

RECENT ADVANCES IN DISCRETE AND DISCRETE-CONTINUOUS MODELING SYSTEMS

Chandra R. Bhat

*The University of Texas at Austin
Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712-0278,
Phone: (512) 471-4535, Fax: (512) 475-8744
Email: bhat@mail.utexas.edu*

Naveen Eluru

*The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
1 University Station, C1761, Austin, TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
Email: naveeneluru@mail.utexas.edu*

ABSTRACT

This chapter discusses the importance of discrete choice and discrete-continuous modeling methods in the cost-benefit evaluation of transportation infrastructure improvement and other related projects. Specifically, discrete choice and discrete-continuous methods constitute the backbone of travel demand models that are used to predict travel characteristics and usage of transportation services under alternative socio-economic scenarios, and for alternative transportation service and land-use configurations. The chapter proceeds to identify some recent developments in the field, with an emphasis on advances in discrete-continuous modeling approaches.

1. BACKGROUND

Transportation planners and engineers have to be able to forecast the response of transportation demand to changes in the attributes of the transportation system and changes in the attributes of the people using the transportation system. Travel demand models are used for this purpose; specifically, travel demand models are used to predict travel characteristics and usage of transport services under alternative socio-economic scenarios, and for alternative transport service and land-use configurations. The need for realistic representations of behavior in travel demand modeling is well acknowledged in the literature. This need is particularly acute today as emphasis shifts from evaluating long-term investment-based capital improvement strategies to understanding travel behavior responses to shorter-term congestion management policies such as alternate work schedules, telecommuting, and congestion-pricing.

Of course, travel demand modeling is also relevant for intercity travel, especially as increasing congestion on intercity highways and at intercity air terminals has raised serious concerns about the adverse impacts of such congestion on regional economic development, national productivity and competitiveness, and environmental quality. Recent studies suggest that intercity travel congestion is likely to grow even further through the next two decades (TIA, 2004; APTA, 2005; US DOT, 2006). To alleviate such current and projected congestion, attention has been focused in recent years on identifying and evaluating alternative proposals to improve inter-city transportation services. Some of these proposals include construction of new (or expansion of existing) express roadways and airports (Koppelman and Bhat, 2006), adding or upgrading conventional rail services (ASA, 2003), and construction of new high-speed ground transportation based on magnetic levitation technology (Persch, 2008).

The large scale nature of the congestion alleviation proposals makes it imperative to undertake a careful *a priori* cost-benefit evaluation with respect to capital investment costs, environmental impacts, job market and economic development impacts, and revenues. Travel demand modeling contributes in an important way to one or more of these elements of a cost-benefit evaluation. For example, in the context of a new transportation intercity travel service being considered, one needs to estimate reliable intercity mode choice models to estimate ridership share on the proposed new (or improved) intercity service and to identify the modes from which existing intercity travelers will be diverted to the new (or upgraded) service. This enables an estimation of the revenues from the fare box as well as the potential reduction in congestion at airports and on roadways. This can be translated to revenues and to environmental impacts due to reductions in roadway/airport congestion. Of course, travel mode choice models can also aid in designing the new service by evaluating the service attributes travelers are most sensitive to.

There are obviously many different kinds of models that may be used in travel demand modeling, though discrete choice models are becoming particularly ubiquitous today for modeling several elements of travel demand such as residential choice, car ownership, activity/trip generation, activity location choice, travel mode choice, and route choice. In particular and regardless of the conceptual paradigm used for modeling (for example, whether a trip-based or an activity-based paradigm is used), discrete choice models and their extensions such as discrete-continuous models constitute the backbone of travel demand models. Thus, the focus of the rest of this chapter will be on discrete choice and discrete-continuous models. Further, because there have been several comprehensive reviews of discrete choice models, we will emphasize more on some recent developments in discrete-continuous modeling in this chapter.

2. DISCRETE CHOICE MODELS

2.1 The Basic Multinomial Logit Model

Most discrete choice models are based on the cross-sectional random utility maximization (RUM) framework of microeconomic theory. The RUM framework assumes that an individual's choice at any choice occasion is a reflection of underlying indirect utilities associated with each of the available alternatives and that the individual selects the alternative which provides her or him the highest utility (or least disutility). The indirect utility that an individual associates with each alternative is not observed to the demand analyst, who then assumes that this utility is composed of two components: (a) a deterministic portion that is based on observed (to the analyst) characteristics of the decision-maker, the choice environment, and the alternative, and (b) an unobserved (to the analyst) component associated with the decision-maker, the choice environment, and/or the alternative. Let the utility U_{qi} that an individual q associates with an alternative i be as follows:

$$U_{qi} = \beta_q' x_{qi} + \varepsilon_{qi} \quad (1)$$

where x_{qi} is a vector of observed variables (including alternative specific constants), β_q is a corresponding coefficient vector which may vary over individuals, and ε_{qi} is an unobserved extreme value random term (while any continuous probability distribution may be chosen for ε_{qi} , it is common to choose the extreme value or Gumbel distribution because it simplifies the probability expressions). Thus, the probability density function and the cumulative distribution function of ε_{qi} are:

$$f(\varepsilon_i) = \frac{1}{\theta_{qi}} e^{-\varepsilon_i / \theta_{qi}} e^{-e^{-\varepsilon_i / \theta_{qi}}} \quad \text{and} \quad F_i(z) = \int_{\varepsilon_i = -\infty}^{\varepsilon_i = z} f(\varepsilon_i) d\varepsilon_i = e^{-e^{-z / \theta_{qi}}},$$

where θ_{qi} is a scale parameter that is related to the variance of ε_{qi} as follows: $Var(\varepsilon_{qi}) = \pi^2 \theta_{qi}^2 / 6$. Also note that the deterministic portion, $\beta_q' x_{qi}$, is linear-in-parameters and is considered as an additive function (for convenience).

The set-up above is very general, and one obtains different model structures based on the assumptions on β_q and the ε_{qi} terms. The simplest model, labeled as the multinomial logit (MNL) model, is obtained by imposing three assumptions on the above set-up. The *first assumption* is that the random components of the utilities of the different alternatives are independent and identically distributed (IID). The assumption of *independence* implies that there are no common unobserved factors affecting the utilities of the various alternatives. This assumption is violated, for example, if a decision-maker assigns a higher utility to all transit modes (bus, train, etc.) in a mode choice modeling context because of the opportunity to socialize, or if the decision maker assigns a lower utility to all the transit modes because of the lack of privacy. In such situations, the same underlying unobserved factor (opportunity to socialize or lack of privacy) impacts on the utilities of all transit modes. The assumption of *identically distributed* (across alternatives) random utility terms implies that the extent of variation in unobserved factors affecting modal utility is the same across all modes (that is $\theta_{qi} = \theta_q$). In general, there is no theoretical reason to believe that this will be the case. For

example, if comfort is an unobserved variable whose values vary considerably for the train mode (based on, say, the degree of crowding on different train routes) but little for the automobile mode, then the random components for the automobile and train modes will have different variances. Both the assumptions of independence and equal error variances across alternatives have significant implications for the competitive structure.

The *second assumption* of the MNL model is that it maintains homogeneity in responsiveness to attributes of alternatives across individuals (*i.e.*, an assumption of response homogeneity). More specifically, the MNL model does not allow sensitivity (or taste) variations to an attribute of alternatives due to unobserved individual characteristics. In equation (1), this implies that $\beta_q = \beta$ ¹. However, unobserved individual characteristics can and generally will affect responsiveness. For example, some individuals by their intrinsic nature may be extremely time-conscious while other individuals may be “laid back” and less time-conscious. Ignoring the effect of unobserved individual attributes can lead to biased and inconsistent parameter and choice probability estimates (see Chamberlain, 1980).

The *third assumption* of the MNL model is that the random utility component of each alternative is independent and identical across individuals. The assumption of independence implies that there are no common unobserved factors across individuals affecting the utility of a particular alternative. This assumption is violated, for example, if the utility for the transit alternative in a mode choice context is higher for individuals residing in a particular area because of such unobserved factors as pedestrian path continuity and good lighting to transit stops. The assumption of identical variance across individuals implies that the extent of variation in unobserved factors affecting an alternative’s utility is the same across individuals (*i.e.*, $\theta_q = \theta$). Further, given the assumptions made thus far, and since the scale of the error term is not identified, θ is set to 1.

The three assumptions discussed above together lead to the simple and elegant closed-form mathematical structure of the MNL. However, these assumptions also leave the MNL model saddled with the “independence of irrelevant alternatives” (IIA) property at the individual level (Luce and Suppes 1965; for a detailed discussion of this property see also Ben-Akiva and Lerman 1985). Thus, relaxing the three assumptions may be important in many choice contexts.

2.2 Advanced Discrete Choice Model Structures

This section discusses four types of advanced discrete choice model structures: (1) The GEV class of models, (2) The mixed multinomial logit (MMNL) class of models, (3) The mixed GEV (MGEV) class of models, and (4) Other mixed discrete choice models.

2.2.1 The GEV Class of Models

The GEV-class of models relaxes the independence from irrelevant alternatives (IIA) property of the multinomial logit model by relaxing the independence assumption between the error terms of alternatives. In other words, a generalized extreme value error structure is used to characterize

¹ Note, however, that if β_q is a linear function of observed exogenous variables, the net result is a situation identical to having $\beta_q = \beta$ with additional interaction terms now included in the x_{qi} vector. Thus, this situation can be accommodated in the MNL model.

the unobserved components of utility of the alternatives as opposed to the univariate and independent extreme value error structure used in the multinomial logit model. There are three important characteristics of all GEV models: (1) The overall variances of the alternatives (*i.e.*, the scale of the utilities of alternatives) are assumed to be identical across alternatives, (2) The choice probability structure takes a closed-form expression, and (3) all GEV models collapse to the MNL model when the parameters generating correlation take values that reduce the correlations between *each pair* of alternatives to zero. With respect to the last point, it has to be noted that the MNL model is also a member of the GEV class, though we will reserve the use of the term “GEV class” to models that constitute generalizations of the MNL model.

