

**On Jointly Analyzing the Physical Activity Participation Levels of Individuals in a Family  
Unit Using a Multivariate Copula Framework**

**Ipek N. Sener**

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: [ipek@mail.utexas.edu](mailto:ipek@mail.utexas.edu)

**Naveen Eluru**

The University of Texas at Austin  
Dept of Civil, Architectural & Environmental Engineering  
1 University Station C1761, Austin TX 78712-0278  
Phone: 512-471-4535, Fax: 512-475-8744  
E-mail: [naveeneluru@mail.utexas.edu](mailto:naveeneluru@mail.utexas.edu)

**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](mailto:bhat@mail.utexas.edu)

\*corresponding author

## **ABSTRACT**

The current paper focuses on analyzing and modeling the physical activity participation levels (in terms of the number of daily “bouts” or “episodes” of physical activity during a weekend day) of all members of a family jointly. Essentially, we consider a family as a “cluster” of individuals whose physical activity propensities may be affected by common household attributes (such as household income and household structure) as well as unobserved family-related factors (such as family life-style and health consciousness, and residential location-related factors). The proposed copula-based clustered ordered-response model structure allows the testing of various dependency forms among the physical activity propensities of individuals of the same household (generated due to the unobserved family-related factors), including non-linear and asymmetric dependency forms. The proposed model system is applied to study physical activity participations of individuals, using data drawn from the 2000 San Francisco Bay Area Household Travel Survey (BATS). A number of individual factors, physical environment factors, and social environment factors are considered in the empirical analysis. The results indicate that reduced vehicle ownership and increased bicycle ownership are important positive determinants of weekend physical activity participation levels, though these results should be tempered by the possibility that individuals who are predisposed to physical activity may choose to own fewer motorized vehicles and more bicycles in the first place. Our results also suggest that policy interventions aimed at increasing children’s physical activity levels could potentially benefit from targeting entire family units rather than targeting only children. Finally, the results indicate strong and asymmetric dependence among the unobserved physical activity determinants of family members. In particular, the results show that unobserved factors (such as residence location-related constraints and family lifestyle preferences) result in individuals in a family having uniformly low physical activity, but there is less clustering of this kind at the high end of the physical activity propensity spectrum

*Keywords:* Copulas, physical activity, family and public health, social dependency, data clustering, activity-based travel analysis

## 1. INTRODUCTION

The potentially serious adverse mental and physical health consequences of obesity have been well documented in epidemiological studies (see, for instance, Nelson and Gordon-Larsen, 2006, and Ornelas *et al.*, 2007). While there are several factors influencing obesity, it has now been established that a low level of physical activity is certainly an important contributing factor (see, Haskell *et al.*, 2007, and Steinbeck, 2008). Besides, earlier studies in the literature strongly emphasize the importance of physical activity even in non-obese and non-overweight individuals from the standpoint of increasing cardiovascular fitness, improved mental health, and decreasing heart disease, diabetes, high blood pressure, and several forms of cancer (USDHHS, 2008; Center for Disease Control (CDC), 2006). But, despite these well acknowledged benefits of physical activity, a high fraction of individuals in the U.S. and other developed countries lead relatively sedentary (or physically inactive) lifestyles. For instance, the 2007 Behavioral Risk Factor Surveillance System (BRFSS) survey suggests that about a third of U.S. adults are physically inactive, while the 2007 Youth Risk Behavior Surveillance survey indicates that about 65.3% of high school students do not meet the current physical activity guidelines.<sup>1</sup>

The low level of physical activity participation in the U.S. population has prompted several research studies in the past decade to examine the determinants of physical activity participation, with the objective of designing appropriate intervention strategies to promote active lifestyles. However, as we discuss later, most of these studies focus on adult physical activity participation or children's/adolescents' physical activity participation, without explicitly considering family-level interactions due to observed and unobserved factors in the physical activity participation levels of all individuals (adults and children/adolescents) of the same family. In this regard, the current paper focuses on analyzing and modeling the physical activity participation levels (in terms of the discrete choice of the number of daily "bouts" or "episodes" of physical activity) of all members of a family jointly. Essentially, we consider a family as a "cluster" of individuals whose physical activity levels may be affected by common household attributes (such as household income and household structure) as well as unobserved family-related factors (such as residential location-related constraints/facilitators of physical activity

---

<sup>1</sup> The current guidelines call for at least 150 minutes a week of moderate-level physical activity (such as jogging, running, mountain climbing, and bicycling uphill) or 75 minutes a week of vigorous-level physical activity (such as brisk walking, bicycling, and water aerobics) for adults. In addition, children and adolescents should participate in at least 60 minutes of physical activity every day, and this activity should be at a vigorous level at least 3 days a week (USDHHS, 2008).

and/or family life-style and health consciousness factors). Ignoring such family-specific interactions due to unobserved factors (also referred to as unobserved heterogeneity in the econometric literature) will, in general, result in inconsistent estimates regarding the influence of covariates and inconsistent probability predictions in discrete choice models (see Chamberlain, 1980 and Hsiao, 1986). This, in turn, can lead to misinformed intervention strategies to encourage physical activity.

The joint generation of physical activity episodes at the household level is also important from an activity-based travel modeling perspective. As discussed by Copperman and Bhat (2007a), much of the focus on activity generation (and scheduling) and inter-individual interactions in the activity analysis field has been on adult patterns. In contrast, few studies have explicitly considered the activity patterns of children, and the interactions of children's patterns with those of adults' patterns, when children are present in the household. If the activity participation of children with adults is primarily driven by the activity participation needs/responsibilities of adults (such as a parent wanting to go to the gym, and tagging along her/his child for the trip), then the emphasis on adults' activity-travel patterns would be appropriate. However, in many instances, it is the children's activity participations, and the dependency of children on adults for facilitating the participations that lead to interactions between adults' and children's activity-travel patterns. Of course, in addition, children can also impact adults' activity-travel patterns in the form of joint activity participation in such activities as shopping, going to the park, walking together, and other social-recreational activities. The joint generation of physical activity episodes in the current paper is consistent with such an emphasis on both adults' and children's activity-travel patterns within a household.

### **1.1 Overview of Earlier Studies on Physical Activity Participation**

The body of work in the area of understanding the determinants of physical activity participation has been burgeoning in the past decade or so in many different disciplines, including child development, preventive medicine, sports medicine, public health, physical activity, and transportation. The intent here is not to provide an exhaustive review of these past studies (some good recent reviews of these works are Wendel-Vos *et al.*, 2005, Allender *et al.*, 2006, Gustafson and Rhodes, 2006, and Ferreira *et al.*, 2007). However, one may make two general observations from past analytic studies. First, almost all of these analytic studies focus on individual physical

activity without recognition that individuals are part of families and that there are potentially strong family interactions in physical activity levels. In this regard, the studies focus on either adults only or children/adolescents only. That is, they have adopted either an “adult-centric” approach focusing on adult physical activity patterns, and used children’s demographic variables (such as presence/number of children in the household) as determinant variables, or a “child-centric” approach focusing on children’s physical activity patterns, and used adults’ (parents’) demographic, attitudinal, and physical activity variables (such as number of adults in the household, support for children’s physical activity, and adults’ physical activity levels) as determinant variables (see Sener and Bhat, 2007 for more details on these approaches; examples of adult-centric studies include Collins *et al.*, 2007, Srinivasan and Bhat, 2008, Dunton *et al.*, 2008, while examples of child-centric studies include Davison *et al.*, 2003, Trost *et al.*, 2003, Cleland *et al.*, 2005, Sener *et al.*, 2008, and Ornelas *et al.*, 2007).<sup>2</sup> While these earlier studies provide important information on the determinants of adults’ or children’s physical activity levels, they do not explicitly recognize the role of the family as a fundamental social unit for the development of overall physical activity orientations and lifestyles. This is particularly important considering parental influence on, and involvement in, children’s physical activities, as well as children’s physical activity needs/desires that may influence parents’ (among other household members) physical activity patterns. Since these effects are likely to be reinforcing (either toward high physical activity levels or low physical activity levels), the appropriate way to consider these family interactions would be to model the physical activity levels of all family members jointly as a package, considering observed and unobserved covariate effects.<sup>3</sup>

---

<sup>2</sup> The works of Trost *et al.* (2003) and Davison *et al.* (2003) are particularly valuable, since they examine different mechanisms through which parents may influence their children’s physical activity pursuits. As identified by Trost *et al.* (2003), these may include genetics, direct modeling (*i.e.*, parents’ own physical activity involvement effects on children’s physical activity levels), provision of time and money resources to support children’s activities, rewarding desirable behaviors and punishing/ignoring undesirable behaviors, parents’ own attitudes and beliefs about the importance of physical activity, and adopting authoritative parenting procedures to encourage children’s physical activity. While most studies in the literature adopt the direct modeling hypothesis, Trost *et al.* (2003) suggest that support-related and parenting beliefs/attitudes are perhaps more important predictors of children’s physical activity levels than direct modeling. Davison *et al.* (2003) indicate that both direct modeling and parental support/parenting practices influence children’s (girls’) physical activity levels.

<sup>3</sup> Note that the clustering effects in physical activity levels among individuals in a family may be due to parental influences and support (or lack of support) for physical activities of children, as discussed earlier. Since parental attitudes and beliefs are likely to impact parental influence, and attitudes/beliefs as well as support mechanisms may be unobserved to the analyst, this could generate dependence in unobserved factors affecting the physical activity levels within a family. However, there are other possible reasons for such family-level clustering. For instance, the quality of physical activity recreation facilities accessible to a family from its residence may be relatively poor, and if this lack of “quality” is difficult to measure/observe, it can be an unobserved deterrent to the physical activity

The second general observation from earlier studies is that they have proposed three broad groups of determinants of individual physical activity within an ecological framework: individual or intrapersonal factors, physical environment factors, and social environment or interpersonal factors (*e.g.* Sallis and Owen, 2002, Giles-Corti and Donovan, 2002, Gordon-Larsen *et al.*, 2005, U.S. Government Accountability Office, 2006; Kelly *et al.*, 2006, Salmon, 2007, and Bhat and Sener, 2009). The category of individual factors includes demographics (such as age, education levels, and gender), and work-related characteristics (employment status, hours of week, work schedule, work flexibility, *etc.*). The category of physical environment factors includes weather, season of year, transportation system attributes (level of service offered by various alternative modes for participation in out-of-home activities), and built environment characteristics (BECs). The final category of social environment factors includes family-level demographics (presence and age distribution of children in the household, household structure, and household income), residential neighborhood demographics, social and cultural mores, attitudes related to, and in support of, physical activity pursuits, and perceived friendliness of one's residential neighborhood. Of these three groups of factors, public health researchers have focused more on the first and third categories of factors (*i.e.*, the individual and social environment factors), particularly as they correlate to participation in such recreational physical activity as sports, walking/biking for leisure, working out at the gym, and unstructured play (see, for instance, Kelly *et al.*, 2006; Salmon, 2007, and Dunton *et al.*, 2008). On the other hand, transportation and urban planning researchers have particularly focused their attention on the first and second category of factors (with limited consideration of the third category in the form of family-level demographics) as they relate to non-motorized mode use for utilitarian activity

---

participation of all individuals in a family. Also, it is not uncommon for families to undertake joint recreational activities, and some families may be more "activity-cohesive" in undertaking recreational pursuits. Such family cohesion effects, when complemented with an overall activity lifestyle orientation, have been shown in earlier qualitative psycho-social and family interaction studies to be positive determinants of the physical activity predispositions of members in a family (see, for example, Ornelas *et al.*, 2007, Springer *et al.*, 2006, Strauss *et al.*, 2001, Allender *et al.*, 2006). If such qualitative indicators of family interaction are unavailable to an analyst, as in the current study, these indicators effectively serve as unobserved facilitators to the physical activity participation of all members of a family. Related to family cohesion, but also a potentially different mechanism for clustering, is family communication intensity. In families with high communication intensity, it is possible that the children affect adults through their acquired (from outside the home) interest or uninterest in physical activities (rather than a one-way impact of parental attitudes on the physical activity levels of all members of the household). This can be another source of clustering effects (see Allender *et al.*, 2006). Overall, the clustering effects can be due to correlated constraints faced by family members (such as residential-location related factors), or correlated lifestyle preferences (such as family cohesion activities) or belief/attitude spillover effects ("rubbing off" of beliefs/attitudes among individuals in a household, moderated by family communication levels), or combinations of these.

purposes (*i.e.* non-motorized forms of travel to participate in an out-of-home activity episode at a specific destination, such as walking/biking to school or to work or to shop; see, for instance, Dill and Carr, 2003, Cervero and Duncan, 2003, and Sener *et al.*, 2009). There have been few studies that consider elements of all three groups of physical activity determinants, and that consider recreational physical activities and non-motorized travel for utilitarian purposes (but see Hoehner *et al.*, 2005 and Copperman and Bhat, 2007a for a couple of exceptions).

## **1.2 The Current Paper in Context and Paper Structure**

In this paper, we contribute to the earlier literature by focusing on the family as a “cluster unit” when modeling the physical activity levels of individuals. In this regard, and because earlier physical activity studies have focused only on adults or only on children, our emphasis is on analyzing physical activity levels of families with one or more parents and children in the household. That is, we examine the determinants of physical activity in the context of family households with children. In doing so, we explicitly accommodate family-level observed and unobserved effects that may influence the physical activity levels of each (and all) individual(s) in the family. Further, we consider variables belonging to all the three groups of individual factors, physical environment factors, and social environment factors. In particular, we incorporate a rich set of neighborhood physical environment variables such as land use structure and mix, population size and density, accessibility measures, demographic and housing measures, safety from crime, and highway and non-motorized mode network measures. However, in the context of social factors, we do not explicitly accommodate physical activity attitudes/beliefs and support systems of individual family members as they influence the physical activity levels of others in the family. This is because our data source does not collect such information, though it is well suited to examine the influence of several other potential determinants. Future studies would benefit from including family-level attitudinal/support variables, while also adopting a family-level perspective of physical activity.

