Transportation Research Part A 46 (2012) 33–47

Contents lists available at SciVerse ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Modeling commuting mode choice jointly with work start time and work duration

Khandker M. Nurul Habib

Department of Civil Engineering, University of Toronto, ON, Canada M5S 1A4

ARTICLE INFO

Article history:
Received 29 April 2010
Received in revised form 2 August 2011
Accepted 7 September 2011

Keywords:
Trivariate model
Peak-period travel demand
Commuting mode choice
Work start time and duration
Activity-based model
Travel demand management policy evaluation

ABSTRACT

This paper presents a joint trivariate discrete–continuous–continuous model for commuters' mode choice, work start time and work duration. The model is designed to capture correlations among random components influencing these decisions. For empirical investigation, the model is estimated using a data set collected in the Greater Toronto Area (GTA) in 2001. Considering the fact that work duration involves medium- to long-term decision making compared to short-term activity scheduling decisions, work duration is considered endogenous to work start time decisions. The empirical model reveals many behavioral details of commuters' mode choice, work start time and duration decisions. The primary objective of the model is to predict workers' work schedules according to mode choice, which is considered a skeletal activity schedule in activity-based travel demand models. However, the empirical model reveals many behavioral details of workers' mode choices and work scheduling. Independent application of the model for travel demand management policy evaluations is also promising, as it provides better value in terms of travel time estimates.

1. Introduction

Work trips in any urban area are always at the center of focus in urban transportation planning and policy analyses. Work trips in aggregation define peak versus off-peak-period traffic flow in the urban transportation network. Although activity-based modeling practice extends beyond peak-period travel, commuting activities are always at the center of all modeling approaches. The idea behind the concept of skeletal activities in an activity-based travel demand modeling framework is to recognize the importance of work activities in defining urban transportation system performance (Habib and Miller, 2006). Today it is becoming increasingly evident that the once common picture of having very sharp peak-period urban traffic flow within a narrow time window due to the collective work trips is being replaced with a conception of traffic flow which is flatter and distributed over longer time windows (Schwanen and Dijst, 2003). Such a phenomenon is widely known as peak spreading (Mayer and Miller, 2001).

The peak spreading phenomenon is a direct result of workers' mode and trip timing choice decisions. The basic purpose of moving away from a peak-period modeling approach and toward 24-h modeling practice is to capture the peak versus off-peak-period tradeoffs in travelers' travel-related decisions. However, even the 24-h activity-based travel demand models cannot properly capture the peak spreading phenomenon (Roorda et al., 2008). Work activity related decisions are particularly relevant to our daily lives. For this reason, work activities and work trips have received considerable attention in the literature (Bhat and Singh, 2000; Bhat, 2001). Such decisions as work mode choice, departure time, and numbers of stops are typically investigated individually using advanced econometric techniques (Bhat, 1996, 2000). However, dealing with

E-mail address: khandker.nurulhabib@utoronto.ca

0965-8564/$ - see front matter
doi:10.1016/j.tra.2011.09.012

work mode choice, start time and duration comprehensively within a unified econometric modeling framework is rare in the literature. It is also worth noting that, in the case of commuting activities, decisions of when to start work, how long to work, what mode to use to reach the work place, etc. are all intricately inter-related (Jara-Diaz, 2003a,b).

From the methodological point of view, most of the advanced techniques available in the literature require discretization of time in some manner (Bhat, 2001; Bhat and Steed, 2002). In such cases, developing the modeling structure always remains at the discretion of the researcher based on the manner in which time is to be discretized. A comprehensive methodological approach is necessary in order to address interrelationships between commuting mode choice, work start time, and work duration. It is necessary to develop a modeling framework that allows the modelers to avoid any arbitrary discretization of time for trip timing decisions. Such an approach can address realistically the peak spreading phenomenon in urban transportation networks while simultaneously facilitating the evaluation of a wide range of alternative policy options in order to reduce peak-period traffic congestion.

Unlike the conventional four-stage model, activity-based travel demand models consider interrelationships of different activity decisions (mode, start time, duration etc.) at the disaggregate level. The typical approach to introducing such interrelationships involves the use of joint probability distributions of the decisions. Still, many operational activity-based models have difficulties with introducing policy sensitivity to the distributions of the decisions that serve as key inputs to the activity schedulers. In TASHA, FAMOS, and ALBATROS for example, base year distributions cross-classified by activity type, person, household, and schedule attributes are used in a decision tree approach or are randomly drawn from observed distributions in order to simulate activity frequencies, start times, and durations for the population (Roorda et al., 2008; Pendyala et al., 2005; Arentze and Timmermans, 2004). Such approaches are insufficient when considering policies and scenarios that have the potential to significantly shift travel trends away from the base year distribution of activity start times. CEMDEP, the econometric modeling system for activity-travel demand, uses econometric models for almost all decisions related to activity-travel demand.

Even in this case the individual econometric models for mode choice, start time, durations, etc. are univariate in nature and thus the approach fails to address the interrelationships between all of these decisions at the estimation stage (Bhat et al., 2004). The activity-based models described by Vovsha and Bradley (2004), on the other hand, are examples of models that include explicit tour-based time-of-day discrete choice models to schedule the travel tours of individuals. Although this approach represents a significant improvement over the use of base year distributions or individual univariate econometric models, limitations remain with in terms of the representation of time as a discrete value. Efforts have been made to incorporate the interrelationship between mode choices and start time of work activity using joint discrete–continuous modeling, but work duration has still been considered as exogenous input to the model (Habib et al., 2009). On the other hand Munizaga et al. (2006) modeled mode choice and duration of work activity jointly, but without considering start time.

The importance of work duration in travel demand is well argued in the literature (Jara-Diaz, 2002, 2003a,b). We earn money by spending time in work activities, whereas we typically spend money by participating in other activities. The worker's work duration is part of a medium- to long-term decision process in contrast to the daily activity scheduling decisions of mode choice, start time, etc. This is a fundamental reality in household-based decision dynamics (Miller, 2005). It is for this reason that work duration should be considered as endogenous to daily activity scheduling decisions, a factor which is figured in by considering work schedule as a skeletal schedule in activity-based travel demand modeling (Roorda et al., 2008). When to start work and which mode to choose are obviously conditional to how long a worker plans to work. However, obviously there are daily adjustments in work start time, work duration, and commuting mode choice decisions also. It is for this reason that an investigation of joint mode choice, work start time, and work duration must consider systematic and random factors influencing these three decisions. At the same time, in order to be consistent with a theoretical understanding of decision dynamics, work duration should be endogenous to the other decisions in the investigation. Research on joint multiple decision modeling has been reported on in the literature (Bhat, 1998a; Munizaga et al., 2008), but investigations of joint mode choice start time and duration of commuters' work activities have been rare. From the practical policy analysis and activity-based travel demand modeling perspective, any effort to model commuters' mode choice, work start time, and duration is crucial.

In this paper, a modeling approach for joint commuting mode choice, work start time, and work duration is developed. The econometric formulation of the presented model ensures correlations among these three decisions and allows for the use of a continuous time modeling approach for start time and duration. In addition, work duration is considered as an endogenous variable in work start time choice. However, this modeling structure can be applied for any discrete–continuous–continuous decision situation. The desirable property of the model is that the likelihood function is of closed form and can be estimated using the conventional maximum likelihood estimation technique. However, the primary contribution of this paper is twofold in nature. Methodologically, it presents a robust structure of dealing endogenous work duration into joint mode choice and start time choice models with continuous hazard specifications for start time and duration. For policy application, it has the potential to test a wide range of travel demand management policies, where policy responses can be expected to affect the tradeoffs between commuting mode choice, work start time and work duration in a continuous time scale. For empirical investigation, the model is estimated using a data set collected in the Greater Toronto Area (GTA).

