A Copula-Based Sample Selection Model of Telecommuting Choice and Frequency

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

and

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

*corresponding author

July 2009

ABSTRACT

The confluence of a need to reduce traffic congestion during the peak periods, as well as reduce vehicle miles of travel due to work-related travel (which contributes to GHG emissions from the transportation sector), has led planning organizations and regional governments to consider several demand management actions, one of them being the promotion of telecommuting. The objective of this study is to contribute to the telecommuting literature by jointly examining the propensity and frequency of workers to telecommute, using a rich set of individual demographics, work-related and occupation characteristics, household demographics, and commute trip/work location characteristics. The data are drawn from the Chicago Regional Household Travel Inventory, collected between 2007 and 2008. From a methodological standpoint, the current study adopts a copula approach that allows the testing of several types of dependency structures between the telecommuting choice and frequency behavioural processes. To our knowledge, this is the first formulation and application in the econometric literature of a copula approach for the case of a binary self-selection mechanism with an ordered-response outcome.

The results clearly indicate that telecommuting choice and the frequency of telecommuting are governed by quite different underlying behavioral processes. In particular, the determinant factors of telecommuting choice and frequency can be different. Further, a factor that has a particular direction of effect on telecommuting choice may have the opposite effect on frequency. Also, the analyst risks the danger of incorrect conclusions regarding dependency in the telecommuting choice and frequency behavioral processes, as well as inconsistent and inefficient parameter estimates, by imposing incorrect dependency structures or assuming independence between the two behavioral processes. Overall, the empirical results indicate the important effects of several demographic and work-related variables on telecommuting choice and frequency, with implications for transportation planning and transportation policy analysis.

Keywords: Telecommuting choice, telecommuting frequency, copula approach, revealed preference analysis, sample selection models, ordered-response structure

1. INTRODUCTION

1.1. Background and Motivation

In May 2006, the U.S. Secretary of Transportation identified traffic congestion as one of the single largest threats to the United States' economic prosperity and way of life. This is reinforced by the most recent Urban Mobility Report by TTI (Schrank and Lomax, 2009), which indicated that the cost of traffic congestion in the U.S. (due to congestion-related delay and wasted fuel) was approximately $87 billion in 2007, an increase of more than 50% from 1997.

Traffic congestion is highest during the morning and evening commute periods, corresponding to the time when workers make the transition from home-to-work or work-to-home. According to the Texas Transportation Institute's (TTI's) mobility report, the congestion-related annual delay per peak period traveler was approximately 36 hours in 2007, up from 14 hours in 1982. The corresponding direct annual cost to a peak period traveler was estimated at $757. This "wasted" cost to the average peak period traveler is an obvious cause of concern in an already struggling economy. At the same time, global climate change, the broad term used to reflect recent global warming trends, has been linked unequivocally to human activity that results in the emission of greenhouse gases. In the U.S., energy-related activities account for three-quarters of total human-generated greenhouse gas (GHG) emissions, mostly in the form of Carbon Dioxide (CO_2) emissions from burning fossil fuels. Recent projections show that the nation's CO_2 emissions would increase from about 5.9 million metric tons in 2006 to 7.4 million metric tons in 2030 if measures are not taken to reduce carbon emissions (NAS, 2008). While about one-half of these emissions come from large stationary sources such as power plants, the transportation sector ranks second and accounts for about one-third of all human generated GHG emissions (EPA, 2007). Further, the transportation sector is one of the most rapidly rising sources of GHG emissions. For example, total U.S. GHG emissions rose 13% between 1990 and 2003, while those from the transportation sector rose 24% during the same period (EPA, 2006).

The confluence of a need to reduce traffic congestion during the peak periods, as well as reduce vehicle miles of travel due to work-related travel (which contributes to GHG emissions from the transportation sector), has led planning organizations and regional governments to consider several demand management actions, one of them being the promotion of telecommuting. Telecommuting, generally defined as "using technologies to work at home or at a location close to home instead of commuting to a conventional work place at the conventional time" (Bagley and

Mokhtarian, 1997), is particularly suited to companies that specialize in occupations requiring high usage of computers and telecommunications. In turn, these companies may realize savings in office space and other office overheads. In fact, a recent study by the General Services Administration (GSA, 2006) reported that the financial benefit a company accrues by allowing its employees to telecommute far outstrips the cost to the company of providing the necessary telecommuting products and services. This finding suggests that instituting telecommuting programs may not only enable planning organizations to reduce traffic congestion/GHG emissions, but also may be an option that many institutions could use to improve their financial bottom line.

Indeed, there is evidence of increasing telecommuting adoption in the U.S. over the past several years. Estimates of the number of U.S. workers in 2000 who telecommuted at least once a month in the U.S. ranged from 17 to 18 million (Jala International, 2000). A more recent study conducted by World at Work (2009) found that the number of U.S. workers who telecommute at least once a month has shown a steady climb to 23.5 million in 2003, 28.7 million in 2006, and 33.7 million in 2008. However, this increase in telecommuting adoption has not necessarily also translated to an increase in the number of days of telecommuting among those who telecommute. In fact, while the number of workers telecommuting has increased by approximately 17% (about 5 million), the number of individuals telecommuting almost every day has decreased by approximately 8% (about 1.2 million) between 2006 and 2008 (World at Work, 2009). These differing and opposite trends in telecommuting adoption and the intensity of adoption (or telecommuting frequency), in conjunction with the potential benefits of telecommuting to the economy and the environment, has led to an increased interest in understanding the underlying processes determining telecommuting choice (or adoption) and telecommuting frequency. The current study contributes to such an understanding by modeling telecommuting choice and telecommuting frequency separately, but jointly. The sample used in the analysis is drawn from the 2008 Chicago Regional Household Travel Inventory (CRHTI), and offers the opportunity to study telecommuting behavior using a very recent revealed preference survey.

The rest of the paper is structured as follows. Section 2 presents a brief overview of the earlier literature on telecommuting and positions the current study within this broader context. Section 3 describes the data collection procedures as well as the sample used in the analysis. Section 4 outlines the modeling methodology employed for the empirical analysis of the current study.

Section 5 presents the empirical results. Finally, Section 6 summarizes important findings from the study and concludes the paper.

2. OVERVIEW OF EARLIER STUDIES AND CURRENT PAPER

In this section, we provide an overview of earlier telecommuting studies to demonstrate the level of interest in the topic and the types of analyses that have been conducted. The intent of the discussion is not to provide an extensive review of the literature, but rather to present important trends in the study of telecommuting (see Tang *et. al.*, 2008 and Walls and Safirova, 2004 for detailed reviews on the subject).

The studies of telecommuting may be broadly classified into three categories: (1) Qualitative studies, (2) Quantitative studies using stated-preference survey data, and (3) Quantitative studies using revealed-preference survey data. The early works on telecommuting adoption were largely qualitative, and focused on examining the motivations and deterrents to telecommuting (see for example, Edwards and Edwards, 1985, Gordon, 1988, and Nilles, 1988). The qualitative discussion on the adoption process has taken new quantitative directions more recently, through the development of adoption frameworks and subsequent operationalizations of probabilistic behavioral models. Such models provide a multivariate picture of the determinants (or deterrents) of telecommuting choice and frequency, and are discussed in more detail below.

The first group of quantitative studies on telecommuting was based on stated preference surveys, ostensibly because the penetration rate of telecommuting in the worker population until the mid-1990s was not adequate to support quantitative modeling using revealed preference data (Mannering and Mokhtarian, 1995). For instance, Bernardino *et al.* (1993) used an ordered probit framework to model the telecommuting willingness of 54 individuals who responded to a survey posted at selected newsgroup sites on the world wide web. Yen and Mahmassani (1994) also used an ordered response framework to examine the stated preference of employees in Austin, Dallas, and Houston to choose to telecommute under various survey-defined hypothetical programmatic scenarios (such as a 5% or 10% increase/decrease in salary in return for telecommuting). Respondents could indicate their willingness to participate in telecommuting in response to each scenario in one of four categories: will not work from home at all, will possibly work from home, will work several days a week from home, and will work from home everyday. Unlike the studies of Bernardino *et al.* and Yen and Mahmassani just discussed, Sullivan *et al.* (1993) estimated a multinomial logit model (rather than an ordered-response model) to analyze telecommuting choice and participation frequency using a stated preference survey of employees of information-oriented firms in Austin, Dallas and Houston. Sullivan *et al*. considered four alternatives for the choice/frequency of telecommuting: "will not telecommute", "possibly will telecommute", "part-time telecommute", and "full-time telecommute". All the above studies, while providing useful insights regarding the stated preferences of individuals to adopt telecommuting, do not adequately examine the actual choices/constraints of individuals that influence telecommuting adoption and frequency. As a result, they are likely to be of limited value for informing the development of policy strategies (Mokhtarian and Salomon, 1996a).

