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

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

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

*corresponding author

July 2009

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) 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 the most important positive determinants of weekend physical activity participation levels are reduced vehicle ownership and increased bicycle ownership. Thus, our results suggest that policies aimed at reducing motorized vehicle ownership and increasing bicycle ownership can serve the dual purpose of reducing traffic congestion (with its consequent benefits) as well as increasing physical activity levels.

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). Specifically, obesity has been established as an important risk factor for health problems such as diabetes, hypertension, coronary heart and related cardiovascular diseases, strokes, some forms of cancer, depression, sleep apnea and anxiety (Swallen *et al.*, 2005 and WHO, 2006). At the same time, there has been a dramatic increase in the prevalence of obesity among U.S. adults, adolescents, and children over the past two decades (Center for Disease Control (CDC), 2009). The 2003-2006 National Health and Nutrition Examination Survey (NHANES) data indicate that about two thirds of adults in the U.S. may be classified as being overweight and almost one-third as being obese (see Ogden *et al.*, 2007; 2008). The same data indicate that 15.6% of children and adolescents (ages 2 through 19 years) may be characterized as being overweight, and 16.3% as being obese.

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 and decreasing heart disease, diabetes, high blood pressure, and several forms of cancer (USDHHS, 2008). Further, physical activity also enhances agility and strength, reduces the need for medical attention, contributes to improved mental health, and decreases depression and anxiety in all individuals (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 one-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.[1]

[1] 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).

The low level of physical activity participation in the U.S. population has prompted several research studies in the past decade that 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/adolescent's 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 family life-style and health consciousness, and residential location-related 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 mis-informed 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. For example, in the case of a parent dropping off a child at soccer practice, it is not the parent's activity but the child's activity, and its location, that determine the temporal and spatial dimensions of the trip (see Sener and Bhat, 2007). Of course, in addition to serve-passenger activities, 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. In this regard, there is increasing recognition that children as young as 6-8 years start developing their own identities and individualities, and social and activity participation needs. They then "work" with their parents to fulfill those needs (see Stefan and Hunt, 2006; CDC, 2005; Eccles, 1999). 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 studies. First, almost all of these 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).[2] 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.

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, 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

[2] 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.

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 purposes 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 and Bhat, 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) individuals 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. Such surveys typically collect information on all types of (out-of-home) episodes of all individuals in sampled households over the course of 1-2 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 allows 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. Further, for our family-level modeling of physical activity, survey data provide information on physical activity participation for each (and all) members of a family.[3] 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. 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 discrete number of physical activity episodes. In the past, several studies have considered such clustered ordered response

[3] 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.

systems by employing a normal mixing distribution to introduce dependencies among the error terms of the underlying propensities, but such an assumption assumes linear and symmetric error dependency patterns. In the current paper, we generalize the structure of the dependency by allowing 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.

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, with the assumption that there is an underlying continuous latent variable whose horizontal partitioning 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 Herriges *et al.*, 2008 and Ferdous *et al.*, 2009 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 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.

A well-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 an example of this method). Under the usual assumption that this cluster-based error term is independent of the remaining error term and the explanatory variables, the analyst can first write the joint probability for each cluster conditional on the cluster error term as the product of the individual probabilities in the cluster, and then obtain the unconditional joint probability for each cluster by integrating out the normally distributed cluster error term. Such a mixing approach or an unobserved individual heterogeneity approach is relatively straightforward, even if it involves integration. However, from a conceptual standpoint, the mixing of error terms generates a joint distribution for each cluster whose form, in general, is anything but straightforward (see O'Brien and Dunson, 2004). Equivalently, the marginal distribution of the individual error terms, and therefore the marginal probabilities in the joint model, do not have an obvious form (except in the case of a random-effects ordered-response probit model). In addition, there is a 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. Such an approach is ideally suited to generate a joint distribution of a cluster outcome 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. 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.

