INVESTIGATING THE ROLE OF SOCIAL NETWORKS IN START TIME AND DURATION OF ACTIVITIES: A TRIVARIATE SIMULTANEOUS ECONOMETRIC MODEL

Khandker M. Nurul Habib. Department of Civil Engineering, University of Toronto, Canada, khandker.nurulhabib@utoronto.ca

Juan-Antonio Carrasco. Department of Civil Engineering, Universidad de Concepción, Chile, j.carrasco@udec.cl

ABSTRACT

In the context of improving our understanding and modeling capabilities of activity scheduling processes in travel behaviour, this paper explores the role of social networks in the start time and duration of social activities. The study is performed using a trivariate joint econometric model, which is capable of capturing the correlation among unobserved influential factors causing endogeneity of these three key decisions. The model captures not only the relevance of socio-demographic variables, but also of the social network dimension of with whom travellers perform social activities. A particular relevant case is the role of travel time to the social activities, which has a positive effect on longer durations and late start times, and which acts as a link between these two basic dimensions (start time and duration) of activity scheduling. The results confirm the relevance of the social context in the episode temporal characteristics, illustrating aspects that future activity-based travel demand models should incorporate these to be able to capture the socializing side of mobility decisions.

Keywords: social networks, duration, start time, social activities, activity-based approach

1. INTRODUCTION

During the last decade, travel demand models have experienced a number of improvements towards a better understanding of peoples' travel behaviour in urban areas. This trend is especially relevant in the context of activity-based models, which have moved towards operational models, explaining increasingly more complex questions about the role of transportation in quality of life (e.g., Roorda *et al*, 2008; Arentze and Timmermans, 2004). A key improvement of activity-based models is the better understanding and modeling of the role of activity scheduling processes in peoples' mobility, in which episode start time and duration constitute key elements. However, as we move forward on understanding these issues, the models' scope needs to be expanded from individual-based perspectives, to more social-based paradigms. It is very important because in this way itcan recognize the fact that a relevant portion of travel motivated by the need of meeting others.

In this context, the objective of this paper is investigating the role of the individuals' social context (social networks) in their activity scheduling decisions. More specifically, the aim of the research is to understanding the role of social networks ("with whom" people interacts) in the two critical scheduling decisions: start time and duration of social activity episodes. The focus on social activities was chosen as a paradigmatic activity type, although the analysis could be expanded without loosing generality.

The data that support this approach corresponds to the first wave of the Toronto Panel Survey (TAPS, 2002-2003; see Roorda and Miller, 2004), which used the instrument CHASE® to collect in detail the activities and scheduling processes of around 400 individuals in a week, identifying with whom these activities were performed with. This latter variable is used to construct the "elicited" respondent's personal network, gathering not only whether the interactions were made with friends of family members, but also richer covariates which capture the structure of the respondent's social network. The structure of the social networks is characterized in terms of size, density, contact variability, number of components, and centralization.

This paper expands a previous effort which also studied the role of "with whom" in start time, duration and social episodes (Habib *et al*., 2008). In that study, the authors investigated two separated bivariate models: one model having "with whom" and start time, and another one having "with whom" and duration as independent variables. Among the results provided by that study, one of the key findings was that, when the social dimension is incorporated explicitly in the start time models, the effect of travel time and travel distances became statistically insignificant. It suggested that it is "with whom" we socialize what would define the social activity scheduling process and not travel time or distance. However, the sequential structure of that analysis relying upon two separate models, was not able to differentiate whether travel distance did not have any relevance in start time choice or whether there was an endogenous effect between start time, duration and travel distance suppressing the systematic effect of travel distance in start time choice due to the with whom dimension. Bivariate model based analysis rightfully identified the insignificant systematic effects of travel distance on social activity scheduling, but overlooks the unobserved endogeneity that creates the correlation between start time, duration and with whom to participate the social activities.

This intriguing result is re-explored in this paper using now a joint trivariate econometric structure, which is not only able to capture the relevant covariate effects, but also the complex relationships between the three different dimensions. In fact, one of its main features is being able to capture not only the covariance relationships between the three dimensions, but also the unobserved endogenous relationships among them. In this way, the model allows the researchers not only to study the endogenous relationship between start time and duration, but also testing to which extend social networks are an endogenous component of the activity scheduling processes. In addition, the joint structure provides good insights that activity scheduling models should have to include the role of social networks explicitly.

Next section of the paper presents the methods employed in this paper. Section three discusses the data structure, with a particular emphasis on the way the social network variables were built. Finally, the empirical results and conclusions are presented.

2. METHODS

2.1. Overview

The methods employed in this paper are designed to capture explicitly the joint relationship between with whom the social activities are performed, the duration of these encounters, and its start time. The model uses an econometric formulation, which includes a discrete choice multinomial logit formulation for the “with whom” choice, and lognormal hazard models for durations and start time choices. Hence, the model can be conceptualized as a joint discrete-continuous-continuous choice model, which does not assume a priori causality among the three dimensions.