The next section of this paper presents the econometric structure of the joint model. Section 3 discusses the data source for empirical investigation. Section 4 discusses the empirical model. The final section summarizes important findings from the research.

2. Econometric modeling framework

2.1. Background

The mode choice decision is inherently a discrete choice decision. On the other hand, start time and duration decisions for any activity can be modeled 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 min or 10 min), 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.

Bhat (1998b) has argued that the specific required work start times according to job type and a close familiarity with commuting travel times and congestion variability can restrict commuters to relatively narrow home-work departure windows within which it is most appropriate to represent time as a continuous variable. The same argument is applicable for work duration. In order to avoid the issues of discretizing time for start time or duration mode, Bhat (1996, 1998b) proposed modeling 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. 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 modeling time in a linear regression model. Time is naturally a continuous variable, given that it is modeled for start time or duration of any activity. If the time is modeled 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 modeling framework which recognizes the continuous nature of time for start time and duration of activities with other discrete attribute decisions.

In this paper we consider start time and duration of work activities as continuous decisions together with discrete commuting mode choice decisions. As the intension is to jointly model commuting mode choice, start time, and duration of work activity, the modeling structure takes the form of a trivariate discrete–continuous–continuous decision structure. Within this structure, the discrete choice component is modeled as a multinomial logit model and the continuous components are modeled as continuous accelerated failure time hazard models. The accelerated failure time hazard modeling 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. However, in this paper our main objective is to consider endogeneity of work duration both in mode choice and start time choice decisions. Work duration is also considered as a variable in the start time model. The next section describes the structure of the model and the estimation process.

2.2. Structure and estimation

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 work activity), and D represents the logarithm of continuous duration of work activity. According to Keifer (1988), the accelerated failure time 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 mode, m, for commuting is expressed by U_m. The utility of mode choice is composed of a systematic utility function (V_m) and an independent and identically gumbel distributed (across the outcomes and individuals) error term, ε_m, 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 σ_S and σ_D variances consecutively. The equations for start time, utility of mode choice and duration can then be expressed as:

$$S = \beta x + \xi \tag{1}$$

$$U_m = V_m + \varepsilon_m \tag{2}$$

$$D = \gamma z + \psi \tag{3}$$

$$\text{where,}\quad \begin{array}{l} \xi \sim N(0, \sigma_S^2) \\ \varepsilon \sim IID\ \text{Gumbel}(0, 1) \\ \psi \sim N(0, \sigma_D^2) \end{array}$$

Beginning with the utility function of mode choice, a multinomial mode choice situation can be expressed in terms of dichotomous situations (Munizaga et al., 2006). Any individual mode, m, is chosen if the utility of this mode is greater than the second maximum utility of all other modes:

$$U_m > \max_{n=1,2,\ldots,I;\quad n\neq m}(U_n) \tag{4}$$

This can be further decomposed as:

$$V_m > \max_{n=1,2,\ldots,I;\quad n\neq m}(U_n) - \varepsilon_m \tag{5}$$

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

$$V_m^* \equiv \max_{n=1,2,\ldots,I;\quad n\neq m}(U_n) - \varepsilon_m \tag{6}$$

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

$$P(V_m^* < V_m) = F(V_m) = \frac{\exp(V_m)}{\exp(V_m) + \sum_{n\neq m}\exp(V_n)} \tag{7}$$

On the other hand, the continuous time hazard models are primarily concerned with the time until the event terminates. For empirical application in this paper, the time until one starts work and the work event are the topics of interest. The time until one starts work can be referred to as the pre-work event end time. The basic formula describing event termination in hazard models is the hazard rate, $\lambda(t)$, 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 $\lambda(t)$ 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 work 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 failure time hazard models. In the proportional hazard model it is assumed that the 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 dependent only on the covariates (Hensher and Mannering, 1994). In the accelerated failure time hazard model, on the other hand, it is assumed that the covariates rescale or accelerate the time directly in the baseline survivor function. The hazard rate varies over time as it is accelerated or decelerated by the covariates. The accelerated failure time 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). Now, assuming an exponential covariates functional form $exp(\beta x)$ and $exp(\gamma z)$ for start time and duration, respectively—the accelerated failure time hazard models for start time and duration work activity can be expressed, (where the subscript $_0$ denotes the baseline), as:

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

As explained by Keifer (1988), these formulations can be exploited to define the accelerated failure time hazard model in the form of linear specifications as mentioned in Eqs. (1) and (3) where the assumed distribution of the error terms, ξ and ψ, defines the form of the corresponding accelerated failure 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 failure time hazard models:

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

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 (for further elaborations please see the Appendix).

$$LL = \sum_{i=1}^{N}\left[\begin{array}{c} \ln\left(\phi\left(\frac{d_i-\gamma_i Z_i}{\sigma_{di}}\right)\right) + \ln\left(\phi\left(\frac{(s_i-\beta_i x_i)-\rho_{dsi}\sigma_{si}J_3(\psi_i)}{\sigma_{si}\sqrt{1-\rho^2_{dsi}}}\right)\right) - \ln\left(\sigma_{si}\sqrt{1-\rho^2_{dsi}}\right) - \ln(\sigma_{di}) - \ln(t_d) - \ln(t_s) \\ + \sum_{mi=1}^{M} mi\ln\left(\Phi\left(\frac{J_2(\varepsilon_i)-\rho_{Dm}(J_3(\psi))-\rho_{mS}\sqrt{1-\rho^2_{Dm}}\left(\frac{(S-\beta x)-\rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho^2_{DS}}}\right)}{\sqrt{\left(1-\rho^2_{Sm}\right)\left(1-\rho^2_{Dm}\right)}}\right)\right) \end{array}\right]$$

Here,

mi is a dummy variable representing choice of specific mode,
ρ is the correlation coefficient with subscript, s, representing continuous start time,
d represents continuous duration and m represents mode.

This paper has developed an estimation code in GAUSS using the BFGS algorithm (Aptech, 2006). We further parameterized the systematic utility function of mode choice model components as functions of individual and mode-specific variables. For empirical estimation, we considered work duration as endogenous to the work start time model. Hence the modeling structure can be expressed as:

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

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. In the formulation of the model described above, the correlation coefficients refer to the correlations among unobserved factors affecting the three decisions where the errors appear with the negative sign (as shown in the Appendix). Hence a positive value of any correlation coefficient indicates negative relationship between the unobserved factors affecting the respective decisions. In the case of a correlation between discrete mode choice and the continuous start time (i.e., time from the reference timeline to the starting point of work activity) decisions, a positive correlation indicates that the unobserved traits leading individuals to choose the specific alternative mode 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 mode choice and the continuous duration decisions, a negative value indicates that the unobserved traits leading individuals to choose the specific alternative mode tend to lead them to choose longer work duration and vice versa. In the case of a correlation between mode-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 work 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 work activity, while at the same time it indicates a reduced hazard rate of the event.