2.2 Copula Basics

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

Bhat and Eluru, 2009; Trivedi and Zimmer, 2007). The precise definition of a copula is that it is a multivariate distribution function defined over the unit cube linking uniformly distributed marginals. Let C be an I-dimensional copula of uniformly distributed random variables U_1, U_2, U_3, …, U_I with support contained in $[0,1]^I$. Then,

$$C_\theta(u_1, u_2, \ldots, u_I) = \Pr(U_1 < u_1, U_2 < u_2, \ldots, U_I < u_I), \tag{1}$$

where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector. A copula, once developed, allows the generation of joint multivariate distribution functions with given marginals. Consider I random variables $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_I$, each with univariate continuous marginal distribution function $F(z_i) = \Pr(\varepsilon_i < z_i)$.[4] 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:

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

The above equation offers a vehicle to develop different dependency patterns for the random variables $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_I$ 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.

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, including comprehensive and non-comprehensive structures, radially symmetric and asymmetric shapes, and asymptotically tail independent and dependent forms (see Nelsen, 2006 and Bhat and Eluru, 2009 for a detailed discussion). The class is very flexible, and easy to construct.

---

[4] 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.

Archimedean copulas are constructed based on an underlying continuous convex decreasing generator function φ from [0, 1] to [0, ∞] 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 φ^{-1} exists. Also, let φ^{-1}be completely monotonic on [0, ∞]. With these preliminaries, we can generate multivariate I-dimensional Archimedean copulas as:

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

where the dependence parameter θ is embedded within the generator function. A whole variety of Archimedean copulas have been identified based on different forms of the generator function. 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 allow only positive associations and equal dependencies among pairs of random variables in their multivariate forms, which is well-suited for cluster analysis where 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):

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

Independence corresponds to $\theta\to 0$. 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:

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

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:

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

Independence is attained in Frank's copula as $\theta \to 0$. 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:

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

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, …, Q), and let i be the index for individuals (i = 1, 2, …, 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, …, 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 Zavonia, 1975):

$$y_{qi}^{*} = \beta' x_{qi} + \varepsilon_{qi}\,,\; y_{qi} = k \text{ if } \psi_k < y_{qi}^{*} \le \psi_{k+1}\,, \tag{8}$$

where x_{qi} is a $(L \times 1)$ vector of exogenous variables for individual i in family q (not including a constant), β is a corresponding $(L \times 1)$ vector of coefficients to be estimated, and ψ_k is the lower bound threshold for count level k ($\psi_0 < \psi_1 < \psi_2 ... < \psi_K < \psi_{K+1}$; $\psi_0 = -\infty$, $\psi_{K+1} = +\infty$).[5] The ε_{qi} terms capture the idiosyncratic effect of all omitted variables for individual i in family q, and are assumed to be independent of β and x_{qi}. The ε_{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 ε_{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:

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

It is important to note above that the level of dependence among individuals of a family can vary across families, as reflected by the θ_q notation for the dependence parameter. As we indicate later, we parameterize this dependence parameter as a function of observed family characteristics in estimation. 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

[5] 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 ψ_{ki}. However, for notational ease, we suppress the subscript i when writing the thresholds.

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 $(m_{q1}, m_{q2}, m_{q3}, ..., m_{qI_q})$ can be written as:

$$P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qI_q} = m_{qI_q}) = \int_{M_q} c_{\theta_q} \left(F(y^*_{q1}), F(y^*_{q2}), ..., F(y^*_{qI_q}) \right) dy^*_{q1} dy^*_{q2} ... dy^*_{qI_q}, \quad (10)$$

where $M_q = \{y^*_{q1}, y^*_{q2}, ..., y^*_{qI_q} : \psi_{(m_{qi})} < y^*_{qi} < \psi_{(m_{qi}+1)}$ for all $i = 1,2,..., I_q\}$ and c_{θ_q} is the copula density. The integration domain M_q is simply the multivariate region of the y^*_{qi} variables $(i = 1,2,..., I_q)$ determined by the observed vector of choices $(m_{q1}, m_{q2}, ..., m_{qI_q})$. 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 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:

$$P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qI_q} = m_{qI_q}) = P(\psi_{m_{q1}} < y^*_{q1} < \psi_{m_{q1+1}}, \psi_{m2} < y^*_{q2} < \psi_{m_{q2+1}}, \ldots \psi_{m_{qI_q}} < y^*_{qI_q} < \psi_{m_{qI_q}+1})$$