The earliest published research effort using revealed preference data for the quantitative evaluation of telecommuting choice/frequency appears to be the one by Olszewski and Mokhtarian (1994). These authors used data obtained from the State of California Telecommuting Pilot Project. Using analysis of variance techniques, the authors examined the influence of demographic and commuting variables on telecommuting frequency (number of telecommuting days per month), among those participating in the pilot project. Thus, the emphasis was solely on the telecommuting frequency dimension, not the choice dimension. The results from the study indicated statistically insignificant effects of age, gender, number of children in the household, and commute distance on telecommuting frequency, though some of these results may simply be an artifact of the limited sample size in the analysis. Subsequent to the Olszewski and Mokhtarian study, Mannering and Mokhtarian (1995) employed a sample of over 433 telecommuters and non-telecommuters from three surveys conducted in 1992 to estimate a multinomial logit model with three possible alternatives: "never telecommute", "infrequently telecommute", and "frequently telecommute". However, the study was limited by the small percentage of telecommuters and a small percentage of frequent telecommuters within the survey sample. Several other studies also focused on the choice of telecommuting, occasionally with some representation of frequency in the broad manner of Mannering and Mokhtarian (1995). The emphasis in these studies was to include specific sets of factors, such as work-related characteristics in Bernardino and Ben-Akiva (1996), subjective personal attitudes and workplace perceptions in Mokhtarian and Salomon (1996b), and a host of motivation-related and constraint-related attitudes/perception variables associated with work, home, travel, and leisure in Mokhtarian and Salomon (1997). Another revealed preference study with a more national focus (rather than the regional focus of the studies just mentioned) is the one by

Drucker and Khattak (2000), who examined the choice of never telecommuting, infrequently telecommuting, and frequently telecommuting using data from the 1995 Nationwide Personal Transportation Survey (NPTS).

Finally, the last few years has seen more research with revealed preference data focusing on both the telecommuting choice as well as a measure of frequency that includes a time frame of reference (such as once a month, once a week, 2-3 times a week, and 4-5 times a week) as opposed to previous broad characterizations as “infrequently” or “frequently” telecommute. Some of these studies also explicitly recognize that the telecommuting choice decision (*i.e.*, whether to telecommute at all or not) and the frequency of telecommuting may be governed by quite different underlying behavioral processes rather than being governed by a single behavioral process. For instance, Popuri and Bhat (2003) were the first to jointly model the choice and frequency decisions. Specifically, they recognized that, while the choice and frequency decisions may not be tied very tightly, they may be related to each other due to observed and unobserved factors. In the latter context, factors such as being techno-savvy or having a general preference to travel less may increase the propensity to telecommute and increase the frequency of telecommuting. Popuri and Bhat’s model results indeed indicate that there is a positive correlation due to unobserved factors in the choice and frequency decisions, and show that failure to accommodate this correlation can lead to inconsistent parameter estimates. However, their data set does not have job-related characteristics (such as industry and occupation categories) that may significantly influence telecommuting. In this regard, Walls *et al.* (2007) examined both the choice and frequency decisions of telecommuting using an extensive set of job-related factors and found substantial influences of these work-related factors. In their study, Walls *et al.* considered the correlation in unobserved factors in the choice and frequency decisions by including a Heckman’s (1979) correction term in the frequency model after estimating the telecommuting binary choice model parameter estimates. They found this correction term to be statistically insignificant, and so estimate independent models of choice and frequency. However, the textbook Heckman’s correction term is valid only for a continuous outcome equation, and not for the ordered response outcome of frequency that Walls *et al.* (2007) employ. The appropriate procedure for the normally distributed underlying processes of choice and frequency that Walls *et al*. assume would be the joint estimation technique of Popuri and Bhat (2003). Finally, Tang *et al.* (2008) examined the effect of objective residential neighborhood built environment factors, as well as subjective perceptions of these factors, on both the adoption and frequency of telecommuting, using a single multinomial logit model (MNL) with the alternatives of non-adoption, 1 day per month adoption frequency, 2-4 days per month, 5-8 days per month, and more than 8 days per month. They also considered ordinal response and count models for frequency, but found these to be less satisfactory than the MNL approach. One limitation of their study is that they consider very few individual/household demographic variables, and no work-related variables (other than commute time).

Overall, the above discussion illustrates the substantial recent interest in jointly analyzing the choice and frequency of telecommuting. The objective of this study is to contribute to this telecommuting literature in several important ways. First, the sample used in this study includes the revealed preference survey responses of 9264 workers from the Chicago region. The sample comprises 1534 telecommuters, which constitutes the largest number of telecommuters in any study so far that we are aware of. The large sample of telecommuters should aid in comprehensively and rigorously "teasing out" the factors that influence the telecommuting adoption and frequency decisions. In fact, the richness of the data allows us to incorporate a variety of variables, including individual demographics, work-related and occupation characteristics, household demographics, and commute trip/work location characteristics. Second, the data sample is obtained from the recently completed 2008 Chicago Regional Household Travel Inventory (CRHTI), thus providing us with the ability to develop a very current perspective of the process driving telecommuting decisions (at least in the Chicago region). In contrast, even the recent studies by Walls *et al.* (2007) and Tang *et al.* (2008) have used relatively dated data from 2002 and 2003, respectively. Third, the survey reduces the ambiguity in the difference between home-based telecommuting and operation of a home-based business by removing individuals who indicated that they were self-employed and worked primarily from home. Thus, the sample of workers considered in the current analysis includes only those who stated expressly that their primary/main work location is a location outside home that they travel to routinely. Finally, from a methodological perspective, we jointly model the choice and frequency of telecommuting rather than independently modeling the two decisions. The failure to capture the jointness among these two inter-related choices can lead to inconsistent parameter estimates and misinformed policy actions, as discussed in Popuri and Bhat (2003). However, we go one step beyond the methodological approach of Popuri and Bhat by using a flexible copula-based approach to characterize the dependency between the error terms in the telecommuting choice and frequency equations. The copula approach allows the testing of several types of dependence structures rather than pre-imposing the very restrictive bivariate normal distribution assumption of Popuri and Bhat (see Bhat and Eluru, 2009 for an extensive discussion of the copula approach).[1]

3. DATA AND SAMPLE DESCRIPTION

3.1. Data Sources

The data used in this study are drawn from the 2008 Chicago Regional Household Travel Inventory (CRHTI), which was sponsored by the Chicago Metropolitan Agency for Planning (CMAP), the Illinois Department of Transportation (IDOT), the Northwestern Indiana Regional Planning Commission, and the Indiana Department of Transportation. The study area of the survey included eight counties in Illinois (Cook, DuPage, Grundy, Kane, Kendall, Lake, McHenry, and Will counties), and three counties in Indiana (Lake, LaPorte, and Porter). The survey was administered using standard postal mail-based survey methods and computer-aided telephone interview (CATI) technology through Travel Tracker Survey to facilitate the organization and storage of the data. A dual sampling frame approach was used, with one sampling frame being the list of land-line telephone numbers in the study area and the other being an address-based frame of all residential addresses that receive U.S. postal mail. This dual approach was used because the sampling frame of land-line telephone numbers has coverage bias toward upper income home owners who have resided in the area for a long time, while the latter address-based frame is less biased and captures low-income, minority, renters, new residents, and cell-only households. But random digit dialing using the sampling frame of telephone numbers is more time and cost-efficient, while the mailing to a random sample from the postal address-based frame is passive, and requires the potential respondent to open the mailing and make contact through return mail or the web or the phone to provide

[1] An important point about the telecommuting choice variable in the study. The Chicago survey asks the following question: "*Does your employer allow you to work from home for pay on a regular basis? This would be in place of driving to a regular work location, something that is commonly referred to as -telework*.-" All those who answered positively to the above question indicated that they telecommuted at least occasionally in the year. This is consistent with the finding of Mokhtarian and Salomon (1997) that almost all individuals who are provided the opportunity to telecommute by their employers will choose to telecommute. One can then argue that the telecommuting choice binary variable in the current study, which is based on the response to the question presented above, may be better viewed as whether an individual chooses an employer who allows telecommuting. But, over the long run, individuals do decide whether to telecommute or not by switching jobs, changing work arrangements, or specializing in occupations more conducive to telecommuting. Thus, one can view the presence of a telecommuting arrangement as a manifestation of basic individual desires and trade-offs related to work and personal characteristics. In this sense, the situation boils down to the choice of the individual to telecommute.

relevant information. Further details of the survey design and implementation methods are available in NuStats (2008).

The survey was conducted expressly to inform the development of regional travel demand models for the Chicago region. It involved the collection of activity and travel information for all household members (regardless of age) during a randomly assigned 1-day or 2-day period (the 1-day period sample focused only on weekdays, while the 2-day period sample targeted two consecutive days including the Sunday/Monday and Friday/Saturday pairs but not the Saturday/Sunday pair). Respondents were asked to provide detailed information on household demographics and individual demographics of each household member, the vehicles owned by the household, and all travel and out-of-home activity episodes for each household member during the assigned survey day. The final sample included information from 14,315 households.