The measure of physical activity we adopt in the current study is the number of out-of-home bouts or episodes (regardless of whether these bouts correspond to recreation or to

walking/biking for utilitarian purposes) on a weekend day as reported in an activity survey.<sup>4</sup> Activity surveys typically collect information on all types of (out-of-home) episodes of all individuals in sampled households over the course of 1 or 2 days. As indicated by Dunton *et al.* (2008), the use of a short-term (1-2 days) self-report reduces memory-related errors compared to other long-term methods of data collection used in the physical activity literature (such as self-reports over a week or a month). Further, survey data allow the consideration of the social context (family characteristics and physical activity levels of family members), while methods that examine the level of use of physical activity environments (such as a park or a playground) do not provide information to consider the social context in any depth. Also, for our family-level modeling of physical activity, survey data provide information on physical activity participation for all members of a family.<sup>5</sup> Finally, the activity survey data used here provide information on residential location, which is used to develop measures of the physical environment variables in the family's neighborhood. Of course, a limitation of activity survey-based data is that some episodes of physical activity, such as free play, in-home physical activity, and incidental physical activity may not be identified well. Further, activity surveys do not provide a measure of the physical activity intensity level. Thus, there are strengths and limitations of using survey data, but such data are ideally suited for family-level cluster analysis of the type undertaken in the current effort.

From a methodological standpoint, the daily number of physical activity episodes of each individual is represented using an ordered response structure, which is appropriate for situations where the dependent variable is ordinal (that is, the dependent variable values have a natural ordering; see Section 2.1 for a description of the ordered-response structure). The jointness between the episodes of different members of the same family is generated by common household demographic and location variables, as well as through dependency among the stochastic error terms of the random latent variables assumed to be underlying the observed

---

<sup>4</sup> The analysis focuses on weekend days because of the high prevalence and duration of participation in physical activities over the weekend days relative to weekdays (see Lockwood *et al.*, 2005), as well as because there is much more joint activity participation within a family (and therefore interactions within a family cluster) on weekend days relative to weekdays (see Srinivasan and Bhat, 2008 and Copperman and Bhat, 2007a). Children, in particular, participate in discretionary activities at much higher levels, and for substantially longer durations, on weekend days compared to weekdays (Stefan and Hunt, 2006).

<sup>5</sup> As we discuss later, the characterization of an activity episode as a physically active one or not is based on the activity type and the type of location (such as bowling alley, gymnasium, shopping mall, *etc.*). Thus, an episode involving recreation activity at a soccer stadium is designated as a physical activity episode. For travel episodes, the episode is designated as physically active if it involves walking or bicycling.



discrete number of physical activity episodes.<sup>6</sup> In the current paper, we allow non-linear and asymmetric error dependencies using a copula structure, which is essentially a multivariate functional form for the joint distribution of random variables derived purely from pre-specified parametric marginal distributions of each random variable. To our knowledge, this is the first formulation and application in the econometric literature of the copula approach for the case of a clustered ordered response model structure.

The rest of this paper is structured as follows. The next section discusses and presents the copula-based clustered ordered-response model structure. Section 3 describes the survey-based data source and sample formation procedures for the empirical analysis. Section 4 discusses the empirical results, and presents the results of a policy-based simulation. Finally, Section 5 summarizes important findings from the study, and concludes the paper.

## **2. MODEL STRUCTURE**

### **2.1 Background**

This paper uses an ordered-response model for analyzing the number of physical activity episodes for each individual. The assumption in this model is that there is an underlying continuous latent variable representing the propensity to participate in physical activity whose partitioning into discrete intervals, based on thresholds on the continuous latent variable scale, maps into the observed set of count outcomes. While the traditional ordered-response model was initially developed for the case of ordinal responses, and while count outcomes are cardinal, this distinction is really irrelevant for the use of the ordered-response system for count outcomes. This is particularly the case when the count outcome takes few discrete values, as in the current empirical case, but is also not much of an issue when the count outcome takes a large number of possible values (see Herrerias *et al.*, 2008 and Ferdous *et al.*, 2010 for detailed discussions).

An important issue, though, is that we have to recognize the potential dependence in the number of physical activity episodes of different members of the same family due to both observed exogenous variables as well as unobserved factors. If there is no dependence based on unobserved factors, one can accommodate the dependence due to observed factors by estimating

---

<sup>6</sup> The analysis in the current paper may be viewed as a reduced form analysis, based off an appropriate (and flexible) econometric structure to deal with the ordinal nature of the daily number of physical activity episodes as well as family-level clustering effects. It is not a structural model based on a formal behavioral process of physical activity generation nor does it explicitly disentangle the many processes that may lead to family-level clustering effects.

independent ordered-response models for each individual in the family after including common exogenous variables. But the dependence due to unobserved family-related factors (such as family life-style and health consciousness, and residential location-related factors) can be accommodated only by jointly modeling the number of episodes of all family members together. This is the classic case of clusters of dependent random variables that has widely been studied and modeled in the transportation and other fields (see Bhat, 2000, Bottai *et al.*, 2006, and Czado and Prokopenko, 2008). In our case, the clusters correspond to family units, although the methodology we present in the current paper can be used for any situation involving clusters.

An established method to deal with unobserved interactions due to cluster effects is a random effects model. In the ordered-response context, this entails adding a common cluster-based normal error term to the latent underlying propensities for each individual in the cluster (see Bhat and Zhao, 2002 for a detailed explanation of the mathematical formulation as well as an empirical example of this method). The main limitation of the random effects model is the restrictive assumption introduced in the dependence structure through the random normal error term. Thus, for instance, in the random effects ordered-response probit model, the joint distribution of error terms is considered multivariate normal, which assumes that the dependence (due to unobserved factors) among the physical activity propensities of family members is radially symmetric. On the other hand, it may be the case that the dependence among the propensities of family members is actually asymmetric; for instance, one may observe family members having a simultaneously low propensity for physical activity participation, but not necessarily family members having a simultaneously high propensity for physical activity participation. That is, unobserved factors that decrease physical activity propensity may “rub off” more among individuals in a family than unobserved factors that increase physical activity propensity. Alternatively, one may have the reverse asymmetry too where family members have a simultaneously high propensity for physical activity propensity, but not a simultaneously low propensity for physical activity propensity.

In the current paper, rather than using the random effects approach, we use a copula approach to accommodate the dependence in physical activity propensity among family members. 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 (see Trivedi and Zimmer, 2007 and Nelsen, 2006). The use of a copula to generate a

joint distribution of a cluster outcome is convenient and flexible for a number of reasons. First, the approach allows testing of a variety of parametric marginal distributions for individual members in a cluster and preserves these marginal distributions when developing the joint probability distribution of the cluster. Second, the copula approach separates the marginal distributions from the dependence structure, so that the dependence structure is entirely unaffected by the marginal distributions assumed. Thus, rank measures of the intra-cluster dependence of the underlying physical activity propensities for members of a family are independent of the marginal distributions used, facilitating a clear interpretation of the dependence structure regardless of the marginal distribution assumed. Third, the clustering context, wherein the level of dependence in the marginal random unobserved terms within a cluster is identical (*i.e.*, exchangeable) across any (and all) pairs of individuals in the cluster, is ideal for the application of a group of copulas referred to as the Archimedean copulas. The Archimedean copulas are closed-form copulas that can be used to obtain the joint multivariate cumulative distribution function of any number of individuals belonging to a cluster. Further, these copulas retain the same form regardless of cluster size, and so it is straightforward to accommodate clusters of varying sizes.<sup>7</sup> Fourth, the Archimedean group of copulas allows testing a variety of radially symmetric and asymmetric joint distributions, as well as testing the assumption of within-cluster independence. Fifth, it is simple to allow the level of dependence within a cluster to vary based on cluster type. For example, the dependence among family members in their latent propensities of physical activity may vary by such family characteristics as family type or income. Finally, the closed-form nature of the model structure resulting from using the Archimedean group of copulas lends itself very nicely to the implementation of a computationally straightforward maximum likelihood procedure for parameter estimation.

---

<sup>7</sup> Technically speaking, one may use a copula approach to allow differential dependence levels among marginal random unobserved terms within a cluster. For instance, it may be argued that the “rubbing off” effects due to unobserved factors (in the context of physical activity participation) are higher between two children in a family than between two adults in a family, or between two adults in a family than between an adult and a child. While such differential dependency patterns within a cluster can be accommodated with specific copula forms (see Bhat and Sener, 2009 and Bhat *et al.*, 2010), they are, in general, quite difficult to accommodate and estimate using maximum likelihood methods. Alternatively, one can estimate models with differential dependency patterns within a cluster using pairwise copulas (*i.e.*, a bivariate copula for each pair of individuals in a family), but such an approach may not have an equivalent multivariate distribution interpretation. The approach we propose and use here is particularly appropriate for cluster-specific effects, where there is an equal level of unobserved dependence between all pairs of entities in a cluster. Such uniform cluster-specific effects are assumed also in the traditional random effects approach discussed earlier.

## 2.2 Copula Basics

The word “copula” was coined by Sklar (1959), and is derived from the Latin word “copulare”, which means to tie, bond, or connect (see Schmidt, 2007). 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 (see Nelsen, 2006, Trivedi and Zimmer, 2007, Bhat and Eluru, 2009). 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 an  $I$ -dimensional copula of uniformly distributed random variables  $U_1, U_2, U_3, \dots, U_I$  with support contained in  $[0,1]^I$ . Then,

$$C_\theta(u_1, u_2, \dots, u_I) = \Pr(U_1 < u_1, U_2 < u_2, \dots, U_I < u_I), \quad (1)$$

where  $\theta$  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  $I$  random variables  $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_I$ , each with univariate continuous marginal distribution function  $F(z_i) = \Pr(\varepsilon_i < z_i)$ .<sup>8</sup> Then, by Sklar’s (1973) theorem, a joint  $I$ -dimensional distribution function of the random variables with the continuous marginal distribution functions  $F(z_i)$  can be generated as follows:

$$\begin{aligned} F(z_1, z_2, \dots, z_I) &= \Pr(\varepsilon_1 < z_1, \varepsilon_2 < z_2, \dots, \varepsilon_I < z_I) = \Pr[U_1 < F(z_1), U_2 < F(z_2), \dots, U_I < F(z_I)] \\ &= C_\theta[u_1 = F(z_1), u_2 = F(z_2), \dots, u_I = F(z_I)]. \end{aligned} \quad (2)$$

The above equation offers a vehicle to develop different dependency patterns for the random variables  $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_I$  based on the copula that is used as the underlying basis of construction.

In the current paper, we use a class of copulas referred to as the Archimedean copulas to generate the dependency between the random variables. The next section briefly discusses the Archimedean class of copulas and presents some specific copulas within this broad family.

---

<sup>8</sup> Note that the univariate marginal distribution functions of the random variables can be different, though we use the more restrictive notation here that the univariate distributions are the same. This is the norm when developing econometric models where the random terms represent individual-level idiosyncratic effects.

### 2.3 Archimedean Copulas

The Archimedean class of copulas is popular in empirical applications, and includes a whole suite of closed-form copulas that cover a wide range of dependency formulations (see Nelsen, 2006 and Bhat and Eluru, 2009 for a detailed discussion). The class is very flexible, and easy to construct, as discussed next.

Archimedean copulas are constructed based on an underlying continuous convex decreasing generator function  $\varphi$  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  $\varphi(0) = \infty$ , so that an inverse  $\varphi^{-1}$  exists. Also, let  $\varphi^{-1}$  be completely monotonic on  $[0, \infty]$ . With these preliminaries, we can generate multivariate I-dimensional Archimedean copulas as:

$$C_\theta(u_1, u_2, u_3, \dots, u_I) = \varphi^{-1}\left(\sum_{i=1}^I \varphi(u_i)\right), \quad (3)$$

where the dependence parameter  $\theta$  is embedded within the generator function. An important characteristic of any multivariate Archimedean copula with the scalar dependence parameter  $\theta$  is that the marginal pairwise distributions between any two random variables (from  $U_1, U_2, U_3, \dots, U_I$ ) is bivariate Archimedean with the same copula structure as the multivariate copula. A whole variety of Archimedean copulas have been identified based on different forms of the generator function  $\varphi$ . In this paper, we will consider four of the most popular Archimedean copulas that span the spectrum of different kinds of dependency structures. These are the Clayton, Gumbel, Frank, and Joe copulas (see Bhat and Eluru, 2009 for graphical descriptions of the implied dependency structures). All these copulas, in their multivariate forms, allow only positive associations and equal dependencies among pairs of random variables, which is well-suited for cluster analysis because we expect positive and equal dependencies among elements within a cluster.

The Clayton copula (Clayton, 1978) has the generator function  $\varphi(t) = (1/\theta)(t^{-\theta} - 1)$ , giving rise to the following I-dimensional copula function (see Huard *et al.*, 2006):

$$C_\theta(u_1, u_2, \dots, u_I) = \left( \left[ \sum_{i=1}^I u_i^{-\theta} \right] - (I-1) \right)^{-1/\theta}, \quad 0 < \theta < \infty. \quad (4)$$

Independence corresponds to  $\theta \rightarrow 0$ . The copula is best suited for strong left tail dependence and weak right tail dependence. That is, it is best suited when individuals in a family show strong tendencies to have low physical activity levels together but not high activity levels together.

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 I-dimensional copula is provided below:

$$C_\theta(u_1, u_2, \dots, u_I) = \exp\left(-\left[\sum_{i=1}^I (-\ln u_i)^\theta\right]^{1/\theta}\right), \quad 1 \leq \theta < \infty. \quad (5)$$

Independence corresponds to  $\theta = 1$ . This copula is well suited for the case when there is strong right tail dependence (strong correlation at high values) but weak left tail dependence (weak correlation at low values). Thus, this copula would be applicable when individuals in a family show strong tendencies to have high physical activity levels together but not low activity levels together.