The general structure of the GEV class of models was derived by McFadden (1978) from the random utility maximization hypothesis, and generalized by Ben-Akiva and Francois (1983). Several specific GEV structures have been formulated and applied within the GEV class, including the Nested Logit (NL) model (for example: Williams, 1977; McFadden, 1978; Daly and Zachary, 1978, Berkovec and Rust, 1985; Hensher and Greene, 2002; Abdel-Aty and Abdelwahab, 2004; Garrow and Koppelman, 2004; Lo *et al.*, 2004), the Paired Combinatorial Logit (PCL) model (Chu, 1989; Koppelman and Wen, 2000), the Cross-Nested Logit (CNL) model (Vovsha, 1997; Papola, 2004; Antonini *et al.*, 2006; Bekhor *et al.*, 2006; Bierlaire, 2006; Hess and Polak, 2006; Abbe *et al.*, 2007; Marzano and Papola, 2008; Robin *et al.*, 2009), the Ordered GEV (OGEV) model (Small, 1987), the Multinomial Logit-Ordered GEV (MNL-OGEV) model (Bhat, 1998a), the ordered GEV-nested logit (OGEV-NL) model (Whelan *et al.*, 2002) and the Product Differentiation Logit (PDL) model (Breshanan *et al.*, 1997). Wen and Koppelman (2001) proposed a general GEV model structure, which they referred to as the Generalized Nested Logit (GNL) model. Swait (2001), independently, proposed a similar structure, which he refers to as the choice set Generation Logit (GenL) model; Swait’s derivation of the GenL model is motivated from the concept of latent choice sets of individuals, while Wen and Koppelman’s derivation of the GNL model is motivated from the perspective of flexible substitution patterns across alternatives. Wen and Koppelman (2001) illustrate the general nature of the GNL model formulation by deriving the other GEV model structures mentioned earlier as special restrictive cases of the GNL model or as approximations to restricted versions of the GNL model. Swait (2001) presents a network representation for the GenL model, which also applies to the GNL model. More recent applications and refinements involving the GNL model include Bekhor and Prashker (2001), Koppelman and Sethi (2005), and Gelhausen and Wilken (2006).

One impediment to the generation of new GEV models, however, is that the conditions developed by McFadden for qualification as a GEV structure are based on a generating function G , which may not map easily into a desired correlation structure. Research by Bierlaire (2002) and Daly and Bierlaire (2006) alleviate this impediment. These two researchers propose a network-based structure to characterize the underlying correlation structure in any choice situation, and show how this network-based representation, if it satisfies some simple conditions (non-emptiness, finiteness, and being circuit-free), can immediately be translated to a model consistent with the GEV structure (this work constitutes a formal and rigorous extension of Swait’s network representation for the GenL model). Newman (2008) enhanced this network structure developed by Daly and Bierlaire (2006) by describing methodologies to correctly normalize the allocation parameters in the network GEV models. Sener *et al.* (2009) and Newman (2009) developed a new form of a Network GEV model that allows the incorporation of error covariance heterogeneity across individuals. Of course, GEV models based on complex network representations, while allowing flexibility in substitution patterns, also entail the

estimation of a substantial number of dissimilarity and allocation parameters. The net result is that the analyst will have to impose informed restrictions on these GEV models, customized to the application context under investigation.

An important point to note here is that GEV models are consistent with utility maximization only under rather strict (and often empirically violated) restrictions on the dissimilarity and allocation parameters (specifically, the dissimilarity and allocation parameters should be bounded between 0 and 1 for global consistency with utility maximization, and the allocation parameters for any alternative should add to 1). The origin of these restrictions can be traced back to the requirement that the variance of the joint alternatives be identical in the GEV models. Also, GEV models do not relax assumptions related to taste homogeneity in response to an attribute (such as travel time or cost in a mode choice model) due to unobserved decision-maker characteristics, and cannot be applied to panel data with temporal correlation in unobserved factors within the choices of the same decision-making agent. However, GEV models do offer computational tractability, provide a theoretically sound measure for benefit valuation, and can form the basis for formulating mixed models that accommodate random taste variations and temporal correlations in panel data.

2.2.2 The MMNL Class of Models

The MMNL class of models, like the GEV class of models, generalizes the MNL model. However, unlike the closed form of the GEV class, the MNL class involves the analytically intractable integration of the multinomial logit formula over the distribution of unobserved random parameters. It takes the structure shown below:

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta | \theta) d(\beta), \text{ where} \quad (2)$$

$$L_{qi}(\beta) = \frac{e^{\beta'x_{qi}}}{\sum_j e^{\beta'x_{qj}}}.$$

P_{qi} is the probability that individual q chooses alternative i , x_{qi} is a vector of observed variables specific to individual q and alternative i , β represents parameters which are random realizations from a density function $f(\cdot)$, and θ is a vector of underlying moment parameters characterizing $f(\cdot)$.

The structure in Equation (2) assumes a continuous distribution for $f(\beta)$. In fact, a discrete distribution can also be used. Such a discrete distribution may take one of two forms. If the entire vector β can take one of S possible values labeled $\beta_1, \beta_2, \dots, \beta_s, \dots, \beta_S$, and the probability of $\beta = \beta_s$ for individual q is π_{qs} , then the appropriate formula is:

$$P_{qi}(\theta) = \sum_s \pi_{qs} L_{qi}(\beta_s), \text{ where} \quad (3)$$

$$L_{qi}(\beta_s) = \frac{e^{\beta_s'x_{qi}}}{\sum_j e^{\beta_s'x_{qj}}}, \text{ and } \sum_s \pi_{qs} = 1 \text{ for all } q.$$

π_{qs} in the equation above can be further parameterized as a function of observable individual attributes using any function that satisfies $\sum_s \pi_{qs} = 1$ (usually a multinomial logit form is used).

In this first form of a discrete distribution for the vector β , the MMNL model becomes equivalent to the latent-class model that has been used in marketing and in transportation (see Kamakura and Russell, 1989; Greene and Hensher, 2003; Bhat, 1997; Gupta and Chintagunta, 1994). A second possible discrete distribution approach is to use a non-parametric form separately for each coefficient in the model. This approach does not impose any prior continuous distribution function, and allows the data to identify the mass points and the associated mixing weights for each coefficient separately. Of course, such a non-parametric distribution specification can lead to convergence problems unless the number of mass points for each coefficient is limited to a small number.

In the rest of this section, we focus on a continuous distribution assumption for $f(\beta)$, since this has been the more dominant assumption under the label of mixed logit. The first applications of the mixed logit structure of Equation (2) appear to have been by Boyd and Mellman (1980) and Cardell and Dunbar (1980). However, these were not individual-level models and, consequently, the integration inherent in the mixed logit formulation had to be evaluated only once for the entire market. Train (1986) and Ben-Akiva *et al.* (1993) applied the mixed logit to customer-level data, but considered only one or two random coefficients in their specifications. Thus, they were able to use quadrature techniques for estimation. The first applications to realize the full potential of mixed logit by allowing several random coefficients simultaneously include Revelt and Train (1998) and Bhat (1998b), both of which were originally completed in the early 1996 and exploited the advances in simulation methods.

The MMNL model structure of Equation (2) can be motivated from two very different (but formally equivalent) perspectives (see Bhat, 2000a). Specifically, a MMNL structure may be generated from an intrinsic motivation to allow flexible substitution patterns across alternatives (error-components structure) or from a need to accommodate unobserved heterogeneity across individuals in their sensitivity to observed exogenous variables (random-coefficients structure) or a combination of the two. Examples of the error-components motivation in the literature include Brownstone and Train (1998), Bhat (1998c), Bekhor *et al.* (2002), Jong *et al.* (2002a, 2002b), Whelan *et al.* (2002), Batley *et al.* (2001a, 2001b), Vichiensan *et al.* (2005), Srinivasan and Mahmassani (2003), Adler *et al.* (2005), Antonini *et al.* (2006), Pinjari and Bhat (2006), Srinivasan and Ramadurai (2006), and Hess *et al.* (2008). The reader is also referred to the work of Walker and her colleagues (Ben-Akiva *et al.*, 2001; Walker, 2002, Walker *et al.*, 2004) and Munizaga and Alvarez-Daziano (2001) for important identification issues in the context of the error components MMNL model. Examples of the random-coefficients structure include Revelt and Train (1998), Bhat (2000b), Hensher (2001), Rizzi and Ortúzar (2003), Amador *et al.* (2005), Brownstone and Small (2005), Hess *et al.* (2005), Small *et al.* (2005), Bhat and Sardesai (2006), Cirillo and Axhausen (2006), Greene *et al.* (2006), Lijesen (2006), Sivakumar and Bhat (2007), Srinivasan and Ramadurai (2006), Gkritza and Mannering (2008), and Milton *et al.* (2008).

A normal distribution is assumed for the density function $f(\cdot)$ in Equation (2) when an error-components structure forms the basis for the MMNL model. However, while a normal distribution remains the most common assumption for the density function $f(\cdot)$ for a random-coefficients structure, other density functions may be more appropriate. For example, a log-normal distribution may be used if, from a theoretical perspective, an element of β has to take the

same sign for every individual (such as a negative coefficient on the travel cost parameter in a travel mode choice model; see Cirillo and Axhausen, 2006). Other distributions that have been used in the literature include triangular and uniform distributions (see Revelt and Train, 2000; Train, 2001; Hensher and Greene, 2003; Amador *et al.*, 2005), the Rayleigh distribution (Siikamaki and Layton, 2001), the censored normal (Cirillo and Axhausen, 2006; Train and Sonnier, 2005), Johnson's SB (Cirillo and Axhausen, 2006; Train and Sonnier, 2005), and nonparametric and semi-parametric distributions (Fosgerau, 2006; Fosgerau and Bierlaire, 2007).²

2.2.3 The Mixed GEV Class of Models

The MMNL class of models is very general in structure and can accommodate both relaxations of the IID assumption as well as unobserved response homogeneity within a simple unifying framework. Consequently, the need to consider a mixed GEV class may appear unnecessary. However, there are instances when substantial computational efficiency gains may be achieved using a MGEV structure. Consider, for instance, Bhat and Guo's (2004) model for household residential location choice. It is possible, if not very likely, that the utility of spatial units that are close to each other will be correlated due to common unobserved spatial elements. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow contiguous alternatives to be correlated. In the MMNL structure, such a correlation structure may be imposed through the specification of a multivariate MNP-like error structure, which will then require multidimensional integration of the order of the number of spatial units (see Bolduc *et al.*, 1996). On the other hand, a carefully specified GEV model can accommodate the spatial correlation structure within a closed-form formulation.³ However, the GEV model structure of Bhat and Guo cannot accommodate unobserved random heterogeneity across individuals. One could superimpose a mixing distribution over the GEV model structure to accommodate such random coefficients, leading to a parsimonious and powerful MGEV structure. Thus, in a case with 1000 spatial units (or zones), the MMNL model would entail a multidimensional integration of the order of 1000 plus the number of random coefficients, while the MGEV model involves multidimensional integration only of the order of the number of random coefficients (a reduction of dimensionality of the order of 1000!).

In addition to computational efficiency gains, there is another more basic reason to prefer the MGEV class of models when possible over the MMNL class of models. This is related to the fact that closed-form analytic structures should be used whenever feasible, because they are always more accurate than the simulation evaluation of analytically intractable structures (see Train, 2003; pg. 191). In this regard, superimposing a mixing structure to accommodate random coefficients over a closed form analytic structure that accommodates a particular desired inter-alternative error correlation structure represents a powerful approach to capture random taste variations and complex substitution patterns. Examples of the MGEV class of models employed in the literature include: Bhat and Guo (2004), Hess *et al.* (2004), Dugundji and Walker (2005), Srinivasan and Athuru (2005), Antonini *et al.* (2006), Bajwa *et al.* (2006), and Lapparant (2006).