The “with whom” decision is modelled as a discrete choice decision, classified into four alternatives:

1. Socialize with family members together with household members
2. Socialize with friends together with household members
3. Socialize with family members without any household member
4. Socialize with friends without any household member

On the other hand, start time and duration decisions for any activity can be modelled as discrete, ordered or continuous decisions. Discretization of time intervals is necessary for both discrete and ordered decision structures. Computationally, discretization of time to model start time and duration may seem easier, but it always leaves scope for aggregation bias in the model. Even if a fine level of temporal resolution is considered (e.g., 5 minutes or 10 minutes), it is difficult to conclusively establish what exact interval size should be used. The times near the boundaries of adjacent discrete intervals are assumed to be distinct choice alternatives, an assumption which is often counter-intuitive. Whatever finer intervals are assumed, alternative adjacent times are nearly identical from the perspective of a commuter but are considered to be distinct choices in a discrete or ordered model for the start time and duration model. As a result, any temporally based policies (e.g. flexible office hour, peak-period congestion pricing, etc.) used in forecasting/prediction must also be applied within the time periods defined during model estimation.

In order to avoid the issues of discretizing time for start time or duration model, Bhat (1996, 1998b) proposed modelling structure as a flexible parameter hazard model combined with discrete choice decision. The multiple duration hazard model proposed by Bhat (1996) apparently does not discretize time but requires ordering of time intervals anyway. The MNL-OGEV model described in Bhat (1998a) overcomes the IIA assumptions between alternative and consecutive time intervals by a GEV structure, but still the time itself is to be ordered based on the researcher's discretion. However, Bhat (1998b) has presented a model of continuous time specification by modelling time in a linear regression model. Time is naturally a continuous variable, given that it is modelled for start time or duration of any activity. If the time is modelled in activity-travel decisions as continuous variables, for application or presentation purposes researchers always have the freedom to discretize it whenever and by whatever means necessary. This is not true, however, the other way around. If the model estimation process requires discretization of time, the researcher loses this freedom. Moreover, it is necessary to develop a modelling framework which recognizes the continuous nature of time for start time and duration of activities with other discrete attribute decisions.

In this paper, start time and duration of social activities are modelled as continuous decisions together with discrete "with whom" decisions. In this way, the modelling structure takes the form of a trivariate discrete-continuous-continuous decision structure. Within this structure, the discrete choice component is modelled as a multinomial logit model and the continuous components are modelled as continuous accelerated life hazard models. The accelerated life hazard modelling approach recognizes the dynamics of start time and duration decisions within the model formulation. The formulation derives correlations between these three decisions without considering any specific sequence of decisions. In addition, the model structure is capable to capture the endogenous role of social duration both in "with whom" and start time choice decisions. The next section describes the structure of the model and the estimation process.

2.2 Structure and Estimation

The paper uses a trivariate econometric model formulation that connects three different types of choices (discrete, continuous and continuous). The model formulation captures the correlations between unobserved factors affecting the choice situations under consideration. The econometric formulation is based on Habib (2010). Let us consider for any individual i, that S represents the logarithm of continuous start time, (i.e., time from the reference time to the point of starting the social activity), and D represents the logarithm of continuous duration of the social activity. According to Keifer (1988), the accelerated life hazard model specification for start time and duration can be expressed as functions of covariates (x and z respectively) multiplied with their corresponding parameters (β and γ, respectively) and additive error terms (ξ and ψ, respectively). On the other hand, let us assume that the utility of choosing a particular kind of people w for socializing is expressed by Uw. The utility of with whom choice is composed of a systematic utility function (Vw) and an independent and identically Gumbel distributed (across the outcomes and individuals) error term, εw, with zero location parameter and unit scale parameter. In the case of the error terms (ξ and ψ) of the hazard models, let us assume that they are normally distributed with zero means and with σS and σD variances consecutively. The equations for start time, utility of with whom choice and duration can then be expressed as:

$$S = \beta x + \xi \quad (1)$$

$$U_w = V_w + \varepsilon_w \quad (2)$$

$$D = \gamma z + \psi \quad (3)$$

where

$$\xi \sim N(0, \sigma_S^2)$$

$$\varepsilon \sim IID\ Gumbel(0,1)$$

$$\psi \sim N(0, \sigma_D^2)$$

Beginning with the utility function of with whom, a multinomial with whom choice situation can be expressed in terms of dichotomous situations (Munizaga et al. 2006). Any individual kind of person w is chosen if their utility is greater than the second maximum utility of all other alternatives:

$$U_w > \max_{z=1,2,.....I;\, z \neq w} \left(U_z\right) \quad (4)$$

This can be further decomposed as:

$$V_w > \max_{z=1,2,.....I;\, w \neq z} \left(U_z\right) - \varepsilon_w \quad (5)$$