The technique of deriving joint likelihood function of mode choice, start time choice, and duration choice is based on Lee's (1983) parametric approach of selectivity bias correction. An alternative option of estimating selectivity bias could be taken into account in the two-stage estimation process explained by Dubin and McFadden (1984). Compared to the method of Dubin and McFadden, Lee's technique imposes an implicit normality assumption on the covariance between outcome and selection indices. Schmertmann (1994) has argued that if the data generating process does not comply with this implicit assumption, Lee's technique may give a biased result. On the other hand, the two-stage estimation technique of Dubin and McFadden does not provide an efficient estimation of variances because of the multicollinearity between regressors. Although ostensibly Dubin and McFadden's approach appears attractive, in this method the non-linear functions of discrete choice explanatory variables are assumed linear in the continuous choice component. Such an implicit assumption may be acceptable for a certain range of values, but it is also likely that this assumption will render unreliable parameter estimates (Bhat and Eluru, 2009).

3. Data and sample

A subset of the 2001 Transportation Tomorrow Survey (TTS), a multimodal travel survey conducted in the Greater Toronto Area (GTA) every five years, has been used for empirical investigation. The TTS survey records the detailed travel records of a random 5% sample of households within the GTA for a single day. Details on the characteristics of the TTS dataset are available in a publication of the Data Management Group (DMG, 2008). At the completion of data preparation, each individual observation (102,975 in total) has a set of available mode choices, an observed work start time, and observed work duration. A series of household and person-specific socio-demographic variables and mode-specific variables are considered for investigation. The mode choice level of service variables have been calculated on a round trip tour basis by summing up the individual values for the home-work-home journeys, while the work start time hazard model covariates have been calculated only for the home-work journey. Auto travel times have been calculated on an hourly basis for all 24 h of the day by assigning all observed zone-to-zone trips to an EMME/2 traffic assignment model. Transit level of service variables were also obtained from EMME/2, but only for the a.m. peak hour, since a detailed off-peak transit network was not available. In order to indentify land-use types in general, the urban density is calculated as the total number of residential unit and job per hectare of built area within the zone. For each individual in the data set, a total of six possible modes are identified. However, the choice set for the mode choice is not constant across all observations. The possible modes are as follows:

1. Auto driver
2. Auto passenger
3. Local transit
4. Local transit park & ride
5. GO transit (a regional transit service) park & ride
6. Walk

Here the individuals without a driver's license do not have the 'auto driver' mode option available. Individuals living in zones without local transit service do not have the 'local transit' mode option available. Individuals living in zones without local transit and without a local transit station with a 'park & ride' facility within a reasonable distance cannot have the 'local transit park & ride' mode option available. Individuals living in zones without a GO transit station within a reasonable distance cannot have the 'GO transit park and ride' mode option available. Individuals with home-to-work distances exceeding five kilometers are considered not to have the 'walk' mode option available.

A number of other modes could be considered, such as GO transit & local transit, only GO transit, local transit & walk, GO transit & walk, etc. However, the relatively small number of observations for such minor modes restricts us to considering only these six major modes. In the case of mode choice decision, one can argue in favor of a nested decision structure, such as 'auto driver' and 'auto passenger' modes nested in a common group, all transit modes nested in a common group, etc. However, the six individual modes considered here have little common with each other. For example, most auto passengers cannot be auto drivers because they either do not have a driver's license or they do not have a car. As a result, 'auto driver' and 'auto passenger' are essentially independent choices from which to choose. In the case of transit, the two modes of transit considered here are entirely different types of modes. Not many individuals have both of the transit modes available to them simultaneously. Hence, the six modes considered for the investigation are essentially independent from each other.

The sample data set is grouped according to the occupation groups:

1. General office/clerical (13,292 observation; 13% of the sample).
2. Manufacturing/construction/trades (22, 361 observations; 22% of the sample).
3. Professional/management/technical (49,385 observation; 48% of the sample).
4. Retail sales and services (17,937 observation; 17% of the sample).

Segmentation of the population according to occupation group allows for the fact that different groups of people will exhibit different commuting behaviors. It would be more logical to perform the segmentation according to income group but, unfortunately, the TTS survey does not collect income information. Accordingly, the generalized and broad occupation groups defined in this study work as surrogate measures of income effects. In this analysis we estimate separate models for each individual occupation group. However, in this paper the trivariate model estimates a work duration model jointly with commuting mode choice and work start time models where it considers work duration as an endogenous variable in the start time model.

4. Empirical analysis

4.1. Model specification

Four joint work start time-commuting mode-work duration models have been estimated for four individual occupation groups. A series of specifications have been tested and the final ones have been reported in this paper. The vast majority of the estimated parameters of the joint models remain statistically significant across all occupation groups, surpassing critical *t*-statistic values at the 5% significance level by a wide margin in most cases. However, some variables with insignificant coefficients are still retained in the models for the purpose of comparison. The goodness of fit of the models is estimated using adjusted rho-square value (Ben-Akiva and Lerman, 1985):

$$\text{Adjusted rho-square} = 1 - \frac{\text{Loglikelihood at convergence} - \text{Number of parameters}}{\text{Loglikelihood of the constant-only model}}$$

Here the constant-only model only includes constants for the start time and duration model components, and the number of parameters indicates the number of parameters in the fully specified model over the number of parameters in the null model. The identification restriction of the joint model is similar to that of the discrete choice model component. Like the discrete choice model component, the number of alternative correlation coefficients, variances, and continuous model components that can be estimated is equal to the total number of discrete alternatives minus one. In this paper, we consider that the variance parameters for both start time and duration model components are generic corresponding to all alternative modes. This assumption does not affect the model result because the variance parameters are acting only as scaling factors for the covariant functions of start time and duration. In the case of correlation coefficients, continuous start time and continuous duration time model components considering the auto driver and auto passenger mode users have similar behaviors, and hence the same coefficients overcome the identification problem. Here it should be clarified that this assumption does not entail that the auto driver and auto passenger modes belong to the same group, a conclusion which would result in a violation of IID assumption in the discrete mode choice component. This assumption refers to the mode users, not the alternative modes, and hence the IID assumption of alternative mode choice is not violated.

4.2. Overall empirical results

The estimated parameters of the empirical model are presented in Table 1. In terms of the goodness of fit measure, the model for professional and general office occupation groups gives the highest adjusted-rho square values. Manufacturing and retail & service group models have the same value of adjusted-rho square. The main reason that the professional and general office groups show the highest values may be the defined start time and duration of these occupation groups. General office and clerical occupations have relatively well-defined work schedules compared to other occupation groups. Similarly, the professional occupation group may also have more defined work schedules than the other two groups—manufacturing and retail & services. The estimated correlation coefficients between unobserved factors influencing mode, start time and duration choice reveal very interesting insights into work activity.

The presence of statistically significant correlation coefficients indicates the presence of unobserved and correlated random factors that cannot be explained by variables available in the data set. It is clear that, with the exception of auto driver and auto passenger modes, the correlations between all other modes versus start time and all other modes versus duration are statistically insignificant. On the other hand, almost all mode-specific start time and duration correlations are highly significant and have the same signs across the modes as well as the occupation groups. This is not the case if we do not consider work duration together with mode choice and start time. In the cases of the joint start time and mode choice models, it can be seen that the correlations between start times and modes are all highly significant (Habib et al., 2009). Moreover, considering work duration as an exogenous variable does not reveal accurately the behavioral relationships.