3.2. Sample Formation and Description

The data assembly process involved several steps. First, the (individual and household) demographic variables and reported activity-travel characteristics were assembled into a single person-level file. Second, since the focus of the study is on telecommuting, only employed individuals were selected from the overall sample. Third, two specific dimensions of each employed individual's work pattern were considered for the current analysis: (1) Telecommuting choice (whether or not person telecommutes – see footnote 1), and (2) Telecommuting frequency (obtained in one of the five categories of "once a year", "a few times a year", "once a month or more", "once a week or more", and "almost everyday"). In the current analysis, we use a binary model for the telecommuting choice component and a five-point ordered-response model for the telecommuting frequency component. Finally, several screening and consistency checks were undertaken to obtain the final sample of 9264 employees.

The data sample for analysis includes 1534 telecommuters (15.9% of the overall sample). This telecommuting percentage is similar to that found in Popuri and Bhat in the New York City area, though it is lesser than the 25% or so telecommuting percentages reported in Walls *et al.* (2007) and Tang *et al.* (2008). This lower percentage in our study is potentially because we are better able to distinguish between telecommuters and home-based business (HBB) workers (*i.e.*, those who work out of home). Tang *et al.* acknowledge that their characterization of telecommuters is likely to be a mix of actual telecommuters and HBB workers. In terms of telecommuting frequency, the split in the sample of telecommuters is as follows: 36 (0.4%) telecommute once a year, 194 (14.6%) telecommute a few times a year, 461 (30.1%) telecommute once or more per month, 649 (42.3%) telecommute once or more per week, and 194 (12.6%) telecommute almost everyday. As expected, those who telecommute do so at least once a month.

In our empirical analysis, we considered several possible sets of variables to explain telecommuting choice and frequency. We do not present an aggregate distribution of each of these variables in the overall sample and in the telecommuting sample because such an examination only provides univariate statistics without controlling for other determinant variables. The appropriate mechanism to study the influence of each variable would be the disaggregate joint model estimated in Section 4.

4. METHODOLOGY

4.1. Model Structure

In our empirical analysis, there are two dependent variables - telecommuting choice, modeled using a binary choice structure, and telecommuting frequency, modeled using an ordered-response structure. These two dependent variables are jointly analyzed using a copula approach that enables flexible dependency in the latent propensities underlying the choice and frequency dimensions. Mathematically, the model system is as follows:

$$t_q^* = \beta' x_q + v_q,\ t_q = 1 \text{ if } t_q^* > 0 \text{ and } t_q = 0 \text{ if } t_q^* < 0$$

$$s_q^* = \gamma' z_q + \eta_q,\ s_q = k \text{ if } \delta_{k-1} < s_q^* < \delta_k,\ k = 1,2,\ldots,K,\ s_q \text{ observed only if } t_q^* > 0, \qquad (1)$$

where q is an index for individuals, k is an index for frequency level, t_q is an observed binary variable indicating whether or not person q chooses to telecommute (t_q =1 if person q telecommutes, 0 otherwise), t_q^* is an underlying continuous variable related to the observed binary variable t_q as shown above, s_q is an observed ordinal variable representing the frequency of telecommuting if individual q telecommutes, s_q^* is a latent continuous variable representing the propensity underlying the telecommuting frequency decision, the δ_q terms represent thresholds that relate s_q^* to the observed variable s_q in the usual ordered-response structure

$(\delta_0 = -\infty, \delta_K = \infty; -\infty < \delta_1 < \delta_2 < \ldots < \delta_{K-1} < \infty)$, x_q and z_q are vectors of explanatory variables (as written in Equation (1), x_q includes a constant, but z_q does not), β and γ are vectors of parameters to be estimated, and v_q and η_q are random error terms, which may take any parametric distribution. In the current study, we examine both logistic and normal marginal distributions for these error terms, and choose the distribution that provides the best data fit. The error terms v_q are assumed to be independent and identically distributed (IID) across individuals *q*, and the error terms η_q are also assumed to be IID across individuals *q*. Further, for the logistic case, a standard logistic distribution is used for the error terms, while, for the normal case, a standard normal distribution is used for the error terms (these standardizations are innocuous normalizations needed for econometric identification). For presentation ease, let the marginal distribution of v_q be *F*(.) and the marginal distribution of η_q be *G*(.).[2]

With the notational preliminaries above, the probability that individual *q* does not telecommute is simply given by:

$$\Pr[t_q = 0] = \Pr[v_q < -\beta' x_q] = F(-\beta' x_q). \tag{2}$$

The probability that the individual *q* telecommutes and does so at a frequency level *k* (*k* = 1,2,…*K*) can be written from Equation (1) as:

$$\begin{aligned}
&\Pr[t_q = 1, s_q = k] = \Pr[\nu_q > -\beta' x_q, \delta_{k-1} - \gamma' z_q < \eta_q < \delta_k - \gamma' z_q] \\
&= \Pr[\nu_q > -\beta' x_q, \eta_q < \delta_k - \gamma' z_q] - \Pr[\nu_q > -\beta' x_q, \eta_q < \delta_{k-1} - \gamma' z_q] \\
&= \Pr[\eta_q < \delta_k - \gamma' z_q] - \Pr[\nu_q < -\beta' x_q, \eta_q < \delta_k - \gamma' z] - \left(\Pr[\eta_q < \delta_{k-1} - \gamma' z_q] - \Pr[\nu_q < -\beta' x_q, \eta_q < \delta_{k-1} - \gamma' z_q]\right) \\
&= G\left(\delta_k - \gamma' z_q\right) - \Pr[\nu_q < -\beta' x_q, \eta_q < \delta_k - \gamma' z_q] - \left(G\left(\delta_{k-1} - \gamma' z_q\right) - \Pr[\nu_q < -\beta' x_q, \eta_q < \delta_{k-1} - \gamma' z_q]\right)
\end{aligned} \tag{3}$$

The above joint probability depends upon the dependence structure between the random variables v_q and η_q. As highlighted before, the incorporation of the dependency effects can be greatly

[2] Thus, in the context of the current analysis, *F*(.) may be the standard logistic cumulative distribution function or the standard normal distribution function. The same is the case with *G*(.). Note that, in the copula approach we use, it is not necessary that both *F*(.) and *G*(.) should be simultaneously logistic (logistic-logistic) or simultaneously normal (normal-normal). Rather, we can also test the normal-logistic and logistic-normal pairings.

facilitated by using a copula approach for modeling joint distributions. In the next section, we identify various copula structures, which accommodate different parametric functional forms for the bivariate dependency surface. This is particularly important since the copula approach does not need the *a priori* specification of the functional form of the dependence surface. Indeed, we can test different functional forms, and select the one that empirically fits the data best. To our knowledge, we are the first to formulate and estimate a copula-based model for the case of a binary self-selection model with an ordinal outcome equation.

4.2. General Bivariate Copula Structure

A copula is a device or function that generates a stochastic dependence relationship (*i.e.*, a multivariate distribution) among random variables with pre-specified marginal distributions (see Trivedi and Zimmer, 2007 or Nelsen, 2006). Specifically, the copula approach separates the marginal distributions from the dependence structure, so that the dependence structure is entirely unaffected by the marginal distributions assumed. This, in turn, provides substantial flexibility in correlating random variables, which may not even have the same marginal distributions. In this regard, recall from Section 4.1 that the random terms v_q and η_q in Equation (1) may have different marginal distributions.

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

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

where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector. A copula, once developed, allows the generation of joint multivariate distribution functions with given marginals. Thus, in the context of the current study, a joint bivariate distribution function of the random variables v_q [with the marginal distribution $F(.)$] and η_q [with the marginal distribution $G(.)$] may be generated as follows (see Sklar, 1973):

$$J(v,\eta) = \Pr(v_q < v, \eta_q < \eta) = \Pr[U_1 < F(v), U_2 < G(\eta)] = C_\theta[u_1 = F(v), u_2 = G(\eta)], \quad (5)$$

where C_θ is a copula function and θ is a dependency parameter (assumed to be scalar), together characterizing the dependency between v_q and η_q. A rich set of bivariate copulas $C_\theta(u_1, u_2)$ are available to generate the dependence between the random variables v_q and η_q, including the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and the Archimedean class of copulas (including the Clayton, Gumbel, Frank, and Joe copulas). For given functional forms of the margins, the precise bivariate dependence profile between the variables v_q and η_q is a function of the copula $C_\theta(u_1, u_2)$ used, and the dependence parameter θ. But, regardless of the margins, the overall nature of the dependence between v_q and η_q is determined by the copula function.