The Frank copula, proposed by Frank (1979), is radially symmetric in its dependence structure like the Gaussian (normal) copula. The generator function is  $\varphi(t) = -\ln[(e^{-\theta t} - 1)/(e^{-\theta} - 1)]$ , and the corresponding copula function is given by:

$$C_\theta(u_1, u_2, \dots, u_I) = -\frac{1}{\theta} \ln\left(1 + \frac{\prod_{i=1}^I (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{I-1}}\right), \quad 0 < \theta < \infty. \quad (6)$$

Independence is attained in Frank's copula as  $\theta \rightarrow 0$ . This copula is suitable for equal levels of dependency in the left and right tails; that is, when individuals either show low physical activity levels together or high activity levels together.

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:

$$C_\theta(u_1, u_2, \dots, u_I) = 1 - \left[1 - \prod_{i=1}^I (1 - (1 - u_i)^\theta)\right]^{1/\theta}, \quad 1 \leq \theta < \infty. \quad (7)$$

The Joe copula is similar to the Gumbel copula, but the right tail positive dependence is stronger. Independence corresponds to  $\theta = 1$ .

## 2.4 Model Formulation

Let  $q$  be an index for clusters (family unit in the current empirical context) ( $q = 1, 2, \dots, Q$ ), and let  $i$  be the index for individuals ( $i = 1, 2, \dots, I_q$ , where  $I_q$  denotes the total number of individuals in family  $q$ , including adults and children; in the current study  $I_q$  varies between 2 and 5). Also, let  $k$  be an index for the discrete outcomes corresponding to the number of weekend day physical activity episodes ( $k = 0, 1, 2, 3, \dots, K$ ). In the usual ordered response framework notation, we write the latent propensity ( $y_{qi}^*$ ) of individual  $i$  in family  $q$  to participate in physical activity as a function of relevant covariates, and then relate this latent propensity to the count outcome ( $y_{qi}$ ) representing the number of weekend physical activity episodes of individual  $i$  in family  $q$  through threshold bounds (see McKelvey and Zavoina, 1975):

$$y_{qi}^* = \beta' x_{qi} + \varepsilon_{qi}, \quad y_{qi} = k \text{ if } \psi_k < y_{qi}^* \leq \psi_{k+1}, \quad (8)$$

where  $x_{qi}$  is a ( $L \times 1$ ) vector of exogenous variables for individual  $i$  in family  $q$  (not including a constant),  $\beta$  is a corresponding ( $L \times 1$ ) vector of coefficients to be estimated, and  $\psi_k$  is the lower bound threshold for count level  $k$  ( $\psi_0 < \psi_1 < \psi_2 \dots < \psi_K < \psi_{K+1}$ ;  $\psi_0 = -\infty$ ,  $\psi_{K+1} = +\infty$ , and  $\psi_1, \psi_2, \dots, \psi_K$  are to be estimated).<sup>9</sup> The  $\varepsilon_{qi}$  terms capture the idiosyncratic effect of all omitted variables for individual  $i$  in family  $q$ , and are assumed to be independent of  $\beta$  and  $x_{qi}$ . The  $\varepsilon_{qi}$  terms are assumed identical across individuals, each with a univariate continuous marginal distribution function  $F(z_{qi}) = \Pr(\varepsilon_{qi} < z_{qi})$ . The error terms can take any parametric marginal distribution, though we confine ourselves to the normal and logistic distributions in the current paper. Due to identification considerations in the ordered-response model, we standardize the univariate distribution functions, so that they are standard normal or standard logistic distributed. However, we allow dependence in the  $\varepsilon_{qi}$  terms across individuals  $i$  in the same family unit  $q$  to allow unobserved cluster effects. This dependency is generated through the use of an Archimedean copula based on Equation (2), where the only difference now is the introduction of the index  $q$  to reflect that the dependence is confined to members of the same family:

---

<sup>9</sup> In the empirical analysis, we allow different thresholds for children and adults. From a strict notation standpoint, this implies that the thresholds should be subscripted as  $\psi_{ki}$ . However, for notational ease, we suppress the subscript  $i$  when writing the thresholds.

$$\begin{aligned} \Pr(\varepsilon_{q1} < z_{q1}, \varepsilon_{q2} < z_{q2}, \dots, \varepsilon_{qI_q} < z_{qI_q}) &= \Pr[U_{q1} < F(z_{q1}), U_{q2} < F(z_{q2}), \dots, U_{qI_q} < F(z_{qI_q})] \\ &= C_{\theta_q} [u_{q1} = F(z_{q1}), u_{q2} = F(z_{q2}), \dots, u_{qI_q} = F(z_{qI_q})]. \end{aligned} \quad (9)$$

It is important to note above that the level of dependence among individuals of a family can vary across families, as reflected by the  $\theta_q$  notation for the dependence parameter. As we indicate later, we parameterize this dependence parameter as a function of observed family characteristics in estimation, which allows us to accommodate different levels of dependency among individuals of different types of families.<sup>10</sup> Technically, one can also use different copula forms (*i.e.*, dependency surfaces) for different families, but, in the current paper, we will maintain the same copula form across all families to keep the estimation tractable (however, note that we test for different copula forms, even if we maintain the same copula form across all families).

## 2.5 Model Estimation

Let  $m_{qi}$  be the actual observed categorical response for  $y_{qi}$  in the sample. Then, the probability of the observed vector of number of episodes across individuals in household  $q$  ( $m_{q1}, m_{q2}, m_{q3}, \dots, m_{qI_q}$ ) can be written as:

$$P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, \dots, y_{qI_q} = m_{qI_q}) = \int_{M_q} c_{\theta_q}(F(y_{q1}^*), F(y_{q2}^*), \dots, F(y_{qI_q}^*)) \left( \prod_{i=1}^{I_q} f_q(y_{qi}^*) \right) dy_{q1}^* dy_{q2}^* \dots dy_{qI_q}^*, \quad (10)$$

where  $M_q = \{y_{q1}^*, y_{q2}^*, \dots, y_{qI_q}^* : \Psi_{(m_{qi})} < y_{qi}^* < \Psi_{(m_{qi+1})} \text{ for all } i = 1, 2, \dots, I_q\}$  and  $c_{\theta_q}$  is the copula density. The integration domain  $M_q$  is simply the multivariate region of the  $y_{qi}^*$  variables ( $i = 1, 2, \dots, I_q$ ) determined by the observed vector of choices ( $m_{q1}, m_{q2}, \dots, m_{qI_q}$ ). The dimensionality of the integration, in general, is equal to the number of individuals  $I_q$  in the family. Thus, if one uses a Gaussian copula, one ends up with integrals of the order of the

---

<sup>10</sup> The use of the notation  $\theta_q$  assumes that the dependency due to unobserved factors is confined to (and identical across) members within a family. In reality, it is possible that the dependency extends beyond members of the same family to members of families within a certain spatial neighborhood and/or within a certain defined social network. Accommodating such generalized multi-level unobserved effects is difficult with Archimedean copulas, but may be achieved using the Gaussian copula combined with a composite marginal likelihood inference approach (see Ferdous *et al.*, 2010, and Spissu *et al.*, 2010). Bhat (2009) has also recently proposed a generalized Gumbel copula within the class of Archimedean copulas that may be used for such multi-level modeling. Overall, the development of flexible copula approaches for the analysis of multi-level modeling is an important area for further methodological research.



number of individuals in the family for the joint probability of the observed combination of the number of activity episodes across individuals in the family. This will need simulation techniques when  $I_q$  is greater than 3. However, in the case of a family-level cluster with identical dependencies between pairs of individuals in the family, one can gainfully employ the Archimedean copulas since they provide closed-form multivariate cumulative distribution functions. In particular, the probability in Equation (10) can be written in terms of  $2^{I_q}$  closed-form multivariate cumulative distribution functions as follows:

$$\begin{aligned}
P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, \dots, y_{qI_q} = m_{qI_q}) &= P(\psi_{m_{q1}} < y_{q1}^* < \psi_{m_{q1+1}}, \psi_{m_{q2}} < y_{q2}^* < \psi_{m_{q2+1}}, \dots, \psi_{m_{qI_q}} < y_{qI_q}^* < \psi_{m_{qI_q+1}}) \\
&= \sum_{a_1=1}^2 \sum_{a_2=1}^2 \dots \sum_{a_{I_q}=1}^2 (-1)^{a_1+a_2+\dots+a_{I_q}} \left[ P(y_{q1}^* < \psi_{m_{q1+a_1-1}}, y_{q2}^* < \psi_{m_{q2+a_2-1}}, \dots, y_{qI_q}^* < \psi_{m_{qI_q+a_{I_q}-1}}) \right] \\
&= \sum_{a_1=1}^2 \sum_{a_2=1}^2 \dots \sum_{a_{I_q}=1}^2 (-1)^{a_1+a_2+\dots+a_{I_q}} \left[ C_{\theta_q}(u_{m_{q1+a_1-1}}, u_{m_{q2+a_2-1}}, \dots, u_{m_{qI_q+a_{I_q}-1}}) \right] \tag{11}
\end{aligned}$$

where  $C_{\theta_q}$  is the one of the four Archimedean copulas discussed in Section 2.3 with an association parameter  $\theta_q$ , and  $u_{m_{qi+a_i-1}} = F(\psi_{m_{qi+a_i-1}} - \beta' x_{qi})$ . The number of cumulative distribution function computations increases rapidly with the number of individuals  $I_q$  in family  $q$ , but this is not much of a problem when the cluster sizes are 6 or less because of the closed-form structures of the cumulative distribution functions. In the current empirical context,  $I_q \leq 5$ . However, in other empirical contexts when there are several individuals in a cluster, one can resort to the use of a composite marginal likelihood approach (see, for instance, the study by Bhat *et al.*, 2010 that employs a combined copula-CML approach to accommodate spatial dependence across observational units).

The association parameter  $\theta_q$  is allowed to vary across families. However, it is not possible to estimate a separate dependence term for each family. So, we parameterize  $\theta_q$  as a function of a vector  $s_q$  of observed family variables, while also choosing a functional form that ensures that  $\theta_q$  for any family  $q$  is within the allowable range for each copula. Thus, we use the form  $\theta_q = \exp(\delta' s_q)$  for the Frank and Clayton copulas, and the form  $\theta_q = 1 + \exp(\delta' s_q)$  for the Gumbel and Joe copulas.

The parameters to be estimated in the model may be gathered in a vector  $\Omega = (\beta', \delta', \psi')$ , where the vector  $\psi$  is the vector of threshold bounds:  $\psi = (\psi_1, \psi_2, \dots, \psi_K)$ . The likelihood function for household  $q$  may be constructed based on the probability expression in Equation (11) as:

$$L_q(\Omega) = P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, \dots, y_{qI_q} = m_{qI_q}). \quad (12)$$

$$\text{The likelihood function is then given by } L(\Omega) = \prod_q L_q(\Omega). \quad (13)$$

The likelihood function above is maximized using conventional maximum likelihood procedures approach. All estimations and computations were carried out using the GAUSS programming language. Gradients of the log-likelihood function with respect to the parameters were coded.

### 3. THE DATA

#### 3.1 The Primary Data Source

The primary source of data is the 2000 San Francisco Bay Area Travel Survey (BATS), which was designed and administered by MORPACE International, Inc. for the Bay Area Metropolitan Transportation Commission (see MORPACE International Inc., 2002). The survey collected detailed information on individual and household socio-demographic and employment-related characteristics from about 15,000 households in the Bay Area. The survey also collected information on all activity and travel episodes undertaken by individuals of the sampled households over a two-day period. For a subset of the sampled households, the two-day survey period included a Friday and a Saturday, or a Sunday and a Monday (however, no household was surveyed on both a Saturday and a Sunday). The current analysis uses the surveyed weekend day (either Saturday or Sunday) of these households. The information collected on activity episodes included the type of activity (based on a 17-category classification system), the name of the activity participation location (for example, Jewish community center, Riverpark plaza, *etc.*), the type of participation location (such as religious place, or shopping mall), start and end times of activity participation, and the geographic location of activity participation.

As discussed earlier, we identified whether an activity episode is physically active or not based on the activity type and the type of participation location at which the episode is pursued,

as reported in the survey.<sup>11</sup> Thus, an episode designated as “recreation” activity by a respondent and pursued at a health club (such as working out at the gym) is labeled as physically active. Similarly, an episode designated as “recreation” activity by a respondent and pursued outdoors (such as walking/running/bicycling around the neighborhood “without any specific destination”) is labeled as being physically active.<sup>12</sup> For the current analysis, we consider only out-of-home activity episodes. In addition, travel episodes to any out-of-home location using non-motorized forms of travel (bicycling and/or walking) are characterized as physical activity episodes. In this regard, each non-motorized travel episode ending at an activity location was characterized as a physical activity episode. For instance, if an individual goes to a grocery shopping center by bike and then returns back home, the individual is considered to have participated in two physical activity episodes.

After categorizing out-of-home episodes into physically active or otherwise, the number of physically active episodes during the weekend day for each individual in each family is obtained by appropriate aggregation. This constitutes the dependent variable in our analysis. Further, while the methodology developed can be used for all types of families, we focus only on families with children in this paper to examine both adults’ and children’s physical activity participations (while also accommodating family-level observed and unobserved effects). In terms of adults, we focus on parents’ physical activity participations and, in terms of children, we focus on the physical activity participation of children between the age of 5 to 15. Further, we restricted ourselves to families with three children or less as they accounted for approximately 97% of families with children.

---

<sup>11</sup> A physically active episode requires regular bodily movement during the episode, while a physically passive episode involves maintaining a sedentary and stable position for the duration of the episode. For example, swimming or walking around the neighborhoods would be a physically active episode, while going to a movie is a physically passive episode.