This can be re-written in terms of a modified variable, V_w^*, as:

$$V_w^* \equiv \max_{z=1,2,.....I;\, z \neq w} \left(U_z\right) - \varepsilon_w \quad (6)$$

Here the expression refers to the condition that the individual kind of people w, is chosen if and only if $V_w > V_w^*$. According to McFadden (1973), in combining the above expression with the distributional assumption ε, the implied marginal distribution of V_w^* can be written as:

$$P(V_w^* < V_w) = F(V_w) = \frac{exp(V_w)}{exp(V_w) + \sum_{z \neq w} exp(V_z)} \quad (7)$$

On the other hand, the continuous time hazard models are primarily concerned with the time until the event terminates. For the empirical application in this paper, the time until one starts the social activity is the topics of interest. The time until one starts the social activity can be referred to as the pre-social event end time. The basic formula describing event termination in hazard models is the hazard rate, *λ(t)*, which is the conditional probability of event termination occurring between time, *t*, and *(t+dt)*, given that the event has not terminated before time, *t*. The hazard rate can be expressed mathematically as: $\lambda(t) = f(t)/(1 - F(t))$, where *F(t)* refers to the cumulative probability distribution function and *f(t)* is the corresponding probability distribution function. Similarly, the survival function *S(t)*, which defines the probability that the event's duration will be greater than or equal to time, *t*, is defined as $S(t) = 1 - F(t)$. Combining the definitions of *λ(t)* and *S(t)*, the hazard function can also be expressed as: $\lambda(t) = f(t)/S(t)$. In the cases of start time and duration of the social activity, we can simply assume specific distributions of *f(t)* and directly estimate the structural parameters of the assumed distributions. (Note that here *t* refers to time in general; in the cases of start time and duration we can use t_s and t_D, respectively, for identification).

However, as we are interested in incorporating covariates in the hazard model, there are two common specifications available: the proportional hazard and accelerated life hazard models. The proportional hazard model assumes that covariates modify the hazard function directly by having a multiplicative effect; the hazard rate is effectively decomposed into one term dependent on time, and another one dependent only on the covariates (Hensher and Mannering, 1994). On the other hand, the accelerated hazard model assumes that the covariates rescale or accelerate the time directly in the baseline survivor function; in this case, the hazard rate varies over time as it is accelerated or decelerated by the covariates. The accelerated life hazard specification is more attractive in our case because one can expect considerable dynamics in the hazard rate influenced by different covariates; the proportional hazard formulation, on the other hand, may not be realistic in such situations (Lee and Timmermans, 2007). Then, assuming an exponential covariates functional form $exp(\beta x)$ and $exp(\gamma z)$ for start time and duration, respectively, the accelerated hazard models for start time and duration social activity can be expressed, (where the subscript $_0$ denotes the baseline), as:

$$
\begin{aligned}
&S(t_S \mid \beta x) = S_0\left[t_S \exp(\beta x)\right] \text{ and } S(t_D \mid \gamma z) = S_0\left[t_D \exp(\gamma z)\right] \\
&\lambda(t_S) = \lambda_0\left[t_S \exp(\beta x)\right]\exp(\beta x) \text{ and } \lambda(t_D) = \lambda_0\left[t_D \exp(\gamma z)\right]\exp(\gamma z)
\end{aligned}
\tag{8}
$$

As explained by Kiefer (1988), these formulations can be exploited to define the accelerated life hazard model in the form of linear specifications as mentioned in Equations (1) and (3) where the assumed distribution of the error terms, *ξ* and *ψ*, defines the form of the corresponding accelerated time hazard model. If we assume that *ξ* and *ψ* are normally distributed, the probability density functions take on the following general form of log-normal accelerated life hazard models:

$$
\begin{aligned}
f(t_S) &= \frac{1}{t_S \sigma_S \sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma_S^2}\left(\ln(t_S) - \beta x\right)\right] = \frac{1}{t_S \sigma_S} \phi\left(\frac{\ln(t_S) - \beta x}{\sigma_s}\right) \\
f(t_D) &= \frac{1}{t_D \sigma_D \sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma_D^2}\left(\ln(t_D) - \gamma z\right)\right] == \frac{1}{t_D \sigma_D} \phi\left(\frac{\ln(t_D) - \gamma z}{\sigma_D}\right)
\end{aligned}
\tag{9}
$$

The hazard model becomes:

$$
\begin{aligned}
\lambda(t_S) &= \frac{1}{t_S \sigma_S} \phi\left(\frac{\ln(t_S) - \beta x}{\sigma_s}\right)\left(1 - \Phi\left(\frac{\ln(t_S) - \beta x}{\sigma_s}\right)\right)^{-1} \\
\lambda(t_D) &= \frac{1}{t_D \sigma_D} \phi\left(\frac{\ln(t_D) - \gamma z}{\sigma_D}\right)\left(1 - \Phi\left(\frac{\ln(t_D) - \gamma z}{\sigma_D}\right)\right)^{-1}
\end{aligned}
\tag{10}
$$

