MODELING INDIVIDUALS' TRAVEL TIME AND MONEY EXPENDITURES

UCD-ITS-RR-00-22

March 2000

by

Cynthia Chen
Institute of Transportation Studies
And
Civil and Environmental Engineering
University of California, Davis 95616, USA
Fax (530) 752-6572
qzchen@ucdavis.edu

and

Patricia L. Mokhtarian
Department of Civil and Environmental Engineering
and
Institute of Transportation Studies
University of California, Davis 95616, USA
Davis, CA 95616
Tel (530) 752-7062
Fax (530) 752-7872
plmokhtarian@ucdavis.edu

Institute of Transportation Studies
One Shields Avenue
University of California
Davis, California 95616
Tel: 530-752-0247 Fax: 530-752-6572
http://www.its.ucdavis.edu/
email: itspublications@ucdavis.edu

Modeling Individuals' Travel Time and Money Expenditures

Cynthia Chen
Institute of Transportation Studies
and
Department of Civil and Environmental Engineering
University of California, Davis 95616 (USA)
(530) 752-6572 (fax)
qzchen@ucdavis.edu

and

Patricia L. Mokhtarian
Institute of Transportation Studies
and
Department of Civil and Environmental Engineering
University of California, Davis 95616 (USA)
(530) 752-7062 (phone), (530) 752-7872 (fax)
plmokhtarian@ucdavis.edu

March, 2000

This work is funded by the DaimlerChrysler Corporation.

Table of Contents

1. Introduction 1

2. Single Linear Equation 2

3. Systems of Equations 6

4. Application of Linear Models and Ordinal Multinomial Models to Relative Desired Mobility 12

5. Duration Analysis 15

6. Utility Maximization Framework 22
6.1. Previous Approaches 22
6.2. Proposed Approach 33

7. Data Needs 40

8. Discussion 44

References 49

Appendix 53
Table 2: Information on Selected US Data Sets 54
Table 3: Components of Selected US Data Sets 55

1. Introduction

In an earlier report, we examined the constancy of travel time and money expenditure by reviewing the empirical evidence at both the aggregate and disaggregate levels (Chen and Mokhtarian, 1999). That report concluded that, although some regularities have been noted at the aggregate level, the considerable variation observed at the disaggregate level does not support the theory of a constant travel time budget. However, individual travel time and money expenditures may be influenced by a number of variables and hence be capable of being modeled with some degree of accuracy. The objective of this report is to explore different approaches to modeling individuals' time and money allocations to travel. Although the focus is on travel, it is important to consider activities as well, due to possible trade-offs between activities and travel. In this report, we consider three categories of activities: mandatory (e.g. paid work), maintenance (e.g., grocery shopping, medical appointments) and discretionary (e.g., social, recreational) activities. These three categories in general encompass all daily activities.

In this report, we consider the ideal study period to be relatively long – for example, a week or a month or even a year. This would hopefully allow us to capture activities and travel that people do not conduct on a daily basis. Examples include long distance business travel and vacation travel. In addition, in using a rather long study period, we avoid the situation where the amount of time allocated to a particular type of activity is zero[1].

[1] In the utility maximization framework, for mathematical tractability it is assumed that the amount of time allocated to each type of activity is greater than zero.

In this report, we discuss five different approaches to disaggregate modeling of time and money allocations to travel and activities: four statistical estimation techniques, and the utility maximization framework within which each of the four statistical techniques may be applied. In Section 2, we present the single linear equation approach, which assumes a single endogenous variable. Where there is more than one endogenous variable, seemingly unrelated regression equations (SUR) or structural equations modeling, described in Section 3, are more appropriate. In Section 4, we discuss the application of linear and ordinal multinomial models to model relative desired mobility. Duration analysis, discussed in Section 5, may also be used to model individuals' time allocation behavior. In Section 6, we present the utility maximization framework and propose a modification of this approach as it has been developed to date. Data needs are discussed in Section 7. Discussion and conclusions come in Section 8.

2. Single Linear Equation

The simplest way to model individuals' expenditure on travel is via the use of a single linear equation:

$$y = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon} ,$$

where

y (n cases × 1) could be time or money expenditure on travel,

$\mathbf{x}$ (n × k variables) are explanatory variables for the corresponding dependent variable,

$\boldsymbol{\beta}$ (k × 1) is a vector of parameters corresponding to the explanatory variables, and

ε (n × 1) is a vector of random disturbance terms for the corresponding equation.

Similar equations could be separately developed for expenditures on each activity type.

As a special type of linear regression, a few studies have used analysis of variance to identify significant factors associated with travel time expenditures. Using data from a 1976 three-weekday trip diary in Munich, Germany, Zahavi and Talvitie (1980) show that household size, car ownership and the interaction between household size and car ownership have statistically significant effects on travel time expenditures.

Kitamura et al. (1992) separately estimated a number of log-linear models in which time expenditure on the *j-th* activity is a function of total time available and other explanatory variables, as shown below[2]:

$$\ln t_j = \theta_j \ln T + \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \ldots + \varepsilon_j ,$$

where

t_j is the time allocated to activity j,

T is the total time available, measured as 24 hours minus work duration,

x*'s* are explanatory variables, and

θ_j and β*'s* are parameters to be estimated.

The main purpose of the model estimation was to examine the hypothesis: $\theta_j = 1$. In other words, whether the ratio of time expenditures allocated to two activities is invariant to the total amount of time available. They found that the results rejected the hypothesis of proportional time allocation to activities.

[2] To deal with cases where $t_j = 0$, a value of 0.5 was added to all cases and the equation was then estimated by the Ordinary Least Squares (OLS) method.

Levinson (1999) estimated single linear equations (using OLS) on daily travel duration for activity *i*. The independent variables included daily frequency and daily duration of activity *i*. The model was estimated for different types of activities including home, work and related, shopping, personal business, school and church, doctor visits, visits to friends and relatives, social/recreation and other activities. The results showed that the activity frequency had a significant and positive effect on travel time allocation to all types of activities. Except for work and related activities, activity duration had a significant effect on travel time expenditure to the corresponding activity type. Travel time expenditure decreased as the amount of time spent on home activities increased; increased as the amount of time spent on all other activities (except for work and related activities) increased. The insignificant relationship between work duration and travel time to work is probably due to the high level of fixity of work duration. In other words, work duration is relatively constant (8 hours a day for most full-time workers) no matter how long one has to travel to work.

One problem with Levinson's model is that if activity duration is endogenous, the OLS estimates will be inconsistent. In reality activity duration is more likely to be endogenous than to be exogenous as activity duration may also be determined by travel time duration. In the case where some regressors are endogenous, the 2 Stage Least Squares (2SLS) method may be used. In an effort to model how individuals allocate their time throughout a day, Ma and Goulias (1998) developed models for travel time expenditures on different types of activities. 2SLS was used to estimate the model due to the expected endogenity of activity duration. Ma and Goulias found that "activity duration is endogenous to travel time only when a person travels to participate in a subsistence activity". In the model of travel time to subsistence (i.e., mandatory) activities (the only estimation model that is presented), both work-related characteristics and person and household-related socio-economic characteristics had significant effects on travel time. Full-time workers and those who lived farther away from work traveled longer than others. However, Ma and Goulias also found that working outside of home decreased the travel time to subsistence activities, a result that appeared to contradict the outcome just mentioned. Travel time to subsistence activities was found to be negatively related to home departure time, amount of time spent on past activity participation and travel on the same day, and number of activities conducted earlier on the same day. Additionally, travel time to subsistence activities was also positively related to the travel time to a previous subsistence activity on the same day and negatively related to the travel time to a previous leisure activity on the same day.