<sup>12</sup> A data-based limitation of the current study is that the data do not allow us to distinguish between individuals who are personally involved in the physical activity and those who are only present during the activity but not “physically” involved in the physical activity. Therefore, for instance, an episode designated as “recreation” activity by a respondent and pursued at a tennis court is labeled as physically active, regardless of whether the individual went to the tennis court to watch some other person play tennis or played tennis himself/herself. Note, however, that individuals who drop off/pick up others from the tennis courts will report their activity type as “pick-up/drop-off” and so this episode will not be considered as a physically active one. Also, there is some possibility that individuals who go to a tennis court and not play tennis will report their activity type as “social” or “resting/relaxing”, in which case these episodes will also not be characterized as “physically active” in our taxonomy.

### 3.2 The Secondary Data Sources

In addition to the 2000 BATS survey data set, several other secondary data sets were used to obtain transportation system attributes and built environment characteristics (within the broad group of physical environment factors discussed in Section 1.1), as well as residential neighborhood demographics (within the broad group of social environment factors in Section 1.1). All these variables were computed at the level of the residential traffic analysis zone (TAZ) of each household.<sup>13</sup> The secondary data sources included land use/demographic coverage data, the 2000 Census of population and household summary files, a Geographic Information System (GIS) layer of bicycle facilities, a GIS layer of highways and local roadways, and GIS layers of businesses. Among the secondary data sets indicated above, the land use/demographic coverage data, LOS data, and the GIS layer of bicycle facilities were obtained from the Metropolitan Transportation Commission (MTC). The GIS layers of highways and local roadways were obtained from the 2000 Census Tiger Files. The GIS layers of businesses were obtained from the InfoUSA business directory.

The transportation system and built environment measures constructed from the secondary data sources include:

1. *Zonal land use structure variables*, including housing type measures (fractions of single family, multiple family, duplex and other dwelling units), land use composition measures (fractions of zonal area in residential, commercial, and other land uses), and a land use mix diversity index computed as a fraction based on the land use composition measures with values between 0 and 1 (zones with a value closer to one have a richer land use mix than zones with a value closer to zero; see Bhat and Guo, 2007 for a detailed explanation on the formulation of this index).
2. *Regional accessibility measures*, which include Hansen-type (Fotheringham, 1983) employment, shopping, and recreational accessibility indices that are computed separately for the drive and transit modes.
3. *Zonal activity opportunity variables*, characterizing the composition of zones in terms of the intensity or the density of various types of activity centers. The typology used for activity centers includes five categories: (a) maintenance centers, such as grocery stores, gas stations,

---

<sup>13</sup> Due to privacy considerations, the point coordinates of each household's residence is not available; only the TAZ of residence of each household is available.

food stores, car wash, automotive businesses, banks, medical facilities, (b) physically active recreation centers, such as fitness centers, sports centers, dance and yoga studios, (c) physically passive recreational centers, such as theatres, amusement centers, and arcades, (d) natural recreational centers such as parks and gardens, and (e) restaurants and eat-out places.

4. *Zonal transportation network measures*, including highway density (miles of highway facilities per square mile), local roadway density (miles of roadway density per square mile), bikeway density (miles of bikeway facilities per square mile), street block density (number of blocks per square mile), non-motorized distance between zones (*i.e.*, the distance in miles along walk and bicycle paths between zones), and transit availability. The non-motorized distance between zones was used to develop an accessibility measure by non-motorized modes, computed as the number of zones (a proxy for activity opportunities) within “x” non-motorized mode miles of the teenager’s residence zone. Several variables with different thresholds for “x” were formulated and tested.

The residential neighborhood demographics constructed from the secondary data sources include:

1. *Zonal population size and employment/population density measures*, including total population, number of housing units, population density, household density, and employment density by several employment categories, as well as dummy variables indicating whether the area corresponds to a central business district (CBD), urban area, suburban area, or rural area.
2. *Zonal ethnic composition measures*, constructed as fractions of Caucasian, African-American, Hispanic, Asian and other ethnic populations for each zone.
3. *Zonal demographics and housing cost variables*, including average household size, median household income, and median housing cost in each zone.

### **3.3 Sample Characteristics**

The final sample used for the analysis comprises 1687 individuals (894 adults and 793 children) from 517 family households residing in nine Counties of the San Francisco Bay Area (Alameda, Contra Costa, San Francisco, San Mateo, Santa Clara, Solano, Napa, Sonoma and Marin). This final sample includes 377 two parent families (73.0% of all families), 85 single mother families (16.4% of all families), and 55 single father families (10.6% of all families). The number of

children in the family varies between one and three children, with the distribution as follows: one child (53.4%), two children (39.8), and three children (6.8 %). The distribution of the number of physically active episodes per weekend day in the entire sample of individuals is: zero episodes (79.8), one episode (17.5%), and two or more episodes (2.7%). The distribution within the sample of adults is zero episodes (80.3%), one episode (16.7%), and two or more episodes (3.0%), while the corresponding distribution within the sample of children is zero episodes (79.2%), one episode (18.4%), and two or more episodes (2.4%). These statistics reveal that there is no substantial difference in the aggregate distribution of the number of weekend day physically active episodes between adults and children.

## **4. MODEL RESULTS**

### **4.1 Variable Specification**

Several different variables within the three broad variable categories of individual factors, physical environment correlates, and social environment determinants were considered in our model specifications. The individual factors included demographics (age, sex, race, driver's license holding, physical disability status, *etc.*) and work-related characteristics (employment status, hours of week, work schedule, and work flexibility, *etc.*); the physical environment factors included weather, season of year, transportation system attributes, and built environment characteristics; and the social environment factors included family-level demographics (household composition and family structure, household income, dwelling type, whether the house is owned or rented, *etc.*) and residential neighborhood demographics (see Section 3.2 for details).

The final model specification was based on a systematic process of eliminating variables found to be statistically insignificant, intuitive considerations, parsimony in specification, and results from earlier studies. Several different variable specifications, functional forms of variables as well as interaction variables were considered for the  $x_{qi}$  vector (that determines exogenous variables affecting physical activity propensity) as well as for the  $s_q$  vector (that captures variations in the level of dependency based on observed family characteristics). The final specification includes some variables that are not highly statistically significant, because of their intuitive effects and potential to guide future research efforts in the field.

## 4.2 Model Specification and Data Fit

The empirical analysis involved estimating models with two different univariate distribution assumptions (normal and logistic) for the random error term  $\varepsilon_{qi}$ , and four different copula structures (Clayton, Gumbel, Frank and Joe) for specifying the dependency between the  $\varepsilon_{qi}$  terms across individuals in each family to represent the family cluster effect. Thus, a total of eight copula-based models were estimated: (1) Normal-Clayton, (2) Normal-Gumbel, (3) Normal-Frank, (4) Normal-Joe, (5) Logistic-Clayton, (6) Logistic-Gumbel, (7) Logistic-Frank, and (8) Logistic-Joe.

In addition, we also estimated two models (one with a normal marginal error term and the other with a logistic marginal error term) that assume independence in physical activity propensity among family members, as well as two models based on the more common methodological approach to accommodate clusters through a family-specific normal mixing error term. To allow a fair comparison between such random-effects models and the copula models, we specified the variance of the random error term in the random-effects models to vary across families based on observed family characteristics (see Bhat and Zhao, 2002, and Bhat, 2000 for such specifications in the past). Such a formulation accommodates heterogeneity across families in the level of association between family members, akin to parameterizing the  $\theta_q$  dependence term in the copula models as a function of the vector  $s_q$  of observed family variables.

To conserve on space, we will only provide the data fit results for the best copula model, the best independent model (from the logistic and the normal distributions for the  $\varepsilon_{qi}$  terms), and the best random-effects model (again from the logistic and normal distributions for the  $\varepsilon_{qi}$  terms). Note that the maximum likelihood estimation of the models with different copulas leads to a case of non-nested models. The most widely used approach to select among competing non-nested copula models is the Bayesian Information Criterion (or BIC; see Quinn, 2007, Genius and Strazzeria, 2008, and Trivedi and Zimmer, 2007, page 65). The BIC for a given copula model is equal to  $-2\ln(L) + B\ln(N)$ , where  $\ln(L)$  is the log-likelihood value at convergence,  $B$  is the number of parameters, and  $N$  is the number of observations. The copula that results in the lowest BIC value is the preferred copula. But, if all the competing models have the same exogenous variables and the same number of thresholds, as in our empirical case, the BIC information selection procedure measure is equivalent to selection based on the largest value of the log-likelihood function at convergence.

Among the copula models, our results indicated that the Logistic-Clayton (LC) model provides the best data fit with a likelihood value of  $-732.844$ .<sup>14</sup> Thus, based on the BIC measure, the LC model provides the best fit. However, the BIC measure does not indicate whether the LC model is statistically significantly better than its competitors. But, since all the copula models have the same value of the log-likelihood at sample shares (that is, when only the thresholds are included in the model), the alternative copula models can be statistically tested using a non-nested likelihood ratio test. In this regard, the difference in the adjusted rho-bar squared ( $\bar{\rho}_c^2$ ) values between the LC model and its closest competitor (which is the Logistic-Frank or LF model) is  $0.0006$ .<sup>15</sup> The probability that this difference could have occurred by chance is less than  $\Phi\{-[-2 \times 0.0006 \times L(C) + (28 - 28)]^{0.5}\}$ . This value, with  $L(C) = -3022.698$ , is almost zero, indicating that the difference in adjusted rho-bar squared values between the LC and the LF models is statistically significant and that the LC model is significantly superior to the LF model. However, note also that, in all the copula models, the dependency parameters were highly statistically significant, with the family-level dependency in unobserved factors varying based on family structure. Specifically, the family-level dependency was different among the three family types of (1) family with both parents, (2) single father family, and (3) single mother family. Between the two independent models, the logistic error term distribution for the margins (*i.e.*, the ordered-response logit or ORL) provided a marginally better fit than the normal error term distribution for the margins (*i.e.*, the ordered-response probit). The log-likelihood value at convergence for the ordered-response logit is  $-916.748$ . Also, between the random effects ordered-response logit (RORL) and the random-effects ordered-response probit (RORP) models, the former (*i.e.*, the RORL model) provided a superior data fit with a convergent log-likelihood value of  $-738.602$ . In both these random-effects models, we also considered variations in the family-level correlation levels across families, and found once again that there was variation based on the same family structure grouping as in the LC model.

---

<sup>14</sup> The likelihood values at convergence for the other copula models were as follows: Logistic-Gumbel ( $-747.75$ ), Logistic-Frank ( $-734.66$ ), Logistic-Joe ( $-752.79$ ), Normal-Clayton ( $740.01$ ), Normal-Gumbel ( $-749.34$ ), Normal-Frank ( $-735.49$ ), and Normal-Joe ( $-754.93$ ).

<sup>15</sup>The adjusted rho-bar squared value  $\bar{\rho}_c^2$  for an ordered-response model is computed as  $\bar{\rho}_c^2 = 1 - [(L(\hat{\beta}) - H) / L(C)]$ , where  $L(\hat{\beta})$  is the log-likelihood at convergence,  $H$  is the number of model parameters excluding the thresholds, and  $L(C)$  is the log-likelihood with only thresholds in the model.



The likelihood ratio test for testing the LC model in this paper with the ORL model is 367.81, which is substantially larger than the critical  $\chi^2$  value with 3 degrees of freedom (corresponding to the three dependency parameters) at any reasonable level of significance, confirming the importance of accommodating dependence in physical activity propensity among family members. The likelihood ratio test for testing the RORL model with the ORL model is 356.29, which again is larger than the critical  $\chi^2$  value with 3 degrees of freedom. The LC and RORL models are non-nested, and may be compared using a non-nested likelihood ratio test (both the LC and RORL models have the same exogenous variables and the same number of thresholds, while differing in the surface shape of the dependency among the error terms of different individuals in a family). Specifically, the difference in the adjusted rho-bar squared ( $\bar{\rho}_c^2$ ) values between the two models is 0.00191. The probability that this difference could have occurred by chance is less than  $\Phi\{-[2 \times 0.00191 \times L(C) + (28 - 28)]^{0.5}\}$ . This value, with  $L(C) = -3022.698$ , is almost zero, indicating that the difference in adjusted rho-bar squared values between the copula-based LC and the RORL models is highly statistically significant and that the copula model is to be preferred over the more traditional random effects model in terms of model fit. Specifically, as we discuss later, the results indicate a clear asymmetry in the dependence relationship among the physical activity propensities of individuals of the same family, an issue that cannot be handled by the random effects approach.

In addition to the model fit on the overall estimation sample, we also evaluated the performance of the ORL, RORL, and LC models on various market segments of the estimation sample (Ben-Akiva and Lerman, 1985 refer to such predictive fit tests as market segment prediction tests). The intent of using such predictive tests is to examine the performance of different models on sub-samples that do not correspond to the overall sample used in estimation. Effectively, the sub-samples serve a similar role as an out-of-sample for validation. The advantage of using the sub-sample approach rather than an out-of-sample approach to validation is that there is no reduction in the size of the sample for estimation. This is particularly an issue in our case because we have only 517 households for estimation. If a model shows superior performance in the subsamples in addition to the overall estimation sample, it is indication that the model indeed provides a better data fit. To evaluate performance of different models within each sub-sample, we use both aggregate and disaggregate measures of fit. At the aggregate level, we compare the mean predicted and actual (observed) number of household-level number of

physical activity episodes per weekend day, using the absolute percentage error (APE) for each of the subsamples. At the disaggregate level, we compute an “out-of-sample” log-likelihood function (OSLLF) approach. The OSLLF is computed by plugging in the sub-sample observations into the log-likelihood function, while retaining the estimated parameters from the overall estimation sample. As indicated by Norwood *et al.* (2001), the model with the highest value of OSLLF is the preferred one, since it is most likely to generate the set of sub-sample observations. The results are provided in Table 1 for segments formed based on three variables: (1) Family income (3 market segments), (2) Household bicycle ownership level (6 market segments), and (3) Family type (3 market segment). The third column provides the mean observed number of household-level physical activity episodes, while the next main column entitled “Aggregate-level fit statistics” provides the mean predicted number of household-level physical activity episodes (and the absolute percentage error or APE in parenthesis) from each of the ORL, RORL, and LC models. The mean predicted number of episodes from the LC model is closer to the true mean for nine of the 12 segments, as evidenced by the APE statistics. Finally, at the disaggregate level, the OSLLF value of the LC model is better than those of the other two models for nine of the 12 segments. All in all, the LC model outperforms the other two models in terms of data fit on the estimation sample as well as on sub-samples of the estimation sample.