This simplifies the formulation of joint likelihood function for the three decisions. The likelihood function of the joint model is of a closed form and can be estimated by using conventional maximum likelihood estimation techniques. The model formulation exploites Lee's (1983) transformation techniques and the loglikelihood function becomes (for details of the derivation of likelihood function, please see Habib, 2010):

$$LL = \sum_{i=1}^{N} \left[\begin{array}{l} ln\left(\varphi\left(\frac{D_i - \gamma_i Z_i}{\sigma_{di}}\right)\right) + ln\left(\varphi\left(\frac{(s_i - \beta_i x_i) - \rho_{Dsi}\sigma_{si}J_3(\psi_i)}{\sigma_{si}\sqrt{1-\rho_{Dsi}^2}}\right)\right) - ln\left(\sigma_{si}\sqrt{1-\rho_{Dsi}^2}\right) - ln(\sigma_{Di}) - ln(t_D) - ln(t_s) \\ + \sum_{wi=1}^{M} wi\, ln\left(\Phi\left(\frac{J_2(\varepsilon_i) - \rho_{Dw}(J_3(\psi)) - \rho_{Sw}\sqrt{1-\rho_{Dw}^2}\left(\frac{(S-\beta x) - \rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho_{DS}^2}}\right)}{\sqrt{(1-\rho_{Sw}^2)(1-\rho_{Dw}^2)}}\right)\right) \end{array} \right]$$

here wi is a dummy variable representing choice of with whom,

ρ is the correlation coefficient with subscript s, representing continuous start time

D represents continuous duration, and w represents with whom

(11)

Here, the term $J(\psi_i)$ indicates equivalent normal variable of with whom choice probability.

$$J_3(\psi) = \Phi^{-1}\left(\frac{\exp(V_w)}{\exp(V_w) + \sum_{z \neq w}\exp(V_z)}\right) \tag{12}$$

The estimation code was developed in the software GAUSS®, using the BFGS algorithm (Aptech, 2006). In order to capture effects of different variables, we further parameterized the systematic utility function of the discrete choice model components as functions of individual and with whom specific variables. Hence the modelling structure can be expressed as:

$$\begin{aligned} D &= \gamma z + \psi \\ S &= \beta x + \beta^{/}(\gamma z + \psi) + \xi = \beta x + \beta^{/}(\gamma z) + (\beta^{/}\psi + \xi) \\ U_m &= V_m + \varepsilon_m \end{aligned} \tag{13}$$

Here β/ is the coefficient of endogenous duration. Although the assumption of endogeneity further complicates the variance-covariance structure of the three decisions, the likelihood function derivation process used in this paper remains the same. Similar evidence is found in Tommaso (1999) for the trivariate continuous-continuous-discrete case and in Pendyala and Bhat (2004) for the bivariate discrete-continuous case. As per the model formulation a positive value of any correlation coefficient indicates negative relationship between the unobserved factors affecting the respective decisions (Habib, 2009). In the case of a correlation between discrete with whom choice and the continuous start time (i.e., time from the reference timeline to the starting point of social activity) decisions, a positive correlation indicates that the unobserved traits leading individuals to choose the specific alternative with whom choice tend to lead them to choose a shorter amount of time from the reference timeline. A shorter amount of time from the reference timeline indicates an earlier departure from home and vice versa. Similarly, for the correlation between the discrete with whom choice and the continuous duration decisions, a negative value indicates that the unobserved traits leading individuals to choose the specific alternative with whom tend to lead them to choose longer social activity duration and vice versa. In the case of a correlation between with whom-specific continuous start time and continuous duration, a positive sign refers to a negative relationship between the unobserved factors affecting start time and duration choice. This indicates that an earlier start time may influence longer social activity duration and vice versa. On the other hand, in the case the covariates in the start time and duration hazard model, the coefficients have a direct positive relationship to time span but an opposite relationship to the hazard rate. For example, a positive coefficient indicates an increase in duration or late starting of the social activity, while at the same time it indicates a reduced hazard rate of the event.

3. DATA

The data that support this approach corresponds to the first wave of the Toronto Panel Survey (TAPS, 2002-2003), which used the instrument CHASE® to collect in detail the activities and scheduling processes of around 400 individuals in a week, identifying with whom these activities were performed with. This latter variable is used to construct the "elicited" respondent's personal network, gathering not only whether the interactions were made with friends of family members, but also richer covariates which capture the structure of the respondent's social network, in terms of size, density, contact variability, number of components, and centralization.