$$= \sum_{a_1=1}^{2} \sum_{a_2=1}^{2} \cdots \sum_{a_{I_q}=1}^{2} (-1)^{a_1+a_2+\ldots a_{I_q}} \left[P(y^*_{q1} < \psi_{m_{q1}+a_1-1}, y^*_{q2} < \psi_{m_{q2}+a_2-1}, \ldots y^*_{qI_q} < \psi_{m_{qI_q}+a_{I_q}-1}) \right]$$

$$= \sum_{a_1=1}^{2} \sum_{a_2=1}^{2} \cdots \sum_{a_{I_q}=1}^{2} (-1)^{a_1+a_2+\ldots a_{I_q}} \left[C_{\theta_q}(u_{m_{q1}+a_1-1}, u_{m_{q2}+a_2-1}, \ldots u_{m_{qI_q}+a_{I_q}-1}) \right] \quad (11)$$

where C_{θ_q} is the one of the four Archimedean copulas discussed in Section 2.3 with an association parameter θ_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 as in Bhat *et al.* (2009).

The association parameter θ_q is allowed to vary across families. However, it is not possible to estimate a separate dependence term for each family. So, we parameterize θ_q as a function of a vector s_q of observed family variables, while also choosing a functional form that ensures that θ_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 ψ is the vector of threshold bounds: $\psi = (\psi_1, \psi_2, \ldots \psi_K)$. The likelihood function for household q may be constructed based on the probability expression in Equation (11) as:

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

The likelihood function is then given by $L(\Omega) = \prod_q L_q(\Omega)$. (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 weekend day, though activity/travel data on only one weekend day was collected (*i.e.*, a subset of households were surveyed on Friday and Saturday or Sunday and Monday, but not Saturday and Sunday). 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.[6] 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. 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

[6] 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.

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.

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.[7] 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:

---

[7] 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.

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, 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 episodes (3.0%), while the corresponding distribution within the sample of children is zero episodes (79.2%), one episode (18.4%), and two 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); 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 examined. 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 ε_{qi}, and four different copula structures (Clayton, Gumbel, Frank and Joe) for specifying the dependency between the ε_{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 parametrizing the θ_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 ε_{qi} terms), and the best random-effects model (again from the logistic and normal distributions for the ε_{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 Strazzera, 2008; 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 -732.573. However, 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 -916.894. 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 -737.624. 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.

The likelihood ratio test for testing between the LC model in this paper and the ORL model is 368.640, which is substantially larger than the critical χ^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 358.540, which again is larger than the critical χ^2 value with 3 degrees of freedom. The LC and RORL models are non-nested, but they can be tested using a non-nested likelihood ratio test. Specifically, the difference in the adjusted rho-bar squared ($\overline{\rho}_c^2$) values between the two models is 0.00167.[8] The probability that this difference could have occurred by chance is less than $\Phi\left\{-[-2\times 0.00167\times L(C)+(32-32)]^{0.5}\right\}$. 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 the following presentation of the empirical results, we focus our attention on the results of the LC model that provides the best data fit.

4.3 Estimation Results

Table 1 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.[9] The first main row of Table 1 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.

---

[8] The adjusted rho-bar squared value $\overline{\rho}_c^2$ for an ordered-response model is computed as $\overline{\rho}_c^{\,2}=1-[(L(\hat{\beta})-H)/L(C)]$, 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.

[9] In the rest of this paper, we will use the terms adults and parents interchangeably, based on the context of the discussion.

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.[10] 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 Trolano, 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 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.*, 2008) 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 an 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,

---

[10] 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.

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 to support the analysis of all these dimensions jointly.

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 then 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 1 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).

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.[11] 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.

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. 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).

[11] The "internet use" variable corresponds to the individuals' internet use over the sampled weekend day for personal reasons such as for browsing (information seeking and shopping), entertainment/games, social e-mail, chat rooms, and banking/financial purposes.

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.[12] However, as shown under "Transportation system and built environment characteristics" in Table 1, 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 accommodate such self-selection effects). The "fraction of multi-family dwelling units" variable effect reveals a higher level of physical activity among children residing in zones with a high percent of multi-family 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 multi-family 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 1 (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

---

[12] 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.