² The reader is referred to Hess and Axhausen (2005), Hess *et al.*, (2005), and Train and Sonnier (2005) for a review of alternative distributional forms and their ability to approximate several different types of true distributions. Also, Sørensen and Nielson (2003) propose a method for determining the best distributional form prior to estimation.

³ The GEV structure used by Bhat and Guo is a restricted version of the GNL model proposed by Wen and Koppelman. Specifically, the GEV structure takes the form of a paired GNL (PGNL) model with equal dissimilarity parameters across all paired nests (each paired nest includes a spatial unit and one of its adjacent spatial units).

2.2.4 Other Mixed Discrete Choice Models

The mixing of a distribution with a closed form analytic expression has application far beyond the MMNL and MGEV structures discussed above. For example, random coefficients can be imposed in an ordered-response multinomial model or a count model. For instance, Eluru and Bhat (2007) formulate a mixed ordered logit model to analyze crash related driver injury severity. The moderating influence of unobserved factors associated with the impact of various attributes affecting injury severity is accommodated by imposing a random coefficients structure in the ordered logit model. The results indicate clear biases in the effects of different variables on injury severity level when unobserved factors moderating the impact of variables are ignored.

3. DISCRETE-CONTINUOUS MODELS

Several consumer demand choices related to travel and other decisions are characterized by a discrete dimension as well as a continuous dimension. Examples of such choice situations include vehicle type holdings and usage, appliance choice and energy consumption, and rent or purchase a home and square footage. In the rest of this chapter, our emphasis will be on recent advancements in modeling discrete-continuous choice situations. In particular, we discuss a flexible copula-based approach for the estimation of joint discrete-continuous systems with dichotomous or polychotomous (or multinomial) discrete endogenous variables. In the remainder of this section, we present the general background of copulas, discuss copula properties and dependence structures, provide an overview of alternative copula structures, and describe a copula application to examine an empirical discrete-continuous choice context.

3.1. General Background

A copula is a multivariate functional form for the joint distribution of random variables derived purely from pre-specified parametric marginal distributions of each random variable. The word copula itself was coined by Sklar, 1959 and is derived from the Latin word “copulare”, which means to tie, bond, or connect (see Schmidt, 2007). Thus, a copula is a device or function that generates a stochastic dependence relationship (*i.e.*, a multivariate distribution) among random variables with pre-specified marginal distributions. The copula approach separates the marginal distributions from the dependence structure, so that the dependence structure is entirely unaffected by the marginal distributions assumed. This provides substantial flexibility in correlating random variables, which may not even have the same marginal distributions. Further, the incorporation of dependency effects in econometric models can be greatly facilitated by using a copula approach for modeling discrete-continuous choices, so that the resulting model can be in closed-form and can be estimated using direct maximum likelihood techniques (the reader is referred to Trivedi and Zimmer, 2007 or Nelsen, 2006 for extensive reviews of copula theory, approaches, and benefits).

The effectiveness of a copula approach has been recognized in the statistics field for several decades now (see Schweizer and Sklar, 1983, Ch. 6), but it is only recently that copula-based methods have been explicitly recognized and employed in the finance, actuarial science, hydrological modeling, and econometrics fields (see, for example, Embrechts *et al.*, 2002; Cherubini *et al.*, 2004; Frees and Wang, 2005; Genest and Favre, 2007; Grimaldi and Serinaldi, 2006; Smith, 2005; Prieger, 2002; Zimmer and Trivedi, 2006; Cameron *et al.*, 2004; Junker and May, 2005; and Quinn, 2007). The precise definition of a copula is that it is a multivariate

distribution function defined over the unit cube linking uniformly distributed marginals. Let C be a K -dimensional copula of uniformly distributed random variables $U_1, U_2, U_3, \dots, U_K$ with support contained in $[0,1]^K$. Then,

$$C_\theta(u_1, u_2, \dots, u_K) = \Pr(U_1 < u_1, U_2 < u_2, \dots, U_K < u_K), \quad (4)$$

where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector. A copula, once developed, allows the generation of joint multivariate distribution functions with given marginals. Consider K random variables $Y_1, Y_2, Y_3, \dots, Y_K$, each with univariate continuous marginal distribution functions $F_k(y_k) = \Pr(Y_k < y_k), k = 1, 2, 3, \dots, K$. Then, by the integral transform result, and using the notation $F_k^{-1}(\cdot)$ for the inverse univariate cumulative distribution function, we can write the following expression for each k ($k = 1, 2, 3, \dots, K$):

$$F_k(y_k) = \Pr(Y_k < y_k) = \Pr(F_k^{-1}(U_k) < y_k) = \Pr(U_k < F_k(y_k)). \quad (5)$$

Then, by Sklar's (1973) theorem, a joint K -dimensional distribution function of the random variables with the continuous marginal distribution functions $F_k(y_k)$ can be generated as follows:

$$\begin{aligned} F(y_1, y_2, \dots, y_K) &= \Pr(Y_1 < y_1, Y_2 < y_2, \dots, Y_K < y_K) = \Pr(U_1 < F_1(y_1), U_2 < F_2(y_2), \dots, U_K < F_K(y_K)) \\ &= C_\theta(u_1 = F_1(y_1), u_2 = F_2(y_2), \dots, u_K = F_K(y_K)). \end{aligned} \quad (6)$$

Conversely, by Sklar's theorem, for any multivariate distribution function with continuous marginal distribution functions, a unique copula can be defined that satisfies the condition in Equation (6).

Copulas themselves can be generated in several different ways, including the method of inversion, geometric methods, and algebraic methods (see Nelsen, 2006; Ch. 3). For instance, given a known multivariate distribution $F(y_1, y_2, \dots, y_K)$ with continuous margins $F_k(y_k)$, the inversion method inverts the relationship in Equation (6) to obtain a copula:

$$\begin{aligned} C_\theta(u_1, u_2, \dots, u_K) &= \Pr(U_1 < u_1, U_2 < u_2, \dots, U_K < u_K) \\ &= \Pr(Y_1 < F_1^{-1}(u_1), Y_2 < F_2^{-1}(u_2), \dots, Y_3 < F_3^{-1}(u_3)) \\ &= F(y_1 = F_1^{-1}(u_1), y_2 = F_2^{-1}(u_2), \dots, y_K = F_K^{-1}(u_K)). \end{aligned} \quad (7)$$

Once the copula is developed, one can revert to Equation (6) to develop new multivariate distributions with arbitrary univariate margins.

A rich set of copula types have been generated using the inversion and other methods, including the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and the Archimedean class of copulas (including the Clayton, Gumbel, Frank, and Joe copulas). These copulas are discussed later in the context of bivariate distributions. In such bivariate distributions, while θ can be a vector of parameters, it is customary to use a scalar measure of dependence. In the next section, we discuss some copula properties and dependence structure concepts for bivariate copulas, though generalizations to higher dimensions are possible.

3.2 Copula Properties and Dependence Structure

Consider any bivariate copula $C_\theta(u_1, u_2)$. Since this is a bivariate cumulative distribution function, the copula should satisfy the well known Fréchet-Hoeffding bounds (see Kwerel,

1988). Specifically, the Fréchet lower bound $W(u_1, u_2)$ is $\max(u_1 + u_2 - 1, 0)$ and the Fréchet upper bound $M(u_1, u_2)$ is $\min(u_1, u_2)$. Thus,

$$W(u_1, u_2) \leq C_\theta(u_1, u_2) \leq M(u_1, u_2). \quad (8)$$

From Sklar's theorem of Equation (6), we can also re-write the equation above in terms of Fréchet bounds for the multivariate distribution $F(y_1, y_2)$ generated from the copula $C_\theta(u_1, u_2)$:

$$\max(F_1(y_1) + F_2(y_2) - 1, 0) \leq F(y_1, y_2) \leq \min(F_1(y_1), F_2(y_2)). \quad (9)$$

If the copula $C_\theta(u_1, u_2)$ is equal to the lower bound $W(u_1, u_2)$ in Equation (8), or equivalently if $F(y_1, y_2)$ is equal to the lower bound in Equation (9), then the random variables Y_1 and Y_2 are almost surely decreasing functions of each other and are called "countermonotonic". On the other hand, if the copula $C_\theta(u_1, u_2)$ is equal to the upper bound $M(u_1, u_2)$ in Equation (8), or equivalently if $F(y_1, y_2)$ is equal to the upper bound in Equation (9), then the random variables Y_1 and Y_2 are almost surely increasing functions of each other and are called "comonotonic". The case when $C_\theta(u_1, u_2) = \Pi = u_1 u_2$, or equivalently $F(y_1, y_2) = F_1(y_1)F_2(y_2)$, corresponds to stochastic independence between Y_1 and Y_2 .

Different copulas provide different levels of ability to capture dependence between Y_1 and Y_2 based on the degree to which they cover the interval between the Fréchet-Hoeffding bounds. Comprehensive copulas are those that (1) attain or approach the lower bound W as θ approaches the lower bound of its permissible range, (2) attain or approach the upper bound M as θ approaches its upper bound, and (3) cover the entire domain between W and M (including the product copula case Π as a special or limiting case). Thus, comprehensive copulas parameterize the full range of dependence as opposed to non-comprehensive copulas that are only able to capture dependence in a limited manner.

To better understand the generated dependence structures between the random variables (Y_1, Y_2) based on different copulas, and examine the coverage offered by non-comprehensive copulas, it is useful to construct a scalar dependence measure between Y_1 and Y_2 that satisfies four properties as listed below (see Embrechts *et al.*, 2002):

- (1) $\delta(Y_1, Y_2) = \delta(Y_2, Y_1)$
- (2) $-1 \leq \delta(Y_1, Y_2) \leq 1$ (10)
- (3) $\delta(Y_1, Y_2) = 1 \Leftrightarrow (Y_1, Y_2)$ comonotonic; $\delta(Y_1, Y_2) = -1 \Leftrightarrow (Y_1, Y_2)$ countermonotonic
- (4) $\delta(Y_1, Y_2) = \delta(G_1(Y_1), G_2(Y_2))$, where G_1 and G_2 are two (possibly different) strictly increasing transformations.