4.2.1. The Gaussian and FGM Copulas

The Gaussian copula is the most familiar of all copulas, and forms the basis for Lee's (1983) sample selection approach. The Gaussian copula takes the following form:

$$C_\theta(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \theta), \tag{6}$$

where $\Phi_2(.,.,\theta)$ is the bivariate cumulative distribution function with Pearson's correlation parameter $\theta(-1 \le \theta \le 1)$. Independence corresponds to $\theta = 0$. The Gaussian copula is comprehensive in its coverage in that it is able to capture the full range of (negative or positive) dependence between two random variables.

The bivariate FGM copula (Morgenstern, 1956, Gumbel, 1960, and Farlie, 1960) takes the following form:

$$C_\theta(u_1, u_2) = u_1 u_2 [1 + \theta(1-u_1)(1-u_2)]. \tag{7}$$

The presence of the θ term $(-1 \le \theta \le 1)$ allows dependence between the uniform marginals u_1 and u_2. Independence corresponds to $\theta = 0$. The FGM copula has a simple analytic form and allows for either negative or positive dependence. However, the FGM copula is not comprehensive in coverage, and can accommodate only relatively weak dependence between the marginals (see Bhat and Eluru, 2009).

Both the Gaussian and FGM copulas assume the property of asymptotic independence. That is, regardless of the level of correlation assumed, extreme tail events appear to be independent in each margin just because the density function gets very thin at the tails (see Embrechts *et al.*, 2002). Further, the dependence structure is radially symmetric about the center point in the Gaussian and FGM copulas. That is, for a given value of θ, the level of dependence is equal in the upper and lower tails. On the other hand, it may be that unobserved factors (such as, say environmental consciousness) that increase telecommuting propensity also increase telecommuting propensity, so that when v_q is highly positive, so is η_q. However, one may not find the same level of strong dependence in the lower end of the (v_q, η_q) spectrum. This implies the case of strong dependency in the right tail, but not in the left tail. Alternatively, one may have the reverse asymmetry too where there is strong dependency in the left tail, but not the right. Or it may be that there is very weak dependency in the two tails, but much stronger dependency in the center of the joint distribution of the propensity to telecommute and to do so frequently (than that implied by the Gaussian or FGM copulas). In general, one does not know *a priori* what kind of dependency structure holds between the unobserved factors influencing the telecommuting choice and frequency decisions. Rather this is an empirical issue to be determined based on which dependency surface fits the data best. In this context, a class of copulas referred to as the Archimedean copulas provide much needed flexibility to test dependency functional forms.

4.2.2. 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 structures, including comprehensive and non-comprehensive copulas, radial symmetry and asymmetry, strong central tendency and weak tail dependency, and asymptotic tail independence and dependence (see Nelsen, 2006 and Bhat and Eluru, 2009 for a detailed discussion). The class is very flexible, and easy to construct.

Clayton Copula

The Clayton copula has the following form (Clayton, 1978):

$$C_\theta(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \ 0 < \theta < \infty. \tag{8}$$

Independence corresponds to $\theta \to 0$. The above copula cannot account for negative dependence. The copula is best suited for strong left tail dependence and weak right tail dependence. That is, it is best suited when individuals who have a low propensity to telecommute (due to unobserved factors) are also likely to telecommute less frequently, but employees who have a high propensity to telecommute are not likely to telecommute more frequently.

Gumbel Copula

The Gumbel copula, first discussed by Gumbel (1960) and sometimes also referred to as the Gumbel-Hougaard copula, has the form provided below:

$$C_\theta(u_1, u_2) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right),\ 1 \le \theta < \infty. \tag{9}$$

Independence corresponds to $\theta = 1$. As with the Clayton copula, the Gumbel copula cannot account for negative dependence. The Gumbel 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).

Frank Copula

The Frank copula, proposed by Frank (1979), is given by:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right),\ -\infty < \theta < \infty. \tag{10}$$

Independence is attained in Frank's copula as $\theta \to 0.$ The copula allows for positive and negative dependence, is comprehensive in its coverage, is radially symmetric in its dependence structure, and imposes the assumption of asymptotic independence. However, the dependence surface of Frank's copula shows very strong central dependency (stronger than the Gaussian copula) and very weak tail dependence (weaker than the Gaussian copula). Frank's copula has been used extensively in empirical applications.

Joe Copula

The Joe copula, introduced by Joe (1993, 1997), has the following copula form:

$$C_\theta(u_1,u_2)=1-\left[(1-u_1)^\theta+(1-u_2)^\theta-(1-u_1)^\theta(1-u_2)^\theta\right]^{1/\theta},\ 1\le\theta<\infty. \tag{11}$$

Independence corresponds to $\theta=1$. The Joe copula is similar to the Gumbel, but the right tail positive dependence is stronger. In fact, from this standpoint, the Joe copula is closer to being the reverse of the Clayton copula than is the Gumbel.

4.3. Model Estimation

The parameters to be estimated in the joint binary choice-ordered response model (that is, telecommuting choice-telecommuting frequency models) include the β vector, the (K-1) δ_k parameters $(\delta_0=-\infty,\delta_K=\infty;-\infty<\delta_1<\delta_2<\ldots<\delta_{K-1}<\infty)$ and the vector γ .

The probability that an individual q telecommutes and does so at a frequency level k (k = 1,2,…K) can be obtained from Equation (3) as follows:

$$\Pr[t_q=1,s_q=k]=G(\delta_k-\gamma'z_q)-G(\delta_{k-1}-\gamma'z_q)-\left[C_\theta(u_{q1},u_{q,k,2})-C_\theta(u_{q1},u_{q,k-1,2})\right], \tag{12}$$

where $u_{q1}=F(-\beta'x_q)$, $u_{q,k,2}=G(\delta_k-\gamma'z_q)$, and $u_{q,k-1,2}=G(\delta_{k-1}-\gamma'z_q)$

Next, let $1[.]$ be an indicator function taking the value of unity if the expression in parenthesis is true and 0 otherwise. Also, define a set of dummy variables M_{qk} as below:

$$M_{qk}=1[t_q=1]\times 1[s_q=k]. \tag{13}$$

Then, the log likelihood function for the copula model takes the following form:

$$\log L=\sum_{q=1}^{Q}\left(1[t_q=0]\times\log[\Pr(t_q=0)]+\sum_{k=1}^{K}M_{qk}\log[\Pr(t_q=1,s_q=k)]\right). \tag{14}$$

All the parameters in the model are estimated by maximizing the log-likelihood function in Equation (14). The model estimation was pursued using the GAUSS matrix programming language.

5. MODEL RESULTS

5.1. Variable Specification

Several variable specifications and functional forms were considered in the model. These included (1) individual demographics, such as age, sex, race, driver's license holding, and physical disability status, (2) work-related and occupation characteristics, such as full time or part-time employment, work schedule flexibility, number of jobs, job industry, and occupation type, 3) household demographics, such as number of adults, number of children, household income, dwelling type, whether the house is owned or rented, and number of phone-lines in the home, and (4) commute trip/work location characteristics, such as usual travel mode when commuting to the out-of-home work location, whether the individual has to pay (or not) to park at the work end, amount of parking costs, home-to-work distance, whether or not the roadway type usually traveled on to work includes a tollway and/or expressway, and tolls paid on the usual route to work. In addition, several interaction effects of the variables were considered.

The final model specification was based on intuitive considerations, insights from previous literature, parsimony in specification, and statistical fit/significance considerations. The final specification includes some variables that are not highly statistically significant, but which are included because of their intuitive effects and potential to guide future research and survey efforts in the field.

5.2. Model Specification and Data Fit

The empirical analysis involved estimating models with two different univariate (*i.e.*, marginal) distribution assumptions (normal and logistic) for the error terms v_q and η_q, and seven different copula structures (independence, Gaussian, FGM, Clayton, Gumbel, Frank, and Joe).[3] As discussed in Section 4, in the copula approach, there is no need to assume that the marginal distributions of the v_q and η_q error terms are simultaneously normal (normal-normal) or logistic (logistic-logistic); instead v_q and η_q terms can have a normal-logistic or logistic-normal distribution. We examined all

[3] Due to space considerations, we are unable to provide additional details on the structures of different copula types. Interested readers are referred to Bhat and Eluru (2009). Also, note that the independence copula, as should be self-explanatory, is a copula that assumes independence. In the notation of Section 4.2, the independence copula corresponds to $C_\theta(u_1,u_2) = u_1u_2$.

these four possible combinations for the error terms v_q and η_q, as well as the seven copula dependency structures, for a total of 28 copula-based models. These are as follows: (1) Normal - Normal Gaussian, (2) Normal-Normal FGM, (3) Normal-Normal Clayton, (4) Normal-Normal Gumbel, (5) Normal-Normal Frank, (6) Normal-Normal Joe, (7) Logistic-Logistic Gaussian, (8) Logistic-Logistic FGM, (9) Logistic-Logistic Clayton, (10) Logistic-Logistic Gumbel, (11) Logistic-Logistic Frank, (12) Logistic-Logistic Joe, (13) Normal-Logistic Gaussian, (14) Normal-Logistic FGM, (15) Normal-Logistic Clayton, (16) Normal-Logistic Gumbel, (17) Normal-Logistic Frank, (18) Normal-Logistic Joe, (19) Logistic-Normal Gaussian, (20) Logistic-Normal FGM, (21) Logistic-Normal Clayton, (22) Logistic-Normal Gumbel, (23) Logistic-Normal Frank, (24) Logistic-Normal Joe, (25) Normal-Normal Independence, (26) Logistic-Logistic Independence, (27) Normal-Logistic Independence, and (28) Logistic-Normal Independence.