Another problem with Levinson's model is that, if a large number of observations in the data set have zero travel time, the OLS estimates will be inconsistent. One ad hoc solution is to add a small positive quantity to all observations, as Kitamura et al. (1992) did. A more rigorous approach is to estimate a tobit model, which is appropriate for cases in which the dependent variable is censored from below and/or above. Flood (1985) used a tobit model approach to account for the problem of a large number of zero observations in modeling the time expenditures on market work, home production and leisure by males and females. He found that the non-labor income decreased the amount of time allocated to market work for both males and females; it however increased males' time allocated to leisure activities. The wage rate had an insignificant effect on males' time allocation, but significantly increased females' time allocated to market work. A high level of education significantly increased females' time allocated to market work, but reduced their time allocated to leisure activities. Like the wage rate, age had an insignificant effect on males' time allocation, but significantly increased females' participation in market work. Being a home-owner did not seem to make a difference in males' time allocation, but significantly reduced females' participation in market work. Larger households placed a burden on both female and male adults; it significantly reduced females' time allocation to market work, increased their home production time, and reduced males' time allocated to leisure activities. The availability of household technology, surprisingly, increased females' time allocation to home production. The presence of children less than 5 years old significantly reduced females' participation in market work and increased their time on home production.

3. Systems of Equations

Within our context, if we were to use the single linear equation approach for both time and money expenditures, and including the three activity types (mandatory, maintenance, and discretionary) as well as travel, we would be modeling eight equations separately. These eight single linear equations may well share some explanatory variables, both observed and unobserved. We may gain some efficiency by modeling them jointly. This is especially the case when the ε's (random disturbance terms) are correlated with each other and/or correlated with explanatory variables. Even if all explanatory variables on the right hand side of the equation are exogenous variables, the ε's could be correlated with each other. This is because the random disturbance terms may not only include factors that are specific to a particular equation, but also factors that are common to more than one equation. In the case where endogenous variables are present on the right hand side, the ε's are not only correlated with each other across equations but also correlated with explanatory variables within the equation.

In the case where all explanatory variables are exogenous but the *ε's* are correlated across equations, the set of equations is called a Seemingly Unrelated Regression Equation (SUR) System, which can be estimated using the Generalized Least Squares (GLS) method. Efficiency over OLS is especially gained by using GLS when the correlation among the *ε's* is substantial. When the correlation is equal to zero, the GLS estimates are identical to the OLS estimates.

In the case where endogenous variables are present on the right hand side, the set of equations is called a Structural Equations System, which can be expressed as:

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \zeta ,$$

where

$\mathbf{y}$ (m variables by 1) is a column vector of endogenous variables and $\mathbf{x}$ (n variables by 1) is a column vector of exogenous variables,

$\mathbf{B}$ (m by m) is a matrix of parameters representing direct causal links between endogenous variables,

Γ (m by n) is a matrix of parameters representing direct causal links of exogenous variables to endogenous variables, and

ζ (m by 1) is a vector of random disturbances, $\mathbf{E}[\zeta] = \mathbf{0}$ and $\mathbf{E}[\zeta\zeta'] = \Sigma \neq \mathbf{0}$.

In structural equations systems, endogenous variables are not only directly influenced by the right-hand variables (both endogenous and exogenous) in its own equation, but also indirectly influenced by variables in other equations (through the influence of those variables on the endogenous variables of those equations). The presence of endogenous variables on the right hand side means that the endogenous variables are correlated with the disturbance terms, in violation of the assumption of OLS. Using OLS to estimate a SES will result in inconsistent estimates. Thus, SESs are estimated using the 3-Stage Least Squares (3SLS) method or Full Information Maximum Likelihood (FIML) method.

For any endogenous variable, the direct causal effects through the endogenous and exogenous variables on the right hand side of its own equation are called direct effects. Therefore, from the equation above, the direct effects of exogenous variables constitute the elements of Γ and the direct effects of endogenous variables are found in the matrix **B**. The effects of variables mediated by other equations in the system are called indirect effects. Mueller (1996) demonstrated that the matrix of indirect effects of endogenous variables on endogenous variables is the sum of an indefinite matrix series:

$$IE_{yy} = \mathbf{B}^2 + \mathbf{B}^3 + \mathbf{B}^4 + \cdots.$$

Similarly, the matrix of indirect effects of exogenous variables x on endogenous variables y can be expressed as follows:

$$IE_{yx} = \mathbf{B}\Gamma + \mathbf{B}^2\Gamma + \mathbf{B}^3\Gamma + \mathbf{B}^4\Gamma + \cdots = (\mathbf{B} + \mathbf{B}^2 + \mathbf{B}^3 + \mathbf{B}^4 + \cdots)\,\Gamma$$

Mueller (1996, p.142) also noted that "for any *recursive* structural equation models involving *NE* latent endogenous variables, the matrix $\mathbf{B}^{\mathbf{NE}}$ (and all subsequent powers of **B**) always will be equal to the **0** matrix, guaranteeing that the series in [the above two equations] converges." The sum of direct and indirect effects on an endogenous variable is called the total effect (Bollen, 1989). As a summary, these direct and indirect effects correspond to entries in the following table.

Table 1: Direct, Indirect and Total Effects of General Structural Equation Models (Mueller, 1996, p. 144)

Effect Component	Exogenous $\rightarrow$ Endogenous	Endogenous $\rightarrow$ Endogenous
Direct (DE)	$\mathbf{\Gamma}$	$\mathbf{B}$
Indirect (IE)	$\mathbf{(I - B)^{-1}\Gamma - \Gamma}$	$\mathbf{(I - B)^{-1} - I - B}$
Total (TE)	$\mathbf{(I - B)^{-1}\Gamma}$	$\mathbf{(I - B)^{-1} - I}$

A number of researchers have used structural equation systems to estimate models of time expenditure on activities and travel. Flood (1985) developed four structural equation systems to examine time expenditures on various activities by male and female adults in the household. The first system for home-related activities consisted of eight single equations, corresponding to home production, leisure, household work, and TV-watching activities by males and females respectively. Estimation of such a system was performed by the 2 Stage Least Squares method (2SLS). The second system modeled time expenditures on market work. Due to a large number of zero observations for market work for females, estimating the system using only observed values of time spent on market work for women would have resulted in biased estimators. Hence, a latent indicator, Z, that can take on any real value and is related to Y, the observed time, by the last equation is used. That is, the second system is expressed as follows:

$$Y_m = \gamma_m Z_f + \mathbf{X_m B_m} + \varepsilon_m ,$$

$$Z_f = \gamma_f Y_m + \mathbf{X_f B_f} + \varepsilon_f ,$$

$$Y_f = \max(0, Z_f) ,$$

where

Y is the dependent variable, observed time spent on market work,

Z is a continuous latent variable, defined by Flood as latent preference for time allocated to market work,

$\boldsymbol{X}$ is a vector of explanatory variables,

$\boldsymbol{B}$ is a vector of parameters,

ε is the random disturbance term,

subscript *m* refers to males, and

subscript *f* refers to females.

Similarly, due to a large number of zero observations for child care and household repair activities, the third (for child care) and the fourth (for home repair) systems are expressed as follows:

$$Z_m = \gamma_m Z_f + \mathbf{X_m B_m} + \varepsilon_m ,$$

$$Z_f = \gamma_f Z_m + \mathbf{X_f B_f} + \varepsilon_f ,$$

$$Y_f = \max(0, Z_f) \text{, and}$$

$$Y_m = \max(0, Z_m) .$$

Estimation of the second, third, and fourth systems is similar to the 2SLS method. First, maximum likelihood was used to estimate reduced forms of the equations in the system to obtain predicted values of Y_m, Z_f in the second and third systems and Z_m and Z_f in the fourth system. Then, replace Y_m, Z_f in the second and third systems and Z_m and Z_f in the fourth system with their predicted values, and then estimate structural parameters by maximum likelihood. Flood found that there was no substantial gain in treating the allocation of time in the household as a system. The results estimated from structural equations systems were essentially the same as those from separate estimation of single linear equations. In general, females' time allocation had no significant effect on males' time allocation behavior. Males' time allocation had a significant effect only on females' leisure and home repair activities.

Golob (1990) examined how travel times by different modes interacted with each other and with car ownership over time using a longitudinal structural equation system with limited and categorical dependent variables. He found that there were significant associations among travel times by different modes and with car ownership. Golob also found significant impacts of exogenous variables on travel times by different modes and car ownership both at the same point in time and in the previous year[3]. The exogenous variables Golob examined included dynamic variables that were measured both at the same point in time and the previous year, and static variables that were measured only at the same point in time. Dynamic variables included two variables related to income, number of persons 18 or older in household, number of persons 12-17 in household, household composed of 2 adults, presence of children less than 12 years old, number of household drivers, presence of 3 or more drivers, and number of household workers. Static variables included four variables related to residential location.