The traditional dependence concept of correlation coefficient ρ (*i.e.*, the Pearson's product-moment correlation coefficient) is a measure of linear dependence between Y_1 and Y_2 . It satisfies the first two of the properties discussed above. However, it satisfies the third property only for bivariate elliptical distributions (including the bivariate normal distribution) and adheres to the fourth property only for strictly increasing linear transformations (see Embrechts *et al.*, 2002 for specific examples where the Pearson's correlation coefficient fails the third and fourth properties; also see Boyer *et al.*, 1999 for limitations of Pearson's correlation coefficient for

asymmetric distributions). These limitations of the traditional correlation coefficient have led statisticians to the use of concordance measures to characterize dependence. Basically, two random variables are labeled as being concordant (discordant) if large values of one variable are associated with large (small) values of the other, and small values of one variable are associated with small (large) values of the other. This concordance concept has led to the use of two measures of dependence in the literature: the Kendall's τ and the Spearman's ρ_s (see Bhat and Eluru, 2009 for a discussion of the Kendall's τ and the Spearman's ρ_s measures). The Kendall's τ and the Spearman's ρ_s measures can be shown to satisfy all the properties listed in Equation (10). In addition, both assume the value of zero under independence and are not dependent on the margins $F_1(\cdot)$ and $F_2(\cdot)$. Hence, it is customary to use these two concordance measures to characterize dependence structures in the copula literature, as opposed to using the familiar Pearson's ρ .

3.3 Alternative Copulas

Several copulas have been formulated in the literature, and these copulas can be used to tie random variables together. We focus on bivariate forms of the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and the Archimedean class of copulas in this discussion. Table 1 provides a concise summary for each of the copulas along with the properties of the Kendall's τ and the Spearman's ρ_s measures.

3.3.1 The Gaussian copula

The Gaussian copula is the most familiar of all copulas, and forms the basis for Lee's (1983) sample selection mechanism. The copula belongs to the class of elliptical copulas, since the Gaussian copula is simply the copula of the elliptical bivariate normal distribution (the density contours of elliptical distributions are elliptical with constant eccentricity). The Gaussian copula takes the following form:

$$C_\theta(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \theta), \quad (11)$$

where $\Phi_2(\cdot, \cdot, \theta)$ is the bivariate cumulative distribution function with Pearson's correlation parameter θ ($-1 \leq \theta \leq 1$). The Gaussian copula is comprehensive in that it attains the Fréchet lower and upper bounds, and captures the full range of (negative or positive) dependence between two random variables. However, it also assumes the property of asymptotic independence. That is, regardless of the level of correlation assumed, extreme tail events appear to be independent in each margin just because the density function gets very thin at the tails (see Embrechts *et al.*, 2002). Conversely, in the Gaussian copula, a positive (negative) correlation gets manifested as clustering along the southwest-to-northeast (northwest-to southeast) plane close to the center point of the joint distribution. However, toward extreme tails, there is scattering or dependence reduction. Further, the dependence structure is radially symmetric about the center point in the Gaussian copula. That is, for a given correlation, the level of

dependence is equal in the upper and lower tails.⁴ On the other hand, extreme tail dependence and asymmetric tail dependence may be important characteristics in bivariate data (even after conditioning each marginal variable in terms of observed covariates).

3.3.2 The Farlie-Gumbel-Morgenstern (FGM) copula

The FGM copula was first proposed by Morgenstern (1956), and also discussed by Gumbel (1958) and Farlie (1960). It has been well known for some time in Statistics (see Conway, 1979; Kotz *et al.*, 2000; Section 44.13). However, until Prieger (2002), it does not seem to have been used in Econometrics. In the bivariate case, the FGM copula takes the following form:

$$C_{\theta}(u_1, u_2) = u_1 u_2 [1 + \theta(1 - u_1)(1 - u_2)]. \quad (12)$$

For the copula above to be 2-increasing (that is, for any rectangle with vertices in the domain of $[0,1]$ to have a positive volume based on the function), θ must be in $[-1, 1]$. The presence of the θ term allows the possibility of correlation between the uniform marginals u_1 and u_2 . Specifically, the density function for the FGM copula is:

$$c_{\theta}(u_1, u_2) = 1 + \theta(1 - 2u_1)(1 - 2u_2). \quad (13)$$

From above, it is clear that, when θ is positive, the density is higher if u_1 and u_2 are both high (both close to 1) or both low (both close to zero). On the other hand, when θ is negative, the density is higher if u_1 is high and u_2 is low, or if u_2 is high and u_1 is low. When θ is zero, it corresponds to independence. Otherwise, depending on whether θ is positive or negative, a positive or negative correlation, respectively, is generated between the continuous variables U_1 and U_2 . Thus, the FGM copula has a simple analytic form and allows for either negative or positive dependence. Like the Gaussian copula, it also imposes the assumptions of asymptotic independence and radial symmetry in dependence structure. However, the FGM copula is not comprehensive in coverage, and can accommodate only relatively weak dependence between the marginals.

3.3.3 The Archimedean class of copulas

The Archimedean class of copulas is popular in empirical applications (see Genest and MacKay, 1986 and Nelsen, 2006 for extensive reviews). This class of copulas includes a whole suite of closed-form copulas that cover a wide range of dependency structures, including comprehensive and non-comprehensive copulas, radial symmetry and asymmetry, and asymptotic tail independence and dependence. The class is very flexible, and easy to construct. Archimedean copulas are constructed based on an underlying continuous convex decreasing generator function φ from $[0, 1]$ to $[0, \infty]$ with the following properties: $\varphi(1) = 0$, $\varphi'(t) < 0$, and $\varphi''(t) > 0$ for all $0 < t < 1$ ($\varphi'(t) = \partial \varphi / \partial t$; $\varphi''(t) = \partial^2 \varphi / \partial^2 t$). Further, in the discussion here, we will assume that

⁴ Mathematically, the dependence structure of a copula is labeled as “radially symmetric” if the following condition holds: $C_{\theta}(u_1, u_2) = u_1 + u_2 - 1 + C_{\theta}(1 - u_1, 1 - u_2)$, where the right side of the expression above is the survival copula (see Nelsen, 2006, page 37). Consider two random variables Y_1 and Y_2 whose marginal distributions are individually symmetric about points a and b , respectively. Then, the joint distribution F of Y_1 and Y_2 will be radially symmetric about points a and b if and only if the underlying copula from which F is derived is radially symmetric.

$\varphi(0) = \infty$, so that an inverse φ^{-1} exists. With these preliminaries, we can generate bivariate Archimedean copulas as:

$$C_{\theta}(u_1, u_2) = \varphi^{-1}[\varphi(u_1) + \varphi(u_2)], \quad (14)$$

where the dependence parameter θ is embedded within the generator function. Note that the above expression can also be equivalently written as:

$$\varphi[C_{\theta}(u_1, u_2)] = [\varphi(u_1) + \varphi(u_2)]. \quad (15)$$

Using the differentiation chain rule on the equation above, we obtain the following important result for Archimedean copulas that will be relevant to the sample selection model discussed in the next section:

$$\frac{\partial C_{\theta}(u_1, u_2)}{\partial u_2} = \frac{\varphi'(u_2)}{\varphi'[C_{\theta}(u_1, u_2)]}, \text{ where } \varphi'(t) = \partial \varphi(t) / \partial t. \quad (16)$$

The density function of absolutely continuous Archimedean copulas of the type discussed later in this section may be written as:

$$c_{\theta}(u_1, u_2) = - \frac{\varphi''(C(u_1, u_2)) \varphi'(u_1) \varphi'(u_2)}{[\varphi'(C(u_1, u_2))]^3}. \quad (17)$$

Another useful result for Archimedean copulas is that the expression for Kendall's τ collapses to the following simple form (see Genest and MacKay, 1986 or Embrechts *et al.*, 2002 for a derivation):

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \quad (18)$$

In the rest of this section, we provide an overview of four different Archimedean copulas: the Clayton, Gumbel, Frank, and Joe copulas.

3.3.3.1 The Clayton copula

The Clayton copula has the generator function $\varphi(t) = (1/\theta)(t^{-\theta} - 1)$, giving rise to the following copula function (see Huard *et al.*, 2006):

$$C_{\theta}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \quad 0 < \theta < \infty. \quad (19)$$

The above copula, proposed by Clayton (1978), cannot account for negative dependence. It attains the Fréchet upper bound as $\theta \rightarrow \infty$, but cannot achieve the Fréchet lower bound. Using the Archimedean copula expression in Equation (18) for τ , it is easy to see that τ is related to θ by $\tau = \theta / (\theta + 2)$, so that $0 < \tau < 1$ for the Clayton copula. Independence corresponds to $\theta \rightarrow 0$.

3.3.3.2 The Gumbel copula

The Gumbel copula, first discussed by Gumbel (1960) and sometimes also referred to as the Gumbel-Hougaard copula, has a generator function given by $\varphi(t) = (-\ln t)^\theta$. The form of the copula is provided below:

$$C_\theta(u_1, u_2) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right), \quad 1 \leq \theta < \infty. \quad (20)$$

Like the Clayton copula, the Gumbel copula cannot account for negative dependence, but attains the Fréchet upper bound as $\theta \rightarrow \infty$. Kendall's τ is related to θ by $\tau = 1 - (1/\theta)$, so that $0 < \tau < 1$, with independence corresponding to $\theta = 1$.

3.3.3.3 The Frank copula

The Frank copula, proposed by Frank (1979), is the only Archimedean copula that is comprehensive in that it attains both the upper and lower Fréchet bounds, thus allowing for positive and negative dependence. It is radially symmetric in its dependence structure and imposes the assumption of asymptotic independence. The generator function is $\varphi(t) = -\ln[(e^{-\theta t} - 1)/(e^{-\theta} - 1)]$, and the corresponding copula function is given by:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right), \quad -\infty < \theta < \infty. \quad (21)$$

Kendall's τ does not have a closed form expression for Frank's copula, but may be written as (see Nelsen, 2006, pg 171):

$$\tau = 1 - \frac{4}{\theta} [1 - D_F(\theta)], \quad D_F(\theta) = \frac{1}{\theta} \int_{t=0}^{\theta} \frac{t}{e^t - 1} dt. \quad (22)$$

The range of τ is $-1 < \tau < 1$. Independence is attained in Frank's copula as $\theta \rightarrow 0$.

3.3.3.4 The Joe copula

The Joe copula, introduced by Joe (1993, 1997), has a generator function $\varphi(t) = -\ln[1 - (1 - t)^\theta]$ and takes the following copula form:

$$C_\theta(u_1, u_2) = 1 - \left[(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_2)^\theta (1 - u_1)^\theta\right]^{1/\theta}, \quad 1 \leq \theta < \infty. \quad (23)$$

The Joe copula is similar to the Clayton copula. It cannot account for negative dependence. It attains the Fréchet upper bound as $\theta \rightarrow \infty$, but cannot achieve the Fréchet lower bound. The relationship between τ and θ for Joe's copula does not have a closed form expression, but takes the following form:

$$\tau = 1 + \frac{4}{\theta} D_J(\theta), \quad D_J(\theta) = \int_{t=0}^1 \frac{[\ln(1 - t^\theta)](1 - t^\theta)}{t^{\theta-1}} dt. \quad (24)$$

The range of τ is between 0 and 1, and independence corresponds to $\theta = 1$.

