai, S., Ballin, M., Bassett, D.R., Buford, T.W., Carnethon, M.R., Chernofsky, A., Dooley, E.E., Ekelund, U., Evenson, K.R., Galuska, D.A., Jefferis, B.J., Kong, L., Kraus, W.E., Larson, M.G., Lee, I.M., Matthews, C.E., Newton, R.L., Jr, Nordström, A., Nordström, P., … Steps for Health Collaborative (2023). Prospective association of daily steps with cardiovascular disease: A harmonized meta-analysis. *Circulation*, 147(2), 122-131.

Paydar, M., and Kamani Fard, A. (2021). The hierarchy of walking needs and the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, 18(14), 7461.

Pearce, M., Garcia, L., Abbas, A., Strain, T., Schuch, F.B., Golubic, R., Kelly, P., Khan, S., Utukuri, M., Laird, Y., Mok, A., Smith, A., Tainio, M., Brage, S., and Woodcock, J. (2022). Association between physical activity and risk of depression: A systematic review and meta-analysis. *JAMA Psychiatry*, 79(6), 550-559.

Pearson, D.G., and Craig, T. (2014). The great outdoors? Exploring the mental health benefits of natural environments. *Frontiers in Psychology*, 5, 1178.

Peterson, R.A., and Cavanaugh, J.E. (2020). Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, 47(13-15), 2312-2327.

Petrin, A., and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1), 3-13.

Przybylowski, A., Stelmak, S., and Suchanek, M. (2021). Mobility behaviour in view of the impact of the COVID-19 pandemic: Public transport users in Gdansk case study. *Sustainability*, 13(1), 364.

Rantanen, T., Eronen, J., Kauppinen, M., Kokko, K., Sanaslahti, S., Kajan, N., and Portegijs, E. (2021). Life-space mobility and active aging as factors underlying quality of life among older people before and during COVID-19 lockdown in Finland-a longitudinal study. *The Journals of Gerontology: Series A*, 76(3), e60-e67.

Rhine, S.L.W., Greene, W.H., and Toussaint-Comeau, M. (2006). The importance of check-cashing businesses to the unbanked: Racial/ethnic differences. *The Review of Economics and Statistics*, 88(1), 146-157.

Rivers, D., and Vuong, Q.H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39(3), 347-366.

Robbennolt, D., Haddad, A.J., Mondal, A., and Bhat, C.R. (2023). Housing choice in an evolving remote work landscape. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Roe, J., Mondschein, A., Neale, C., Barnes, L., Boukhechba, M., and Lopez, S. (2020). The urban built environment, walking and mental health outcomes among older adults: A pilot study. *Frontiers in Public Health*, 8, 575946.

Sallis, J.F., Adlakha, D., Oyeyemi A., Salvo, D. (2023). Public health research on physical activity and COVID-19: Progress and updated priorities. *Journal of Sport and Health Science*, 12, 553-556.

Schwiebert, J. (2013). Sieve maximum likelihood estimation of a copula-based sample selection model. Leibniz University Hannover, Institute of Labor Economics, https://conference.iza.org/conference_files/SUMS_2013/schwiebert_j8731.pdf, accessed January 9, 2024.

Shaer, A., and Hagshenas, H. (2021). Evaluating the effects of the COVID-19 outbreak on the older adults' travel mode choices. *Transport Policy*, 112, 162-172.

Smith, M.S., Loaiza-Maya, R., and Nott, D.J. (2020). High-dimensional copula variational approximation through transformation. *Journal of Computational and Graphical Statistics*, 29(4), 729-743.

Szabo, Z., Liu, X., and Xiang, L. (2020) Semiparametric sieve maximum likelihood estimation for accelerated hazards model with interval-censored data. *Journal of Statistical Planning and Inference*, 205, 175-192.

Terza, J.V., Basu, A., and Rathouz, P.J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3), 531-543.

Van Acker, V., Mokhtarian, P.L., and Witlox, F. (2014). Car availability explained by the structural relationships between lifestyles, residential location, and underlying residential and travel attitudes. *Transport Policy*, 35, 88-99.

Van Wee, B. (2009). Self-Selection: A key to a better understanding of location choices, travel behaviour and transport externalities? *Transport Reviews*, 29(3), 279-292.

Van den Berg, P., Kemperman, A., De Kleijn, B., Borgers, A. (2016). Ageing and loneliness: The role of mobility and the built environment. *Travel Behavior and Society*, 5, 48-55.

Vetrovsky, T., Cupka, J., Dudek, M., Kuthanova, B., Vetrovska, K., Capek, V., and Bunc, V. (2017). Mental health and quality of life benefits of a pedometer-based walking intervention delivered in a primary care setting. *Acta Gymnica*, 47(3), 138-143.

Vytlacil, E., and Yildiz, N. (2007). Dummy endogenous variables in weakly separable models. *Econometrica*, 75(3), 757-779.

Wan, F., Small, D., and Mitra, N. (2018). A general approach to evaluating the bias of 2-stage instrumental variable estimators. *Statistics in Medicine*, 37(12), 1997-2015.

Wang, J., Yang, Y., Peng, J., Yang, L., Gou, Z., Lu, Y. (2021). Moderation effect of urban density on changes in physical activity during the coronavirus disease 2019 pandemic. *Sustainable Cities and Society*, 72, 103058.

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.

Wei, Y., Tang, Y., and McNicholas, P.D. (2019). Mixtures of generalized hyperbolic distributions and mixtures of skew-t distributions for model-based clustering with incomplete data. *Computational Statistics and Data Analysis*, 130, 18-41.

Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters*, 69(3), 309-312.

Won, J., Alfini, A.J., Weiss, L.R., Michelson, C.S., Callow, D.D., Ranadive, S.M., Gentili, R.J., and Smith, J.C. (2019). Semantic memory activation after acute exercise in healthy older adults. *Journal of the International Neuropsychological Society*, 25(6), 557-568.

Wooldridge, J.M. (2015). Control function methods in applied econometrics. *The Journal of Human Resources*, 50(2), 420-445.

Xu, X., and Lee, L-F. (2018). Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. *Journal of Econometrics*, 203(1), 96-112.

Yamada, M., Kimura, Y., Ishiyama, D., Otobe, Y., Suzuki, M., Koyama, S., Kikuchi, T., Kusumi, H., and Arai, H. (2020). Effect of the COVID-19 epidemic on physical activity in community-dwelling older adults in Japan: A cross-sectional online survey. *The Journal of Nutrition, Health & Aging*, 24(9), 948-950.

Yeo, I.K., and Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954-959.

Yildiz, N. (2013). Estimation of binary choice models with linear index and dummy endogenous variables. *Econometric Theory*, 29(2), 354-92.

Zimmerman, D.W. (1998). Invalidation of parametric and nonparamteric statistical tests by concurrent violation of two assumptions. *Journal of Experimental Education*, 67(1), 55-68.

Zhou, Q., Hu, T., and Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518), 664-672.