TAPS provides useful empirical data to support the activity scheduling process modeling pursued in Toronto's activity-based models (Miller and Roorda, 2003). The specific survey was conducted in that city between 2002 and 2003, with 271 households (426 adults) participating in the weeklong survey. The CHASE instrument was designed to collect information about activities in both planning and execution stages. For this seven-day activity diary survey, the participants are required to record the individual activity information prior to the starting of the day. The CHASE program tracks the activity information that is added first, and then modified, deleted and executed over time. The first time added information represents the agenda formation, which undergoes modification or sometimes deletion for scheduling. The final scheduled observations include the information regarding scheduling pressure. CHASE divides all activities into nine major groups, of which social is one; social activities in this paper are those self-classified by the respondents as well as those that correspond to "going to restaurants" and "having meals at home", were at least one non-household member participates. Further details about each of these categories can be found in Doherty et al (2004). CHASE collects a variety of attributes related to the activity type, the actor of the activity and the household within which the actor resides. In addition to this general information, some specific information about the activity is collected by actively prompting the respondent in an *End of Week Review* (EWR). EWR systematically queries stated spatial and temporal flexibilities, normal duration, and frequency of the activity type of concern. A detailed description of this EWR component of CHASE is available in Doherty et al. (2004). After cleaning some observations, 294 individuals in 208 households were selected for analyses. Total number of individual social episodes for these 294 people was 1223. 124 of the sampled persons are male and 170 are female. For further details about these data the reader can consult Doherty et al. (2004).

In addition to socio-economic and activity specific variables, the role of the respondent's social networks in activity-travel decisions are incorporated explicitly. From an empirical perspective, an efficient way of addressing the impact of the social context is by considering the individual's personal network information. Personal networks constitute a useful approach to study the relevance of the social structure in activity-travel decisions. A personal network is "my network" for any given individual. In personal network analysis, the respondent individual is referred to as an "ego" and all of the people with whom he/she interacts are referred to as "alters". The number of "alters" within an ego's personal network indicates the size of the network. Measures of the size of a personal network vary according to the purpose of study. For example, McCarty (2002) estimated networks of about 250 ties in an American sample. However, as scope conditions get more specific, the number of network members decreases.

There is a long tradition about methods to gather personal networks, mainly from sociology (Marsden, 1990; 2005), but also - more recently – from the travel behavior field (Larsen et al., 2006; Carrasco et al.¸ 2008). Most of these methods explicitly "elicit" the personal network from certain prompt questions, and then gather information about travel behavior patterns. However, an alternative method – used in this paper – consists of gathering the personal networks from the respondents' diary contacts (Fu, 2007) or, in other words, from with whom individuals interacted within a certain period.

Following that approach, personal networks are devised in this paper based on information of the people (alters) with whom a respondent (ego) socializes. Constructing these personal networks is feasible since the CHASE instrument records the specific persons with whom the respondent socialized in each episode. For the purposes of this paper, alters are divided in two categories: "family", which correspond to both close and extended family members; and the "friends", which correspond to all the other people who are not family members. At the same time, CHASE also records whether the alters are household members or not.

The process of building the elicited respondent's social network is illustrated below. Table 1 shows potential social activities that the respondent may have engaged, with their corresponding start time, duration and with whom dimensions. This latter information can be processed in order to elicit a the respondent's social network emerging from the diary as presented in Figure 1, and where the presence of links between nodes represent whether specific alters shared any activity. In this way, the respondent's personal network is built using the information about the contacts of each social activity; this process allows the study to construct variables at the network level, related with the network's size and structure.

Using the previous information, the following personal network variables were built and studied:

- Total number of family members, total number of friends, total number of alters with whom the respondent socialized that week. All these variables measure the overall size of the respondent's personal network in the period studied.
- Proportion of family members and proportion of friends. This is the ratio between the number of friends and the total number of people with whom the respondent had social activities. These variables measure the relative importance of each role in the respondent's contact; that is, whether the respondent is more family- or friend-oriented, or whether there is a balance in the kinds of contacts s/he has.
- Variability of with whom. For each alter, a variability index is constructed calculating the ratio between the number of social episodes s/he had with the respondent and the total number of episodes that the respondent performed during the week. The average variability index of all the alter members from the respondent's personal network corresponds to the variable variability of with whom. This variable measures the "variety seeking" on social contacts, serving as a proxy of the fragmentation of the respondent's personal network. A number close to one involves a low variability of people, that is, the respondent tends to have social episodes mostly with the same people for all the episodes, whereas a number close to zero involves a high variability of people, where most of the social episodes involved different alters.