Besides the data fit superiority of the LC model, our results also show that the LC model provides more efficient estimates. In particular, the average of the trace of the covariance matrix of parameter estimates is 0.00136 for the LC model, 0.00664 for the RORL model estimated coefficients, and 0.00377 for the ORL model, indicating the higher standard errors (by 175-390%) from the RORL and the ORL models relative to the preferred LC model.<sup>16</sup> That is, the recognition of family dependence leads to substantially improved econometric efficiency.

In the following presentation of the empirical results, we focus our attention on the results of the LC model that provides the best data fit.

---

<sup>16</sup> The covariance matrix of the RORL model will provide higher values just because the coefficients estimated from the RORL model are larger in magnitude compared to the ORL and LC models (because the random effects in the RORL model increases the total error variance to a value beyond 1, while the ORL and LC models normalize the error term variance to 1). However, we normalized the coefficients in the RORL model by taking the weighted mean (across family types based on the shares of each family type) of the error variance, and computed the trace value of the implied covariance matrix of the normalized RORL coefficients. This allows an apples-to-apples comparison of the trace values across the ORL, RORL, and LC models.

Table 1 Measures of Fit

Sample details	Number of households	Mean observed number of household-level physical activity episodes	Aggregate-level fit statistics			Disaggregate-level fit statistics		
			ORL	RORL	LC	“Out-of-sample” log-likelihood function (OSLLF)		
						ORL	RORL	LC
<b>Full sample</b>	517	0.7485	0.7498 (0.17%)	0.7529 (0.59%)	0.7473 (0.16%)	-916.748	-738.602	-732.844
<b>Family Income</b>								
Less than 60K	121	0.6529	0.6395 (2.05%)	0.6172 (5.47%)	0.6594 (1.00%)	-191.837	-140.805	-139.079
Between 60K and 90K	209	0.6794	0.6685 (1.60%)	0.6597 (2.90%)	0.6686 (1.59%)	-343.958	-279.977	-282.021
Greater than 90K	187	0.8877	0.9120 (2.74%)	0.9449 (6.44%)	0.8922 (0.51%)	-380.953	-317.820	-311.744
<b>Household bicycle ownership</b>								
0	49	0.4694	0.6503 (38.54%)	0.6671 (42.12%)	0.6003 (27.89%)	-61.812	-43.566	-45.863
1	55	0.6182	0.5805 (6.10%)	0.5850 (5.37%)	0.5852 (5.34%)	-86.636	-70.599	-69.022
2	89	0.5730	0.6437 (12.34%)	0.6589 (14.99%)	0.6343 (10.70%)	-132.726	-110.842	-108.799
3	108	0.7315	0.7040 (3.76%)	0.7089 (3.09%)	0.6898 (5.70%)	-186.284	-147.809	-147.435
4	136	0.8382	0.8497 (1.37%)	0.8447 (0.78%)	0.8348 (0.41%)	-272.895	-210.816	-206.721
5 or more	80	1.0750	0.9370 (12.84%)	0.9288 (13.60%)	1.0033 (6.67%)	-176.395	-154.970	-155.004
<b>Family type</b>								
Two parent	377	0.7613	0.7849 (3.10%)	0.7648 (0.46%)	0.7605 (0.11%)	-700.618	-557.709	-555.895
Single mother	85	0.5765	0.4999 (13.29%)	0.5714 (0.88%)	0.5518 (4.28%)	-115.405	-96.506	-94.161
Single father	55	0.9273	0.8953 (3.45%)	0.9516 (2.62%)	0.9593 (3.45%)	-100.725	-84.387	-82.787

### 4.3 Estimation Results

Table 2 presents the estimation results for the LC model. The coefficients provide the effects of variables on the latent propensity of an individual to participate in weekend out-of-home physically active episodes. For ease in presentation, we indicate the effects of independent variables separately on adults (*i.e.*, parents) and children, though the estimation is undertaken for all individuals together, while also accommodating unobserved dependencies in the physical activity propensities of individuals within a family.<sup>17</sup> The first main row of Table 2 provides estimates of the threshold values (for parents and children). These do not have any substantive interpretation; rather, they simply serve to translate the latent propensity into the observed ordered categories of the number of physical activity participations.

#### 4.3.1 Individual Factors

The effects of individual characteristics indicate the influence of the parents' age on both parents' and children's physical activity propensities. In particular, we find important interaction effects of sex and age in the physical activity propensity of adults. This is interesting, since many earlier studies examine the impact of sex and age as two separate variables or focus only on women (see, for example, Weuve *et al.*, 2004, and King *et al.*, 2005). However, our results suggest that there are important interaction effects between age and sex in adults' physical activity propensity.<sup>18</sup> In particular, our results indicate no statistically significant differences in weekend day physical activity propensity between male and female adults until the age of 35 years. On the other hand, most earlier studies indicate that male adults tend to be more physically active compared to female adults at almost any age (see, for example, Schulz and Schoeller, 1994, Azevedo *et al.*, 2007, and Troiano *et al.*, 2008). Further, according to our results, the propensity for weekend physical activity is lower for males who are 35 years of age or more relative to their younger counterparts (less than 35 years of age), while, for females in family

---

<sup>17</sup> In the rest of this paper, we will use the terms adults and parents interchangeably, based on the context of the discussion.

<sup>18</sup> Note that we tried various threshold age values to capture the age-related effects in our specification, but the thresholds of 35 years and 45 years provided the best fit. This dummy variable specification was better than a continuous age specification and a specification that considered non-linear spline effects. For male adults, there was literally no difference in the coefficients for the "35-45" years and "over 45 years" age categories. So, we have a single coefficient for these two categories for males. For females, there were larger differences in the two age categories. Thus, even though not statistically different at the 0.05 level of significance, we retained different coefficients on the two age categories for females.

**Table 2 Estimation Results for the Number of Out-of-Home Weekend Physically Active Activity Episodes**

Variable	Adults (Parents)		Children (aged 5-15)	
	Parameter	t-stat	Parameter	t-stat
<b>Threshold parameters</b>				
Threshold 1	3.084	4.68	2.702	4.02
Threshold 2	5.138	6.86	5.187	7.13
<b>Individual factors</b>				
Male adult (Father) between 35 -45 years	-1.297	-3.20	-1.586	-3.69
Male adult (Father) over 45 years	-1.297	-3.20	-1.586	-3.69
Female adult (Mother) between 35 -45 years	2.137	4.06	1.822	3.57
Female adult (Mother) over 45 years	1.848	3.95	1.704	3.87
Child's age	-	-	-0.044	-1.56
Adult's internet use	-0.295	-1.26	-	-
<b>Physical environment factors</b>				
<i>Season and activity day</i>				
Winter	-0.428	-1.31	-	-
Sunday	-0.580	-2.73	-0.635	-2.84
<i>Transportation system and built environment characteristics</i>				
Bicycling facility density (miles of bike lanes per square mile)	0.073	2.03	0.106	2.75
Fraction of multi family dwelling units	-	-	0.479	1.03
Presence of physically inactive recreation centers (such as theaters, amusement parks, inactive clubs (e.g. video games or cards))	-	-	-0.387	-1.39
<b>Social environment factors</b>				
<i>Family-level demographics</i>				
Two-parent families	0.422	1.60	-	-
Presence of children aged less than 5 years	1.565	2.57	-	-
Family income greater than 90k	0.283	1.27	0.484	2.13
Own household	-0.655	-2.31	-0.425	-1.55
Number of motorized vehicles	-0.227	-1.62	-	-
Number of bicycles	-	-	0.121	2.10
<i>Residential neighborhood demographics</i>				
Fraction of Caucasian American population	0.632	1.24	-	-
Fraction of African-American population	-	-	-2.783	-1.34

households, the propensity is higher for individuals who are 35 years or more relative to their younger counterparts (less than 35 years of age). Hawkins *et al.* (2009) find a similar result of increased physical activity among women in middle ages (40-59 years) relative to their younger peers, but this holds only for Hispanic women in their sample. As importantly, the implication of our results is that women who are 35 years of age or over have a higher propensity to participate in physically active episodes relative to their male counterparts. Of course, one should keep in mind that the measure of physical activity in our study (as in Dunton *et al.*, 2008 and Sener *et al.*, 2009) is the number of physical activity bouts on a weekend day as reported in a general activity survey, while several earlier studies have considered time expended in physical activity over longer stretches of time (such as a week or a longer period of time) using focused physical activity surveys or objective measurements of physical activity. Overall, there is a clear need for a joint analysis of different dimensions of physical activity, including types of physical activity bouts, time investments and number of bouts, where bouts occurred and time-of-day of bouts, weekend day versus weekday patterns, as well as with-whom bouts occurred. Understanding the role of demographics and other variables on each and all of these physical activity dimensions can provide important information for effective intervention strategies. While the field is moving toward such comprehensive analyses of physical activity (see, for example, Dunton *et al.*, 2008 and Sener *et al.*, 2008), the challenge is to obtain reliable data and develop methods to support the analysis of all these dimensions jointly. This is an important direction for future research in the physical activity area.

Parental age also has an important effect on children's physical activity propensity, though, once again, the effect is different for mothers and fathers. Children in families with young fathers (less than 35 years of age) have a higher physical activity propensity relative to children in families with older fathers, while children in families with young mothers have a lower physical activity propensity relative to children in families with older mothers. Taken together with the impact of parental age on parental physical activity, these results perhaps suggest that children explicitly model their parents' physical activity participation so that children in households with one or both physically active parents are more likely to be physically active. Overall, the results indicate that the highest levels of physical activity across all individuals in a family (parents and children) tend to be in two-parent families with young fathers (less than 35 years of age) and older mothers (35 years of age or more), while the lowest

levels of physical activity are in two-parent families with the father over 35 years of age and the mother less than 35 years of age. Previous studies (see, for example, Davison *et al.*, 2003) have suggested that mothers and fathers support and shape the physical activity participation of children in quite different ways, with fathers taking more of an explicit modeling role (a more hands-on physical activity-embracing role) and mothers taking more of a logistics support role (driving children to coaching camps and related physical activity opportunity locations). It would be interesting in future studies to examine if such differential support roles of parents in influencing children's physical activity participation are somehow being manifested in the parental age-based effects found in this study. In any case, the results suggest that policy interventions aimed at increasing children's physical activity levels could potentially benefit from targeting entire family units rather than targeting only children.

The effect of the child's age variable in Table 2 indicates that older children have a lower propensity to partake in physical activities. This is a result that is consistent with the findings of earlier studies (see, for example, Sallis *et al.*, 2000, and Sener *et al.*, 2008). While there may be several reasons for this result, one reason may be that, as children get older, they gravitate more toward unstructured social activities rather than structured sports activities and unstructured free play (Copperman and Bhat, 2007b). It is interesting to note here that we did not find any statistically significant effect of the child's age on parents' physical activity propensity.

Finally, within the category of individual characteristics, adults who use the internet during the weekend day are less likely to partake in physical activity compared to adults who do not use the internet.<sup>19</sup> This result may be a reflection of overall sedentary inclinations or lesser time availability for physically active pursuits in the day (due to getting "sucked up" in social conversations or internet browsing or e-mail checking). While only marginally significant, this result emphasizes the need to balance the positive aspects of internet connectivity with the potentially detrimental effect on physical activity lifestyles (see also Kennedy *et al.*, 2008).

In addition to the variables discussed above, we also examined the effects of work-related factors on physical activity propensity of family members. But we did not find any statistically significant impacts even at the 15% level.

---

<sup>19</sup> The "internet use" variable corresponds to the individuals' internet use over the sampled weekday for personal reasons such as for browsing (information seeking and shopping), entertainment/games, social e-mail, chat rooms, and banking/financial purposes.

### ***4.3.2 Physical Environment Factors***

In the group of physical environment factors, the first set of variables corresponds to season and activity day variables. The season variables suggest a lower propensity among adults to participate in weekend physical activities during the cold winter months relative to other times of the year (though this effect is not significant at the 0.05 significance level). Such seasonal variations have been found in other studies of adult physical activity participation (see Tucker and Gilliland, 2007, Sener and Bhat, 2007, and Pivarnik *et al.*, 2003). This may be attributed to the discomfort in participating in outdoor physically active pursuits during the winter season in the San Francisco Bay area, though this result is perhaps not transferable to areas with a rich set of winter sports activities such as skiing or skating. Interestingly, we did not find such similar season effects for children's physical activity participation. The activity day variable indicates lower physical activity propensity among both parents and children on Sundays compared to Saturdays, presumably because of the time investment in religious and social activities on Sundays. Further, as indicated in some other studies, Sundays serve the purpose of "rest" days at home before the transition to school or work the next day (see, for instance, Bhat and Gossen, 2004).