For the auto driver and auto passenger modes the negative correlations between unobserved factors influencing mode and start time indicates that the unobserved factors influencing auto users also influences workers to start late in the day. This evidence is significant for the professional, retail & service and, manufacturing occupation groups. Similarly, the negative correlations between unobserved factors influencing mode and duration choice indicate that the unobserved factors influencing auto mode choice also influence longer work duration in the day. This evidence is true for the professional and retail & service occupation groups, whereas the positive correlations between the unobserved factors corresponding to mode-specific start time and duration indicate that the unobserved factors influencing workers to start early in the day also influence them to work longer hours, and vice versa. This finding is consistent with the daily total time budget constraints. It is worth noting, then, that there are significant trivariate tradeoffs involved in decisions among auto users of different occupation groups that are not captured by systematic variables. It is clear that random factors influencing auto users to start late also influence them to work shorter hours compared to the other mode users. Such behavioral understandings would not be captured if univariate models for mode, start time, and duration were used.

The significant mode-specific start time and duration correlation coefficients vary across the modes and across the occupation groups considerably. The same sign appearing across the modes and occupation groups reveals the generic relationship by which unobserved factors influencing late start time leave less amount of time for work, given that we have limited time budget per day. However, this effect is the highest for the GO park & ride mode. In general, local transit park & ride and GO park & ride modes induce high correlations between start time and duration for all occupation groups that cannot be explained by systematic variables. The explanation is the rigid transit scheduling constraints involved in these two types of modes as a result of the competition for limited parking spaces and the lack of all-day service. However, we need to be careful when interpreting the values of individual correlation coefficients since they capture general trends in the links. Also, there are many competing factors at work in each of the different occupation groups, including vastly different work shift timing policies and locational considerations that affect the availability and level of service offered by each mode.

Table 1
Estimated model coefficients.

Variable	Professional		General Office		Service/Retail		Manufacturing	
Mode	Param	*t*-Stat	Param	*t*-Stat	Param	*t*-Stat	Param	*t*-Stat
Mode choice model component								
Constant:								
Auto driver	−0.6957	−3.21	−1.4976	−4.77	−3.7841	−14.90	−3.8998	−13.10
Local transit	3.6252	43.82	3.4625	26.44	2.7653	25.81	1.8532	17.63
Local transit park & ride	–	–	–	–	−2.0881	−4.66	−1.7593	−6.54
Walk/bike	1.4803	3.38	2.2835	14.24	–	–	–	–
Full time employment:								
Auto driver	0.3290	5.45	0.2203	2.95	0.3623	6.36	0.1677	1.90
Local transit park & ride	0.9080	9.43	0.5697	3.77	1.0526	2.50	–	–
GO park & ride	0.7619	3.37	–	–	–	–	–	–
Total cost = in-vehicle + parking:								
Auto driver, auto passenger	−0.0593	−18.63	−0.0616	−9.87	−0.0723	−10.92	−0.0624	−8.39
Fare per km distance:								
Local transit, local transit park & ride, GO park & ride	−0.4878	−12.43	−0.4641	−6.81	−0.2449	−5.70	−0.3433	−6.37
Park & ride auto cost:								
Local transit park & ride, GO park & ride	−0.2264	−6.00	−0.4917	−11.50	−0.2596	−5.16	−0.2412	−3.78
Auto in-vehicle travel time (min):								
Auto driver, auto passenger	−0.0363	−31.12	−0.0333	−16.22	−0.0249	−10.35	−0.0166	−6.25
Local transit park & ride, GO park & ride	−0.0250	−6.18	–	–	–	–	–	–
Transit in-vehicle travel time (min):								
Local transit, local transit park & ride, GO park & ride	−0.0129	−19.83	−0.0095	−9.06	−0.0086	−6.90	−0.0081	−6.55
Waiting time (min):								
Local transit, local transit park & ride, GO park & ride	−0.1584	−40.12	−0.1330	−24.34	−0.0924	−19.65	−0.0762	−17.93
Walking time (min):								
Local transit, local transit park & ride, GO park & ride	−0.0330	−23.91	−0.0321	−15.05	−0.0223	−12.14	−0.0188	−10.67
Home-work place distance in km:								
Walk/bike	−0.8194	−37.15	−1.1553	−19.62	−0.9312	−25.57	−0.7000	−22.63
More than 1 vehicle in household:								
Auto driver	0.6679	14.18	0.5954	8.78	0.5669	9.15	0.6493	11.80
Local transit	−1.0202	−17.75	−1.0755	−13.62	−1.1793	−16.72	−1.0823	−14.89
Local transit park & ride	0.4996	5.13	0.3161	2.13	0.4914	1.90	0.5818	1.88
GO park & ride	0.3578	3.83	–	–	–	–	–	–
Walk/bike	−0.8047	−9.36	−1.0409	−6.59	−0.7584	−7.62	−0.3073	−2.52
Logarithm of age in years:								
Auto driver	0.7482	13.24	0.8783	10.38	1.4809	20.47	1.3428	16.76
GO park & ride	0.3964	6.03	0.4470	9.02	–	–	–	–
Walk/bike	0.2016	1.71	–	–	0.6145	17.64	–	–
Gender: male:								
Auto driver	0.8027	21.01	0.9434	11.35	0.9457	15.46	1.5045	26.88
Local transit	0.4193	8.92	0.4544	4.85	0.1173	1.68	0.4721	7.33
Walk/bike	1.0047	13.65	1.0742	6.54	0.8110	8.34	1.3951	12.70
Median zonal income in thousands:								
Local transit	−0.0021	−4.69	−0.0021	−2.53	−0.0032	−3.97	–	–
Walk/bike	−0.0040	−4.96	–	–	−0.0050	−4.35	–	–
Home zone urban density:								
Auto driver	−0.0011	−7.89	–	–	–	–	–	–
Work zone urban density:								
Local transit	0.0005	19.34	0.0004	9.29	0.0007	11.66	0.0008	10.47
Walk/bike	0.0008	15.41	0.0011	8.77	0.0011	8.95	0.0016	10.38
Work start time model component								
Constant:								
Auto driver, auto passenger	6.7915	500.90	6.9633	311.57	7.3190	317.57	7.0181	246.43
Local transit	6.8262	469.75	6.9645	291.55	7.3258	287.89	7.0148	223.42
Local transit park & ride	6.6869	352.14	6.7984	266.32	7.1500	189.00	6.8193	134.65
GO park & ride	6.7124	252.97	6.8877	132.31	7.0760	197.77	6.7695	137.45
Walk/bike	6.7974	478.13	6.9937	261.40	7.3599	281.37	7.0135	228.23
Full time employment:								
All modes	−0.0671	−17.85	−0.0788	−14.65	−0.1053	−18.18	−0.0729	−8.23

Table 1 (*continued*)