Fujii et al. (1997, cited by Kitamura et al., 1997) developed a structural equation system analyzing trade-offs between time expenditures on activities and travel. They found that a 10-minute reduction of commute time would increase average total out-of-home activity duration by 1.88 minutes, average total in-home activity duration by 7.11 minutes, and average total travel time by 0.36 minutes. The number of home-based trip chains after returning home from work would increase about 30%, from 0.03 to 0.04.

Golob and McNally (1997) estimated a structural equation model system examining the trade-off in time expenditure on different activities (work, maintenance, and discretionary) and corresponding travel to each type of activity, separately by females and males residing in the same household. Similar to the study by Golob (1990), they not only found significant associations among dependent variables, but also significant impacts of exogenous variables on dependent variables[4].

Lu and Pas (1999) examined the interaction between individuals' activity participation and travel behavior. They found that daily travel time increased with the amount of time spent on maintenance and out-of-home activities, but decreased with the amount of time spent on in-home activities. As for socio-demographics, total daily travel time was positively related to age, income, and number of workers, and negatively related to number of vehicles and number of children. The likely explanation for the relationship to number of vehicles is that households with fewer vehicles must rely more on slower transit and walk modes, resulting in longer travel times.

4. Application of Linear Models and Ordinal Multinomial Models to Relative Desired Mobility

Theoretical and empirical work in this area to date has focused on analyzing observed travel time and money expenditures. Mokhtarian and Salomon (forthcoming), on the other hand, hypothesized the existence of a "desired or ideal travel time budget", which varies at the disaggregate level as a function of personality, lifestyle, travel-related attitudes, stage in lifecycle, and other socio-economic and demographic variables. In practice, it would be difficult to obtain a quantitative measure of the ideal travel time budget because respondents may not

---

[3] For detailed results, please refer to Chen and Mokhtarian (1999).

articulate such a concept to themselves. However, in a recent data collection effort carried out by UC Davis, a measure of relative desired mobility was obtained. Specifically, in the survey, respondents were asked to indicate whether they would prefer to travel much less, less, about the same, more, or much more compared to what they do now[4]. This measure was obtained separately for short-distance and long-distance travel "overall", and by purpose and mode categories.

It would be useful to model this relative desired mobility, to increase our understanding of people's likely reaction to developments (whether technological, policy-based or personal) that make it easier or harder to travel. For example, people who want to travel much more than they do now would react differently than those who want to travel much less, to the increased availability of new urban mixed-use developments. The simplest modeling method is to use the single equation approach in which we set one of the observed measures of relative desired mobility as the dependent variable and link this dependent variable to a set of independent variables. Alternatively, several or all of the mode-, purpose-, and distance-specific measures may be modeled simultaneously as a set of seemingly unrelated regressions. It is also possible to construct a conceptual model in which some of the explanatory variables for relative desired mobility (such as travel liking, and current mobility) are themselves functions of other variables and therefore endogenous. In this case, structural equations modeling would be the appropriate approach.

One problem with any of these approaches is that in prediction, the models permit predicted values of relative desired mobility outside the range of observed responses. The five-category

[4] The time frame to which the word "now" referred was not indicated.

ordered-response variable is being treated as an unrestricted continuous variable. A more rigorous way to model such variables is via ordered multinomial models. Consider the following equation for a latent variable $y*$:

$$y* = x'\beta + u\,,$$

where

$y*$ is a latent variable, which can be interpreted as the amount people want to travel relative to their current amount, measured on a continuous scale, and

u is a random disturbance term.

Thus, the observed outcome (i.e., the measured RDM) depends on how large $y*$ is. We can express that outcome as follows:

$$y = \begin{cases} \textit{much less}, & \textit{if } \; y* \le \alpha_1 \\ \textit{less}, & \textit{if } \; \alpha_1 < y* \le \alpha_2 \\ \textit{about the same}, & \textit{if } \; \alpha_2 < y* \le \alpha_3 \\ \textit{more}, & \textit{if } \; \alpha_3 < y* \le \alpha_4 \\ \textit{much more}, & \textit{if } \; y* > \alpha_4, \end{cases}$$

where

y is the observed measure of relative desired mobility, and

α_1, α_2, α_3, and α_4 are cut points to be estimated. Since the origin of the $y*$ is arbitrary, it is often convenient to set, say, $\alpha_1 = 0$ and estimate the remaining $\alpha\,s$ relative to that base. In our case, however, it would be natural to set the midpoint $\frac{\alpha_2 + \alpha_3}{2} = 0.$ This would allow interpretation of the $\alpha\,s$ as the relative amounts of travel corresponding to each qualitative label on the scale, where "more" would be positive and "less" would be negative. Thus, for example, it could be determined whether "much more" and "more" are closer together in the respondents' minds than "more" and "about the same" or than "less" and "much less", etc.

Two common examples of ordered multinomial models are ordered probit and ordered logit models. The ordered probit model is obtained by assuming that *u* takes a standard normal distribution while the ordered logit model is obtained by assuming that *u* takes a logistic distribution.

5. Duration Analysis

Expenditures of time on activities and travel may also be modeled via duration analysis. A key element in duration analysis is the specification of the hazard rate, "a rate at which spells are completed after duration t, given that they last at least until t" (Greene, 1993). Suppose the random variable *T*, the duration of the spell, has a continuous probability distribution *f(t)*, where *t* is the realization of *T*. Its cumulative probability distribution can be expressed as follows:

$$F(t) = \int_0^t f(s)ds = \text{Prob}(T \leq t)\,.$$

The probability that spells last at least *t* is given by the following function:

$$S(t) = 1 - F(t) = \text{Prob}(T \geq t)\,.$$

Then, the hazard rate function can be expressed as follows:

$$\lambda(t) = \lim_{\Delta \to 0} \frac{\text{Prob}(t \leq T \leq t + \Delta \mid T \geq t)}{\Delta}$$

$$= \frac{f(t)}{S(t)}\,.$$

The hazard rate function $\lambda(t)$ can take many forms, depending on the type of distribution assumed for *f(t)*. Commonly assumed distributions include exponential, Weibull, and log-logistic. For the exponential distribution, the hazard rate $\lambda(t) = \gamma$, where γ is a constant. In other words, the hazard function is memoryless; the rate at which the spell is completed does not depend on the duration of the spell. For the Weibull distribution, the hazard rate $\lambda(t) = \gamma \alpha t^{\alpha-1}$, where $\gamma > 0$ and $\alpha > 0$. Depending on the values of γ and α, the hazard rate function can be either monotonically increasing or decreasing, with the exponential distribution resulting as the special case when $\alpha = 1$. For the log-logistic distribution, the hazard rate $\lambda(t) = \gamma \alpha t^{\alpha-1} / (1 + t^{\alpha} \gamma)$, where $\gamma > 0$ and $\alpha > 0$. For $\alpha > 1$, the hazard function first increases with duration t and then decreases. For $0 < \alpha < 1$, the hazard function first decreases with duration and then increases. For $\alpha = 1$, the hazard function monotonically decreases with t.

The estimation of the hazard rate function can be done either parametrically or non-parametrically. In the parametric method, the duration density function is assumed to be $f(t,\theta)$, where t is the duration and θ refers to parameters to be estimated. The log-likelihood function may be expressed as $L = \prod_{i=1}^{n} f(t_i, \theta)$ for a sample of n completed spells. Given an assumed functional form of λ, consistent parameters can be estimated via the usual maximum likelihood procedure.

Sometimes, not only the duration t, but also other explanatory variables, affect the hazard function. For example, the hazard rate may be affected by the socio-economic characteristics of the individual. Kiefer (1988) summarized a number of specifications in which explanatory variables can be included. The simplest one is the proportional hazard model, which is expressed as follows:

$$\lambda(t, x, \beta, \lambda_0) = \Phi(x, \beta)\lambda_0(t),$$

where

$\lambda_0(t)$ is the baseline hazard function, corresponding to $\Phi(\cdot) = 1$.

Commonly, $\Phi(x, \beta)$ is specified as $\exp(x'\beta)$. With the proportional hazard model specification, the effect of explanatory variables is to multiply the baseline hazard function by a factor. In other words, the effect of an explanatory variable on duration is constant.