The Bayesian Information Criterion (BIC) is employed to select the best copula model among the first 24 competing non-nested copula models (see Quinn, 2007, Genius and Strazzera, 2008, Trivedi and Zimmer, 2007, page 65), since the traditional likelihood ratio test for comparing these alternative copula-based models is not applicable. The BIC for a given copula model is equal to $-2\ln(L)+K\ln(Q)$, where $\ln(L)$ is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The copula that results in the lowest BIC value is the preferred copula. However, since all the competing models in the current analysis have the same exogenous variables and the same number of thresholds and constants, the BIC information selection procedure measure is equivalent to selection based on the largest value of the log-likelihood function at convergence.

Among the first 24 copula models, the Normal-Normal Frank (NNF) model provided the best data fit, with a corresponding Kendall's measure of dependency of 0.22 and a likelihood value of -5119.092.[4] The positive dependence between the v_q and η_q terms is intuitive, indicating that unobserved factors (such as feeling more productive working from home or preferring to work without others around) that increase an employee's propensity to telecommute also increase the

[4] Kendall's measure of dependency (τ) transforms the copula dependency parameter (θ) into a number between -1 and 1 (see Bhat and Eluru, 2009). For the Frank copula, $\tau = 1-\frac{4}{\theta}\left[1-\frac{1}{\theta}\int_{t=0}^{\theta}\frac{t}{e^t-1}dt\right]$. Independence is attained in Frank's copula as $\theta \rightarrow 0$.

employee's inclination to telecommute frequency. Among the final four independence copula models, the Normal-Normal Independence (NNI) model provided the best data fit, with a likelihood value of -5122.240. Since both the NNF and the NNI models have the same margins for both v_q and η_q, they can be compared using a likelihood ratio test (the NNI model, which is equivalent to independent models of telecommuting choice and frequency, is obtained by restricting the dependence parameter in the NNF model to zero, as discussed in Section 4.2.2). The chi-squared test statistic is 6.30, very strongly rejecting the null hypothesis of independence between the telecommuting choice and frequency equations at close to the 0.01 level of significance for one degree of freedom. Interestingly, the log-likelihood value at convergence for the classic textbook structure (see Lee, 1983) that assumes a normal-normal Gaussian (NNG) model structure is -5121.320, with a corresponding Kendall's measure of dependency of 0.10.[5] The likelihood ratio statistic for the test between the NNG and NNI models is only 1.84. Thus, one is unable to reject the null hypothesis of independence between telecommuting choice and frequency at the usual levels of significance used in hypothesis testing. The implication is clear. One can get inappropriate results regarding the dependency between two random variables just because of the imposition of a specific parametric form for the dependency. In the current empirical context, using the typical bivariate normal distributional assumption between the telecommuting choice and frequency equations provides the incorrect result that there is no statistically significant dependency, while using the Frank copula indicates the clear presence of dependency and provides a statistically superior fit. Of course, due to sample selection issues, incorrect results about the presence and nature of dependency in the telecommuting choice/frequency model system can, and in general will, lead to inconsistent estimates of the telecommuting frequency model parameters. Thus, one has to empirically test alternative profiles of dependency and select the most appropriate one.[6]

[5] For the Gaussian copula, $\tau = \frac{2}{\pi}\arcsin(\theta)$. Independence is attained in Gaussian's copula when $\theta = 0$.

[6] The Frank's copula allows a stronger central clustering of data points and lesser clustering at the edges relative to the Gaussian copula. In the current empirical context, this means that individuals are likely to be clustered around the medium-medium levels of the two-dimensional telecommuting propensity-telecommuting frequency inclination spectrum, and less so at the low-low end or the high-high end of the spectrum.

5.3. Estimation Results

To conserve on space, we only present the results for the best NNF model.[7] The results are presented in Table 1. The highly significant negative constant in the binary telecommuting choice model is simply a reflection of the large share of non-telecommuters in the sample. The thresholds at the top of Table 1 for the ordered-response frequency model do not have any substantive interpretations. They simply serve the purpose of mapping the latent propensity into the observed frequency levels. Unlike the binary telecommuting choice model, we did not include a separate constant term in the ordered-response telecommuting frequency model because of identification consideration (note that we have already included four threshold parameters in the model). Also note that, for dummy exogenous variables, the category that does not appear in the table is the base category. This base category is explicitly identified in the text discussion below.

5.3.1. Individual Demographics

The first set of exogenous variables in Table 1 corresponds to individual demographics. The effects of the female-related variables indicate that women are less likely to telecommute compared to men if they reside in households with no children. However, when there are children in the household, the tendency of women not to telecommute (relative to men) reduces, though they are still likely to telecommute less than men. These results are consistent with the findings in the literature (see, for instance, Mannering and Mokhtarian, 1995, Drucker and Khattak, 2000, and Popuri and Bhat, 2003). As indicated by Tang *et al.* (2008), the lower telecommuting propensity of women relative to men may be because men occupy jobs with "more autonomy and bargaining power", as well as jobs that need telecommunications expertise. The higher telecommuting propensity of women in households with children (relative to women in children in households with no children) is presumably because of child-care responsibilities. The age-related effects suggest a lower propensity among young adults less than 30 years of age (relative to their older peers) to telecommute and telecommute frequently, perhaps because older, experienced, employees are more able to exercise personal choices regarding work arrangements (see Mokhtarian and Meenakshisundaram, 2002 and Walls *et al.,* 2007 for similar results). Education is clearly a very important factor that positively influences the choice of

[7] The estimates from the other copula models and the independent model were, as one would expect, different from those obtained from the NNF model. Further, the standard errors of the telecommuting frequency model estimates were, in general, smaller than those from the other models, indicating efficiency benefits as well from using the NNF structure.

telecommuting and the frequency of telecommuting, another recurring finding in the literature (the base category for the education variables in the table corresponds to an education level below a bachelor's degree). Finally, the results show a positive propensity to telecommute among employees with a driver's license. This is a result also obtained in Drucker and Khattak (2000), but needs further exploration to analyze the underlying reasons.

5.3.2. Work-Related and Occupation Characteristics

In the category of individual work-related characteristics, full-time employed individuals (≥30 hours per week) are more likely to have a telecommuting arrangement than those working part-time (<30 hours per week). It may be argued that employers are in general less willing to allow part-time employees to telecommute (because these individuals are already showing up to work only partly in the week), explaining the positive effect of full-time employment on telecommuting propensity. The result here is at odds with Tang *et al.* (2008) and Popuri and Bhat (2003), who find that part-time employees are more likely to telecommute than full-time employees. These authors suggest that individuals who are motivated by familial/other responsibilities to work part-time may also seek jobs that have flexible arrangements such as telecommuting. However, some of these effects are perhaps being captured by the work flexibility and occupation characteristics in our current study, while these other studies did not control for work flexibility or occupation. However, among those who telecommute, the results in Table 1 reveal that full-time employed individuals telecommute less frequently than part-time employed individuals. It is possible that full-time employed individuals have more obligations to be at work frequently, leading to the negative effect of full-time employment on the frequency dimension (see also Tang *et al.*, 2008 and Yeraguntla and Bhat, 2005).

Individuals with flexible work schedules are more likely (than individuals with no work schedule flexibility) to telecommute and telecommute frequently. One would anticipate that individuals who want work flexibility will look for jobs that provide them both temporal flexibility (as captured in the work schedule flexibility variables) as well as spatial flexibility (*i.e.,* telecommuting options). Thus, the positive association between work schedule flexibility and telecommuting propensity/frequency is to be expected. The propensity of telecommuting also increases with an increase in the number of jobs, presumably a reflection of trying to manage time more efficiently by working at home and saving work-related travel time to multiple work locations.