Variable	Professional		General Office		Service/Retail		Manufacturing	
Mode	Param	*t*-Stat	Param	*t*-Stat	Param	*t*-Stat	Param	*t*-Stat
Work duration in hours:								
All modes	−0.0354	−44.47	−0.0443	−31.93	−0.0474	−35.89	−0.0549	−33.33
Total travel time in hours:								
Auto driver, auto passenger	−0.1507	−39.10	−0.1680	−19.38	−0.2232	−20.46	−0.1674	−16.49
Local transit	−0.1144	−16.65	−0.1159	−10.55	−0.1105	−7.32	−0.1131	−7.41
Local transit park & ride	−0.0321	−2.48	–	–	–	–	–	–
GO park & ride	−0.0711	−4.51	−0.1000	−3.06	–	–	–	–
Distance in km between home to work location								
Walk/bike	–	–	−0.0156	−2.30	−0.0156	−2.54	–	–
Dwelling type: house:								
All modes	−0.0113	−5.32	−0.0108	−2.76	−0.0054	−1.14	−0.0327	−6.97
Household size:								
All modes	0.0038	5.89	0.0063	5.05	0.0067	4.48	0.0145	10.44
Logarithm of age in years:								
All modes	−0.0640	−20.91	−0.0889	−16.97	−0.1633	−28.15	−0.0857	−12.80
Gender = male:								
All modes	–	–	0.0213	5.66	0.0249	6.15	−0.0301	−6.33
Home zone urban density/100:								
All modes	0.0044	4.94	0.0046	2.13	0.0037	1.43	0.0146	4.67
Work zone urban density/100:								
All modes	0.0006	3.80	0.0009	2.86	0.0012	2.18	0.0044	5.66
Free parking available in work zone:								
Auto driver	0.0203	5.23	–	–	0.0299	2.65	–	–
Work duration model component								
Constant:								
Auto driver, auto passenger	6.0967	185.06	5.8339	385.11	5.7393	316.87	5.7585	116.31
Local transit	6.1870	172.39	5.8585	372.09	5.7976	307.49	5.7713	118.14
Local transit park & ride	6.0063	91.80	5.8636	191.75	5.8570	89.50	5.7835	64.12
GO park & ride	6.0707	116.60	5.7492	89.79	5.9696	90.81	5.8535	65.26
Walk/bike	6.1036	159.08	5.7473	195.71	5.7051	194.50	5.6458	93.70
Full time employee:								
All modes	0.2911	30.66	0.2867	29.62	0.3152	30.79	0.3114	20.60
Total travel time in hours:								
Auto driver, auto passenger	0.1363	13.11	0.1687	9.66	0.0830	3.67	0.0478	2.60
Local transit	−0.0337	−1.78	–	–	–	–	–	–
Local transit park & ride	0.1518	2.56	–	–	–	–	–	–
GO park & ride	0.0872	3.03	0.1112	2.56	–	–	–	–
Dwelling type = house:								
All modes	−0.0202	−3.47	−0.0324	−4.06	−0.0313	−3.12	−0.0317	−3.59
Household size:								
All modes	0.0099	5.67	0.0075	3.00	0.0185	6.06	0.0213	8.27
No. of vehicle in household:								
All modes	–	–	–	–	–	–	−0.0084	−2.06
Logarithm of age in years:								
All Modes	−0.0821	−9.86	–	–	–	–	0.0213	1.74
Gender = male:								
All Modes	–	–	0.0206	2.71	–	–	0.0221	2.54
Median zonal Income in thousands:								
All modes	−0.0002	−3.99	−0.0001	−2.12	−0.0003	−3.74	−0.0002	−2.86
Home zone urban density/100:								
All modes	0.0048	1.94	0.0143	3.29	0.0116	2.18	0.0140	2.47
Work zone urban density/100:								
All modes	0.0038	8.44	0.0021	3.23	0.0034	2.93	–	–
Ancillary parameters								
Variance of work start time model	0.1827	263.54	0.1894	154.41	0.2700	176.17	0.2938	171.11
Variance of work duration model	0.4995	313.62	0.3855	164.30	0.5641	189.55	0.5320	211.58

(*continued on next page*)

Table 1 (*continued*)

Variable	Professional		General Office		Service/Retail		Manufacturing	
Mode	Param	*t*-Stat	Param	*t*-Stat	Param	*t*-Stat	Param	*t*-Stat
Correlation coefficients:								
Mode choice-start time auto driver, auto passenger	−0.0548	−1.89	–	–	−0.1026	−2.99	−0.1646	−4.14
Mode choice-duration auto driver, auto passenger	−0.0961	−3.42	–	–	−0.1064	−3.20	–	–
Start time-duration								
Auto driver, auto passenger	0.1868	19.63	0.1794	12.62	0.2263	17.66	0.2727	22.38
Local transit	0.1418	3.91	–	–	–	–	0.2709	6.59
Local transit park & ride	0.7370	41.95	0.7500	26.42	0.8282	28.02	0.8751	32.28
GO park & ride	0.8330	111.76	0.8173	50.67	0.9034	62.83	0.8738	33.57
Walk/bike	0.0988	2.44	0.1820	3.96	0.1189	3.18	0.2820	5.28
No. of observation	49,392		13,511		17,944		22,431	
Log likelihood	−650,675		−176,797		−247,035		−306,968	
Adjusted rho-square	0.16		0.16		0.12		0.12	
VOTS – auto users	**37**		**33**		**20**		**15**	

The correlations discussed here indicate the correlation between unobserved factors influencing the respective decisions. Hence, the findings from the discussions on such estimated correlation coefficients of the empirical models have significant policy implications in terms of identifying factors that affect peak-period traffic congestion. Any travel demand management policy that can influence work hour and work start time can also significantly influence choice of whether or not to use an automobile. High correlations in the cases of start time and duration of GO park & ride and local transit park & ride modes are demonstrative of the strict scheduling patterns of these mode users. Any policies making the park & ride stations more accessible or providing for higher off-peak-period service can influence people to avoid peak-period travel.

4.3. Mode choice model

The estimated mode choice model components are presented in Table 1. Alternative specific constants reveal that the auto driver mode has a very low baseline utility (neglecting all other variables influencing mode choice) compared to the auto passenger and local transit park & ride modes. Although the baseline preference is low, mobility, accessibility and the many other variables used in the model specification influence most commuters to utilize the auto mode. The employment status (full-time versus part-time) of the worker also has an influence on commuting mode choice. Full-time workers are more likely to choose other motorized modes than walking. Within motorized modes, park & ride types of modes are more likely to be chosen than the other modes. This finding is also an indicator of increasing urban sprawling in the GTA. Many full-time workers live in suburban areas where park & ride-type modes are the most attractive. The variable, total number of intermediate stops in home-work-home tour is accommodated in the model in order to test the trip chaining effects in mode choice decisions. It is clear that the commuters making a higher number of intermediate stops in their home-work-home tours are highly likely to choose the auto driver mode because of its flexibility in use.

In the case of cost, for auto driver and auto passenger modes the in-vehicle cost is added to parking cost to define the total cost. For transit-related modes, transit fair components are specified as cost per unit linear distance (home-to-work location). For park & ride modes additional total auto costs related to park & ride use are specified separately. In the cases of the auto driver and auto passenger modes the parking cost is added to in-vehicle cost to obtain a reasonable value for the implied Value of Travel Time Savings (VOTS). On the other hand, in the case of transit-related modes, total in-vehicle cost cannot be calculated in the same way as auto in-vehicle cost because of the nature of the transit fare system in the GTA. Both GO and local transit in the GTA have a flat fare system, and so, in order to capture the sensitivity of the in-vehicle cost variable, fare per unit distance (km) has proven to be the most accurate specification. For local transit, in-vehicle travel time and wait time are considered separately. For transit park & ride modes, in addition to in-vehicle travel time, wait time and the accessing auto in-vehicle travel time are considered separately. Intuitively, all cost and all travel time coefficients are negative in the model. Because of the separate cost variable specifications for transit and auto modes, we cannot compare cost coefficients across the modes; as a result, the VOTS calculation is concerned with auto modes only. Separate VOTS are calculated for individual occupation groups. The VOTS is the highest for the professional group ($37 per hour) followed by the VOTS for general office occupation group ($32 per hour), retail & service occupation group ($21 per hour) and manufacturing occupation group ($16 per hour). It is very interesting to note that the VOTS for the different occupation groups improves significantly if we consider work duration as endogenous to the mode choice and start time decisions (Transport Canada, 2006). This finding is supported by the argument of Munizaga et al. (2006) that, in the case of work activity, work duration should be considered within the mode choice decision framework. However, this paper goes one step further by integrating the work duration as endogenous to the joint mode choice and start time decision framework. The superiority of this modeling framework is proven by the pattern of correlation coefficients between unobserved and random influencing factors of the three decisions as discussed above.