3.4 Copula Based Application of Household Residential Location Choice and Vehicle Miles Travelled

In the preceding sections, we presented an overview of copulas and different types of copulas. In this section we provide a discussion of a copula based application to residential location choice and daily vehicle miles of travel (VMT). The material in this section is drawn from Bhat and Eluru (2009).

There has been considerable interest in the land use-transportation connection in the past decade, motivated by the possibility that land-use and urban form design policies can be used to control, manage, and shape individual traveler behavior and aggregate travel demand. A central issue in this regard is the debate whether any effect of the neighborhood on travel demand is causal or merely associative (or some combination of the two; see Bhat and Guo, 2007). To explicate this, consider a cross-sectional sample of households, some of whom live in a neo-urbanist neighborhood and others of whom live in a conventional neighborhood. A neo-urbanist neighborhood is one with high population density, high bicycle lane and roadway street density, good land-use mix, and good transit and non-motorized mode accessibility/facilities. A conventional neighborhood is one with relatively low population density, low bicycle lane and roadway street density, primarily single use residential land use, and auto-dependent urban design. Assume that the VMT of households living in conventional neighborhoods is higher than the VMT of households residing in neo-urbanist neighborhoods. The question is whether this difference in VMT between households in conventional and neo-urbanist households is due to “true” effects of the built environment, or due to households self-selecting themselves into neighborhoods based on their VMT desires. For instance, it is at least possible (if not likely) that unobserved factors that increase the propensity or desire of a household to reside in a conventional neighborhood (such as overall auto inclination, a predisposition to enjoying travel, safety and security concerns regarding non-auto travel, *etc.*) also lead to the household putting more vehicle miles of travel on personal vehicles. If this self selection is not accounted for, the difference in VMT attributed directly to the variation in the built environment between conventional and neo-urbanist neighborhoods can be mis-estimated. On the other hand, accommodating for such self-selection effects can aid in identifying the “true” causal effect of the built environment on VMT.

The situation just discussed can be cast in the form of Roy’s (1951) endogenous switching model system (see Maddala, 1983; Chapter 9), which takes the following form:

$$\begin{aligned}
 r_q^* &= \beta'x_q + \varepsilon_q, & r_q &= 1 \text{ if } r_q^* > 0, & r_q &= 0 \text{ if } r_q^* \leq 0, \\
 m_{q0}^* &= \alpha'z_q + \eta_q, & m_{q0} &= 1[r_q = 0]m_{q0}^* \\
 m_{q1}^* &= \gamma'w_q + \xi_q, & m_{q1} &= 1[r_q = 1]m_{q1}^*
 \end{aligned} \tag{25}$$

The notation $1[r_q = 0]$ represents an indicator function taking the value 1 if $r_q = 0$ and 0 otherwise, while the notation $1[r_q = 1]$ represents an indicator function taking the value 1 if $r_q = 1$ and 0 otherwise. The first selection equation represents a binary discrete decision of households to reside in a neo-urbanist built environment neighborhood or a conventional built environment neighborhood. r_q^* in Equation (25) is the unobserved propensity to reside in a conventional neighborhood relative to a neo-urbanist neighborhood, which is a function of an ($M \times 1$)-column vector x_q of household attributes (including a constant). β represents a

corresponding $(M \times 1)$ -column vector of household attribute effects on the unobserved propensity to reside in a conventional neighborhood relative to a neo-urbanist neighborhood. In the usual structure of a binary choice model, the unobserved propensity r_q^* gets reflected in the actual observed choice r_q ($r_q = 1$ if the q th household chooses to reside in a conventional neighborhood, and $r_q = 0$ if the q th household decides to reside in a neo-urbanist neighborhood). ε_q is usually a standard normal or logistic error term capturing the effects of unobserved factors on the residential choice decision.

The second and third equations of the system in Equation (25) represent the continuous outcome variables of log(vehicle miles of travel) in our empirical context. m_{q0}^* is a latent variable representing the logarithm of miles of travel if a random household q were to reside in a neo-urbanist neighborhood, and m_{q1}^* is the corresponding variable if the household q were to reside in a conventional neighborhood. These are related to vectors of household attributes z_q and w_q , respectively, in the usual linear regression fashion, with η_q and ξ_q being random error terms. Of course, we observe m_{q0}^* in the form of m_{q0} only if household q in the sample is observed to live in a neo-urbanist neighborhood. Similarly, we observe m_{q1}^* in the form of m_{q1} only if household q in the sample is observed to live in a conventional neighborhood.

The potential dependence between the error pairs (ε_q, η_q) and (ε_q, ξ_q) has to be expressly recognized in the above system, as discussed earlier from an intuitive standpoint.⁵ The classic econometric estimation approach proceeds by using Heckman's or Lee's approaches or their variants (Heckman, 1974, 1976, 1979, 2001; Greene, 1981; Lee, 1982, 1983; Dubin and McFadden, 1984). Lee's full information maximum likelihood approach has seen more application in the literature relative to the other approaches because of its simple structure, ease of estimation using a maximum likelihood approach, and its lower vulnerability to the collinearity problem of two-step methods. But Lee's approach is also critically predicated on the bivariate normality assumption on the transformed normal variates in the discrete and continuous equation, which imposes the restriction that the dependence between the transformed discrete and continuous choice error terms is linear and symmetric. The copula-based approach, described in the preceding section, allows us to test alternative bivariate distributional assumptions for the discrete and continuous error terms. The model estimation approach to implementing alternative functional forms is described subsequently.

3.4.1 Model Estimation

In the current section, we discuss the maximum likelihood estimation approach for estimating the parameters of Equation system (25) with different copulas.

Let the univariate standardized marginal cumulative distribution functions of the error terms $(\varepsilon_q, \eta_q, \xi_q)$ in Equation (25) be $(F_\varepsilon, F_\eta, F_\xi)$, respectively. Assume that η_q has a scale parameter of σ_η , and ξ_q has a scale parameter of σ_ξ . Also, let the standardized joint

⁵The reader will note that it is not possible to identify any dependence parameters between (η_q, ξ_q) because the vehicle miles of travel is observed in only one of the two regimes for any given household.

distribution of (ε_q, η_q) be $F(.,.)$ with the corresponding copula $C_{\theta_0}(.,.)$, and let the standardized joint distribution of (ε_q, ξ_q) be $G(.,.)$ with the corresponding copula $C_{\theta_1}(.,.)$.

Consider a random sample size of Q ($q=1,2,\dots,Q$) with observations on $(r_q, m_{q0}, m_{q1}, x_q, z_q, w_q)$. The switching regime model has the following likelihood function:

$$L = \prod_{q=1}^Q \left[\frac{1}{\sigma_\eta} \cdot f_\eta \left(\frac{m_{q0} - \alpha' z_q}{\sigma_\eta} \right) \cdot \frac{\partial}{\partial u_{q2}^0} C_{\theta_0}(u_{q1}^0, u_{q2}^0) \right]^{1-r_q} \times \left[\frac{1}{\sigma_\xi} \cdot f_\xi \left(\frac{m_{q1} - \gamma' w_q}{\sigma_\xi} \right) \left\{ 1 - \frac{\partial}{\partial u_{q2}^1} C_{\theta_1}(u_{q1}^1, u_{q2}^1) \right\} \right]^{r_q}, \quad (26)$$

where $u_{q1}^0 = F_\varepsilon(-\beta' x_q)$, $u_{q2}^0 = F_\eta \left(\frac{m_{q0} - \alpha' z_q}{\sigma_\eta} \right)$, $u_{q1}^1 = u_{q1}^0$, $u_{q2}^1 = F_\xi \left(\frac{m_{q1} - \gamma' w_q}{\sigma_\xi} \right)$.

Any copula function can be used to generate the bivariate dependence between (ε_q, η_q) and (ε_q, ξ_q) , and the copulas can be different for these two dependencies (*i.e.*, C_{θ_0} and C_{θ_1} need not be the same). Thus, there is substantial flexibility in specifying the dependence structure, while still staying within the maximum likelihood framework and not needing any simulation machinery. In the current paper, we use normal distribution functions for the marginals $F_\varepsilon(.)$, $F_\eta(.)$ and $F_\xi(.)$, and test various different copulas for C_{θ_0} and C_{θ_1} .

3.4.2 Empirical Analysis

The data used for this analysis is drawn from the 2000 San Francisco Bay Area Household Travel Survey (BATS) designed and administered by MORPACE International Inc. for the Bay Area Metropolitan Transportation Commission (MTC). The dependent variables of interest in the current effort are (1) household residential location characterized as neo-urbanist and conventional neighborhoods (see Pinjari *et al.*, 2008 for details) and (2) household VMT.

The empirical analysis involved estimating models with the same structure for (ε_q, η_q) and (ε_q, ξ_q) , as well as different copula-based dependency structures. This led to 6 models with the same copula dependency structure (corresponding to the six copulas discussed in Section 3.3), and 24 models with different combinations of the six copula dependency structures for (ε_q, η_q) and (ε_q, ξ_q) . We also estimated a model that assumed independence between ε_q and η_q , and ε_q and ξ_q .

The maximum-likelihood estimation of the sample selection model with different copulas leads to a case of non-nested models. The most widely used approach to select among the competing non-nested copula models is the Bayesian Information Criterion (or BIC; see Trivedi and Zimmer, 2007, page 65). The BIC collapses to a comparison of the log-likelihood values across different models. The log-likelihood values for the five best copula dependency structure combinations are: (1) Frank-Frank (-6842.2), (2) Frank-Joe (-6844.2), (3) FGM-Joe (-6851.0), (4) Independent-Joe (-6863.7), and (5) FGM-Gumbel (-6866.2). It is evident that the log-likelihood at convergence of the Frank-Frank and Frank-Joe copula combinations are higher

compared to the other copula combinations. Between the Frank-Frank and Frank-Joe copula combinations, the former is slightly better. The log-likelihood value for the structure that assumes independence (*i.e.*, no self-selection effects) is -6878.1. All the five copula-based dependency models reject the independence assumption at any reasonable level of significance, based on likelihood ratio tests, indicating the significant presence of self-selection effects. Interestingly, however, the log-likelihood value at convergence for the classic textbook structure that assumes a Gaussian-Gaussian copula combination is -6877.9, indicating that there is no statistically significant difference between the Gaussian-Gaussian (G-G) and the independence-independence (I-I) copula structures. Clearly, the traditional G-G copula combination indicates the absence of self-selection effects. However, this is simply an artifact of the normal dependency structure, and is indicative of the kind of incorrect results that can be obtained by placing restrictive distributional assumptions.