An important empirical contribution of the current study is the variety of occupation types incorporated in the models. The base for introducing the occupation dummy variables in our specification includes manufacturing, transportation, retail and other occupations (for ease, we will refer to the base category as MATRE). We chose these categories as the base since it is quite likely that those in these occupations will need to travel everyday to their work location. The results in Table 1 indicate statistically significant differences among individuals in different occupations in their telecommuting propensity and frequency. Workers in the communications area are more likely to telecommute and to do so frequently relative to those in the MATRE category. Further, employees in service-related occupations, in general, also have a higher telecommuting propensity than the MATRE occupation category, sometimes also reinforced by higher telecommuting frequency. The only exceptions are for employees in educational services, and health care or social assistance. This is indeed quite expected, since the jobs of workers in these latter two service professions naturally require face-to-face interactions with students and those who need health care/social assistance, respectively. But among those who are able to telecommute in these two professions, the frequency of telecommuting is higher than in the MATRE occupations. Finally, individuals working for the government are the least likely to telecommute. Individuals working for the government may need to be involved in quick coordination/organization responses in uncertain conditions, and are likely to participate in interactive knowledge and information based activities. Such work-related characteristics and activities are likely to be facilitated by face-to-face contact and interactions with colleagues and others (see Storper and Venables, 2004). Further, it is possible that government employees may not be able to work from home because of the need to work with sensitive information that can be accessed only in their secure work location environment.

5.3.3. Household Demographics

The results of household demographics show that individuals in households with more workers have a higher preference to adopt telecommuting and to telecommute frequently relative to households with fewer workers. As household income increases, individuals are significantly more likely to telecommute, a finding that is consistent with many previous studies (see Mannering and Mokhtarian, 1995, Bernardino and Ben-Akiva, 1996, and Bhat and Popuri, 2003). This may be attributed to more control over work location-related and work timing choices as one “climbs the work ladder”. The effect of the “number of household vehicles” variable is interesting, and suggests a lower telecommuting propensity and frequency among individuals in households with more vehicles. This result differs from those of Drucker and Khattak (2000) and Popuri and Bhat (2003). However, it may simply be a reflection of individuals who telecommute choosing to own fewer vehicles. Future studies should examine the potential jointness in the choices of the number of vehicles and telecommuting. Finally, the availability of a fax machine at home increases both the telecommuting choice and the frequency propensity, presumably due to better access to telecommunications at home.

5.3.4. Commute-Trip/Work Location Characteristics

A general caveat regarding the effect of commute trip/work location characteristics on telecommuting choice and frequency. All of these attributes are potentially endogenous to the choice/frequency of telecommuting, although almost all earlier studies, like the current study, have considered such variables as exogenous to the choice of telecommuting. For instance, commute distance or commuting time has been used as an exogenous variable in several earlier studies. The implicit assumption is that individuals decide on their work location (and, therefore, the job they want), as well as their residential location, before evaluating telecommuting options. However, it is certainly possible that individuals first decide that they want to telecommute, and then find a job (and a corresponding work location) that satisfies their needs. A person may then deliberately choose to locate herself/himself quite far away from the work place, because s/he need not go in to work everyday. More broadly speaking, an argument could be made that all work-related decisions (including telecommuting, work schedule flexibility, full time versus part-time, and perhaps even occupation type) and residential location choice decisions should be modeled in one single joint model system that also implicitly determines the choice of a work location and commute trip attributes.[8] But, in the process of practical modeling, the analyst needs to make informed judgments and assumptions regarding what may be considered exogenous variables. We suggest that an area of future research should be to examine the various choices surrounding work characteristics, residential location decisions, and telecommuting, to provide meaningful guidance regarding which

[8] In their review, Walls *et al.* (2007) raise the endogeneity issue in the specific context of Popuri and Bhat's (2003) use of mode choice to work variables (and availability of a fax machine/multiple phone lines at home) to explain telecommuting choice. However, we are surprised that they should choose to target a single paper and only selected variables to discuss the endogeneity issue. We submit that endogeneity of variables is a potential problem in all published works in the area of telecommuting choice and frequency.

variables may be considered more endogenous than others. The results of such a research pursuit can be gainfully employed in the specification of telecommuting choice/frequency models. In the rest of this section, we discuss the effects of commute trip/work-related characteristics on telecommuting choice/frequency, though the caveat just discussed should be kept in mind.

The first variable under commute trip/work-related characteristics in Table 1 corresponds to the direct one-way home-to-work commute distance of employees. As expected, individuals whose (one-way) commute distance is longer than 25 miles are more likely to telecommute (and telecommute with high frequency) compared to individuals with a (one-way) commute distance less than 25 miles (see also Mokhtarian and Meenakshisundaram, 2002). The next variable suggests that the frequency of telecommuting decreases if the roadway type normally traveled on to work includes an expressway, probably due to less-stressful driving conditions on expressways than on other arterial streets. The positive influence of non-car modes of travel (walk/bicycle/transit) to work is consistent with Tang *et al.*'s (2008) finding that individuals with pro-bike and pro-transit views have a higher propensity to telecommute relative to others. One explanation is that individuals who bicycle/walk/use transit to reach work are environmentally conscious, and see telecommuting as another means to reduce auto travel. Next, vehicle availability for work positively influences the frequency of telecommuting. Although this effect is consistent with most of the literature in the field, the reason for this positive relationship needs further exploration in future studies. Finally, individuals who make several non-work trips on the workday are more likely to telecommute, while those who have to pay to park at work have a higher frequency of telecommuting than those who do not have to pay to park (we also examined the effects of the amount of any parking costs and tolls usually paid on the commute trip, but both of these policy-relevant variables did not turn out to be statistically significant even at the 0.15 level of significance).

5.4. Elasticity Effects

The parameters on the exogenous variables in Table 1 do not directly provide a sense of the absolute magnitude of the effects of variables. One can compute the aggregate "elasticity" effects of each dummy exogenous variable (such as the "female" dummy variable) or ordinal exogenous variable (such as "number of household workers") in Table 1 to obtain an estimate of these magnitude effects. These aggregate elasticity effects can be computed for each ordinal level of telecommuting frequency. However, this procedure results in a rather large set of numbers. To simplify the presentation, we have chosen to assign cardinal values to each of the ordinal levels of telecommuting frequency, and then compute the elasticity effects of exogenous variables on the expected total number of days per month of telecommuting. The cardinal value assignments for the telecommuting ordinal frequency levels in the model are as follows: (1) telecommuting once a year (k = 1 in the notation of Section 4.1): 1/12 = 0.083 telecommuting days per month, (2) a few times a year (k = 2): 4/12 = 0.333 telecommuting days per month, (3) once a month or more (k = 3): 12/12 = 1 telecommuting day per month, (4) once a week or more (k = 4): 4*12/12 = 4 telecommuting days per month, (5) almost every day (k = 5): 22* 12/12 = 22 telecommuting days per month. With these assignments, the expected value of the number of telecommuting days per month for individual q (d_q) using Equation (3) in Section 4.1 is:

$$E(d_q) = \sum_{k=1}^{5} c_k \times \Pr[t_q = 1, s_q = k], \tag{15}$$

where c_k is the cardinal value assignment corresponding to telecommuting ordinal frequency level k. Note that the expected value above is a function of variables in both the vectors x_q and z_q (see Equation (3)). If there are common variables in x_q and z_q (such as age, employment level, and occupation characteristics in our empirical specification), these variables will impact the expected value of the number of telecommuting days per month both through the telecommuting choice binary model and the telecommuting frequency ordered response model.

To compute the aggregate-level “elasticity” effect of a dummy exogenous variable, we change the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. We then sum the shifts in the expected aggregate number of telecommuting days per month in the two subsamples after reversing the sign of the shifts in the second subsample, and compute the effective percentage change in the expected total number of telecommuting days per month across all individuals in the sample due to a change in the dummy variable from 0 to 1. To compute the aggregate level elasticity effect of an ordinal variable, we increase the value of the variable by 1 and compute the percentage change in the expected total number of telecommuting days per month across all individuals in the sample.

Table 2 provides the elasticity effects. The first entry in the table indicates that the number of telecommuting days per month for women with no children is, on average, about 11.8% less than the number of telecommuting days per month for men. On the other hand, the second entry shows that the telecommuting days per month for women with children is, on average, about 3.73% less than the number of telecommuting days per month for men. Other entries may be similarly interpreted. The results reveal that employees with flexible work schedules (especially if they are fully flexible) and employees working in real-estate, rental or leasing services are substantially more likely to telecommute frequently than those with no work schedule flexibility and employees in the MATRE (manufacturing, transportation, retail and other) occupation category, respectively. These variables have the highest impacts on the number of days of telecommuting per month. Other variables with substantial positive impacts include one-way commute distance, having to pay to park at work (relative to free parking at work), being in occupations related to management of companies or enterprises (relative to being in the MATRE occupation category), holding a graduate degree (relative to an education level lower than an undergraduate degree), using a non-motorized mode to get to work (relative to the use of a motorized mode to get to work), and having a fax machine at home. For all the variables identified above (except for the "pay to park at work" variable), the high positive impact is because these variables positively influence both the choice and frequency model components of telecommuting (see Table 1). Further, the magnitudes of the estimated parameters on these variables in each model component are quite high relative to the estimated parameters on other variables. For the "pay to park at work" variable, the net effect on number of telecommuting days per month is quite substantial (even though it does not affect the telecommuting choice component) because it has a high positive effect in the frequency component of the model system. Finally, the results show that the number of non-work trips on the work-day, being a female with children (relative to being a male), being a full-time employee (relative to being a part-time employee), and the number of jobs held have a much smaller impact on the number of telecommuting days per month relative to other explanatory variables.