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 1 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 δ 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 θ_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.[13] For the estimated association parameters, the values of the Kendall's τ 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 ε_{qi} of physical activity propensity for any two individuals in the same family q, based on family type. 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. The figures also show the higher (lower) dependency at the lower end of the physically activity spectrum for single mother (single father) families relative to two-parent families.

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 physical activity levels. Such

[13] See Bhat and Eluru (2009) for a description of this dependency measure. The traditional dependence concept of correlation coefficient ρ 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 τ, 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, $\tau = \theta / (\theta + 2)$.

a strategy constitutes a good "capture" mechanism to bring educational campaigns to those who may benefit most from such campaigns.[14]

To summarize, the discussion above clearly 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 provide relatively poorer data fits. These models also provide biased estimates that are quite different from those obtained from the Logistic-Clayton (LC) model, as we discuss in the next section.

4.3.5 Aggregate Impacts of Variables

In this section, we examine the magnitude of the influence of variables on the number of out-of-home weekend physical activity episode participations. The results are presented 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. Further, we also confine our attention to selected variables to focus attention. These variables (and the change in the variables that are examined) are as follows: (1) increase in adult internet usage (10% increase) , (2) increase in bicycling facility density in the zone of each household (10% increase), (3) increase in the fraction of multi-family dwelling units in the zone of residence of each household (10% increase), (4) decrease in vehicle ownership for each household (decrease by one vehicle, except for zero vehicle households whose car ownership level is left unchanged), (5) increase in bicycle ownership for each household (increase by one bicycle), (6) increase in the fraction of Caucasian-American population in the zone of residence of each household (increase by 10%), and (7) increase in the fraction of African-American population in the zone of each household in the sample (increase by 10%). To examine the impact of these changes, we compute an effective percentage change (between the base case and

[14] 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.

the case with the change in each variable) in the expected aggregate share of adults and children participating in weekend physical activity episodes.

Table 2 presents the results. Several important observations may be made from the table. First, it is clear from the table that the most important determinants of weekend physical activity participation are vehicle ownership (for adults) and bicycle ownership (for children). This is interesting, since it catapults policies aimed at reducing motorized vehicle ownership and increasing bicycle ownership as 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. Obviously, the continued collaboration of transportation and public health professionals is important in designing and obtaining traction for implementing such policies. Second, there is a clear 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. 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. For instance, the ORL and RORL models underestimate the influence of an increase in adults' internet use at home on adult physical activity levels, as well as underestimate the influence of reduced household vehicle ownership levels on adult physical activity levels. Thus, use of the

ORL and RORL models would lead (inappropriately) to the reduced consideration of educational policies targeting the public regarding the potential detrimental effects of internet use at home (from a physical activity standpoint), and reduced emphasis on economic and other policies aimed at reducing vehicle ownership levels.

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 the most important determinants of weekend physical activity participation are vehicle ownership (for adults) and bicycle ownership (for children). Thus, our results suggest that policies aimed at reducing motorized vehicle ownership and increasing bicycle ownership can serve the dual purpose of reducing traffic congestion (and its consequent benefits) as well as 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 are grateful to Lisa Macias for her help in formatting this document.

REFERENCES

Allender, S., G. Cowburn, C. Foster C (2006) Understanding Participation in Sport and Physical Activity among Children and Adults: A Review of Qualitative Studies. *Health Educ Res***,** 21, 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.

Bhat, C.R. (2000) A Multi-Level Cross-Classified Model for Discrete Response Variables. *Transportation Research Part B*, 34, 567-582.

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 Gossen, R. (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 (2009) A Flexible Spatially Dependent Discrete Choice Model: Formulation and Application to Teenagers' Weekday Recreational Activity Participation. Technical Paper, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin.

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. *Child Development*.

Center for Disease Control (CDC) (2006) Youth Risk Behavior Surveillance – United States, 2005. Morbidity and Mortality Weekly Report, 55(SS-5).

Center for Disease Control (CDC) (2009) U.S. Obesity Trends 1985–2007. Available at: http://www.cdc.gov/obesity/data/trends.html.