As expected, home-to-work distance has a negative influence on the utility of the walk mode. Parameter signs of household automobile were also found to have conformed to expectation. A higher number of household automobiles positively influences auto and park & ride-type mode utility and negatively influences local transit and walk mode utility. Older people are more likely to use auto, GO park & ride, and walking mode than younger people. Males are more likely to use auto driver, local transit, and walking mode than females. As the TTS data set does not have household income information, we use zonal average or median income as surrogate measures of income. In our case it seems that zonal median income better represents the income effect in terms of statistical significance than the zonal average income.

Higher zonal median income shows negative influence on the utility of the auto driver, local transit and walking modes. The explanation is that the wealthy people living in higher-income neighbourhoods primarily reside in suburban areas where park & ride-type modes are more attractive. The parameter estimates for the professional occupation group indicate that individuals are less likely to drive if they reside in higher-density areas, while the results for other occupation groups indicate that this variable does not have any significant effect. The effect of work zone urban density is consistent across the occupation groups, indicating that the workers working in higher-density urban areas are more likely to use local transit or walking mode over other auto-oriented modes.

4.4. Work start time model

Table 1 shows the estimated covariate parameters of the work start time hazard model. Here, the variance of lognormal distribution is considered to be the same across the occupation groups, but alternative mode-specific constants are estimated. Although the alternative specific constants do not seem to vary significantly across the modes, given the small value of estimated variance, the smaller variation in alternative specific constant will induce a larger variation in work start time or departure time. Constants are almost the same for individual occupation groups, but they vary across the occupation groups. It is clear that the retail & service and manufacturing occupation groups start later than professional and general office occupation groups. Commuters with full-time employment status are likely to start earlier than commuters with a part-time work status. Work duration has a highly significant negative coefficient in the model, indicating that people who work longer start earlier and vice versa; (this is consistent with the explanation of correlation coefficient values discussed in the previous sub-section). The high statistical significance of this parameter also recognizes the fact that work duration is endogenous to mode choice and start time decisions.

Considering the coefficients of mode-specific travel time, it is clear that the travel time has a consistent negative sign across the modes. Similarly, home-to-work location distance has the negative sign for walking mode, confirming the intuitive assumption that individuals start earlier if the travel time requirement is longer. Although the mode-specific travel time coefficients are all of the same order of magnitude, an interesting result is that the auto travel time parameter has the highest negative value of all modes. This implies that non-auto users are less sensitive to travel time than are auto users. The best explanation for this phenomenon would be that non-auto users are more certain about the travel time than are auto users and hence they are less sensitive to travel time. Considering household-level socio-demographic variables, it is clear that the household size influences delays in work starting, which is demonstrative of the higher number of household obligations involved with a larger household size. In terms of dwelling type, it is clear that people living in houses rather than apartments are more likely to start earlier, which reflects the type of residential area and spatial clustering of similar types of workers. In terms of age, it is clear that older people start earlier than younger people. Males start earlier than females, suggesting that females are more involved in household activities in the morning than are males. Both home zone and work zone urban density shows the effect of higher urban density influencing a late start of work. It may be difficult to explain such a relationship with work start time and zonal urban density. A possible explanation of this is the balancing effect of this land-use variable with alternative mode-specific constant terms of the hazard models. This is primarily due to the fact that detailed knowledge of official workplace hours and degrees of work hour flexibility are unavailable at the individual level in the TTS dataset.

4.5. Work duration model

The work duration hazard model is presented in Table 1. This model also includes alternative specific constant terms, each with common variance across the modes. Although the constant terms are almost the same across the modes, they vary across the occupation groups. The values of the constant terms reflect the relative work durations of the corresponding occupation groups. It is clear that commuters in the professional occupation group work the longest hours. Unlike with the start time hazard model, most covariates have a positive sign in the work duration model and the positive constant terms indicate baseline work duration. The model also shows that full-time workers work longer hours than part-time workers, which is a general trend in the GTA.

As expected, travel time has positive effect for all modes. Longer travel time leads to longer work duration for all mode users. However, travel time variables have been found to be insignificant for park & ride-type modes for the general office, retail & service and manufacturing occupation groups. As such, it is arguable whether or not we should consider travel time as a variable in work duration model or not. Critics may argue that travel time is defined by home and work location of the workers and has little to do with work duration. We would argue, however, that someone will be more willing to spend a longer time traveling if she/he needs to work longer hours. As discussed above, work duration is part of the medium- to long-term decision process, and it can also be taken into account in home and work location choice. It would have been best to consider wage rate or income as the key variable in work duration model. However, neither individual nor household income data is available in the data set, we believe that commuting travel time would act as a surrogate measure in this case. In addition to travel time as a surrogate measure for income, zonal median income is considered as a variable in the work duration model. This variable reflects the fact that higher-income workers work shorter hours than lower-income people. In the case of land-use, it is clear that people working or living in areas with higher urban density work longer hours, which is consistent across the occupation groups.

In considering household- and personal-level variables, it is clear that older people work shorter hours than younger people. Males work longer hours than females. A greater household size influences workers to earn more and hence work longer. The total number of household automobiles becomes significant only for the manufacturing occupation group. The data has exhibited that people of this occupation group work shorter hours if they have a higher number of household automobiles. A higher number of automobiles per household among workers in the manufacturing group indicates a higher income/wage rate and a higher income/wage rate, in turn, indicates shorter hours.

5. Summary and conclusions

In this paper we have formulated and estimated a joint model of mode choice, start time and duration, which recognizes the natural continuous feature of activity timing decisions. The model addresses the endogeneity of activity duration within mode choice and starting time decisions. The econometric modeling structure represents a generic discrete–continuous–continuous situation which can be applicable to many complex decision process modeling approaches. For empirical investigation, the model is applied to work activity using a data set collected in the Greater Toronto Area in 2001. The empirical investigation reveals many behavioral insights into activity scheduling and mode choice decisions. Most importantly, it becomes clear that the real relationship between commuting mode choice and trip timing can be indentified if mode choice, start time, and duration are modeled jointly.