Table 1: Illustration of scheduling and social variables

Day	**Start Time**	**Duration**	**Activity**	**With Whom**
1	12:00	60 min.	Lunch at home	Family: Mary (M), Francis (F)
1	13:00	150 min.	Social (host)	Family: Mary (M). Friends: Albert (A), Peter (P)
2	14:00	90 min.	Soccer play	Friends: Albert (A), Peter (P), Joseph (J)
2	18:00	45 min.	Dinner	Friend: Peter (P)
3	15:00	180 min.	Social (visit)	Family: Francis (F)
4	15:00	60 min.	Social (host)	Family: Mary (M), Francis (F). Friends: Peter (P)
4	18:00	75 min.	Dinner	Family: Mary (M)

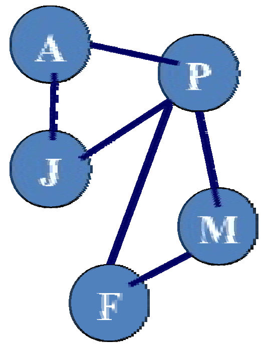

Figure 1: Example of emerging elicited social network from activity scheduling in Table 1

4. EMPIRICAL RESULTS

The model results are shown in Table 2. The first relevant aspect to remark is that the pair wise correlations between the variables start time, duration, and "with whom" are statistical significant, showing that the model captures relevant endogenous unobserved factors that influence the three choices.

Table 2: Model Results

Utility Function of Multinomial Logit "With-Whom" Choice			
Variable name	*Parameter*	*Std. err.*	*Prob.*
"With-Whom" Alternative 1: Family and household members			
Logarithm of Age	-1.877	0.341	0.000
Male Dummy	0.786	0.638	0.218
Income in 10000 CAD	0.121	0.067	0.071
Household Size	-0.162	0.208	0.438
Proportion of Friends in Network	5.365	1.029	0.000
Adult with Partnership	1.662	0.760	0.029
"With-Whom" Alternative 2: Friend and household members			
Constant	4.643	0.698	0.000
Logarithm of Age	-0.452	0.179	0.012
Income in 10000 CAD	0.025	0.020	0.197
Household Size	-0.335	0.054	0.000
Proportion of Family Member in Network	-0.578	0.287	0.044
Variability of Social Network	-0.562	0.143	0.000
"With-Whom" Alternative 3: Family without household members			
Constant	2.688	0.963	0.005
Logarithm of Age	-0.532	0.254	0.036
Male Dummy	-0.293	0.098	0.003
Household Auto Ownership Level	0.063	0.074	0.395
Income in 10000 CAD	0.061	0.028	0.030
Household Size	-0.394	0.069	0.000
Proportion of Friends in Network	3.004	0.326	0.000
Household Head	-0.251	0.137	0.068
Covariate Function of Start Time: Lognormal Hazard Model			
Constant_With-Whom_Alt 2	-0.065	0.073	0.375
Constant_With-Whom_Alt 3	-0.304	0.075	0.000

Constant_With-Whom_Alt 4	-0.310	0.084	0.000
Total People Involved	-0.010	0.012	0.411
Duration Flexibility (Dummy)	-0.044	0.028	0.115
Travel Time in Hours	0.125	0.034	0.000
Household Auto Ownership Level	0.060	0.022	0.007
No. of Children at Home	0.017	0.017	0.299
Home Maker_At Home Job	-0.045	0.059	0.445
Logarithm of Age	-0.063	0.021	0.002
Weekly Frequency	-0.007	0.003	0.030

Table 2 (cont'd): Model results

Covariate Function of Duration Lognormal Hazard Model			
Variable Name	*Parameter*	*Std. err.*	*Prob.*
Constant_With-Whom_Alt 2	0.311	0.132	0.019
Constant_With-Whom_Alt 3	0.439	0.134	0.001
Constant_With-Whom_Alt 4	0.209	0.149	0.161
Total People Involved	0.142	0.025	0.000
Duration Flexibility (Dummy)	0.269	0.051	0.000
Travel Time in Hours	0.161	0.056	0.004
Household Auto Ownership Level	0.066	0.036	0.069
No. of Children at Home	-0.078	0.030	0.009
Home Maker_At Home Job	0.080	0.120	0.503
Total Alters in the Network	0.013	0.009	0.125
Logarithm of Age	-0.079	0.038	0.037
Male	-0.034	0.052	0.510
Weekly Frequency	-0.044	0.006	0.000
Correlation Between Start Time and "With-Whom"			
Corr_St-Ww2	0.988	0.005	0.000
Correlation Between Duration and "With-Whom"			
Corr_Dur-Ww2	-0.418	0.199	0.035
Correlation Between Duration and Start Time, Alternative"With-Whom"			
Corr_St-Dur_Ww2	-0.081	0.042	0.051
Corr_St-Dur_Ww4	-0.493	0.060	0.000
Variance of Continuous Hazard Start Time , Alternative "With-Whom"			
VarSt1	0.246	0.050	0.000
VarSt2	0.602	0.019	0.000
VarSt3	0.556	0.021	0.000
VarSt4	0.635	0.036	0.000
Variance of Continuous Hazard Duration, Alternative "With-Whom"			
VarDur1	0.471	0.097	0.000
VarDur1	0.862	0.023	0.000
VarDur1	0.859	0.033	0.000
VarDur1	0.905	0.051	0.000