We tested several transportation system and built environment variables, though most of these did not turn out to be statistically significant even at the 15% level of significance.<sup>20</sup> However, as shown under "Transportation system and built environment characteristics" in Table 2, both adults and children in households residing in areas with high bicycle facility density (as measured by miles of bicycle lanes per square mile in the residential traffic analysis zone) are more likely to participate in physically active pursuits relative to individuals in other households. Of course, this result (and the rest of the effects in the transportation system/built environment variable category) should be viewed with some caution since we have not considered potential residential self selection effects. That is, it is possible that highly physically active families self-select themselves into zones with built environment measures that support their active lifestyles (see Bhat and Guo, 2007 and Bhat and Eluru, 2009 for methodologies to

---

<sup>20</sup> This may be a reflection of the use of a traffic analysis zone (TAZ) as a spatial unit of resolution for computing transportation system and built environment attributes, which is admittedly rather coarse. Future studies should consider more micro-scale measures to represent transportation system and built environment variable effects, but we are constrained to use the TAZ in this study because residence locations were tagged only to TAZs due to privacy considerations.



accommodate such self selection effects; combining such methodologies with the copula methodology proposed here for accounting for family clustering effects is left for future research). The “fraction of multifamily dwelling units” variable effect reveals a higher level of physical activity among children residing in zones with a high percent of multifamily dwelling units. This may be a reflection of more opportunities for joint physical activity participation with peers and other individuals in neighborhoods with a high share of multifamily units. Finally, the presence of physically inactive recreation centers in a zone reduces the physical activity propensity of children residing in that zone (though this effect is only marginally significant).

#### **4.3.3 Social Environment Factors**

The family demographics effects in Table 2 (within the category of social environment factors) show that adults in two-parent families have a higher propensity to participate in physically active episodes over the weekend day relative to families with only one parent, perhaps because of increased opportunities for joint participation in out-of-home adult physical activity participation or because one of the parents can look after children at home while the other participates in physical activity. The results also indicate the higher physical activity propensity of parents with young children (less than 5 years of age) relative to parents of older children (5 years or more). This may be related to the increased demands and reliance of older children on their parents for logistics and related support to participate in activities based on their own independent needs (see Stefan and Hunt, 2006, CDC, 2005, Eccles, 1999), leaving less time for parents to pursue physical activities. Both parents and children in high income families (with an annual income of more than \$90,000) have a higher propensity (than low income families) for physical activities, presumably due to fewer financial restrictions to travel to, and participate in, physical activities (see Parks *et al.*, 2003, and Day, 2006). On the other hand, the results in Table 2 indicate a lower weekend physical activity participation propensity among individuals (adults and children) residing in their own houses relative to individuals residing in non-owned houses. Finally, as the number of motorized vehicles in the family increases, adults (but not children) are less likely to engage in physical activity episodes, while, as the number of bicycles in the household increases, children (but not adults) are more likely to engage in physical activity episodes. Of course, a caution here is that this may be an associative effect rather than a causal

effect. That is, rather than fewer cars/more bicycles engendering more physical activity, it may be that households with physically active individuals choose to own fewer cars/more bicycles.

The neighborhood race composition effects under neighborhood residential demographics do show a general trend of higher (lower) physical activity propensity among adults (children) residing in neighborhoods with a high share of Caucasian-American households (African-American households) relative to adults (children) residing in other neighborhoods. As indicated by Rai and Finch (1997), physical activity in the population has generally been a “white” domain. Gordon-Larsen *et al.* (2005, 2006) also suggest that the lower physical activity propensity among children in predominantly African-American neighborhoods may be because of poor neighborhood quality and lack of good recreational centers.

#### ***4.3.4 Dependence Effects***

The estimated copula-based clustered ordered response model incorporates the jointness between physical activity episodes of family members not only through observed factors but also based on unobserved factors. As indicated earlier, the Clayton copula turned out to provide the best fit. The association parameter is parameterized in the Clayton copula as  $\theta_q = \exp(\delta's_q)$ , where the  $\delta$  vector is estimated. As indicated earlier, in our estimations, the  $s_q$  vector included three dummy variables: (1) family with both parents, (2) single mother family, and (3) single father family. The implied Clayton association parameter  $\theta_q$  for these three family types and their corresponding standard errors (computed using the familiar delta method; see Greene, 2003, page 70) are as follows: Family with both parents: 1.866 (0.155), single mother family: 2.158 (0.467), and single father family: 1.413 (0.478). All of these parameters are very highly statistically significant (relative to the value of ‘0’, which corresponds to independence), indicating the strong dependence among the unobserved physical activity determinants of family members.

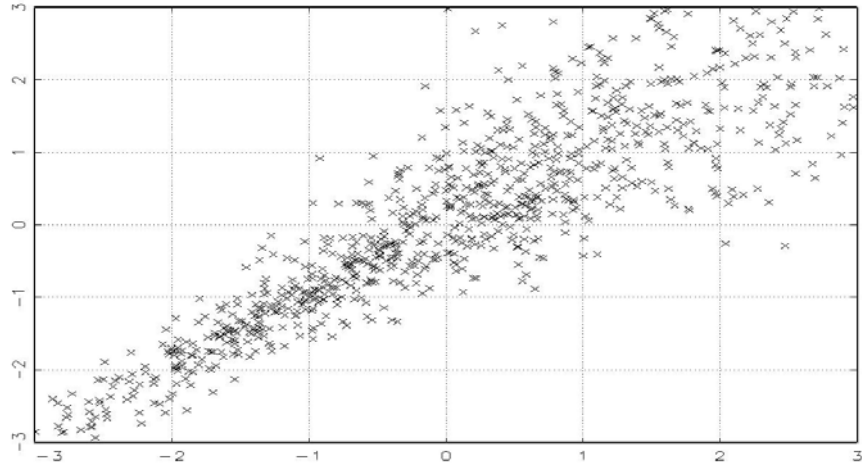
Another common way to quantify the dependence in the copula literature is to compute the Kendall's measure of dependence.<sup>21</sup> For the estimated association parameters, the values of the Kendall's  $\tau$  are (standard errors are in parenthesis): Family with both parents: 0.483 (0.021), single mother family: 0.519 (0.054), and single father family: 0.414 (0.082).

The dependence form of the Clayton copula implies that the dependency in unobserved components across family members in the propensity to participate in physically active episodes is strong at the left tail, but not at the right tail. Figure 1 plots the dependency scatterplot of the relationship between the unobserved components  $\varepsilon_{qi}$  of physical activity propensity for any two individuals in the same family  $q$ , based on family type.<sup>22</sup> As can be observed, the results indicate that individuals in a family tend to have uniformly low physical activity (tighter clustering of data points at the low end of the physical activity spectrum), but there is lesser clustering of individuals in a family toward the high physical activity propensity spectrum. In other words, the dependence among the physical activity propensities of family members is asymmetric, with a stronger tendency of family members to simultaneously have low physical activity levels than to simultaneously have high physical activity levels. Equivalently, it is easier for a family to lapse into a sedentary lifestyle because of the sedentary lifestyle of one of its members, while families do not come out of a sedentary lifestyle as easily just because of the active lifestyle of one of its members. From an education-based intervention standpoint to promote physical activity, the result that there is strong clustering within individuals in a family at the low physical activity spectrum end is encouraging. It suggests that a cost effective strategy would be to identify individuals who have a low physical activity level, then trace the individual back to her/his household, and target the entire family unit, all of whose members are likely to have low

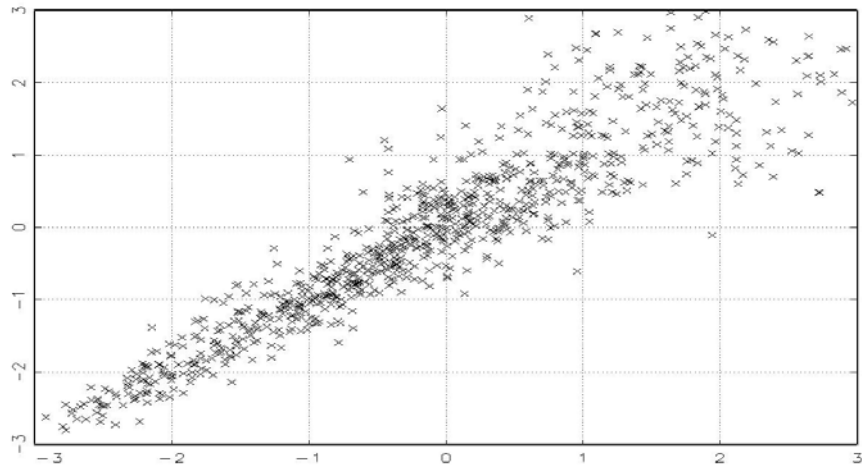
---

<sup>21</sup> See Bhat and Eluru (2009) for a description of this dependency measure. The traditional dependence concept of correlation coefficient  $\rho$  is not informative for asymmetric distributions, and has led statisticians to use concordance measures. 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 the Kendall's  $\tau$ , which is in the range between 0 and 1, assumes the value of zero under independence, and is not dependent on the margins. For the Clayton copula,  $\tau = \theta / (\theta + 2)$ .

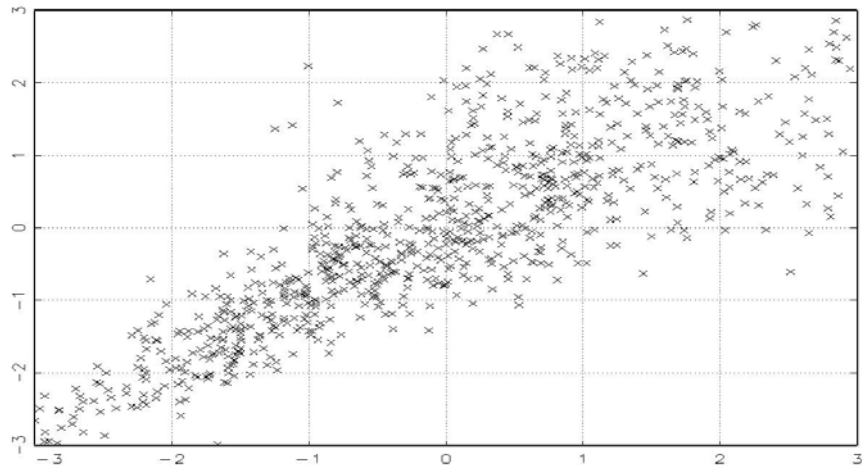
<sup>22</sup> For instance, Figure 1(a) represents the dependency scatterplot of the relationship between the unobserved components ( $\varepsilon_{qi}$ ) of physical activity propensity of two individuals (represented by each axis) residing in the same two-parent family. Note that the physical activity propensities  $y_{qi}^*$  are latent; thus, the scatterplots of  $\varepsilon_{qi}$  are based on the implied copula dependence shape that leads to the best model fit to the observed data. In our case, this is the Clayton copula, with the shapes being a function of the estimated Kendall's  $\tau$  value. The dependency relationships presented in Figure 1 will be the same for any two individuals within the same family, since the association parameter  $\theta_q$  varies across families, not between members of the same family.



(1a)



(1b)



(1c)

**Figure 1 Logistic-Clayton Copula Plots across Family Types**

(1a) Two-parents families ( $\tau = 0.483$ ); (1b) Single mother families ( $\tau = 0.519$ ); (1c) Single father families ( $\tau = 0.414$ )

physical activity levels. Such a strategy constitutes a good “capture” mechanism to bring educational campaigns to those who may benefit most from such campaigns.<sup>23</sup> More generally, the asymmetric “spillover” or “rubbing off” effect suggests that family-level information dissemination and targeting strategies to move away from sedentary lifestyles may be more effective than individual-level strategies to promote active lifestyles. The figures also show the higher (lower) dependency (especially at the lower end of the physical activity spectrum) for single mother (single father) families relative to two-parent families. This suggests a need to focus particularly on single mother households, and provide such families information regarding the potentially adverse effects of sedentary lifestyles.

To summarize, the discussion above illustrates that the dependency effects within a family (in the propensity to participate in physical activity) are asymmetric and statistically significant. A model that does not consider dependence between individuals in a family (*i.e.*, the simple ordered response model) and a model that accommodates only a restrictive normal dependency form are unable to consider flexible and asymmetric dependence patterns, while the copula-based approach is able to do so. These models also provide inconsistent estimates, as we discuss in the next section.

#### ***4.3.5 Aggregate Impacts of Variables***

The parameters on the exogenous variables in Table 2 do not directly provide the magnitude of the effects of the variables on the number of out-of-home weekend physical activity participations. To do so, we compute the aggregate-level “elasticity effects” of each variable. In particular, to compute the aggregate-level elasticity of a dummy exogenous variable (such as the “male adult (father) between 35-45 years” variable), we compute the expected aggregate share of individuals participating in each number of activity episodes in the “base case” and the corresponding share in the “scenario case” after increasing the number of male individuals between 35-45 years by 10% (with an appropriate decrease in the base category of male individuals younger than 35 years). We then compute an effective percentage change in the

---

<sup>23</sup> The statement here is not intended to be patronizing in any way to those who have low physically active levels. In fact, many individuals with low physically active levels may already know a substantial amount of statistics about the potential benefits of regular physical activity (to themselves and to society as a whole), and may be making informed choices. But, as in all promotional campaigns of services/products, one of the important tasks is to efficiently identify the population groups who are current “non-consumers” (*i.e.*, those who do not partake much in physical activity levels in the empirical context of the current paper) and attempt to “convert” them. The statement should be viewed in this light.

expected aggregate share of individuals participating in each number of activity episodes due to a change from the base case to the scenario case. On the other hand, to compute the aggregate level elasticity effect of an ordinal variable (such as number of motorized vehicles), we increase (or decrease) the value of the variable by 1 and compute a percentage change in the expected aggregate share of individuals participating in each number of activity episodes. Finally, the aggregate-level “arc” elasticity effect of a continuous exogenous variable (such as fraction of African-American population) is obtained by increasing the value of the corresponding variable by 10% for each individual in the sample, and computing a percentage change in the expected aggregate share of individuals participating in each number of activity episodes. While the aggregate level elasticity effects are not strictly comparable across the three different types of independent variables (dummy, ordinal, and continuous), they do provide order of magnitude effects.