Work activity is considered as a skeletal activity in almost all activity-based travel demand models. Skeletal activity attributes (start time, duration, etc.) are generated and scheduled first in order to develop the skeleton of travel-activity patterns. Making activity-based travel demand models sensitive to the subtle travel demand management policies, (e.g. flexible office hours, peak-period congestion pricing, traffic engineering measures to improve travel time of public transit, better inter-modal integration to encourage park & ride by the suburban residents etc.) requires that the modeling approach for developing the skeletal activity-travel pattern should address the behavioral tradeoffs involved in decisions regarding the component of the skeleton pattern. The empirical investigation of this paper essentially models the whole skeletal framework of a worker's daily activity schedule (considering only work as their skeletal activity), but at the same time it accommodates the tradeoffs involved in skeleton formation with respect to commuting mode choice. However, beyond than its consideration merely as part of an overall activity-based travel demand modeling framework, independent application of this model is very suitable to evaluating travel demand management policies targeting workers' peak-period commuting trips. As the model considers work duration jointly within start time and model choice decisions, it will model both a.m. and p.m. peak-period commuting trips.

In conclusion, it should be noted that the modeling structure developed in this paper has far wider applicability than to work activity only. In particular, the formulation corresponds to a joint discrete–continuous–continuous choice model in the discrete choice literature. This model can be utilized for many other travel-related decisions (e.g., modeling vehicle type choice and usage: when to change vehicles and how much to use vehicles; shopping location choice and timing of shopping activity, etc.); for land-use-related decisions (e.g., home type choice, amount of investment, and duration of stay); for marketing context (e.g., product brand choice, quantity to buy and usage etc.), etc.

Acknowledgements

This research has been partially funded by NSERC and the University Start-Up grant. The author acknowledges discussions with Profs. Eric Miller and Nick Day. An earlier version of paper was presented at the 11th IATBR conference (December 2009). Comments and suggestions of audiences in 11th IATBR conference helped to improve the quality of the paper.

Appendix A

For the continuous variables with lognormal hazard specification, the probability of observing any specific values can be written as (Johnson et al., 1994)

$$\Pr(t_S) = \frac{1}{t_S \sigma_S} \phi\left(\frac{S - \beta x}{\sigma_S}\right) \tag{A1}$$

$$\Pr(t_D) = \frac{1}{t_D \sigma_D} \phi\left(\frac{D - \gamma z}{\sigma_D}\right) \tag{A2}$$

Now, as the random components of start time and duration model equations as well as of the utility function of mode choice have completely specified marginal distributions, they all individually can be transformed into equivalent standard normal variables. As per Lee (1983):

$$\begin{aligned}\xi^* &= J_1(\xi) = \frac{S-\beta x}{\sigma_S} = \Phi^{-1}[F(\xi)] \\ \varepsilon^* &= J_2(\varepsilon) = \Phi^{-1}[F(V_m)] \\ \psi^* &= J_3(\psi) = \frac{D-\gamma z}{\sigma_D} = \Phi^{-1}[F(\psi)]\end{aligned} \tag{A3}$$

In order to recognize the correlations between these three decisions, let us assume that these three variables are trivariate normally (TVN) distributed:

$$TVN[J_1(\xi), J_2(\varepsilon), J_3(\psi); \rho_{Sm}, \rho_{mD}, \rho_{SD}]$$

Here ρ indicates correlation coefficients and the subscripts S, D, and m represent start time, duration, and mode choice decisions, respectively. We can easily simplify the TVN distribution as composed of a series of bivariate normal distributions (BVN), expressed as:

$$TVN[J_1(\xi), J_2(\varepsilon), J_3(\psi); \rho_{Sm}, \rho_{mD}, \rho_{SD}] = BVN[J_1(\xi), J_3(\psi); \rho_{SD}] \cap BVN[J_2(\varepsilon), J_3(\psi); \rho_{mD}] \cap BVN[J_1(\xi), J_2(\varepsilon); \rho_{Sm}]$$

Rewriting the specifications, we obtain:

$$\begin{aligned}&\xi = S - \beta x \\ &\psi = D - \gamma Z \\ &\int_{-\infty}^{V_m} f(\varepsilon, (S-\beta x), (D-\gamma Z)) = \frac{\partial}{\partial\xi\partial\psi} TVN(\varepsilon, \xi, \psi; \rho_{Sm}, \rho_{mD}, \rho_{SD})\end{aligned} \tag{A4}$$

The likelihood (L) of observing a specific value of S_1 and D_1 and choosing a specific mode, m, by any individual, i, is the joint probability of $L_i = \Pr(S = S_1) \cap \Pr(D = D_1) \cap \Pr(Mode = m)$. Addressing the correlations between these three variables, we can write:

$$L_i = \frac{\partial}{\partial\xi\partial\psi} TVN(\varepsilon, \xi, \psi; \rho_{sm}, \rho_{md}, \rho_{sd}) \tag{A5}$$

In order to obtain the final mathematical formulation of the joint likelihood function of the three decisions, we can apply successively the transformation of variable approach articulated by Lee (1983).

Starting with the duration model specification, D, the probability density function is:

$$g(\psi) = \frac{1}{t_D\sigma_D}\phi(J_3(\psi)) = \frac{1}{t_D\sigma_D}\phi\left(\frac{D-\gamma Z}{\sigma_D}\right) \tag{A6}$$

Updating the mean and variance of ε with respect to ψ gives $\mu_{m|D} = \rho_{Dm}(J_3(\psi))$ and

$\sigma_{m|D} = \sqrt{1-\rho_{Dm}^2}$ respectively. Similarly, updating the mean and variance of ξ with respect to ψ gives $\mu_{s|D} = \rho_{Ds}\sigma_s(J_3(\psi))$ and $\sigma_{s|D} = \sigma_s\sqrt{1-\rho_{Ds}^2}$ respectively.

Now for the start time, S, the transformed probability density function is:

$$g(\xi) = \frac{1}{t_S\sigma_{S|D}}\phi\left(\frac{(S-\beta x)-\mu_{S|D}}{\sigma_{S|D}}\right) = \frac{1}{t_S\sigma_S\sqrt{1-\rho_{DS}^2}}\phi\left(\frac{(S-\beta x)-\rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho_{DS}^2}}\right) \tag{A7}$$

Updating the mean and variance of ε with respect to modified mean and variance of ξ gives:

$$\mu'_m = \rho_{Dm}(J_3(\psi)) + \rho_{mS}\sqrt{1-\rho_{Dm}^2}\left(\frac{(S-\beta x)-\mu_{S|D}}{\sigma_{S|D}}\right) = \rho_{Dm}(J_3(\psi)) + \rho_{mS}\sqrt{1-\rho_{Dm}^2}\left(\frac{(S-\beta x)-\rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho_{DS}^2}}\right) \tag{A8}$$

$$\sigma'_m = \sqrt{1-\rho_{Sm}^2}\sqrt{1-\rho_{Dm}^2} = \sqrt{(1-\rho_{Sm}^2)(1-\rho_{Dm}^2)} \tag{A9}$$

Moreover, the transformed cumulative probability function of ε becomes:

$$G(\varepsilon) = \Phi\left(\frac{J_2(\varepsilon)-\mu'_m}{\sigma'_m}\right) = \Phi\left(\frac{J_2(\varepsilon) - \rho_{Dm}(J_3(\psi)) - \rho_{mS}\sqrt{1-\rho_{Dm}^2}\left(\frac{(S-\beta x)-\rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho_{DS}^2}}\right)}{\sqrt{(1-\rho_{Sm}^2)(1-\rho_{Dm}^2)}}\right) \tag{A10}$$