6. CONCLUSIONS AND IMPLICATIONS

Commute-related vehicular travel is a major cause of peak period traffic congestion in urban areas, as well as contributes significantly to transportation sector-based greenhouse gas emissions. This has led researchers and planning agencies to explore travel demand management strategies to reduce commute-related vehicular travel. One such strategy is to reduce commute person miles of travel (and, therefore, vehicle miles of travel) through the promotion of telecommuting. In the current paper, we contribute to the existing telecommuting literature by jointly analyzing the choice and the frequency of telecommuting, using data from the 2007/2008 Chicago Regional Household Travel Inventory (CRHTI). The CRHTI is a particularly rich data source for our analysis because it provides a sample of more than 1500 telecommuters working in several different occupations and job classifications, in a field that has been dogged by the paucity of "comprehensive data sets [that] address telecommuting behavior across a wide range of individuals holding jobs with different employers" (Walls *et al*., 2007). Further, we are able to reduce the ambiguity in the difference between home-based telecommuting and the operation of a home-based business by removing individuals who indicated that they were self-employed and worked primarily from home. Thus, the sample of over 1500 telecommuters is in a sense "true" telecommuters rather than an unknown amalgam of telecommuters and home-based workers. Overall, the richness of the data and the clarity in the characterization of telecommuters allow us to incorporate a variety of variables, including individual demographics, work-related and occupation characteristics, household demographics, and commute trip/work location characteristics to examine the determinants of telecommuting choice and frequency. Another important feature of the CRHTI is that it offers a very current perspective of the process driving telecommuting decisions. In an age of rapid developments and innovations in communication technologies, it only stands to reason that the forces shaping and influencing the ability/choice to telecommute evolve quickly, underscoring the need to undertake telecommuting analysis using recently collected telecommuting patterns. Finally, the current paper examines the telecommuting choice and frequency dimensions in a unified framework using a flexible copula-based approach that overcomes restrictive *a priori* dependency assumptions imposed on the choice and frequency decisions (such as assuming independence or a bivariate normal distribution assumption). To our knowledge, this is the first formulation and application in the econometric literature of a copula approach for the case of a binary self-selection mechanism with an ordered-response outcome.

The empirical results indicate the important effects of several demographic and work-related variables. First, the results clearly indicate that telecommuting choice and the frequency of telecommuting may be governed by quite different underlying behavioral processes rather than being governed by a single behavioral process. In particular, the determinant factors of choice and frequency can be different. Thus, according to our results, gender and presence of children in the household impacts the telecommuting choice decision, but not the frequency dimension. Further, a factor that has a particular direction of effect on telecommuting choice may have the opposite effect on frequency. For instance, our results indicate that full-time employment is positively associated with the choice of telecommuting, but negatively associated with the frequency of telecommuting. Second, unobserved factors that predispose an individual to choose to telecommute also increase the individual's telecommuting frequency. But the results also emphasize that pre-imposing a specific dependency structure between the telecommuting choice and frequency decisions can lead to inappropriate conclusions regarding the presence and extent of dependency. In the current paper, we found that using the typical bivariate normal distribution assumption between the telecommuting choice and frequency equations provides the incorrect conclusion of no statistically significant dependency, while using the Frank copula indicates the clear presence of dependency. Further, the influence of exogenous variables from models assuming different dependency structures are different from one another, and the standard errors of the telecommuting frequency model estimates were, in general, smaller from the best-fit Frank copula structure than those from other structures. Overall, one risks the danger of incorrect conclusions regarding dependency in the telecommuting choice and frequency behavioral processes, as well as inconsistent and inefficient parameter estimates, by imposing incorrect dependency structures. It behooves the analyst to empirically test alternative profiles of dependency (*i.e.*, copulas) and select the most appropriate one. Third, work schedule flexibility and occupation type are important determinants of telecommuting choice and frequency. In particular, workers whose schedules are fully flexible and who are in the real estate, rental, or leasing occupations are much more likely to telecommute than their peers. Fourth, several factors related to the commute trip and work location influence telecommuting choice and frequency. For instance, our results suggest that individuals who have to pay to park at the work place are more frequent telecommuters than those who do not have to pay to park. Also, those who usually bicycle, walk, or use transit to reach their work place are also more likely to telecommute. Of course, these commute mode choice decisions may be related to built environment attributes at the residence end and/or at the work end, so they may be proxying for built environment effects. Future studies would benefit from the consideration of a comprehensive set of built environment variables, in addition to the many categories of variables included here. As indicated earlier, it would also be helpful to examine the many choices surrounding work characteristics, residential location decisions, and telecommuting to provide meaningful guidance regarding which variables may be considered endogenous and which exogenous in telecommuting choice/frequency modeling.

The empirical results have implications for transportation planning analysis, especially because of the projected changes in demographic and employment-related variables (such as age, households with and without children, and work characteristics) in the U.S. population. The models estimated in this paper can be used to assess the impacts of these changes. The model results can also be used to target specific employee groups, and employer groups based on industry sector, to increase the extent of telecommuting. Companies can use the results to predict how many employees would show up to work on any given workday, which may help plan for office space and parking space. Companies and planning agencies can also evaluate the effects of imposing parking fees at the work place. Finally, the predictions from the model system developed in this paper can feed into larger-scale activity-based travel demand modeling systems that use work-related decisions of individuals as a "peg" around which to schedule other activities and travel.

ACKNOWLEDGMENTS

The authors would like to acknowledge Renee Alsup for assistance with identifying, compiling, and providing earlier papers on telecommuting.

REFERENCES

Bagley, M.N., and P.L. Mokhtarian (1997) Analyzing the Preference for Non-exclusive Forms of Telecommuting: Modeling and Policy Implications. *Transportation*, 24(3), 203-226.

Bernardino, A.T., and Ben-Akiva, M., and I. Salomon (1993) Stated Preference Approach to Modeling the Adoption of Telecommuting. *Transportation Research Record* 1413, TRB, National Research Council, Washington, D.C., 22-30.

Bernardino, A.T., M. Ben-Akiva (1996) Modeling the Adoption of Telecommuting: Comprehensive framework. *Transportation Research Record*, 1552, 161-170.

Bhat, C.R., and N. Eluru (2009) A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling. *Transportation Research Part B*, 43(7), 749-765.

Chicago Regional Household Travel Inventory (2009). Available at: http://www.cmap.illinois.gov/TravelTrackerData.aspx

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.

Drucker, J, and A.J. Khattak (2000) Propensity to Work from Home: Modeling Results from the 1995 Nationwide Personal Transportation Survey. *Transportation Research Record 1706*, TRB, National Research Council, Washington, D.C., 108-117.

Edwards, P. and S. Edwards (1985) *Working from Home – Everything You Need to Know About Living and Working Under the Same Roof.* J.P. Tarcher, Los Angeles, California.

Embrechts, P., McNeil, A. J., Straumann, D. (2002) Correlation and dependence in risk management: Properties and pitfalls. In M. Dempster (ed.) *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, 176-223.

EPA (2006) Greenhouse Gas Emissions from the U.S. Transportation Sector 1990–2003, Office of Transportation and Air Quality (6401A) EPA 420 R 06 003 March 2006. Available at: www.epa.gov/OMS/climate/420r06003.pdf.

EPA (2007) Climate Change: Human-Related Sources and Sinks of Carbon Dioxide. Environmental Protection Agency and U.S. Department of Energy, 2007. Available at: http://www.epa.gov/climatechange/emissions/co2_human.html.

Farlie, D. J. G. (1960) The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* 47(3-4), 307-323.

Frank, M. J. (1979) On the simultaneous associativity of F(x, y) and x + y - F(x, y). *Aequationes Mathematicae* 19(1), 194-226.

General Services Administration (GSA) (2006) *Telework Technology Cost Study*. US General Services Administration. www.gsa.gov/teleworklibrary.

Genius, M., Strazzera, E. (2008) Applying the copula approach to sample selection modeling. *Applied Economics* 40(11), 1443-1455.

Gordon, G.E. (1998) The Dilemmas of Telework: Technology vs. tradition. In *Telework: Present Situation and Future Development of a New Form of Work organization*, edited by W.B. Korte, W.J. Steinle and S. Robinson, North-Holland, 1988.