The results are presented in Table 3 for the standard ordered-response logit (ORL) model, the random effects ordered-response model (RORL) and the LC models. To reduce clutter, we simplify the effects from the ordered models to a simple binary effect of variables on the share of adults (parents) and children participating in physical activity episodes. Also, to obtain standard deviations of the estimated magnitude effects, we undertake a bootstrap procedure using 26 draws of the coefficients (on the exogenous variables) based on their estimated sampling distributions. The mean magnitude effect across these 26 draws is in the column labeled “Mean” and the standard deviation of the magnitude effect is in the column labeled “Std. Dev.”. The numbers in the “mean” and “std. dev.” columns may be interpreted as the mean and standard deviation estimates, respectively, of the percentage change in the share of adults and children participating in one or more physically active recreational episodes during the weekend day. For instance, the first number “-11.94” with a standard deviation of “1.83” corresponding to the “male adult (father) between 35-45 years” variable in the ORL model indicates that the share of adults participating in active recreation decreases by about 12% (with a standard deviation of this effect being 1.83%) if the percentage of male adults between 35-45 years increases by 10% (with a corresponding decrease in the percentage of male adults below 35 years of age). On the other hand, the number “-13.51” with a standard deviation of “1.5” (under the “children” column for the ORL model) implies that the share of children participating in active recreation decreases by about 13.5% (with a standard deviation of 1.5%) if the percentage of male adults between 35-45

Table 3 Impact of Change in Individual, Physical, and Social Environment Factors

		% Change in Expected Aggregate Share of Individuals participating in physically Active Episodes											
		ORL				RORL				LC			
Variable	Formulation of the Change on the Variable	Adults		Children		Adults		Children		Adults		Children	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
<b>Individual factors</b>													
Male adult (Father) between 35 -45 years	Increased by 10%	-11.94	1.83	-13.51	1.50	-8.42	1.67	-9.14	1.32	-13.36*	2.05	-14.98*	2.04
Male adult (Father) over 45 years	Increased by 10%	-11.94	1.83	-13.51	1.50	-8.42	1.67	-9.14	1.32	-13.36*	2.05	-14.98*	2.04
Female adult (Mother) between 35 -45 years	Increased by 10%	19.80	2.14	14.74	1.62	14.62	2.03	11.18	1.33	21.55*	2.48	16.74*	2.01
Female adult (Mother) over 45 years	Increased by 10%	18.11	2.69	15.57	1.93	14.13	2.45	11.97	1.86	18.61	2.92	15.77 <sup>+</sup> *	2.23
Child's age	Increased by 1	-	-	-2.21	0.59	-	-	-1.04	0.37	-	-	-1.99	0.53
Adult's internet use	Increased by 10%	-0.28	0.38	-	-	-1.17	0.28	-	-	-1.25 <sup>+</sup>	0.46	-	-
<b>Physical environment factors</b>													
<i>Season and activity day</i>													
Winter	Increased by 10%	-2.82	0.70	-	-	-1.66	0.52	-	-	-1.77	0.62	-	-
Sunday	Increased by 10%	-2.65	0.50	-2.32	0.45	-1.64	0.42	-1.70	0.36	-3.40*	0.60	-3.12*	0.54
<i>Transportation system and built environment characteristics</i>													
Bicycling facility density (miles of bike lanes per square mile)	Increased by 10%	2.43	0.29	2.50	0.31	1.60	0.26	2.13	0.28	1.72 <sup>+</sup>	0.24	2.37	0.34
Fraction of multi family dwelling units	Increased by 10%	-	-	1.63	0.27	-	-	1.11	0.17	-	-	1.23	0.26
Presence of physically inactive recreation centers (such as theaters, amusement parks, inactive clubs (e.g. video games))	Increased by 10%	-	-	-4.40	0.69	-	-	-1.40	0.40	-	-	-1.64 <sup>+</sup>	0.60
<b>Social environment factors</b>													
<i>Family-level demographics</i>													
Two-parent families	Increased by 10%	4.39	0.43	-	-	3.50	0.41	-	-	3.46	0.48	-	-
Presence of children aged less than 5 years	Increased by 10%	13.92	3.01	-	-	14.63	3.47	-	-	16.71	3.15	-	-
Family income greater than 90k	Increased by 10%	2.24	0.44	4.08	0.47	2.14	0.48	3.05	1.40	2.72	0.55	3.75	0.54
Own household	Increased by 10%	-2.01	0.48	-2.99	0.55	-1.24	0.45	-0.34	0.38	-3.65 <sup>+</sup> *	0.70	-1.85*	0.60
Number of motorized vehicles	Decreased by 1	8.87	3.16	-	-	6.71	2.36	-	-	10.77	3.88	-	-
Number of bicycles	Increased by 1	-	-	14.85	1.19	-	-	9.42	1.23	-	-	9.84 <sup>+</sup>	1.12
<i>Residential neighborhood demographics</i>													
Fraction of Caucasian-American population	Increased by 10%	5.29	0.55	-	-	3.56	0.55	-	-	3.59 <sup>+</sup>	0.58	-	-
Fraction of African-American population	Increased by 10%	-	-	-1.02	0.18	-	-	-0.78	0.14	-	-	-0.53 <sup>+</sup>	0.19

<sup>+</sup>Coefficient is statistically different from the corresponding ORL coefficient at the 90% confidence level

\*Coefficient is statistically significantly different from the corresponding RORL coefficient at the 90% confidence level

years increases by 10%. Similarly, the number “-2.21” with a standard deviation of “0.59” corresponding to the “child’s age” variable in the ORL model reflects that an increase by 1 year for all children leads to about a 2.2% decrease (with a standard deviation of 0.59%) in the share of children participating in physically active recreation, while the number “2.43” (standard deviation of 0.29) for the effect of the “Bicycling facility density” implies that the share of adults participating in active recreation increases by 2.43% due to a 10% increase in the miles of bicycle lanes per square mile in each residence zone.

Several important observations may be made from Table 3. First, the physical environment variables (middle rows of the table) have a smaller (and inelastic) effect on physical activity participation relative to sociodemographic variables (the top and bottom rows of the table). This is consistent with other studies in the literature that indicate that, while the built environment may be engineered to increase physical activity, the ability to do so is rather limited (see, for instance, Copperman and Bhat, 2007a, Goodell and Williams, 2007, and TRB, 2005). Among the individual factors, the age of the father and mother have a substantial impact on the physical activity levels of all members of a family. In the group of family-level demographics, the presence of very young children and the number of motorized vehicles are important determinants of the physical activity levels of adults in a family, while the number of bicycles is an important determinant of the physical activity levels of children in a family. The important effects of vehicle ownership (for adults) and bicycle ownership (for children) catapults policies aimed at reducing motorized vehicle ownership and increasing bicycle ownership as potentially important ones to consider not only from the standpoint of reducing traffic congestion and greenhouse gas emissions, but also from the perspective of improving public health. However, the caveat mentioned earlier needs to be emphasized again; that is, this relationship of motorized vehicle ownership and bicycle ownership with physical activity may be an associative one rather than a causal one. Second, there is an impact of the fraction of Caucasian-American population in a zone on the physical activity levels of adults in that zone, though the reasons for this finding are not obvious. Is it that recreational opportunities and facilities (some of which are not captured in the built environment variables considered in this study) are better in zones with a high Caucasian-American population, as suggested by Gordon-Larsen *et al.* (2005, 2006), or are there other reasons for the differences? Additional qualitative investigation into this finding should provide valuable insights. Third, adding bicycle lanes and increasing bicycle facility density does



increase physical activity levels in both adults and children, even though the usual caveat has to be added that the directionality of this influence needs to be examined carefully. In particular, whether this influence is a causal effect of bicycle facility density on physical activity levels or simply a self-selection effect of highly physically active-oriented individuals locating themselves in areas with good bicycle facilities is an open question (see Bhat and Guo, 2007 and Pinjari *et al.*, 2008 for additional discussions of this issue). Finally, there are differences in the effects of variables between the ORL, RORL, and LC models. In the column corresponding to the LC model results, we identify those magnitude estimates from the LC model that are statistically different from the corresponding magnitude estimates from the ORL model (identified by a “+” next to the LC coefficient) and from the RORL model (identified by a “\*” next to the LC coefficient). A 90% level of confidence is used to determine statistically significant differences. The bootstrap-based standard deviation estimates of coefficient estimates are used in the computation. As one can notice, there are eight variable effects that are statistically different between the LC and ORL models, and nine variable effects that are statistically different between the LC and the RORL models. This, combined with the better data fit of the LC model, points to the inconsistent effects from the ORL and RORL models. Overall, the results underscore the importance of testing different copula structures for accommodating family dependencies to avoid the risks of inappropriate covariate influences and inconsistent predictions of the number of out-of-home weekend physically active activity episodes. Interestingly, our results suggest that it is possible that not accommodating clustering effects at all (that is, ignoring dependency) could be better from the standpoint of estimating consistent variable elasticity effects relative to accommodating clustering effects using an inappropriate dependency surface. This observation is based on the fewer mean estimates in Table 3 that are significantly different between the LC and ORL models compared to between the LC and RORL models.

## 5. CONCLUSION

This paper presents a copula-based model to examine the physical activity participation levels of individuals, while also explicitly accommodating dependencies due to observed and unobserved factors within individuals belonging to the same family unit. In the copula-based approach, the model structure allows the testing of various dependency forms, including non-linear and asymmetric dependencies among family members. For instance, family members may be likely

to have simultaneously low propensities for physical activity but not simultaneously high propensities, or high propensities together but not low propensities together. In the current paper, we focus on the Archimedean class of copulas, a class that is ideally suited to the clustering context where the level of dependence in the marginal random unobserved terms within a cluster is identical (*i.e.*, exchangeable) across any (and all) pairs of individuals in the cluster.

The measure of physical activity we adopt in the current study is the number of out-of-home physical activity bouts or episodes (regardless of whether these bouts correspond to recreation or to walking/biking for utilitarian purposes) on a weekend day as reported by respondents in the 2000 San Francisco Bay Area Survey. Accordingly, we use an ordered-response structure to analyze physical activity levels, while testing various multivariate copulas. The empirical results indicate that the Logistic-Clayton (LC) model specification provides the best data fit. That is, individuals in a family tend to have uniformly low physical activity, but there is lesser clustering of individuals in a family toward the high physical activity propensity spectrum. This result suggests that a cost effective “capture” mechanism to bring educational campaigns to those who may benefit most from such campaigns would be to identify individuals who have a low physical activity level, then trace the individual back to her/his household, and target the entire family unit, all of whose members are likely to have low physical activity levels.

A number of individual factors, physical environment factors, and social environment factors are considered in the empirical analysis. The results indicate that physical environment factors are not as important in determining physical activity levels as individual and social environment factors. Also, decreased vehicle ownership (for adults) and increased bicycle ownership (for children) are important positive determinants of weekend physical activity participation. These results should be carefully examined as they might be useful in developing policies aimed at not only reducing traffic congestion (and its consequent benefits), but also increasing physical activity levels. In addition, individual factors (demographics, work characteristics, internet use at home), physical environment variables (season and activity-day variables, as well as built environment measures), and social environment factors (family-level demographics and residential neighborhood demographics) are other important determinants of physical activity participation levels.

In closing, we have proposed a copula structure to accommodate clustering effects in ordinal response models, and applied the methodology to a study of physical activity participation levels of individuals as part of their families. A rich set of potential determinants of the number of out-of-home weekend day physical activity episodes is considered. However, we do not accommodate physical activity attitudes/beliefs and support systems of individual family members as they influence the physical activity levels of others in the family. This is because our data source does not collect such information. Future studies would benefit from including such family-level attitudinal/support variables, while also adopting a family-level perspective of physical activity as in the current study.

### **ACKNOWLEDGEMENTS**

This research was partially funded by a Southwest Region University Transportation Center grant. The authors acknowledge the helpful comments of four anonymous reviewers on an earlier version of the paper. The authors are grateful to Lisa Macias for her help in formatting this document.

## REFERENCES

- Allender, S., G. Cowburn, C. Foster (2006) Understanding Participation in Sport and Physical Activity among Children and Adults: A Review of Qualitative Studies. *Health Education Research*, 21(6), 826-835.
- Azevedo, M.R., C.L.P. Araujo, F.F. Reicher, F.V. Siqueria, M.C. da Silva, and P.C. Halla (2007) Gender Differences in Leisure-time Physical Activity. *International Journal of Public Health*. 52(1), 8-15.
- Ben-Akiva, M., and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.
- Bhat, C.R. (2000) A Multi-Level Cross-Classified Model for Discrete Response Variables. *Transportation Research Part B*, 34(7), 567-582.
- Bhat, C.R. (2009) A New Generalized Gumbel Copula for Multivariate Distributions. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, August 2009.
- 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. Gossen (2004) A Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice. *Transportation Research Part B*, 38(9), 767-787.
- 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 I.N. Sener (2009) A Copula-Based Closed-Form Binary Logit Choice Model for Accommodating Spatial Correlation Across Observational Units. *Journal of Geographical Systems*, 11(3), 243-272
- Bhat, C.R., and H. Zhao (2002) The Spatial Analysis of Activity Stop Generation. *Transportation Research Part B*, 36(6), 557-575.
- Bhat, C.R., I.N. Sener, and N. Eluru (2010) A Flexible Spatially Dependent Discrete Choice Model: Formulation and Application to Teenagers' Weekday Recreational Activity Participation. *Transportation Research Part B*, 44(8-9), 903-921.
- Bottai, M., N. Salvati, and N. Orsini (2006) Multilevel Models for Analyzing People's Daily Movement Behavior. *Journal of Geographical Systems*, 8(1), 97-108.
- Center for Disease Control (CDC) (2005) Positive parenting tips for healthy child development. Department of Health and Human Services, National Center on Birth Defects and Developmental Disabilities.
- Center for Disease Control (CDC) (2006) Youth risk behavior surveillance-United States, 2005. *Morbidity and Mortality Weekly Report* 55, No. SS-5, Department of Health and Human Services.
- Cervero, R. and M. Duncan (2003) Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area. *American Journal of Public Health*, 93(9), 1478-1483.