Table 3: Model Goodness of Fit Measures

Loglikelihood of Full Model	-8093.25
Loglikelihood of Null Model	-9630.65
Rho-Square	0.16

In the case of systematic effects for the discrete choice ***"with whom"***, the negative sign on all the alternatives with respect to the alternative of reference (socializing with friends but without any household of family member) shows a tendency of younger people to socialize more with their peer friends than with their family alters, all else equal. On the contrary, people with higher income shows a higher tendency for activities with family members, especially with specialized interactions, such as only friends or only friends and household members. In terms of social network variables, the proportion of friends in the network is a positive variable with respect to the alternatives which involve family (with or without household members). Conversely, having a higher proportion of family members in the network has negative effect in the probability to choose friends as alternatives. These two results are consistent with previous findings from the literature which show a "saturation" effect (Carrasco and Miller, 2009), which involves that people with more friends will tend to choose their few – but possible very relevant – family alters, all else equal. Another relevant social network variable is the negative effect of the variability of social network in the friend and household member alternative, which shows that people with more variability of social relations are those who have more propensities to choose friends for their social activities. Finally, other socio-demographic variables which are relevant on the "with whom" utility functions are gender, household size, lifecycle (adult with partnership), and being household head.

In the case of ***duration***, social activities will tend to be longer when there are more people from the respondent's social network involved. Similarly, duration is positively related with the total number of alters from the overall set of the respondent's social activities; that is, people with more social contacts will tend to spend longer social activities. Social activities tend to be shorter for older people and for those who have children at home, the latter effect possible due to scheduling constraints. People who work at home – and possibly with more flexible schedules will tend to have longer social activities, as well as those who have a car. Finally – and consistently with the earlier study by Habib et al. (2008) – social activity duration is positively related with travel time, result consistent with the literature and the intuition regarding the relationship between the utility of performing the activity and the cost (travel time) of performing it.

In the case of ***start time***, there are no social network significant variables, but socio-demographics and a level of service variable. Intuitively, more duration flexibility involves the propensity to start later the social activity. Also, having a car involves earlier social activities as well as having more children at home. If more people from the social network are involved in the social activity, start times tend to be later, possibly due to the more difficult scheduling coordination process.

Contrary to the previous study with these data (Habib et al., 2008), travel time is a significant and positive variable for start times, which means that longer travel time distances involve earlier start times. The contrast with the earlier study shows the relevance of using a jointly trivariate econometric study to represent the joint relationship between with whom, duration, and start times, and which could incorporate the unobserved endogenous effects between them. In fact, the earlier study by Habib et al. (2008), was not capable of explicitly incorporating the relationship between duration and start time, and the role of travel times on this relationship, given that considered two separate bivariate models of with whom / duration and with whom / start time. In that sense, the result presented on this paper shows that travel times are a crucial link between duration and start times on social activities, when unobserved effects are taken into account.

5. CONCLUSIONS

This paper has presented an econometric trivariate model, which studies the relationship between two key activity scheduling variables (duration and start time) and the role of social networks, studied through the social contacts of the decision makers (with whom). The model expands an earlier study (Habib et al., 2008) by incorporating the endogenous unobserved factors which help to understand the interrelation between the three variables of study. A particularly remarkable result is the positive role of travel times on both duration and start times, showing that travel time is a relevant link between both variables, and that is does matter even in social activities. In addition, the relevant role of social networks – shown through the statistical significant correlation with duration and start times – emphasizes the idea that activity-based models which need to be explicit on the relevance of this social dimension. In fact, these empirical results – and others from the previous literature discussed earlier in this paper – show that models that build the travellers' activity scheduling need to consider their social networks since they shape key aspects such as duration and start times.

Overall, this study presents a model that not only captures a rich set of variables explaining activity scheduling – such as social network size, contact variability, and duration flexibility – but also the different relationships between social networks and activity scheduling decisions, both in terms of correlations and endogenous relations. The results confirm the relevance of the social context in the episode temporal characteristics, illustrating aspects that future activity-based travel demand models should incorporate to be able to capture the socializing side of mobility decisions.

A key caveat with respect to this result is due to the scope of the social networks that are considered in this paper. As discussed before, the model presented here considers only those social contacts from the "elicited" social network, that is, which emerge from daily activities, and with a scope of one week (constrained by the data set). As a consequence, the role of other relevant social contacts in activity scheduling decisions – either from an emotional or instrumental viewpoint – are not reflected in the results here, and should be integrated in future modelling efforts.