Gumbel, E.J. (1960) Bivariate Exponential Distributions. *Journal of the American Statistical Association*, 55(292), 698-707.

Heckman, J. J. (1979) Sample selection bias as a specification error, *Econometrica*, 47(1), 153-161.

Jala International, Inc. (2000) *Telecommuting Forecast for the United States of America*. 2000. http://www.jala.com/ustcforecast.htm.

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.

Lee LF (1983) Generalized Econometric Models with Selectivity. *Econometrica* 51: 507-512.

Mannering, J.S., and P.L. Mokhtarian (1995) Modeling the Choice of Telecommuting Frequency in California: An Exploratory Analysis. *Technological Forecasting and Social Change* 49(1), 49-73.

Mokhtarian, P.L., and R. Meenakshisundaram (2002) Patterns of Telecommuting Engagement and Frequency. *Prometheus*, Vol. 20, No. 1, pp. 21-37.

Mokhtarian, P.L. and I. Salomon (1996a) Modeling the Choice of Telecommuting: 2. A Case of the Preferred Impossible Alternative. *Environment and Planning A*, Vol. 28, pp. 1859-1876.

Mokhtarian, P. L. and I. Salomon (1996b) Modeling the Choice of Telecommuting: 3. Identifying the Choice Set and Estimating Binary Choice Models for Technology-based Alternatives. *Environment and Planning* A, Vol. 28, pp. 1877-1894.

Mokhtarian, P. L., and I. Salomon (1997) Modeling the Desire to Telecommute: The Importance of Attitudinal Factors in Behavioral Models. *Transportation Research Part A*, 31(1), 35-50.

Morgenstern, D. (1956) Einfache beispiele zweidimensionaler verteilungen. *Mitteilingsblatt fur Mathematische Statistik* 8(3), 234-235.

NAS (2008) *What You Need to Know About Energy?* National Academy of Sciences, The National Academies.

Nelsen, R. B. (2006) An Introduction to Copulas (2nd ed.), Springer-Verlag, New York, 2006.

Nilles, J.M. (1988) Traffic Reductions By Telecommuting: A Status Review and Selected Bibliography. *Transportation Research Part A*, 22(4), 301-317.

NuStats (2008) Cicago Regional Household Travel Inventory Draft Final Report. Prepared for Chicago Metropolitan Agency for Planning. Available at: http://www.cmap.illinois.gov/TravelTrackerData.aspx.

Olszewski, P., and P.L. Mokhtarian (1994) Telecommuting Frequency and Impacts for State of California Employees. *Technological Forecasting and Social Change*, Vol. 45, pp. 275-286.

Popuri, Y.D., and C.R. Bhat (2003) On Modeling Choice and Frequency of Home-Based Telecommuting. *Transportation Research Record 1858*, TRB, National Research Council, Washington, D.C., 55-60.

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

Schrank, D., and T. Lomax (2009) The 2009 Urban Mobility Report. Texas Transportation Institute, The Texas A&M University System. July 2009.

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

Storper, M.,and A.J. Venables (2004) Buzz: Face-to Face Contact and the Urban Economy. *Journal of Economic Geography*, 4(4), 351-370.

Sullivan, M.A., H.S. Mahmassani and J-R. Yen (1993) Choice Model of Employee Participation in Telecommuting Under a Cost-Neutral Scenario. *Transportation Research Record 1413*, TRB, National Research Council, Washington, D.C., 31-41.

Tang, W., Mokhtarian, P., and S. Handy (2008) The Role of Neighborhood Characteristics in the Adoption and Frequency of Working at Home: Empirical Evidence from Northern California. *Institute of Transportation Studies.* http://repositories.cdlib.org/itsdavis/UCD-ITS-RR-08-21

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

Walls, M., and Safirova, E. (2004) A Review of the Literature on Telecommuting and its Implications for Vehicle Travel and Emissions. RFF Discussion Paper 04-44, September.

Walls, M., Safirova, E., and Y. Jiang (2007) What Drives Telecommuting? Relative Impact of Worker Demographics, Employer Characteristics, and Job Types. *Transportation Research Record 2010,* TRB, National Research Council, Washington, D.C., 111-120.

World at Work. Telework Trendlines (2009) WorldatWork, The Total Rewards Association, 2009. http://www.workingfromanywhere.org/news/Trendlines_2009.pdf.

Yen, J-R. and H.S. Mahmassani (1994) The Telecommuting Adoption Process: Conceptual Framework and Model Development. Report 60055-1, prepared for the Southwest Region University Transportation Center, Center for Transportation Research, The University of Texas at Austin, Austin, Texas.

Yeraguntla, A. and Bhat, C. R (2005) Classification taxonomy and empirical analysis of work arrangements. *Transportation Research Record 1926*, TRB, National Research Council, Washington, D.C., 233-241.

LIST OF TABLES

Table 1. Estimation Results for Employees' Telecommuting Choice and Frequency Models

Explanatory Variables	Joint Model - *NNF*			
	Telecommuting Choice Model		Telecommuting Frequency Model	
	Estimate	t-stat	Estimate	t-stat
Constant	-3.010	-21.33	-	-
Threshold 1	-	-	-1.403	-4.61
Threshold 2	-	-	-0.349	-1.18
Threshold 3	-	-	0.598	2.29
Threshold 4	-	-	1.885	8.45
Individual Demographics				
Female	-0.129	-3.15	-	-
Female with children	0.059	1.98	-	-
Age less than 30 years	-0.219	-3.32	-0.207	-1.82
Education: Bachelor's or Undergraduate degree	0.310	6.43	-	-
Education: Graduate degree	0.429	8.51	0.117	1.94
Driver license	0.326	2.93	-	-
Work-related and Occupation Characteristics				
Full-time employment (>30 hours/week)	0.272	5.07	-0.213	-2.52
Partially flexible	0.954	17.81	0.244	2.15
Fully flexible	1.687	28.02	0.681	5.28
Number of jobs	0.047	1.30	-	-
Occupation				
Communications	0.474	6.51	0.255	2.43
Service-based				
Service – Finance and insurance	0.227	3.38	-	-
Service – Real estate, rental, or leasing	0.595	4.46	0.604	3.67
Service – Professional, scientific, or technical service	0.314	5.86	-	-
Service – Management of companies, or enterprises	0.278	2.56	0.218	1.45
Service – Arts, entertainment, or recreation	-	-	0.292	1.57
Service – Educational services	-0.087	-1.23	0.221	2.25
Service – Health care or social assistance	-0.107	-1.71	0.282	3.01
Government	-0.157	-2.04	-	-
Household Demographics				
Number of household workers	0.053	1.86	0.076	1.70
Household income between 75K-100K	0.286	5.32	-	-
Household income greater than 100K	0.462	9.66	-	-
Number of household vehicles	-0.050	-2.10	-0.051	-1.57
Fax at home	0.170	3.17	0.277	3.69
Commute-trip/Work location Characteristics				
One-way commute distance more than 25 miles	0.232	6.06	0.390	7.10
Commute trip made on an expressway	-	-	-0.247	-2.28
Walk/bike to work	0.196	2.35	0.267	2.30
Transit to work	0.209	4.35	-	-
Vehicle available for work	-	-	0.257	4.20
Number of non-work trips on the work-day	0.017	2.06	-	-
Pay to park at work?	-	-	0.474	2.19
Number of Observations	9624			
Dependency parameter estimate (t-stat)	2.086 (2.56)			
Log-likelihood at convergence	-5119.092			

Table 2. Elasticity Effects for Expected Number of Telecommuting Days per Month

Explanatory Variables	**Joint Model** ***NNF***
Individual Demographics	
Female with no children	-11.80
Female with children	-3.73
Age less than 30 years	-36.22
Education: Bachelor's or Undergraduate degree	28.80
Education: Graduate degree	54.26
Driver license	28.03
Work-related and Occupation Characteristics	
Full-time employment (>30 hours/week)	4.08
Partially flexible	106.29
Fully flexible	350.16
Number of jobs	4.31
Occupation	
Communications	87.66
Service-based	
Service – Finance and insurance	21.37
Service – Real estate, rental, or leasing	184.24
Service – Professional, scientific, or technical service	29.86
Service – Management of companies, or enterprises	57.60
Service – Arts, entertainment, or recreation	33.91
Service – Educational services	15.30
Service – Health care or social assistance	20.04
Government	-13.96
Household Demographics	
Number of household workers	13.48
Household income between 75K-100K	26.53
Household income greater than 100K	43.58
Number of household vehicles	-9.49
Fax at home	50.89
Commute-trip/Work location Characteristics	
One-way commute distance more than 25 miles	63.75
Commute trip made on an expressway	-23.61
Walk/bike to work	54.20
Transit to work	19.53
Vehicle available for work	28.10
Number of non-work trips on the work-day	1.55
Pay to park at work?	58.33