Another model specification in which explanatory variables can be included is called the accelerated lifetime model, which is specified as follows:

$$\lambda(t, x, \beta) = \lambda_0[t\Phi(x, \beta)]\Phi(x, \beta),$$

where

$\lambda_0(\cdot) = -d \ln S_0 / dt$,

$S(t, x, \beta) = S_0[t\Phi(x, \beta)]$, and

S_0 is the baseline survival function.

The accelerated lifetime model essentially rescales the time axis by $\Phi(\cdot)$. Kiefer (1988) commented that proportional hazard model specifications allow fairly general transformations of the duration variable but restrict the error distribution to only the type I extreme value distribution, whereas the accelerated lifetime hazard model specifications allow fairly general specifications of the error distribution but restrict the transformation of the duration variable.

Neither proportional hazard model specifications nor accelerated model specifications allow for interaction between the explanatory variables and the duration *t*, which sometimes may be too restrictive. Within our context, one may hypothesize, for example, that the effect of age on duration of travel time may become stronger with the length of the spell. To remedy this problem, the model specification for the hazard rate function may be expressed as follows (Kiefer, 1988):

$$\lambda(t, x, \beta) = \exp[g(t, x, \beta)].$$

The above specification allows an explanatory variable to have a different effect on duration at one point in time than at another.

There have been several applications of duration models to travel behavior analysis. Hamed and Mannering (1993) used a hazard rate function with a Weibull distribution to model travelers' postwork home-stay duration. They found that the home-stay duration was positively related to number of workers in the household, and negatively related to the number of children in the household. If the individual arrived home between 9:00 am and 4:00 pm, the chance of participating in activities outside of home was greater than if the individual arrived at home at other times. If the individual arrived home between 6:00 pm and 8:00 pm, the chance of participating in activities outside of home was less than if the individual arrived home at other times. The estimated duration parameter was less than one, suggesting that the longer an individual stays at home, the less likely that he will participate in an activity outside of home.

In modeling the duration of shopping during the return home trip from work, Bhat (1996a) compared proportional hazard models with a Weibull baseline specification and with a non-parametric baseline specification. Within each specification, he also compared among models without heterogeneity, with gamma heterogeneity, and with non-parametric heterogeneity. He found that the parametric baseline specification provided biased estimates. Control of heterogeneity did not alleviate the problem of biased estimates, though ignoring heterogeneity did not underestimate the duration dependence. In conclusion, Bhat recommended using a non-parametric baseline model specification and testing for various distributions to control for heterogeneity, in preference to arbitrarily choosing a particular parametric baseline model specification.

The above discussion of duration analysis only concerns spells with a single exit. This may be undesirable in some situations under which spells can end in a number of ways. For example, the spell of a particular activity such as paid work could end at the start of a recreational activity or a shopping activity. The hazard rate for the transition from paid work to recreational activity may well be very different from that for the transition from paid work to shopping activity. Competing risk models have been developed to deal with spells with more than one exit. The hazard rate function is expressed as follows (Ettema et al., 1995):

$$\lambda_k(t) = \lim_{\Delta \to 0} \frac{\text{Prob}(t \le T \le t + \Delta, D_k = 1 \mid T \ge t)}{\Delta},$$

where

$\lambda_k(t)$ is the rate at which the spells will end at the *kth* exit, and

D_k is the dummy variable indicating whether exit k is chosen or not.

The proportional hazard and accelerated lifetime versions of competing risk models can be expressed as follows:

$\lambda_k(t, x, \beta) = \Phi_k(x_k, \beta_k)\lambda_{0k}(t)$ for the proportional hazard model, and

$\lambda_k(t, x, \beta) = \lambda_{0k}[t\Phi_k(x_k, \beta_k)]\Phi_k(x_k, \beta_k)$ for the accelerated lifetime model.

In estimation, either a non-parametric or a parametric distribution for the *k-th* specific baseline hazard rate function may be specified.

Competing risk models have also been applied in travel behavior modeling. Ettema et al. (1995) used competing risk models to model the activity duration and the type of activity for the new engagement. They compared both a generic model specification (a model that was generic to all types of activity) and an activity type-specific model specification. They found that the performance of the generic model was not as good as the activity type-specific model specification in terms of the goodness of fit ratios.

Bhat (1996b) estimated a joint model of outcome and outcome-specific hazards to model the duration of shopping and social/recreation activities of workers during the evening commute home. In the sample, the individual may choose to go directly home, to participate in shopping activities before returning home, or to participate in social/recreational activities before returning home. Bhat estimated two versions of the model: one assuming independence between activity type choice and activity duration and the other accommodating the potential correlation between activity type choice and activity duration. The parameter estimates for the activity type choice model were almost identical for both versions of the model. Older age increased the probability of choosing shopping activities, but decreased the probability of choosing recreational activities. Compared to females, males were more likely to participate in recreational activities than in shopping activities. The presence of children under eleven years old decreased the probabilities of choosing both shopping and recreational activities. Higher household income increased the probability of choosing recreational activities, but had no impact on shopping activities.

Availability of automobile and being able to depart from work before 4:00 pm increased the probability of choosing recreational activities. Long work duration increased the probability of going directly home.

For duration models for shopping and recreational activities, the parameter estimates agreed in sign, but differed in magnitude. The duration of recreational activities was positively related to being male, household income, and returning young adult (1 if the individual is an employed adult living with one or both parents) and negatively related to work duration. The duration of shopping activities was positively related to returning young adult and departure from work before 4:00 pm and negatively related to driving alone to work and work duration. Bhat also noted that the model accommodating the potential correlation between activity type choice and duration was better than the model assuming independence between the two because the estimated correlation coefficient was found to be significantly different from zero.

In an effort to model how individuals allocate their time throughout a day, Ma and Goulias (1998) developed a number of competing risk duration models in the form of an accelerated lifetime specification to model activity duration and probabilities associated with various activity types (including subsistence, maintenance, and leisure activities). Ma and Goulias (1998) argued that the traditional competing risk model has the assumption of independence between activity type and the time that the activity will terminate, which is not realistic. For estimation, they adopted the two-step approach of Cardell (1997; cited in Ma and Goulias, 1998), in which they first estimated a multinomial logit model for the probabilities of activity types in which to engage, followed by an activity duration model that included a log-sum term from the multinomial logit model.

6. Utility Maximization Framework

6.1. Previous Approaches

Under the utility maximization framework, individuals are assumed to make choices in order to maximize an underlying utility function. Utility maximization is usually not unrestricted; rather the utility is maximized subject to constraints (e.g., a budget constraint). In microeconomics, a consumer's selection of quantities of goods and services subject to a budget constraint is commonly expressed as follows:

Max.

$$U(\mathbf{X})$$

subject to:

$$\mathbf{PX} \leq y\ ,$$

where

U is the consumer's utility for the vector $\mathbf{X}$,

$\mathbf{X}$ is a vector of quantities of the goods and services in the choice set,

$\mathbf{P}$ is the corresponding vector of prices for the goods and services in the choice set, and

y is a scalar representing the total possible expenditure.

A number of empirical studies have been conducted to describe individuals' allocation behavior. Kitamura (1984) examined how individuals allocate time among various activities. He formulated the problem as follows:

Max.

$$U(t_1, t_2, ..., t_J) = \sum_{j=1}^{J} \xi_j V_j(t_j, x_j)$$

subject to:

$$\sum_{j=1}^{J} t_j = T\,,$$

$$t_j \geq 0,\ j = 1, 2, ..., J\,,$$

where

U is the total utility of the individual's time allocation behavior,

t_j is the amount of time spent on the *j-th* activity,

ξ_j is an unknown and random weight of the utility for activity j in calculating the total utility, $\xi_j > 0$,

V_j is the utility derived from the *j-th* activity,

x_j is a vector of exogenous variables characterizing the *j-th* activity, and

T is the total amount of time available.