Hence, if we write the likelihood function of the joint decisions for only the individual, i, it becomes:

$$L_i = \Pr(S = s)\Pr(\varepsilon \leqslant J_2(\varepsilon))\Pr(D = d)$$
$$= \frac{1}{t_D\sigma_d}\phi\left(\frac{d-\gamma Z}{\sigma_d}\right)\Phi\left(\frac{J_2(\varepsilon)-\mu'_m}{\sqrt{(1-\rho^2_{Sm})(1-\rho^2_{Dm})}}\right)\frac{1}{t_S\sigma_s\sqrt{1-\rho^2_{ds}}}\phi\left(\frac{(s-\beta x)-\rho_{ds}\sigma_S J_3(\psi)}{\sigma_s\sqrt{1-\rho^2_{ds}}}\right) \quad \text{(A11)}$$

Considering M alternatives available for the discrete choice, the likelihood function becomes:

$$L_i = \prod_{m=1}^{M}\left(\begin{array}{l}\frac{1}{t_d\sigma_d}\phi\left(\frac{d-\gamma Z}{\sigma_d}\right)\left(\Phi\left(\frac{J_2(\varepsilon)-\rho_{Dm}(J_3(\psi))-\rho_{mS}\sqrt{1-\rho^2_{Dm}}\left(\frac{(S-\beta x)-\rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho^2_{DS}}}\right)}{\sqrt{(1-\rho^2_{Sm})(1-\rho^2_{Dm})}}\right)\right)^{mi} \\ \frac{1}{t_s\sigma_s\sqrt{1-\rho^2_{ds}}}\phi\left(\frac{(s-\beta x)-\rho_{ds}\sigma_s J_3(\psi)}{\sigma_s\sqrt{1-\rho^2_{ds}}}\right)\end{array}\right) \quad \text{(A12)}$$

$mi = 1$ if m alternative is chosen 0 otherwise

Now, if the sample of population under investigation has the total of N individuals, the log likelihood function becomes:

$$LL = \sum_{i=1}^{N}\left[\begin{array}{c}\ln\left(\phi\left(\frac{d_i-\gamma_i Z_i}{\sigma_{di}}\right)\right)+\ln\left(\phi\left(\frac{(s_i-\beta_i x_i)-\rho_{dsi}\sigma_{si}J_3(\psi_i)}{\sigma_{si}\sqrt{1-\rho^2_{dsi}}}\right)\right)-\ln\left(\sigma_{si}\sqrt{1-\rho^2_{dsi}}\right)-\ln(\sigma_{di}t_d t_s) \\ +\sum_{mi=1}^{M} mi\ln\left(\Phi\left(\frac{J_2(\varepsilon_i)-\rho_{Dm}(J_3(\psi))-\rho_{mS}\sqrt{1-\rho^2_{Dm}}\left(\frac{(S-\beta x)-\rho_{DS}\sigma_S J_3(\psi)}{\sigma_S\sqrt{1-\rho^2_{DS}}}\right)}{\sqrt{(1-\rho^2_{Sm})(1-\rho^2_{Dm})}}\right)\right)\end{array}\right] \quad \text{(A13)}$$

This formulation is based on the homogenous population assumption. However, considering heterogeneity in mode choice and/or start time and duration choice model will introduce multiple integrations in this likelihood function. In that case simulated likelihood estimation will be necessary.

References

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

Arentze, T.A., Timmermans, H.J.P., 2004. A learning-based transportation oriented simulation system (ALBATROS). Transportation Research Part B 7, 613–633.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge, Massachusetts.

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., 2000. Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. Transportation Science 34 (2), 228–238.

Bhat, C.R., 2001. Modelling the commuting activity-travel pattern of workers: formulation and empirical analysis. Transportation Science 35 (1), 61–79.

Bhat, C.R., 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.

Bhat, C.R., Guo, J.Y., Srinivasan, S., Sivakumar, A., 2004. A comprehensive econometric microsimulator for daily activity-travel patterns (CEMDEP). Transportation Research Record 1894, 57–66.

Bhat, C.R., Singh, S.K., 2000. A comprehensive daily activity-travel model system for workers. Transportation Research Part A 34 (1), 1–22.

Bhat, C.R., Steed, J., 2002. A continuous-time model choice for urban shopping trip. Transportation Research Part B 36, 207–224.

Data Management Group (DMG), 2008. Transportation Tomorrow Survey. Joint Program in Transportation, University of Toronto. <http://www.jpint.utoronto.ca/dmg/tts.html>, June 2009.

Dubin, J., McFadden, D., 1984. An econometric analysis of residential electrical appliance holdings and consumption. Econometrica 52, 345–362.

Habib, K.M.N., Miller, E.J., 2006. Modelling workers' skeletal travel activity schedule. Transportation Research Record 1985, 88–97.

Habib, K.M.N., Day, N., Miller, E.J., 2009. Modelling joint trip timing and mode choice: lessons learned from Greater Toronto Area. Transportation Research Part A 43, 639–653.

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

Jara-Diaz, S., 2002. The goods/activity framework for discrete travel choices: Indirect utility and value of time. In: Mahmassani, H.S. (Ed.), Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges. Elsevier Science Ltd.

Jara-Diaz, S., 2003a. On the goods-activities technical relations in the time allocation theory. Transportation 30, 245–260.

Jara-Diaz, S. (2003b). Modelling activity duration and travel choice a common microeconomic framework. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

Johnson, N.L., Kotz, S., Balakrisnan, N., 1994. Continuous Univariate Distributions, vol. 1. John Wiley & Sons Inc, New York.

Keifer, N.M., 1988. Econometric duration data and hazard function. Journal of Economic Literature XXVI, 646–679.

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

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

Mayer, M.D., Miller, E.J., 2001. Urban Transportation Planning: A Decision Oriented Approach. McGraw Hill.

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

Miller, E.J., 2005. An integrated framework of modelling short- and long-run household decision making. In: Timmermans, H.J.P. (Ed.), Progress in Activity Based Analysis. Elsevier, New York.

Munizaga, M.A., Correia, R., Jara-Diaz, S., Ortuzar, J.D., 2006. Valuing time with a joint mode choice-activity model. International Journal of Transport Economics XXXIII (2), 193–210.
Munizaga, M., Jara-Diaz, S., Greeven, P., Bhat, C.R., 2008. Econometric calibration of the joint time assignment – mode choice model. Transportation Science 42 (2), 208–219.
Pendyala, R.M., Bhat, C.R., 2004. An exploration of relationship between timing and duration of maintenance activity. Transportation 31, 429–456.
Pendyala, R.M., Kitamura, R., Kikuchi, A., Yamamoto, S., Fuji, S., 2005. Florida activity mobility simulator (FAMOS): overview and preliminary validation results. Transportation Research Record 1921, 123–130.
Roorda, M.J., Miller, E.J., Habib, K.M.N., 2008. Validation of TASHA: A 24-h activity scheduling microsimulation. Transportation Research Part A 42, 360–375.
Schmertmann, C.P., 1994. Selectivity bias correction methods in polychotomous sample selection models. Journal of Econometrics 60, 101–103.
Schwanen, T., Dijst, M., 2003. Time windows in workers' activity patters: empirical evidence from the Netherlands. Transportation 30, 261–283.
Tommaso, M.L.D., 1999. A trivariate model of participation, fertility and wages: the Italian case. Cambridge Journal of Economics 23, 623–640.
Transport Canada Environmental Affairs, 2006. The cost of urban congestion in Canada.
Vovsha, P., Bradley, M., 2004. Hybrid discrete choice departure-time and duration model for scheduling travel tours. Transportation Research Record 1894, 46–56.