Cervero, R. and Duncan, M (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, 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, L. Blizzard (2005) Parental Exercise is Associated with Australian Children's Extracurricular Sports Participation and Cardiorespiratory Fitness: A Crosssectional Study. *Int J Behav Nutr Phys Act,* 2, 3

Collins, B. and 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., 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, L.L. Birch LL. (2003) Parents' Activity-Related Parenting Practices Predict Girls' Physical Activity. Med Sci Sport Exerc, 35(9), 1589-95.

Day K. (2006) Active Living and Social Justice: Planning for Physical Activity in Low-income, Black, and Latino Communities. *J Am Plann Assoc,* 72, 88-99.

Dill, J. and Carr, T. (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 (2009) A Multivariate Ordered Response Model System for Adults' Weekday Activity Episode Generation by Activity Purpose and Social Context. Technical paper, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin, July.

Ferreira, I, K. Horst, W. Wendel-Vos, S. Kremers, F. van Lenthe, J. Brug (2007) Environmental Correlates of Physical activity in Youth – A Review and Update. *Obes Rev*, 8, 129-154.

Fotheringham, A.S. (1983) Some Theoretical Aspects of Destination Choice and their Relevance to Production-Constrained Gravity Models. *Environment and Planning*, 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., Strazzera, E., 2008. Applying the Copula Approach to Sample Selection Modeling. *Applied Economics* 40(11), 1443-1455.

Giles-Corti, B. and R.J. Donovan (2002) Socioeconomic Status Differences in Recreational Physical Activity levels and Real and Perceived Access to a Supportive Physical Environment. *Preventive Medicine*, 35, 601-611.

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*. 5th 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, Rhodes R (2006) Parental Correlates of Physical Activity in Children and Early Adolescents. *Sports Medicine*, 36, 79-97.

Haskell, W.L., Lee, I-M, Pate, R.R., Powell, K.E., Blair, S.N., Franklin, B.A., Macera, C.A., Heath, G.W., Thompson, P.D., and Bauman, A. (2007) Physical Activity and Public Health: Updated Recommendations for Adults from the ACSM and the AHA. *Circulation*, 116, 1081-1093.

Hawkins, M.S, K.L. Storti, C.R. Richardson, W.C. King, S.J. Strath, R.G. Holleman, 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, 282-298.

Hoehner C.M., L.K.B. Ramirez, M.B. Elliot, S. Handy, R. Brownson (2005) Perceived and Objective Environmental Measures and Physical Activity among Urban Adults. *Am J Prev Med,* 28(2S2), 105-116.

Hsiao, C. (1986) *Analysis of Panel Data*. Cambridge University Press, Cambridge.

Huard, D., Evin, G., Favre, A.-C., 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. *Arch. Dis. Child*, 91, 35–38.

Kennedy T.L.M., Smith A., Wells A.T., Wellman B. (2008) Networked families. Pew Internet & American Life Project, http://www.pewinternet.org/.

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.

McKelvey, R.D., Zavoina, W. (1975) A statistical model for the analysis of ordinal-level dependent variables. *Journal of Mathematical Sociology* 4, 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 Heath Risk Behaviors. *Pediatrics*, 117(4), 1281-1290.

O'Brien, S.M. and D.B. Dunson (2004) Bayesian multivariate logistic regression, *Biometrics*, 60(3), 739-746.

Ogden, C.L., M.D. Carroll, M.A. McDowell, K.M. Flegal (2007) Obesity among Adults in the United States-No Statistically Significant Change since 2003-2004. *Center for Disease Control (CDC) Report,* Data Brief Number 1, November, 8 pp.

Ogden, C.L., M.D. Carroll, K.M. Flegal (2008) High Body Mass Index for Age among U.S. Children and Adolescents, 2003-2006. *JAMA*, 299(20), 2401-2405.

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, 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*. San Francisco: Jossey Bass.

Sallis, J.F., J.J. Prochaska, W.C. Taylor (2000) A Review of Correlates of Physical Activity of Children and Adolescents. *Med Sci Sports Exerc*, 32 963-975

Salmon, J., M.L. Booth, P. Phongsavan, N. Murphy, and A. Timperio (2007). Promoting Physical Activity Participation among Children and Adolescents. *Epidemiologic Reviews*.