Assuming $V_j(t_j, x_j) = \gamma_j f_j(x_j) \ln t_j$, where $\gamma_j > 0$ is an unknown but constant scale factor, and setting $\partial L / \partial t_j = 0$ (where L is the Lagrangian function), the optimal time allocation to the *j-th* activity can be derived as follows:

$$t^* = \frac{\xi_j \gamma_j f_j(x_j)}{\sum_{i=1}^{J} \xi_i \gamma_i f_i(x_i)} T\,,\ \ i, j = 1, 2, \ldots, J,$$

which can be further written as:

$$\ln(t_j^* / t_i^*) = \ln(\gamma_j f_j(x_j)) - \ln(\gamma_i f_i(x_i)) + \ln \xi_j - \ln \xi_i .$$

Suppose $x_j > 0, j = 1, 2, \ldots, J$, and let $f_j(x_j) = \prod_{k=1}^{K_j} x_{jk}^{\alpha_{jk}}$, $j = 1, 2, \ldots, J$; $\xi_j = e^{\eta_j}$, $j = 1, 2, \ldots, J$; and $(\ldots, \eta_j, \ldots) \sim MVN(0, \Sigma_\eta)$. Also suppose that for some base activity i, its utility function can be expressed as follows, by setting $\gamma_i f_i(x_i) = 1$:

$$V_i(t_i, x_i) = \ln t_i .$$

Let the first activity be the base activity for the normalization. Then $\ln(t_j^* / t_1^*)$ can be expressed as follows:

$$\ln(t_j^* / t_1^*) = R_j + \alpha_j^{'} X_j + \varepsilon_j , j = 1, 2, \ldots, J,$$

where

$$R_j = \ln \gamma_j ,$$

$$\alpha_j^{'} = (\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jK_j}) ,$$

$$X_j = (\ln x_{j1}, \ldots, \ln x_{jK_j})^{'} ,$$

$\varepsilon_j = \eta_j - \eta_1$, and

$(\varepsilon_1, \ldots, \varepsilon_J) \sim MVN(0, \Sigma_\varepsilon)$.

The above formulation can be estimated using the least squares method. The formulation assumes that all $t_j > 0$, which may not be a realistic assumption when J is not small or the study period is short. In other words, it is entirely possible for an individual not to perform some activities at all during the study period. This naturally leads to a situation where the dependent variable (time allocation to a particular activity) is censored at zero. Assuming $J = 2$ (where the two types of activities are mandatory and discretionary), Kitamura formulated a tobit model that accommodated zero time allocated to one of the activities. The estimation results using the 1977 Baltimore Travel Demand Data Set showed that both work-related variables and socio-economic variables were significant. More specifically, having a work location within the city of Baltimore or arriving at work after 9 A.M. reduced the time allocated to discretionary activities. Although not verified by comparing the commute times of those who allocated little time and those who allocated much time to discretionary activities, the significance of these two variables may suggest relatively long commutes by the former group either due to long distances or slow traffic speeds. The use of an automobile for the work trip had a positive impact on the time allocated to discretionary activities. Work duration had a negative effect on the time allocated to discretionary activities, which was quite expected. Time allocated to discretionary activities seemed to vary by day of week, with Friday being the highest among weekdays. In terms of socio-economic characteristics, availability of cars in the household and number of nonworkers increased the amount of time allocated to discretionary activities. Males spent more time on discretionary activities than females. Time allocated to discretionary activities decreased significantly with age and number of children in the household. Women with children between 5 to 15 years old spent more time on discretionary activities than did others, probably due to participating in activities with their children.

Mathematically, Kitamura's model can be readily applied to modeling travel time expenditure in our context. Conceptually, there is a problem in doing that. There may be relationships between activity duration and travel time expenditure that cannot be determined freely by the individual. For example, given home and work locations that are fixed over the short run, for a particular individual, an eight-hour work duration requires one hour's commute every day; no matter how much the individual likes or hates the commute, he can't change it over the short run. Such a linkage between activity allocation and travel allocation was not addressed in Kitamura's model as it was developed to model activity duration only.

Flood (1985) modeled the amount of time spent on home production, leisure, sleep and personal care, and market work activities by male and female adults in the same household. As the time spent on market work can be obtained from the total time available minus the sum of the time spent on the other three activities, Flood's model excluded the variable for the amount of time spent on market work. The household was assumed to maximize total utility, given by the following form:

Max.

$$U(t_{m1}, t_{m2}, t_{m3}, t_{f1}, t_{f2}, t_{f3}, y)$$

subject to:

$$py \le w_m (T - \sum_{i=1}^{3} t_{mi}) + w_f (T - \sum_{i=1}^{3} t_{fi}) + \mu ,$$

$$\sum_{i=1}^{3} t_{mi} \le T ,$$

$$\sum_{i=1}^{3} t_{fi} \le T ,$$

$$t_{k,i}, y \ge 0 \qquad k = m, f; \quad i = 1,2,3 ,$$

where

t_{m1}, t_{m2}, t_{m3} are time allocations to home production, leisure, and sleep and personal care activities by the male,

t_{f1}, t_{f2}, t_{f3} are time allocations to home production, leisure, and sleep and personal care activities by the female,

p is the price of consumption,

y is the total household consumption of goods and services,

w_m, w_f are the male and female wage rates,

T is the total time available per day, that is 24 hours, and

μ is the household non-labor income.

Following Becker (1965), Flood assumed $p = 1$ for all households and then converted the monetary budget constraint to one in terms of full income, expressed as follows:

$$y + w_m \sum_{i=1}^{3} t_{mi} + w_f \sum_{i=1}^{3} t_{fi} \leq Y \equiv w_m T + w_f T + \mu \,.$$

By using the indirect translog utility function (Christensen et al., 1975) and Roy's identity, Flood was able to derive closed-form demand functions. The explanatory variables included in Flood's model were mainly socio-economic characteristics related to the individual and the household. He found that the presence of children had a significant impact on females' time allocation behavior: with the presence of young children in the household, females spent almost two more hours on home production, 25 minutes less on sleeping/personal care, and an hour and 20 minutes less working. The effect on female time allocation of having one additional household member was the same in sign to that of the presence of young children, but of lesser magnitude. The largest effect of having an additional member was on the female's time allocation to home production, which increased by 42 minutes. Being a home-owner increased time allocation to home production and leisure for both males and females, and females' time allocation to sleep/personal care. However it reduced time allocated to market work for both males and females, and males' time for sleep/personal care. Compared to other variables, age and education had a minimal effect on time allocation behavior. Age had a negative effect on males' market work time but a positive effect on males' sleep/personal care time. Higher education increased females' market work time, but decreased females' leisure time and males' market work time. The negative relationship between education level and males' market work time was quite unexpected. At the same education level, females spent more time on market work than did their male counterparts.

The concern with Flood's model is similar to that of Kitamura's model, if it were to be applied in our context. Neither Kitamura nor Flood accounted for time allocation to travel. Kraan (1996), on the other hand, formulated a model that contained terms measuring the total distance traveled, including the distance to work, and total travel time (expressed as distance divided by average speed). Her model is expressed as follows:

Max.

$$T^{\beta} \cdot d^{\gamma} \cdot f^{\vartheta} \cdot T_H^{\omega} \cdot G^{\chi}; \quad \beta, \gamma, \vartheta, \omega, \chi \in (0,1)$$

subject to:

$$T + \frac{d}{v} + T_H = T_{tot} - T_w,$$

$$c_T \cdot T + c_d \cdot d + c_f \cdot f + G = Y + w \cdot T_w,$$

$$T, d, f, T, G \geq 0,$$

where

T is the total time spent on out-of-home/non-work activities (maintenance and leisure),

d is the total distance traveled, including the distance to work,

f is the frequency of all out-of-home/non-work activities,

T_H is the total time spent on in-home/non-work activities,

G is the total amount of money spent on consumption goods and services other than travel and out-of-home/non-work activities,

v is average speed,

T_{tot} is the total time budget, measured as 24 hours minus the hours needed for sleep,

T_w is the total amount of time spent on working,

c_T, c_d, c_f are unit costs of T, d, and f,

w is the wage rate,

Y is unearned income including benefits and interest, etc., and

β, γ, ϑ, ω, and χ are unknown parameters to be estimated.