- Chamberlain, G. (1980) Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, 47(1), 225-238.
- 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.
- Cleland, V., A. Venn, J. Fryer, T. Dwyer, and L. Blizzard (2005) Parental Exercise is Associated with Australian Children's Extracurricular Sports Participation and Cardiorespiratory Fitness: A Cross-Sectional Study. *International Journal of Behavioral Nutrition and Physical Activity*, 2(3).
- Collins, B.S., A.L. Marshall, and Y. Miller (2007) Physical Activity in Women with Young Children: How can We Assess "Anything that's not Sitting"? *Women and Health*, 45(2), 95-116.
- Copperman, R.B., and C.R. Bhat (2007a) An Analysis of the Determinants of Children's Weekend Physical Activity Participation. *Transportation*, 34(1), 67-87.
- Copperman, R.B., and C.R. Bhat (2007b) An Exploratory Analysis of Children's Daily Time-Use and Activity Patterns Using the Child Development Supplement (CDS) to the US Panel Study of Income Dynamics (PSID). *Transportation Research Record*, 2021, 36-44.
- Czado, C., and S. Prokopenko (2008) Modeling Transport Mode Decisions Using Hierarchical Binary Spatial Regression Models with Cluster Effects. *Statistical Modeling*, 8(4), 315-345.
- Davison, K.K, T.M. Cutting, and L.L. Birch (2003) Parents' Activity-Related Parenting Practices Predict Girls' Physical Activity. *Medicine & Science in Sports & Exercise*, 35(9), 1589-95.
- Day, K. (2006) Active Living and Social Justice: Planning for Physical Activity in Low-income, Black, and Latino Communities. *Journal of American Planning Association*, 72(1), 88-99.
- Dill, J., and T. Carr (2003) Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters will Use Them – Another look. Paper presented at the 82nd Annual Meeting of the Transportation Research Board, Washington DC.
- Dunton G.F., D. Berrigan, R. Ballard-Barbash, B.I. Graubard and A.A. Atienza (2008) Social and Physical Environments of Sports and Exercise Reported among Adults in the American Time Use Survey. *Preventive Medicine*, 47(5), 519-524.
- Eccles, J.S. (1999) The Development of Children Ages 6 to 14. *The Future of Children*, 9(2), 30-44.
- Ferdous, N., N. Eluru, C.R. Bhat, and I. Meloni (2010) A Multivariate Ordered Response Model System for Adults' Weekday Activity Episode Generation by Activity Purpose and Social Context. *Transportation Research Part B*, 44(8-9), 903-921.
- Ferreira, I., K. Horst, W. Wendel-Vos, S. Kremers, F. van Lenthe, and J. Brug (2007) Environmental Correlates of Physical activity in Youth – A Review and Update. *Obesity Reviews*, 8(2), 129-154.

- Fotheringham, A.S. (1983) Some Theoretical Aspects of Destination Choice and their Relevance to Production-Constrained Gravity Models. *Environment and Planning A*, 15(8), 1121-1132.
- Frank, M.J. (1979) On the Simultaneous Associativity of  $F(x, y)$  and  $x + y - F(x, y)$ . *Aequationes Mathematicae*, 19(1), 194-226.
- Genius, M., and E. Strazzeria (2008) Applying the Copula Approach to Sample Selection Modeling. *Applied Economics*, 40(11), 1443-1455.
- Giles-Corti, B., and R.J. Donovan (2002) The Relative Influence of Individual, Social and Physical Environment Determinants of Physical Activity. *Social Science and Medicine*, 54(12), 1793-1812.
- Goodell, S., and C.H. Williams (2007). The Built Environment and Physical Activity: What is the Relationship? Policy Brief No. 11, The Synthesis Project, Robert Wood Johnson Foundation.
- Gordon-Larsen, P., R.G. McMurray, and B.M. Popkin (2005) Determinants of Adolescent Physical Activity and Inactivity Patterns. *Pediatrics*, 105(6), E83.
- Gordon-Larsen, P., M. Nelson, P. Page, and B.M. Popkin (2006) Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity. *Pediatrics*, 117(2), 417-424.
- Greene, W.H. (2003). *Econometric Analysis*. 5<sup>th</sup> edition, Prentice Hall, Macmillan, New York.
- Gumbel, E.J. (1960) Bivariate Exponential Distributions. *Journal of the American Statistical Association*, 55(292), 698-707.
- Gustafson, S., and R. Rhodes (2006) Parental Correlates of Physical Activity in Children and Early Adolescents. *Sports Medicine*, 36(1), 79-97.
- Haskell, W.L., I.M. Lee, R.R. Pate, K.E. Powell, S.N. Blair, B.A. Franklin, C.A. Macera, G.W. Heath, P.D. Thompson, and A. Bauman (2007) Physical Activity and Public Health: Updated Recommendations for Adults from the ACSM and the AHA. *Circulation*, 116(9), 1081-1093.
- Hawkins, M.S, K.L. Storti, C.R. Richardson, W.C. King, S.J. Strath, R.G. Holleman, and A.M. Kriska (2009) Objectively Measures Physical Activity of U.S. Adults by Sex, Age, and Racial/Ethnic Groups: Cross-Sectional Study. *International Journal of Behavioral Nutrition and Physical Activity*, 6(31).
- Herriges, J.A., D.J. Phaneuf, and J.L. Tobias (2008) Estimating Demand Systems when Outcomes are Correlated Counts. *Journal of Econometrics*, 147(2), 282-298.
- Hoehner C.M., L.K.B. Ramirez, M.B. Elliot, S. Handy, and R. Brownson (2005) Perceived and Objective Environmental Measures and Physical Activity among Urban Adults. *American Journal of Preventive Medicine*, 28(2S2), 105-116.
- Hsiao, C. (1986) *Analysis of Panel Data*. Cambridge University Press, Cambridge.
- 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.
- Kelly, L.A., J.J. Reilly, A. Fisher, C. Montgomery, A. Williamson, J.H. McColl, J.Y. Paton, and S. Grant (2006) Effect of Socioeconomic Status on Objectively Measured Physical Activity. *Archives of Disease in Childhood*, 91(1), 35-38.
- Kennedy, T.L.M., A. Smith, A.T. Wells, and B. Wellman (2008) Networked Families. Pew Internet & American Life Project, <http://www.pewinternet.org/Reports/2008/Networked-Families.aspx>
- King, K., S. Belle, J. Brach, L. Simkin-Silverman, T. Soska, and A. Kriska (2005) Objective Measures of Neighborhood Environment and Physical Activity in Older Women. *American Journal of Preventive Medicine*, 28(5) 461-469.
- Lockwood, A., S. Srinivasan, and C.R. Bhat (2005) An Exploratory Analysis of Weekend Activity Patterns in the San Francisco Bay Area. *Transportation Research Record*, 1926, 70-78.
- McKelvey, R.D., and W. Zavoina (1975) A Statistical Model for the Analysis of Ordinal-Level Dependent Variables. *Journal of Mathematical Sociology*, 4(Summer), 103-120.
- MORPACE International, Inc., 2002. Bay Area Travel Survey Final Report, March.
- Nelsen, R.B. (2006) *An Introduction to Copulas* (2nd ed). Springer-Verlag, New York.
- Nelson, M.C., and P. Gordon-Larsen (2006) Physical Activity and Sedentary Behavior Patterns are Associated with Selected Adolescent Health Risk Behaviors. *Pediatrics*, 117(4), 1281-1290.
- Norwood, B., P. Ferrier, and J. Lusk (2001) Model Selection Criteria Using Likelihood Functions and Out-of-Sample Performance. Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, Missouri, April 23-24.
- Ornelas, I.J., K.M. Perreira, and G.X. Ayala (2007) Parental Influences on Adolescent Activity: A Longitudinal Study. *The International Journal of Behavioral Nutrition and Physical Activity*, 4(3).
- Parks, S.E., R.A. Housemann, and R.C. Brownson (2003) Differential Correlates of Physical Activity in Urban and Rural Adults of Various Socioeconomic Backgrounds in the United States. *Journal of Epidemiology and Community Health*, 57(1), 29-35.
- Pinjari, A.R., N. Eluru, C.R. Bhat, R.M. Pendyala, and E. Spissu (2008) Joint Model of Choice of Residential Neighborhood and Bicycle Ownership: Accounting for Self-Selection and Unobserved Heterogeneity. *Transportation Research Record*, 2082, 17-26
- Pivarnik, J.M., M. Reeves, and A.P. Rafferty (2003) Seasonal Variation in Adult Leisure-Time Physical Activity. *Medicine & Science in Sports & Exercise*, 35(6), 1004-1008
- 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

- Rai, D., and H. Finch (1997) Physical Activity 'From Our Point of View'. Health Education Authority, London.
- Sallis, J.F., and N. Owen (2002). Ecological Models of Health Behavior. In K. Glanz, B.K. Rimer, and F.M. Lewis (eds.) *Health Behavior and Health Education: Theory, Research, and Practice*, third ed., 462-484, Jossey-Bass, A Wiley Imprint, San Francisco, CA.
- Sallis, J.F., J.J. Prochaska, and W.C. Taylor (2000) A Review of Correlates of Physical Activity of Children and Adolescents. *Medicine & Science in Sports & Exercise*, 32(5), 963-975
- Salmon, J., M.L. Booth, P. Phongsavan, N. Murphy, and A. Timperio (2007) Promoting Physical Activity Participation among Children and Adolescents. *Epidemiologic Reviews*, 29(1), 144-159
- Schmidt, T. (2007) Coping with Copulas. In J. Rank (ed.) *Copulas - From Theory to Application in Finance*, 3-34, Risk Books, London.
- Schulz, L.O., and D.A. Schoeller (1994) A Compilation of Total Energy Expenditures and Body Weights in Healthy Adults. *American Journal of Clinical Nutrition*, 60, 676-68.
- Sener, I.N., and C.R. Bhat (2007) An Analysis of the Social Context of Children's Weekend Discretionary Activity Participation. *Transportation*, 34(6), 697-721.
- Sener, I.N., R.B. Copperman, R.M. Pendyala, and C.R. Bhat (2008) An Analysis of Children's Leisure Activity Engagement: Examining the Day of Week, Location, Physical Activity Level, and Fixity Dimensions. *Transportation*, 35(5), 673-696.
- Sener, I.N., N. Eluru, and C.R. Bhat (2009) Who are Bicyclists? Why and How Much are they Bicycling? *Transportation Research Record*, 2134, 63-72.
- 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(6), 449-460.
- Spissu, E., N. Eluru, I.N. Sener, C.R. Bhat, and I. Meloni (2010) A Cross-Clustered Model of Home-Based Work Participation Frequency During Traditionally Off-Work Hours. *Transportation Research Record*, forthcoming.
- Springer, A.E., S.H. Kelder, D.M. Hoelscher (2006) Social Support, Physical Activity and Sedentary Behavior Among 6th-grade Girls: A Cross-Sectional Study. *International Journal of Behavioral Nutrition and Physical Activity*, 3(8).
- Srinivasan, S., and C.R. Bhat (2008) An Exploratory Analysis of Joint-Activity Participation Characteristics Using the American Time Use Survey. *Transportation*, 35(3), 301-328.
- Steinbeck, K.S. (2008) The Importance of Physical Activity in the Prevention of Overweight and Obesity in Childhood: A Review and an Opinion. *Obesity Reviews*, 2(2), 117-130.
- Stefan, K.J., and J.D. Hunt (2006) Age-based Analysis of Children in Calgary, Canada. Presented at the 85<sup>th</sup> Annual Meeting of the Transportation Research Board, Washington, D.C., January.



- Strauss, R., D. Rodzilsky, G. Burack, and M. Colin (2001) Psychosocial Correlates of Physical Activity in Healthy Children. *Archives of Pediatrics & Adolescent Medicine*, 155(8), 897-902.
- Transportation Research Board (TRB) (2005) Does the Built Environment Influence Physical Activity? Examining the Evidence. TRB Special Report 282, The National Academies.
- Trivedi, P.K., and D.M. Zimmer (2007) Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, 1(1), Now Publishers.
- Troiano, R.P., D. Berrigan, K.W. Dodd, L.C. Masse, T. Tilert, and M. McDowell (2008) Physical Activity in the United States Measures by Accelerometer. *Medicine & Science in Sports & Exercise*, 40(1), 181-188.
- Trost, S.G., J.F. Sallis, R.R. Pate, P.S. Freedson, W.C. Taylor, and M. Dowda (2003) Evaluating a Model of Parental Influence on Youth Physical Activity. *American Journal of Preventive Medicine*, 25(4), 277-282
- Tucker, P., and J. Gilliland (2007) The Effect of Season and Weather on Physical Activity: A Systematic Review. *Public Health*, 121(12), 909-922.
- U.S. Department of Health and Human Services (USDHHS) (2008) 2008 Physical Activity Guidelines for Americans. Available at: <http://www.health.gov/paguidelines/pdf/paguide.pdf>
- U.S. Government Accountability Office (GAO) (2006) Childhood Obesity: Factors Affecting Physical Activity. Report GAO-07-260R, Childhood Obesity and Physical Activity, Congressional Briefing. Available at: <http://www.gao.gov/new.items/d07260r.pdf>.
- Wendel-Vos, W., M. Droomers, S. Kremers, J. Brug, and F. van Lenthe (2005) Potential Environmental Determinants of Physical Activity in Adults. In *Environmental Determinants and Interventions for Physical Activity, Nutrition and Smoking: A Review*. Edited by: Brug J, van Lenthe F. Erasmus University Medical Centre, Rotterdam.
- Weuve, J., J.H. Kang, J.E. Manson, M.M.B. Breteler, J.H. Ware, and F. Grodstein (2004) Physical Activity Including Walking, and Cognitive Function in Older Women. *JAMA-The Journal of the American Medical Association*, 292(12), 1454-1461.