In the following presentation of the empirical results, we focus our attention on the results of the Independent-Independent (or I-I copula) specification that ignores self-selection effects entirely and the Frank-Frank (or F-F copula) specification that provides the best data fit. Table 2 provides the results, which are discussed below.

3.4.2.1 Binary choice component

The results of the binary discrete equation of neighborhood choice provide the effects of variables on the propensity to reside in a conventional neighborhood relative to a neo-urbanist neighborhood. The parameter estimates indicate that younger households (*i.e.*, households whose heads are less than 35 years of age) are less likely to reside in conventional neighborhoods and more likely to reside in neo-urbanist neighborhoods, perhaps because of higher environmental sensitivity and/or higher need to be close to social and recreational activity opportunities (see also Lu and Pas, 1999). Households with children have a preference for conventional neighborhoods, potentially because of a perceived better quality of life/schooling for children in conventional neighborhoods compared to neo-urbanist neighborhoods. Also, as expected, households who own their home and who live in a single family dwelling unit are more likely to reside in conventional neighborhoods.

3.4.2.2 Log(VMT) continuous component for neo-urbanist neighborhood regime

The estimation results corresponding to the natural logarithm of vehicle miles of travel (VMT) in a neo-urbanist neighborhood highlight the significance of the number of household vehicles and number of full-time students. As expected, both of these effects are positive. In particular, log(VMT) increases with number of vehicles in the household and number of students. The effect of number of vehicles is non-linear, with a jump in log(VMT) for an increase from no vehicles to one vehicle, and a lesser impact for an increase from one vehicle to 2 or more vehicles (there were only two households in neo-urbanist neighborhoods with 3 vehicles, so we are unable to estimate impacts of vehicle increases beyond 2 vehicles in neo-urbanist neighborhoods). Interestingly, we did not find any statistically significant effect of employment and neighborhood characteristics, in part because the variability of these characteristics across households in neo-urbanist zones is relatively small.

The copula dependency parameter between the discrete choice residence error term and the log(VMT) error term for neo-urbanist households is highly statistically significant and negative for the F-F model. The θ estimate translates to a Kendall's τ value of -0.26. The

negative dependency parameter indicates that a household that has a higher inclination to locate in conventional neighborhoods would travel less than an observationally equivalent “random” household if both these households were located in a neo-urbanist neighborhood (a “random” household, as used above, is one that is indifferent between residing in a neo-urbanist or a conventional neighborhood, based on factors unobserved to the analyst). Equivalently, the implication is that a household that makes the choice to reside in a neo-urbanist neighborhood is likely to travel more than an observationally equivalent random household in a neo-urbanist environment, and much more than if an observationally equivalent household from a conventional neighborhood were relocated to a neo-urbanist neighborhood. This may be attributed to, among other things, such unobserved factors characterizing households inclined to reside in neo-urbanist settings as a higher degree of comfort level driving in dense, one-way street-oriented, parking-loaded, traffic conditions.

The lower travel tendency of a random household in a neo-urbanist neighborhood (relative to a household that expressly chooses to locate in a neo-urbanist neighborhood) is teased out and reflected in the high statistically significant negative constant in the F-F copula model. On the other hand, the I-I model assumes, incorrectly, that the travel of households choosing to reside in neo-urbanist neighborhoods is independent of the choice of residence. The result is an inflation of the VMT generated by a random household if located in a neo-urbanist setting.

3.4.2.3 Log(VMT) continuous component for conventional neighborhood regime

The household socio-demographics that influence vehicle mileage for households in a conventional neighborhood include number of household vehicles, number of full-time students, and number of employed individuals. As expected, the effects of all of these variables are positive. The household vehicle effect is non-linear, with the marginal increase in log(VMT) decreasing with the number of vehicles. In addition, two neighborhood characteristics – density of vehicle lanes and accessibility to shopping – have statistically significant effects on log(VMT) in the conventional neighborhood regime. Both these effects are negative, as expected.

The dependency parameter in this segment for the F-F model is highly statistically significant and positive. The θ estimate translates to a Kendall’s τ value of 0.36. The positive dependency indicates that a household that has a higher inclination to locate in conventional neighborhoods is likely to travel more in that setting than an observationally equivalent random household. Again, the I-I model ignores this residential self-selection in the estimation sample, resulting in an over-estimation of the VMT generated by a random household if located in a conventional neighborhood setting (see the higher constant in the I-I model relative to the F-F model corresponding to the conventional neighborhood VMT regime).

4. CONCLUSION

This chapter has presented an overview of discrete choice models structures including multinomial logit model, generalized extreme value (GEV) models, mixed multinomial logit (MMNL) models, mixed generalized extreme value models and other mixed discrete choice models. The formulations presented are quite flexible although estimation using the maximum likelihood technique requires the evaluation of multi-dimensional integrals (in the MMNL and MGEV models). Since, there have been several comprehensive reviews of discrete choice models, the current chapter has emphasized more on some recent developments in discrete-

continuous modeling in this chapter. In particular, the chapter presents a discussion of the recently formulated copula-based methodology for modeling discrete-continuous choices.

The approach is based on the concept of a “copula”, which is a multivariate functional form for the joint distribution of random variables derived purely from pre-specified parametric marginal distributions of each random variable. The copula concept has been recognized in the statistics field for several decades now, but it is only recently that it has been explicitly recognized and employed in the econometrics field. The reasons for the recent interest in the copula approach for modeling discrete-continuous choices can be summarized as follows: First, the copula approach does not entail any more computational burden than the commonly used Lee’s approach (Lee, 1983). Second, the approach allows the analyst to stay within the familiar maximum likelihood framework for estimation and inference, and does not entail any kind of numerical integration or simulation machinery. Third, the approach allows the marginal distributions in the discrete and continuous equations to take on any parametric distribution, just as in Lee’s method. Finally, under the copula approach, Lee’s coupling method is but one of a suite of different types of couplings that can be tested.

In addition to the general overview of copulas, the current chapter provides a discussion of copula properties and several alternate copula dependency structures employed in the econometrics and statistics literature. Further, to illustrate the copula approach, we discuss a copula based approach to model residential neighborhood choice and daily household vehicle miles of travel (VMT) using the 2000 San Francisco Bay Area Household Travel Survey (BATS). The self-selection hypothesis in the current empirical context is that households select their residence locations based on their travel needs, which implies that observed VMT differences between households residing in neo-urbanist and conventional neighborhoods cannot be attributed entirely to built environment variations between the two neighborhoods types. This example is based on the result of Bhat and Eluru (2009).

To summarize, the recent advancements in econometrics should facilitate the application of behaviorally rich structures in transportation-related discrete and discrete-continuous choice modeling in the years to come.

REFERENCES

- Abbe, E., M. Bierlaire, and T. Toledo (2007). Normalization and correlation of cross-nested logit models. *Transportation Research Part B*, 41(7), 795-808.
- Abdel-Aty, M., and H. Abdelwahab (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. *Accid. Anal. Prev.* 36, 447-456.
- Adler, T., C.S. Falzarano, and G. Spitz (2005). Modeling service trade-offs in air itinerary choices. *Transportation Research Record*, 1915, 20-26.
- Amador, F.J., R.M. Gonzalez, and J.D. Ortuzar (2005). Preference heterogeneity and willingness to pay for travel time savings. *Transportation*, 32, 627-647.
- Antonini, G., M. Bierlaire, and M. Weber (2006). Discrete Choice Models of Pedestrian Walking Behavior. *Transportation Research Part B*, 40, 667-687.
- APTA (2005). Connecting America: Making the Case for Intercity Rail Commissioned by the American Public Transportation Association (APTA)'s Committee on Intercity Corridor Development, Prepared by Reconnecting America, November 2005.
- ASA (2003). Austin San Antonio Intermunicipal Community Rail District 2003 Annual Report.
- Batley, R., T. Fowkes, D. Watling, G. Whelan, A. Daly, and E. Hato (2001a). Models for analysing route choice, Paper presented at the 33rd Annual Conference of the Universities Transport Studies Group Conference, University of Oxford.
- Batley, R., T. Fowkes, G. Whelan, and A. Daly (2001b). Models for choice of departure time, Paper presented to the European Transport Conference, Association of European Transport, University of Cambridge.
- Bajwa, S., S. Bekhor, M. Kuwahara, and E. Chung (2006). Discrete choice modeling of combined mode and departure time. Paper presented at the 11th International Conference on Travel Behavior Research, Kyoto, August.
- Bekhor, S., and J.N. Prashker (2001). Stochastic User Equilibrium Formulation for Generalized Nested Logit Model. *Transportation Research Record*, 1752, 84-90.
- Bekhor, S., M.E. Ben-Akiva, and M.S. Ramming (2002). Adaptation of logit kernel to route choice situation. *Transportation Research Record*, 1805, 78-85.
- Bekhor, S., M.E. Ben-Akiva, and M.S. Ramming (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144 (1), 235-247
- Ben-Akiva, M., and B. Francois (1983). μ -Homogeneous generalized extreme value model. Working paper, Department of Civil Engineering, MIT, Cambridge, MA.
- Ben-Akiva, M. and S.R. Lerman (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press, Cambridge, MA.
- Ben-Akiva, M., D. Bolduc, and M. Bradley (1993). Estimation of travel model choice models with randomly distributed values of time. *Transportation Research Record*, 1413, 88-97.
- Ben-Akiva, M., D. Bolduc, and J. Walker (2001). Specification, estimation and identification of the logit kernel (or continuous mixed logit) model. Working Paper, Department of Civil Engineering, MIT.
- Berkovec, J., and J. Rust (1985). A nested logit model of automobile holdings for one vehicle households. *Transportation Research Part B*, 19(4), 275-285
- Bierlaire, M. (2002). The network GEV model. Presented at the 2002 Swiss Transport Research Conference (STRC), Monte Verita.

- Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, 144(1), 287-300
- Bhat, C.R. (1997). Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel. *Transportation Research Part B*, 31(1) 11-21.
- Bhat, C.R. (1998a). An analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B*, 32(6) 361-371.
- Bhat, C.R. (1998b). Accommodating variations in responsiveness to level-of-service variables in travel mode choice modeling. *Transportation Research Part A*, 32(7), 495-507.
- Bhat, C.R. (1998c). Accommodating flexible substitution patterns in multidimensional choice modeling: Formulation and application to travel mode and departure time choice. *Transportation Research Part B*, 32(7), 455-466.
- Bhat, C.R. (2000a). Flexible model structures for discrete choice analysis. In D.A. Hensher and K.J. Button (eds.) *Handbook of Transport Modelling*, 71-90, Pergamon, New York.
- Bhat, C.R. (2000b). Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation Science*, 34(2) 228-238.
- Bhat, C.R., and J. Guo (2004). A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research Part B*, 38(2), 147-168.
- Bhat, C.R., and J.Y. Guo (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 N. Eluru (2009). A Copula-Based Approach to Accommodate Residential Self-Selection in Travel Behavior Modeling. *Transportation Research Part B*, 43(7), 749-765.
- Bhat, C.R., and R. Sardesai (2006). The impact of stop-making and travel time reliability on commute mode choice, *Transportation Research Part B*, 40(9), 709-730.
- Boyer, B., M. Gibson, and M. Loretan (1999). Pitfalls in tests for changes in correlation. International Finance Discussion Paper 597, Board of Governors of the Federal Reserve System.
- Bolduc, D., B. Fortin, and M. Fournier (1996). The effect of incentive policies on the practice location of doctors: A multinomial probit analysis. *Journal of Labor Economics*, 14(4), 703-732.
- Boyd, J., and J. Mellman (1980). The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis. *Transportation Research Part A*, 14, 367-378.
- Bresnahan, T.F., S. Stern, and M. Trajtenberg (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, 28(0) S17-S44.
- Brownstone, D., and K.A. Small (2005). Valuing time and reliability: Assessing the evidence from road pricing demonstrations, *Transportation Research Part A*, 39(4), 279-293.
- Brownstone, D., and K. Train (1998). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89, 109-129.
- Cameron, A.C., T. Li, P. Trivedi, and D. Zimmer (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *The Econometrics Journal* 7(2), 566-584.
- Cardell, S., and F. Dunbar (1980). Measuring the societal impacts of automobile downsizing, *Transportation Research Part A*, 14, 423-434.

- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.
- Cherubini, U., E. Luciano, and W. Vecchiato (2004). *Copula Methods in Finance*. John Wiley & Sons, Hoboken, NJ.
- Chu, C. (1989). A paired combinational logit model for travel demand analysis. In *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, 4, 295-309, Western Periodicals, Ventura, CA.
- Cirillo, C., and K.W. Axhausen (2006). Evidence on the distribution of values of travel time savings from a six-week diary. *Transportation Research Part A*, 40, 444-457.
- Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of family tendency in chronic disease incidence. *Biometrika* 65(1), 141-151.
- Conway, D.A. (1979). Multivariate distributions with specified marginals. Technical Report #145, Department of Statistics, Stanford University.
- Daly, A., and M. Bierlaire (2006). A general and operational representation of generalised extreme value models. *Transportation Research Part B*, 40(4), 285-305.
- Daly, A., and S. Zachary (1978). Improved multiple choice models. In D.A. Hensher and M.Q. Dalvi (eds.) *Determinants of Travel Choice*, 335-357, Saxon House, Westmead
- Dubin, J., and D. McFadden (1984). An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica*, 52, 345-362.
- Dugundji, E.R., and J.L. Walker (2005). Discrete choice with social and spatial network interdependencies. *Transportation Research Record*, 1921, 70-78.
- Eluru, N., and C. Bhat (2007). A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. *Accident Analysis and Prevention*, 39(5), 1037-1049.
- Embrechts, P., A.J. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: Properties and pitfalls. In M. Dempster (ed.) *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, 176-223.
- Farlie, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* 47(3-4), 307-323.
- Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. *Transportation Research Part B*, 40(8), 688-707.
- Fosgerau, M., and M. Bierlaire (2007). A practical test for the choice of mixing distribution in discrete choice models. *Transportation Research Part B*, 41(7), 784-794.
- Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, 19(1), 194-226.
- Frees, E.W., and P. Wang (2005). Credibility using copulas. *North American Actuarial Journal* 9(2), 31-48.
- Garrow, L.A., and F.S. Koppelman (2004). Multinomial and Nested Logit Models of Airlines Passengers' No-Show and Standby Behaviour. *Journal of Revenue and Pricing Management*, 3(3), 237-253.
- Gelhausen, M.C., and D. Wilken (2006). Airport and Access Mode Choice - A Generalized Nested Logit Model Approach. Proceedings of the 10th Air Transport Research Society Conference 2006 in Nagoya, Japan, 1-30.
- Genest, C., and A.-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12(4), 347-368.

- Genest, C., and R.J. MacKay (1986). Copules archimediennes et familles de lois bidimensionnelles dont les marges sont donnees. *The Canadian Journal of Statistics* 14(2), 145-159.
- Gkritza, K., and F.L. Mannering (2008). Mixed logit analysis of safety-belt use in single-and multi-occupant vehicles. *Accident Analysis and Prevention*, 40(2), 443-451
- Greene, W. (1981). Sample selection bias as a specification error: A comment. *Econometrica* 49(3), 795-798.
- Greene, W.H., and D.A. Hensher (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B*, 37(8) 681-698.
- Greene, W.H., D.A. Hensher, and J. Rose (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B*, 40, 75-92.
- Grimaldi, S., and F. Serinaldi (2006). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources* 29(8), 1155-1167.
- Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York.
- Gumbel, E.J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292), 698-707.
- Gupta, S., and P.K. Chintagunta (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, 31(1) 128-136.
- Heckman, J. (1974). Shadow prices, market wages and labor supply. *Econometrica*, 42(4), 679-694.
- Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *The Annals of Economic and Social Measurement*, 5(4), 475-492.
- Heckman, J.J. (1979). Sample selection bias as a specification error, *Econometrica*, 47(1), 153-161.
- Heckman, J.J. (2001). Microdata, heterogeneity and the evaluation of public policy. *Journal of Political Economy*, 109(4), 673-748.
- Hensher, D.A. (2001). Measurement of the valuation of travel time savings. *Journal of Transport Economics and Policy*, 35(1), 71-98.
- Hensher, D.A., and W.H. Greene (2002). Specification and Estimation of the Nested Logit Model: Alternative Normalization. *Transportation Research Part B*, 36, 1-17.
- Hensher, D.A. and W.H. Greene (2003). The mixed logit model: The state of practice. *Transportation*, 30(2), 133-176.
- Hess, S., and K.W. Axhausen (2005). Distributional assumptions in the representation of random taste heterogeneity. Presented at the 5th Swiss Transport Research Conference (STRC), Monte Verita, March 2005.
- Hess, S., and J.W. Polak (2006). Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area. *Transportation Research Part E*, 42(2), 63-81.
- Hess, S., M. Bierlaire, and J.W. Polak (2004). Development and application of a mixed cross-nested logit model. Proceedings of the XXIth European Transport Conference.
- Hess, S., M. Bierlaire, and J.W. Polak (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A*, 39, 221-236.
- Hess, S., J.M. Rose, and D.A. Hensher (2008). Asymmetric preference formation in willingness to pay estimates in discrete choice models. *Transportation Research Part E*, 44(5), 847-863.

- Huard, D., G. Evin, and A.-C. Favre (2006). Bayesian copula selection. *Computational Statistics & Data Analysis*, 51(2), 809-822.
- Joe, H. (1993). Parametric families of multivariate distributions with given marginals. *Journal of Multivariate Analysis*, 46(2), 262-282.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- Jong, G. de, M. Pieters, and A. Daly (2002a). Testing models for time of day choice with individual-specific effects. Working Paper, RAND Europe, Leiden.
- Jong, G. de, A. Daly, M. Pieters, C. Vellay, M. Bradley, and F. Hofman (2002b). A model for time of day and mode choice using error components logit. Working Paper, RAND Europe, Leiden.
- Junker, M., and A. May (2005). Measurement of aggregate risk with copulas. *The Econometrics Journal*, 8(3), 428-454.
- Kamakura, W.A., and G. Russell (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26(4) 379-390.
- Koppelman, F.S., and C.R. Bhat (2006). A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. Prepared for U.S. Department of Transportation Federal Transit Administration.
- Koppelman, F., and V. Sethi (2005). Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior. *Transportation Research Part B*, 39(9), 825-853.
- Koppelman, F., and C. Wen (2000). The paired combination logit model: Properties, estimation and application. *Transportation Research Part B*, 34(2) 75-89.
- Kotz, S., N. Balakrishnan, and N.L. Johnson (2000). *Continuous Multivariate Distributions, Vol. 1, Models and Applications*, 2nd edition. John Wiley & Sons, New York.
- Kwerel, S.M. (1988). Frechet bounds. In S. Kotz, N. L. Johnson (eds.) *Encyclopedia of Statistical Sciences*, Wiley & Sons, New York, 202-209.
- Lapparent, M. (2006). The choice of a mode of transportation for home to work trips in the French Parisian region: application of mixed GEV models within non linear utility functions. Presented at the 11th International Conference on Travel Behavior Research, Kyoto, August.
- Lee, L.-F. (1982). Some approaches to the correction of selectivity bias. *Review of Economic Studies*, 49(3), 355-372.
- Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(2), 507-512.
- Lijesen, M.G. (2006). A mixed logit based valuation of frequency in civil aviation from SP-data. *Transportation Research Part E*, 42(2), 82-94.
- Lo, H.K., C.-W. Yip, and Q.K. Wan (2004). Modeling Competitive Multi-modal Transit Services: A Nested Logit Approach. *Transportation Research Part C*, 12(3-4), 251-272.
- Lu, X.L., and E.I. Pas (1999). Socio-demographics, activity participation, and travel behavior. *Transportation Research Part*, 33(1), 1-18.
- Luce, R., and P. Suppes (1965). Preference, utility and subjective probability, In R. Luce, R. Bush, and E. Galanter (eds) *Handbook of mathematical psychology*, Vol. 3, Wiley New York.
- Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, U.K.