Kraan's utility function included five terms: time allocated to out-of-home/non-work activities, total distance traveled, frequency of out-of-home/non-work activities, time allocated to in-home/non-work activities, and total amount of money spent on consumption goods and services. Her formulation sets the marginal utility with respect to each of these five arguments as positive and diminishing. This is based on the assumption that ceteris paribus, one would prefer to have more of each of these five arguments. For example, everything else being equal, one would prefer to spend more time on out-of-home/non-work activities, or travel to a farther destination, or perform more out-of-home/non-work activities, or spend more time on in-home/non-work activities or spend more money on consumption goods and services. The monetary budget constraint did not include those costs incurred with in-home/non-work activities such as utility bills; these in-home costs may be viewed as being combined with the cost of consumption G.

Through the above formulation, Kraan was able to derive closed-form non-linear demand functions for T, d, f, T_H, and G. Due to the unavailability of data, the empirical application of the above model was conducted by forgoing the monetary budget constraint. In her empirical application, the ϑ parameter restricting the frequency of all out-of-home, non-work activities was set to zero and consequently the decision variable f dropped out of the model. Under these restrictions, Kraan derived the following simple linear equations for time allocated to out-of-home/non-work activities, total travel time, and time allocated to in-home/non-work activities:

$$T = \frac{\beta}{\beta + \gamma + \omega} \cdot T_{tot},$$

$$\frac{d}{v} = t_t = \frac{\gamma}{\beta + \gamma + \omega} \cdot T_{tot},$$

$$T_H = \frac{\omega}{\beta + \gamma + \omega} \cdot T_{tot},$$

where

t_t is the total travel time, calculated as the ratio of distance to speed.

Without loss of generality, the slope parameters can be multiplied by any positive constant α. Setting $\alpha = \beta + \gamma + \omega$, the demand functions can be expressed as follows:

$$T = c_{out} + \beta \cdot T_{tot} + \varepsilon_1,$$

$$t_t = c_{tt} + \gamma \cdot T_{tot} + \varepsilon_2,$$

$$T_H = c_H + \omega \cdot T_{tot} + \varepsilon_3,$$

where

c_{out}, c_{tt}, and c_H are constants added to the demand function, and

ε_1, ε_2, and ε_3 are random disturbances in the demand functions.

Estimation of the above three demand functions was performed by minimizing $\sum \varepsilon^2$ with respect to $\beta, \gamma, \omega, c_{out}, c_H$, and c_{tt}, where $\sum \varepsilon^2$ can be expressed as follows:

$$\sum_k \varepsilon^2 = \sum_k (T^k - c_{out} - \beta \cdot T_{tot}^k)^2 + \sum_k (T_H^k - c_H - \omega \cdot T_{tot}^k)^2 + \sum_k (t_t^k - c_{tt} - \gamma \cdot T_{tot}^k)^2 ,$$

where k is the index for individual in the sample. The minimization was performed through ordinary least squares. After β, γ, and ω were estimated, they were normalized to ensure the sum to be equal to one.

In estimation, Kraan used the Netherlands Time Budget Survey Data of 1990. She estimated the demand functions for the entire sample and all activities, for the entire sample and only discretionary activities, and for various population groups by all activities and only discretionary activities. In the estimation that involved all activities and all subjects in the sample, she found that time allocations to out-of-home/non-work activities, in-home/non-work activities, and travel increased with the total time budget (measured as 24 hours minus the hours needed for sleep). The increase was the largest for out-of-home/non-work activities. The increases for in-home activities and total travel time were similar in terms of their magnitude. In the estimation that involved only discretionary activities and all subjects in the sample, Kraan found that the largest and positive effect of the total time budget was on in-home/non-work activities. Time allocated to out-of-home/non-work activities also increased with the total time budget, but (in contrast to the model including all activities) total travel time decreased with the total time budget.

For the estimation that compared different population groups and for all activities, Kraan divided the sample into six clusters based on employment status, including full-time workers, part-time workers, students, housewives, pensioners, and unemployed. Demand functions were estimated for each of these clusters. Additional demand functions were estimated for students living on their own (a subgroup within the students cluster) and for single workers, which includes single workers from both full-time workers and part-time workers. Kraan found that except for time allocated to in-home/non-work activities for single workers and total travel time for single workers, all estimated slope coefficients were significant at a 95% confidence level. Kraan also estimated the demand functions for only discretionary activities for the same of set of clusters. Again, she found that all slope coefficients were significant at a 95% confidence level, meaning a significant impact of the total time budget on all types of time allocation for all types of population groups.

Although incorporating time allocation to travel, Kraan's model has some limitations. First, like Kitamura's model, although in a different form, Kraan assumed a positive but diminishing marginal utility with respect to time allocated to out-of-home/nonwork activities, in-home/non-work activities, frequency of all out-of-home/nonwork activities, and total distance traveled. On the other hand, in Flood's model, the utility function is a flexible functional form that is approximated by a second order Taylor Series expansion. Obviously, a flexible functional form is preferable to a more constrained one, especially in an area of study that is still exploratory. The second problem of Kraan's model is shared with Kitamura's and Flood's model approaches if they were used to model travel time allocation. Kraan's model, although incorporating travel time, did not address the linkage that might exist between activity duration and travel duration. There is the purely accounting relationship that all time expenditures must sum to the total amount of time available, but there is no explicit acknowledgement of the relationship that out-of-home activities are necessarily accompanied by some amount of travel.

6.2. Proposed Approach

To incorporate the relationship between time allocated to both activities and travel, we propose a utility maximization framework following Evans (1972):

Max.

$$V(a_w, a_m, a_d, a_t, G)$$

subject to:

$$a_w + a_m + a_d + a_t = \tau ,$$

$$c_m a_m + c_d a_d + c_t a_t + G = w \cdot a_w + Y ,$$

$$a_t \geq b_w a_w + b_m a_m + b_d a_d , \qquad b_w, b_m, b_d \geq 0,$$

where

a_w is the time spent on working,

a_m is the time spent on maintenance activities,

a_d is the time spent on discretionary activities,

a_t is the time spent on travel,

G is the cost of other goods and services consumed,

τ is the total time available,

c_m is the unit cost of maintenance activities,

c_d is the unit cost of discretionary activities,

c_t is the unit cost of travel,

w is the wage rate,

Y is all unearned income including dividends, interest, etc., and

b_w, b_m, and b_d are the number of units of time spent on work, maintenance, and discretionary activities, respectively, associated with one unit of travel time.

In the above formulation, the first constraint is the time constraint while the second constraint is the monetary constraint. In the last constraint, we assume a linear inequality relating the time allocated to activities and the travel to engage in those activities. The linear specification is probably quite a simplification of reality, nevertheless it serves as a first step toward modeling the relationship between activity duration and travel time expenditure. This constraint hypothesizes that each unit of time spent on activity i requires at least b_i units of travel time. If the individual derives only negative utility from travel, he will not spend more than the required minimum on travel. If the individual also derives positive utility from travel (e.g., one may well derive positive utility from driving through Yosemite National Park), he may spend more than the required minimum on travel. In the discussion of this approach, we will set this constraint to an equality. The equality constraint, for the case where the individual wants to spend exactly the required minimum amount of time on travel, represents the boundary condition. When the equality constraint is applied even in the case where the individual wants to spend more than the required minimum amount of time on travel, the positive utility derived from travel is reflected in inflated estimates of the b's.

Using the Lagrange multiplier approach to constrained maximization, and assuming a_w, a_m, a_d and $a_t > 0$ (which is required to ensure the solution does not fall on a boundary), the relevant first order conditions are as follows:

$$V_w = \lambda - \mu w + kb_w,$$

$$V_m = \lambda + \mu\, c_m + kb_m,$$

$$V_d = \lambda + \mu\, c_d + kb_d\,,$$

$$V_t = \lambda + \mu\, c_t + k\,,$$

$$V_G = \mu\,,$$

$k \geq 0$; *either* $k = 0$ *or* $a_t = b_w a_w + b_m a_m + b_d a_d\,,$

where

V_w, V_m, V_d, and V_t, are marginal utilities of time expenditures on work, maintenance, discretionary activities and travel,

V_G is the marginal utility of monetary expenditure on other goods and services,

λ is the marginal utility of relaxing the time constraint by one unit, called the marginal utility of time,

μ is the marginal utility of relaxing the monetary constraint by one unit, called the marginal utility of money, and

k is the marginal utility of relaxing the constraint on the individual's allocation of time to travel by one unit.