Schulz, L.0. and D.A. Schoeller, D. A. (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). An Analysis of Bicyclists and Bicycling Characteristics: Who, Why, and How Much are they Bicycling?. Forthcoming, *Transportation Research Record.*

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika* 9, 449-460.

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 KJ, Hunt JD (2006) Age-based analysis of children in Calgary, Canada. Presented at the 85th Annual Meeting of the Transportation Research Board, Washington, D.C., January.

Swallen, K.C, E.N. Reither, S.A. Haas, A.M. Meier (2005) Overweight, Obesity and Health-Related Quality of Life among Adolescents: The National Longitudinal Study of Adolescent Health. *Pediatrics*, 115, 340-347.

Trivedi, P. K., Zimmer, D. M., 2007. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics,* 1(1), Now Publishers.

Trolano, R.P., D. Berrigan, K. 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, M. Dowda (2003) Evaluating a Model of Parental Influence on Youth Physical Activity. *Am J Prev Med,* 25, 277-282

Tucker, P. and J. Gilliland (2007) The Effect of Season and Weather on Physical Activity: A Systematic Review. *Public Health,* 121, 909-922.

United States Department of Health and Human Services (USDHHS) (2008) 2008 Physical Activity Guidelines for Americans. Available at: http://www.health.gov/paguidelines/pdf/paguide.pdf

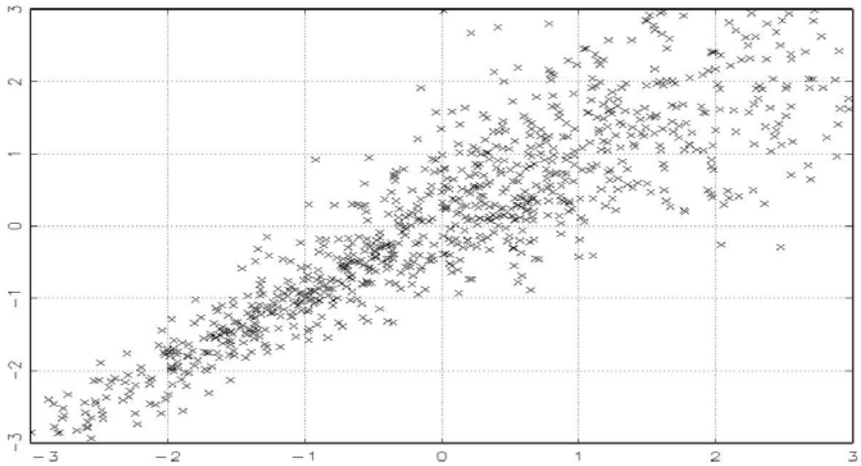
Wendel-Vos, W., M. Droomers, S. Kremers, J. Brug F. van Lenthe (2005) Potential Environmental Determinants of Physical Activity in Adults. In *Environmental Determinants and Interventions for Physical Activity, Nutrition and Smoking: A Review*. Edited by: Brug J, van Lenthe F. Erasmus University Medical Centre, Rotterdam.

Weuve, J., J.H. Kang, J.E. Manson, M.M.B. Breteler, J.H. Ware, F. Grodstein (2004) Physical Activity Including Walking, and Cognitive Function in Older Women. *JAMA*, 292(12), 1454-1461.

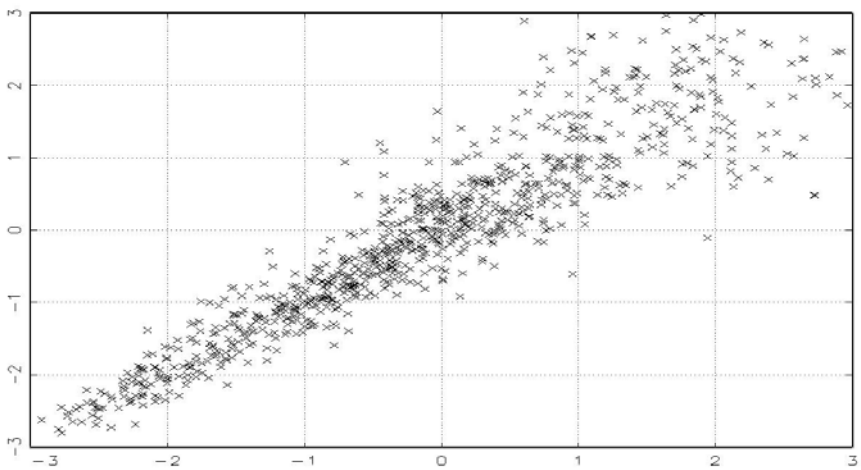
World Health Organization (WHO) (2006) Obesity and Owerweight. Available at: http://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf.