6. REFERENCES

Aptech Systems 2006. GAUSS User's Manual. Maple Valley, CA.

Arentze, T. and H. Timmermans (2004), "A learning-based transportation oriented simulation system," Transportation Research Part B, 38, 613-33.

Bhat, C. R. (1996). "A generalized multiple duration proportional hazard model with an application to activity behaviour during the evening work-to-home commute." Transportation Research Part B, 30, 465-480.

Bhat, C. R. (1998a). "A model of post home arrival activity participation behaviour." Transportation Research Part B, 32(6), 384-400.

Bhat, C. R. (1998b). "Analysis of travel mode and departure time choice for urban shopping trips." Transportation Research Part B, 32(6), 361-371.

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

Carrasco, J.A., B. Hogan, B. Wellman, and E. J. Miller (2008), "Collecting social network data to study social activity-travel behaviour: An egocentred approach," Environment and Planning B, 35 (6), 961-80.

Carrasco, J.A. and E. J. Miller (2009), "The social dimension in action: A multilevel, personal networks model of social activity frequency," Transportation Research A, 43 (1), 90-104.

Doherty, S, E. Nemeth, M. Roorda, and E. J. Miller (2004), "Design and assessment of the Toronto Area Computerized Household Activity Scheduling Survey," Transportation Research Record: Journal of the Transportation Research Board, 1894, 140-49.

Dubin, J. and McFadden, D. (1984). "An econometric analysis of residential electrical appliance holdings and consumption." Econometrica, 52:,345-362.

Fu, Y-C. (2005), "Measuring personal networks with daily contacts: a single-item survey question and the contact diary," Social Networks, 27 (3): 169-86.

Habib, K.M.N., (2010). "Joint Model of Commuting Mode Choice, Work Start Time and Duration". The paper under review for Transportation Reseach Part A.

Habib, K.N., J.A. Carrasco, and E. J. Miller (2008), "Social context of activity scheduling: Discrete-continuous model of relationship between "with whom," Transportation Research Record: Journal of the Transportation Research Board, 2076, 81-87.

Hensher, D. A. and Mannering, F. L. (1994). "Hazard-based duration models and their application to transport analysis." Transport Reviews, 14(1), 63-82.

Keifer, N. M. (1988). "Econometric duration data and hazard function." Journal of Economic Literature, XXVI, 646-679.

Larsen, J., J. Urry, and K. W. Axhausen (2006), Mobilities, Networks, Geographies. Aldershot: Ashgate Publishing Limited.

Lee, L. F. (1983). "Notes and comments on generalized econometric models with selectivity." Econometrica, 51(2), 507-512.

Lee, B. and Timmermans, H. J. P. (2007). "A latent class accelerated hazard model of activity episode durations." Transportation Research Part B, 41, 426-447.

Marsden, P. (1990), "Networks data and measurement," Annual Review of Sociology, 16, 435-63.

Marsden, P. (2005), "Recent developments in network measurement," in Models and Methods in Social Network Analysis, P. Carrington and J. Scott and S. Wasserman, Eds. New York: Cambridge University Press.

McFadden, D. (1973). "Conditional logit analysis of qualitative choice behavior." In Frontiers in Econometrics (Zaremmbka G.S., Ed.) Academic Press, New York.

McCarty, C. (2002), "Structure in personal networks," Journal of Social Structure, 3 (1), http://www.cmu.edu/joss/content/articles/volume3/McCarty.html.

Munizaga, M. A., Correia, R., Jara-Diaz S., and Ortuzar, J. D. (2006). "Valuing time with a joint mode choice-activity model." International Journal of Transport Economics, XXXIII(2), 193-210.

Pendyala, R. M. and Bhat, C. R. (2004). "An exploration of relationship between timing and duration of maintenance activity." Transportation, 31, 429-456.

Schmertmann, C. P. (1994). "Selectivity bias correction methods in polychotomous sample selection models." Journal of Econometrics, 60, 101-103.

Tommaso, M. L. D. (1999). "A trivariate model of participation, fertility and wages: The Italian case." Cambridge Journal of Economics, 23, 623-640.

Roorda, M.J., and Miller, E.J. (2004), "Toronto Activity Panel Survey: demonstrating the benefits of a multiple instrument panel survey". Paper presented at the 7th International Conference on Travel Survey Methods, Costa Rica.

Roorda, M., Miller, E.J. and Habib, K.M.N. (2008) "Validation of TASHA: A 24-hour Travel Activity Microsimulation Model" Transportation Research Part A, 42: 360-375.

ACKNOWLEDGEMENTS

The data collection was financed by the Social Sciences and Humanities Research Council of Canada (SSHRCC), Major Collaborative Research Initiative (MCRI). This work was also facilitated by a scholar visit of the first author in Chile, funded by the Chilean National Fund for Sciences and Technology, FONDECYT, grant Nº 11070075.