Marginal utilities of time allocated to work, maintenance, and discretionary activities and travel are therefore functions of λ, μ, k, the corresponding b, and the corresponding unit cost or wage rate. The marginal utility of other consumption G is equal to the marginal utility of money. The last condition is a Kuhn-Tucker condition. $k > 0$ if the individual spends exactly the minimum amount of time on travel and $k = 0$ if the individual spends more than the minimum amount of time on travel. If the individual is willing to spend more than the minimum amount of time on travel, then any small increase or decrease in the minimum amount of time he must spend on travel will not alter his utility level.

In order to derive the demand functions, let us first examine the constraint: $a_w + a_m + a_d + a_t = \tau$, which can be rewritten as: $a_t = \tau - a_w - a_m - a_d$. In other words, we only need to solve the demand functions for a_w, a_m, and a_d. As noted earlier, we will also set the inequality constraint: $a_t \geq b_w a_w + b_m a_m + b_d a_d$, to an equality constraint: $a_t = b_w a_w + b_m a_m + b_d a_d$. Substituting a_t in the constraint: $c_m a_m + c_d a_d + c_t a_t + G = w \cdot a_w + Y$, we obtain the following equation:

$$(-w + c_t b_w)\, a_w + (c_m + c_t b_m)\, a_m + (c_d + c_t b_d)\, a_d + G = Y .$$

Let

$$p_w = -w + c_t b_w ,$$

$$p_m = c_m + c_t b_m ,$$

$$p_d = c_d + c_t b_d \text{, and}$$

$$p_G = 1 .$$

We then can re-write the constraint as: $p_w a_w + p_m a_m + p_d a_d + p_G G = Y$. Following the notation of Becker (1965), we may term p_w, p_m, and p_d as full prices of maintenance, discretionary and work activities, that is, the cost of the activity itself plus the cost of the required associated travel time. The full price of the work activity is negative, representing the net income earned by that activity. p_G is the price of other consumption goods, which is set to be 1. This revised constraint conforms to the usual monetary budget in classical microeconomics problems.

Following the approach of Deaton and Muellbauer (1980), any arbitrary cost function can be approximated by the following function, provided that $\sum_i \alpha_i = 1, \sum_j \gamma_{kj}^* = \sum_k \gamma_{kj}^* = \sum_j \beta_j = 0$:

$$\log c(u,p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_j \gamma^*_{kj} \log p_k \log p_j + \mu\beta_0 \prod_k p_k^{\beta_k}$$

where

$\log c(u,p)$ is the logarithm of the cost function,

u is the utility level, $0 \leq u \leq 1$,

p is a vector of prices for various goods and services, and

$\alpha_0, \alpha_k, \gamma^*_{kj}, \mu, \beta_0, \beta_k$ are parameters.

Any cost function has a fundamental property: $\partial c(u,p)/\partial p_i = q_i$, where q_i is the quantity of the *i-th* good or service or the duration of performing the *i-th* activity (the $\alpha\, s$ in our notation). $\partial c(u,p)/\partial p_i = q_i$ can be re-written as: $\dfrac{\partial \log c(u,p)}{\partial \log p_i} = \dfrac{p_i q_i}{c(u,p)} = w_i$, where w_i is the budget share of good i. From this property, Deaton and Muellbauer (1980) derived demand functions for the budget share of good i, called the Almost Ideal Demand System (AIDS). As our formulation of the model has conformed to the classical microeconomic problem, we can now apply the AIDS system in our context. The demand functions for a_w, a_m, a_d and G in the share form can be derived as follows:

$$\frac{p_w a_w}{Y} = \alpha_w + \gamma_{ww} \log p_w + \gamma_{wm} \log p_m + \gamma_{wd} \log p_d + \beta_w \log(Y/P),$$

$$\frac{p_m a_m}{Y} = \alpha_m + \gamma_{mw} \log p_w + \gamma_{mm} \log p_m + \gamma_{md} \log p_d + \beta_m \log(Y/P),$$

$$\frac{p_d a_d}{Y} = \alpha_d + \gamma_{dw} \log p_w + \gamma_{dm} \log p_m + \gamma_{dd} \log p_d + \beta_d \log(Y/P),$$

$$\frac{G}{Y} = \alpha_G + \gamma_{Gw} \log p_w + \gamma_{Gm} \log p_m + \gamma_{Gd} \log p_d + \beta_G \log(Y/P),$$

where

$$\log P = \alpha_0 + \alpha_w \log p_w + \alpha_m \log p_m + \alpha_d \log p_d + \frac{1}{2}\gamma_{ww}(\log p_w)^2 + \frac{1}{2}\gamma_{mw} \log p_m \log p_w +$$
$$\frac{1}{2}\gamma_{dw} \log p_d \log p_w + \frac{1}{2}\gamma_{wm} \log p_w \log p_m + \frac{1}{2}\gamma_{mm}(\log p_m)^2 + \frac{1}{2}\gamma_{dm} \log p_d \log p_m$$
$$+ \frac{1}{2}\gamma_{wd} \log p_w \log p_d + \frac{1}{2}\gamma_{md} \log p_m \log p_d + \frac{1}{2}\gamma_{dd}(\log p_d)^2, \text{ and}$$

$$\gamma_{ij} = \frac{1}{2}(\gamma_{ij}^* + \gamma_{ji}^*).$$

Setting $\gamma_{mw} = \gamma_{wm}$, $\gamma_{dw} = \gamma_{wd}$, and $\gamma_{dm} = \gamma_{md}$, $\log P$ can be written as follows:

$$\log P = \alpha_0 + \alpha_w \log p_w + \alpha_m \log p_m + \alpha_d \log p_d + \frac{1}{2}\gamma_{ww}(\log p_w)^2 + \gamma_{mw} \log p_m \log p_w +$$
$$\gamma_{dw} \log p_d \log p_w + \frac{1}{2}\gamma_{mm}(\log p_m)^2 + \gamma_{dm} \log p_d \log p_m + \frac{1}{2}\gamma_{dd}(\log p_d)^2.$$

In the above system of demand functions, parameters to be estimated include α_0, α_w, α_d, α_m, α_G, β_w, β_d, β_m, β_G, γ_{ww}, γ_{mw}, γ_{dw}, γ_{mm}, γ_{dm}, γ_{dd}, γ_{Gw} γ_{Gm}, and γ_{Gd}. Variables whose values are known include a_w, a_d, a_m, and G as well as Y. For p's, the c's are known but the b's are unknown. Therefore, in order to estimate the parameters listed above, one must first estimate the b's such that the p's become known. The b's may be estimated by the following function:

$$a_t = b_w a_w + b_m a_m + b_d a_d + \varepsilon.$$

Therefore, estimation of the above system of demand functions can be performed in two separate steps. In the first step, one estimates the b's and then, using estimated values of the b's in the demand system, other parameters of interest can be estimated.

In the actual estimation of the model, specification of the monetary constraint can be quite a problem. The wage rate is easy to determine but the unit costs of maintenance and discretionary activities and travel are difficult to determine due to the tremendous variation in terms of costs from individual to individual. The difficulty is also aggravated by the lack of data on the costs of different activities and travel. Most time use studies only collect information on subjects' time use, not their monetary expenditures. Therefore, in the actual estimation of the model, one may have to forgo the monetary constraint entirely (as Kraan did), in which case, the model formulation is expressed as follows:

Max.