LIST OF FIGURES

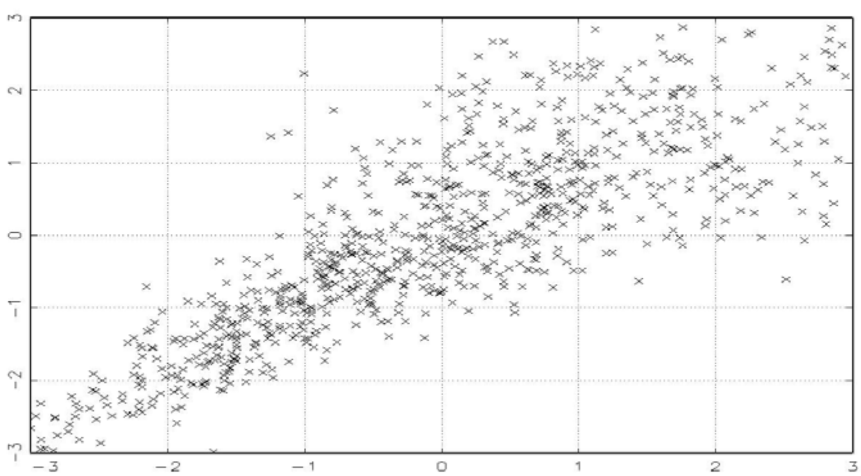
LIST OF TABLES



(1a)



(1b)



(1c)

Figure 1. Logistic-Clayton Copula Plots across Family Types (1a) Two-parents families (τ =0.483; (1b) Single mother families (τ =0.519); (1c) Single father families (τ = 0.414).

Table 1. Estimation Results for the Number of Out-of-Home Weekend Physically Active Activity Episodes

	Adults (Parents)		Children (aged 5-15)	
Variable	**Parameter**	**t-stat**	**Parameter**	**t-stat**
Threshold parameters				
Threshold 1	3.084	4.68	2.702	4.02
Threshold 2	5.138	6.86	5.187	7.13
Individual factors				
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	-	-
Physical environment factors				
Season and activity day				
Winter	-0.428	-1.31	-	-
Sunday	-0.580	-2.73	-0.635	-2.84
Transportation system and built environment characteristics				
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
Social environment factors				
Family-level demographics				
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
Residential neighborhood demographics				
Fraction of Caucasian American population	0.632	1.24	-	-
Fraction of African-American population	-	-	-2.783	-1.34

Table 2. Impact of Change in Individual, Physical and Social Environment Factors

Variable	Formulation of the Change on the Variable	Aggregate change in the percentage probability of expected physical activity participation					
		ORL		RORL		LC	
		Adults	Children	Adults	Children	Adults	Children
Individual factors							
Parental internet use	Increased by 10%	-1.046	-	-1.691	-	-2.177	-
Physical environment factors							
Built environment characteristics							
Bicycling facility density (miles of bike lanes per square mile)	Increased by 10%	2.441	2.497	1.646	2.206	1.585	2.232
Fraction of multi-family dwelling units	Increased by 10%	-	1.418	-	0.993	-	0.965
Social environment factors							
Family-level demographics							
Household vehicle ownership	Decreased by 1	15.297	-	11.682	-	18.762	-
Household bicycle ownership	Increased by 1	-	14.947	-	9.273		9.283
Residential neighborhood demographics (spacing issues)							
Fraction of Caucasian-American population	Increased by 10%	4.945	-	3.260	-	3.012	-
Fraction of African-American population	Increased by 10%	-	-1.422	-	-1.022	-	-0.941