- Marzano, V., and A. Papola (2008). On the covariance structure of the cross-nested logit model. *Transportation Research Part B*, 42(2), 83-98.
- McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, 672, 72-77.
- Milton, J., V. Shankar, and F. Mannering (2008). Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention*, 40(1), 260-266.
- Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungsblatt fur Mathematische Statistik*, 8(3), 234-235.
- Munizaga, M., and R. Alvarez-Daziano (2001). Mixed logit versus nested logit and probit. Working Paper, Departamento de Ingeniera Civil, Universidad de Chile.
- Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed., Springer-Verlag, New York.
- Newman, J.P. (2008). Normalization of network generalized extreme value models. *Transportation Research Part B*, 42(10), 958-969
- Newman, J.P. (2009). Systematically heterogeneous covariance in network GEV models. Report TRANSP-OR 090511 Transport and Mobility Laboratory Ecole Polytechnique Fédérale de Lausanne.
- Papola, A. (2004). Some Developments on the Cross-Nested Logit Model. *Transportation Research Part B*, 38(9), 833-851.
- Persch, J.A. (2008). America's fastest train moves ahead. Msnbc.com, retrieved 10/06/2009
- Pinjari, A.R., and Bhat, C.R. (2006). On the nonlinearity of response to level of service variables in travel mode choice models. *Transportation Research Record*, 1977, 67-74.
- Pinjari, A. R., Eluru, N., Bhat, C. R., Pendyala, R. M., Spissu, E. (2008). Joint model of choice of residential neighborhood and bicycle ownership: Accounting for self-selection and unobserved heterogeneity. *Transportation Research Record* 2082, 17-26.
- Prieger, J.E. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of Applied Econometrics* 17(4), 367-392.
- Quinn, C. (2007). The health-economic applications of copulas: Methods in applied econometric research. Health, Econometrics and Data Group (HEDG) Working Paper 07/22, Department of Economics, University of York.
- Revelt, D., and K. Train (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level, *Review of Economics and Statistics*, 80, 647-657.
- Revelt, D., and K. Train (2000). Customer-specific taste parameters and mixed logit. Working Paper No. E00-274, Department of Economics, University of California, Berkeley.
- Rizzi, L.I., and J.D. Ortúzar (2003). Stated preference in the valuation of interurban road safety. *Accident Analysis and Prevention*, 35(1), 9-22.
- Robin, T., G. Antonini, M. Bierlaire, and J. Cruz (2009). Specification, estimation and validation of a pedestrian walking behavior model. *Transportation Research Part B*, 43(1), 36-56.
- Roy, A.D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, New Series 3(2), 135-146.
- Schmidt, T. (2007). Coping with copulas. In J. Rank (ed.) *Copulas - From Theory to Application in Finance*, 3-34, Risk Books, London.
- Schweizer, B., and A. Sklar (1983). *Probabilistic Metric Spaces*. North-Holland, New York.
- Sener, I.N., R.M. Pendyala, and C.R. Bhat (2009). Accommodating Spatial Correlation Across Choice Alternatives in Discrete Choice Models: An Application to Modeling Residential

- Location Choice Behavior. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.
- Siikamaki, J., and Layton, D. (2001). Pooled models for contingent valuation and contingent ranking data: valuing benefits from biodiversity conservation, Working Paper, Department of Agricultural and Resource Economics, University of California, Davis.
- Sivakumar, A., and Bhat, C.R. (2007). A comprehensive, unified, framework for analyzing spatial location choice. *Transportation Research Record*, 2003, 103-111.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229-231.
- Sklar, A., (1973). Random variables, joint distribution functions, and copulas. *Kybernetika* 9, 449-460.
- Small, K. (1987). A discrete choice model for ordered alternatives. *Econometrica*, 55(2) 409-424.
- Small, K., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists' preferences for travel time and reliability: Implications for road pricing. *Econometrica*, 73(4), 1367-1382.
- Smith, M.D. (2005). Using copulas to model switching regimes with an application to child labour. *Economic Record*, 81(S1), S47-S57.
- Sørensen, M. V., and O. A. Nielsen (2003). MSL for Mixed Logit Model Estimation—On Shape of Distributions. *Proc., European Transport Conference*, Strasbourg, France..
- Srinivasan, K.K., and S.R. Athuru (2005). Analysis of within-household effects and between-household differences in maintenance activity allocation. *Transportation*, 32, 495-521.
- Srinivasan, K.K., and H.S. Mahmassani (2003). Analyzing heterogeneity and unobserved structural effects in route-switching behavior under ATIS: A dynamic kernel logit formulation. *Transportation Research Part B*, 37(9), 793-814.
- Srinivasan, K.K., and Ramadurai, G. (2006). Dynamics and variability in within-day mode choice decisions: role of state dependence, habit persistence, and unobserved heterogeneity. Presented at the Transportation Research Board 85th Annual Meeting, Washington D.C.
- Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research B*, 35(7) 643-666.
- TIA (2004). Domestic Research: Travel Volume and Trends. Retrieved from <http://www.tia.org/Travel/tvt.asp>
- Train, K. (1986). *Qualitative Choice Analysis*, MIT Press, Cambridge.
- Train, K. (2001). A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. Working Paper, Department of Economics, University of California, Berkeley.
- Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.
- Train, K., and G. Sonnier (2005). Mixed logit with bounded distributions of correlated partworths. In R. Scarpa and A. Alberini (Eds) *Applications of Simulation Methods in Environmental and Resource Economics*, Kluwer Academic Publishers, Boston, MA.
- Trivedi, P.K., and D.M. Zimmer (2007). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics* 1(1), Now Publishers.

- USDOT (2006). America on the Go: Long Distance Transportation Patterns: Mode Choice. US Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, May.
- Vichiensan, V., K. Miyamoto, and Y. Tokunaga (2005). Mixed logit model framework with structuralized spatial effects: a test of applicability with area unit systems in location analysis. *Journal of Eastern Asia Society for Transportation Studies*, 6, 3789-3802.
- Vovsha, P. (1997). Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Research Record*, 1607, 6-15.
- Walker, J.L. (2002). Mixed logit (or logit kernel) model: dispelling misconceptions of identification. *Transportation Research Record*, 1805, 86-98.
- Walker, J., Ben-Akiva, M., Bolduc, D. (2004). Identification of the logit kernel (or mixed logit) model. Presented at the 10th International Conference on Travel Behavior Research, Lucerne, Switzerland, August
- Wen, C-H., and F.S. Koppelman (2001). The generalized nested logit model. *Transportation Research Part B*, 35(7), 627-641.
- Whelan, G., R. Batley, T. Fowkes, and A. Daly (2002). Flexible models for analyzing route and departure time choice. European Transport Conference Proceedings, Association for European Transport, Cambridge.
- Williams, H. (1977). On the formation of travel demand models and economic evaluation measures of user benefits. *Environment and Planning A*, 9, 285-344.
- Zimmer, D.M., and P.K. Trivedi (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business and Economic Statistics*, 24(1), 63-76.

Table 1 Characteristics of Alternative Copula Structures†

Copula	Dependence Structure Characteristics	Archimedean Generation Function $\psi(t)$	$\psi'(t)$	θ range and value for index	Kendall's τ and range	Spearman's ρ_s and range
Gaussian	Radially symmetric, weak tail dependencies, left and right tail dependencies go to zero at extremes	Not applicable	Not applicable	$-1 \leq \theta \leq 1$ $\theta = 0$ is independence	$\frac{2}{\pi} \arcsin(\theta)$ $-1 \leq \tau \leq 1$	$\frac{6}{\pi} \arcsin\left(\frac{\theta}{2}\right)$ $-1 \leq \rho_s \leq 1$
FGM	Radially symmetric, only moderate dependencies can be accommodated	Not applicable	Not applicable	$-1 \leq \theta \leq 1$ $\theta = 0$ is independence	$\frac{2}{9}\theta$ $-\frac{2}{9} \leq \tau \leq \frac{2}{9}$	$\frac{1}{3}\theta$ $-\frac{1}{3} \leq \rho_s \leq \frac{1}{3}$
Clayton	Radially asymmetric, strong left tail dependence and weak right tail dependence, right tail dependence goes to zero at right extreme	$\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$	$t^{-\theta-1}$	$0 < \theta < \infty$ $\theta \rightarrow 0$ is independence	$\frac{\theta}{\theta + 2}$ $0 < \tau < 1$	No simple form $0 < \rho_s < 1$
Gumbel	Radially asymmetric, weak left tail dependence, strong right tail dependence, left tail dependence goes to zero at left extreme	$\varphi(t) = (-\ln t)^\theta$	$-\frac{\theta}{t}(-\log t)^{\theta-1}$	$1 \leq \theta < \infty$ $\theta = 1$ is independence	$1 - \frac{1}{\theta}$ $0 \leq \tau < 1$	No simple form $0 \leq \rho_s < 1$
Frank	Radially symmetric, very weak tail dependencies (even weaker than Gaussian), left and right tail dependencies go to zero at extremes	$\varphi(t) = -\ln \left[\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right]$	$\frac{\theta}{1 - e^{\theta t}}$	$-\infty < \theta < \infty$ $\theta \rightarrow 0$ is independence	See Equation (22) $-1 \leq \tau \leq 1$	$1 - \frac{12}{\theta}(D_1 \theta) - D_2(\theta)^*$ $-1 \leq \rho_s \leq 1$
Joe	Radially asymmetric, weak left tail dependence and very strong right tail dependence (stronger than Gumbel), left tail dependence goes to zero at left extreme	$\varphi(t) = -\ln[1 - (1-t)^\theta]$	$\frac{-\theta(1-t)^{\theta-1}}{1 - (1-t)^\theta}$	$1 \leq \theta < \infty$ $\theta = 1$ is independence	See Equation (24) $0 \leq \tau < 1$	No simple form $0 \leq \rho_s < 1$

$$* D_k(\theta) = \frac{k}{e^k} \int_{t=0}^{\theta} \frac{t^k}{(e^t - 1)} dt$$

†Source: Bhat and Eluru (2009)

Table 2 Estimation Results of the Switching Regime Model‡

Variables	Independence-Independence Copula		Frank-Frank Copula	
	Parameter	t-stat	Parameter	t-stat
Propensity to choose conventional neighborhood relative to neo-urbanist neighborhood				
Constant	0.201	4.15	0.275	5.72
Age of householder < 35 years	-0.131	-2.35	-0.143	-2.75
Number of children (of age < 16 years) in the household	0.164	4.62	0.161	4.59
Household lives in a single family dwelling unit	0.382	6.79	0.337	6.28
Own household	0.597	10.37	0.497	8.81
Log of vehicle miles of travel in a neo-urbanist neighborhood				
Constant	-0.017	-0.16	-0.638	-5.48
Household vehicle ownership				
Household Vehicles = 1	2.617	21.50	2.744	24.26
Household Vehicles ≥ 2	3.525	25.44	3.518	27.40
Number of full-time students in the household	0.183	2.13	0.112	1.41
Copula dependency parameter (θ)	--	--	-2.472	-6.98
Scale parameter of the continuous component	1.301	40.62	1.348	34.31
Log of vehicle miles of travel in a conventional neighborhood				
Constant	0.379	2.28	0.163	1.08
Household vehicle ownership				
Household Vehicles = 1	3.172	21.77	3.257	25.43
Household Vehicles = 2	3.705	25.32	3.854	29.92
Household Vehicles ≥ 3	3.931	25.92	4.102	30.41
Number of employed individuals in the household	0.229	7.24	0.208	6.66
Number of full-time students in the household	0.104	5.06	0.131	6.27
Density of bicycle lanes	-0.023	-3.08	-0.024	-3.24
Accessibility to shopping (Hansen measure)	-0.024	-7.34	-0.027	-8.19
Copula dependency parameter (θ)	--	--	3.604	7.22
Scale parameter of the continuous component	0.891	75.78	0.920	63.59
Log-likelihood at convergence	-6878.1		-6842.2	

‡ Source: Bhat and Eluru (2009)