$$V(a_w, a_m, a_d, a_t)$$

subject to:

$$a_w + a_m + a_d + a_t = \tau,$$

$$a_t = b_w a_w + b_m a_m + b_d a_d.$$

The two constraints can be consolidated into one by substituting the second constraint into the first time constraint. The consolidated constraint can be expressed as follows:

$$(1 + b_w) \cdot a_w + (1 + b_m) \cdot a_m + (1 + b_d) \cdot a_d = \tau.$$

The consolidated constraint again conforms to the classical microeconomic problem, by setting $p_w^{'} = 1 + b_w$, $p_m^{'} = 1 + b_m$, and $p_d^{'} = 1 + b_d$. Similarly, the demand functions for a_w, a_m, and a_d in the share form can be derived as follows:

$$\frac{p_w^{'} a_w}{\tau} = \alpha_w + \gamma_{ww} \log p_w^{'} + \gamma_{wm} \log p_m^{'} + \gamma_{wd} \log p_d^{'} + \beta_w \log(\tau / P'),$$

$$\frac{p_m^{'} a_m}{\tau} = \alpha_m + \gamma_{mw} \log p_w^{'} + \gamma_{mm} \log p_m^{'} + \gamma_{md} \log p_d^{'} + \beta_m \log(\tau / P'),$$

$$\frac{p_d^{'} a_d}{\tau} = \alpha_d + \gamma_{dw} \log p_w^{'} + \gamma_{dm} \log p_m^{'} + \gamma_{dd} \log p_d^{'} + \beta_d \log(\tau / P'),$$

where

$$\log P' = \alpha_0 + \alpha_w \log p_w^{'} + \alpha_m \log p_m^{'} + \alpha_d \log p_d^{'} + \frac{1}{2}\gamma_{ww}(\log p_w^{'})^2 + \frac{1}{2}\gamma_{mw} \log p_m^{'} \log p_w^{'} +$$
$$\frac{1}{2}\gamma_{dw} \log p_d^{'} \log p_w^{'} + \frac{1}{2}\gamma_{wm} \log p_w^{'} \log p_m^{'} + \frac{1}{2}\gamma_{mm}(\log p_m^{'})^2 + \frac{1}{2}\gamma_{dm} \log p_d^{'} \log p_m^{'} +$$
$$\frac{1}{2}\gamma_{wd} \log p_w^{'} \log p_d^{'} + \frac{1}{2}\gamma_{md} \log p_m^{'} \log p_d^{'} + \frac{1}{2}\gamma_{dd}(\log p_d^{'})^2.$$

Similarly, by setting $\gamma_{mw} = \gamma_{wm}$, $\gamma_{dw} = \gamma_{wd}$, and $\gamma_{dm} = \gamma_{md}$, we obtain the following function.

$$\log P' = \alpha_0 + \alpha_w \log p_w^{'} + \alpha_m \log p_m^{'} + \alpha_d \log p_d^{'} + \frac{1}{2}\gamma_{ww}(\log p_w^{'})^2 + \gamma_{mw} \log p_m^{'} \log p_w^{'} +$$
$$\gamma_{dw} \log p_d^{'} \log p_w^{'} + \frac{1}{2}\gamma_{mm}(\log p_m^{'})^2 + \gamma_{dm} \log p_d^{'} \log p_m^{'} + \frac{1}{2}\gamma_{dd}(\log p_d^{'})^2.$$

In the above system of demand functions, parameters to be estimated include α_0, α_w, α_d, α_m, β_w, β_d, β_m, γ_{ww}, γ_{mw}, γ_{dw}, γ_{mm}, γ_{dm}, and γ_{dd}. Variables whose values are known include a_w, a_d, and a_m. Similarly, in order to estimate the above demand system, the p's must be known. This implies that the b's must be estimated before the estimation of the demand system. Again, we suggest that the b's first be estimated from the function: $a_t = b_w a_w + b_m a_m + b_d a_d + \varepsilon$. Then, the parameters in the demand system can be estimated using estimated values of the b's.

7. Data Needs

Although differing in the way that models are estimated, for each of the approaches we have described, we would expect travel time and activity duration and expenditure to be a function of the same set of variables. Specifically, for application of any of these methodologies, ideally the following set of variables is needed:

- duration of travel over a study period,

- duration of activities over a study period,
- money allocation to travel,
- money allocation to activities,
- expenditures on other goods and services,
- personal and household characteristics,
- transportation network-related characteristics, and
- other variables including personality, lifestyle, and attitudinal variables.

A measure of the duration of travel over a study period is obviously unavoidable if one is interested in travel time allocation. Similarly, if one is also interested in monetary expenditure on travel, the cost of the observed travel needs to be measured. Collection of information on activities is mainly due to the belief in the existence of a linkage between activities and travel. In fact, a number of empirical studies have verified the existence of such a linkage. From the resource perspective, the linkage between activities and travel exists because every one of us faces finite budgets in terms of time and money. From the conceptual perspective, the linkage exists because engaging in certain activities comes with a travel “overhead”. Empirical evidence has also shown that variables identifying personal and household characteristics as well as transportation network-related characteristics are important in individuals’ travel time and money allocation behavior. Other variables such as attitudes variables may also play an important role in travel time and money allocation and warrant further investigation.

In a typical activity diary, duration of activities is usually measured and duration of travel can be derived. Similarly, in a typical trip diary, duration of travel is usually measured and duration of activities can be derived. Generally, a trip diary also collects out-of-pocket travel-related expenditures, such as parking fees, transit fares, and tolls. The operating cost of a personal vehicle may be calculated based on the mileage. As for activity-related costs, neither activity or trip diaries usually collect this information. The same applies to expenditures on other goods and services. In other words, to examine a complete picture of travel time and money allocation, a new data collection effort may be needed to collect information on activity-related costs and expenditures on other goods and services. Information on personal and household characteristics is usually collected along with either an activity or trip diary. Information on transportation network-related characteristics can be obtained from land use and travel surveys. If the researchers are interested in testing the significance of other variables (e.g., attitudinal variables), data collection on these variables may need to be initiated as they are not generally measured in travel or activity diary studies.

Although having access to all of the variables listed above would be ideal, the lack of some of them would not necessarily invalidate a modeling effort. For example, even though data on money allocation may not be available, it is still productive to analyze travel time allocation. Many interesting research questions on travel time allocation exist and these research inquiries well deserve a modeling effort. For example, one might want to investigate the applicability of duration models in our context. Or, one might want to simply apply our utility maximization framework with real-world data sets and test the theoretical framework.

Tables 2 and 3 in the Appendix list four selected available data sets in the US. The Nationwide Personal Transportation Survey (NPTS) collected information on individuals throughout the US while the other three are regional household surveys. All of these would permit (with varying levels of accuracy) the estimation of time expenditures on travel and activities. All but the NPTS have data on travel costs; none have data on activity costs. All contain some data on personal and household characteristics. Transportation network characteristics could be inferred for the three regional data sets. The NPTS and the Puget Sound data sets also contain a limited amount of attitudinal data.

The Puget Sound Transportation Survey is a panel survey that initially started in 1989. In a panel survey, information on sample households and individuals is collected at multiple times throughout a study period that usually lasts multiple years. As time progresses, the households and individuals who drop out are replaced by newly-recruited households and individuals with similar characteristics. Use of panel surveys has many advantages in travel behavior analysis and these advantages are readily applicable in our context.

By examining multiple measurements for the same observation unit, many unobserved factors can be controlled and thus more precise measurement of behavioral changes can be obtained (Kitamura, 1990). For example, typical cross-sectional studies might attribute the differences in travel behavior to age differences while in fact the difference should be attributed to a generation effect. Or if there were a period effect (e.g., the effect of oil embargo years on travel behavior), a typical cross-section survey cannot detect it.

Panel survey analysis can also be very useful in forecasting. The validity of applying results from a cross-sectional data set to forecast the future must be based on the following conditions

(Kitamura, 1990). First, “behavioral changes are instantaneous.” Second, “behavioral changes are symmetric, or reversible.” And last, the “behavioral relation is stationary (invariant over time).” Evidence from recent literature and our own observations of daily lives casts serious doubts on these conditions. Behavioral change over time is a gradual, dynamic adaptation to the stimuli and this process may involve time lags and asymmetry. Panel data sets can be used to model these dynamic behavioral changes more precisely than cross-sectional data sets.

The many advantages of panel data sets do not come without drawbacks. When using panel data sets in modeling, researchers must also handle problems such as attrition (households/individuals who drop out in later waves) and panel conditioning (the responses in later waves are influenced by responses in early waves). And these problems usually imply that more complicated modeling procedures ought to be used. Therefore, the decision on whether to use a panel data set must be weighed carefully in the modeling effort.

8. Discussion

In this report, we have discussed how a single linear equation, structural equations modeling, and duration analysis can be applied to model travel time and money expenditures, and how the first two of these techniques plus ordinal multinomial models can be applied to model relative desired mobility, as measured in another UC Davis study. Any one of these techniques can be used to estimate travel time expenditures in the context of micro-simulation. As an emerging approach to regional travel demand forecasting, micro-simulation models the entire activity and travel pattern of individuals within the region for the study period (e.g., a day). Once the activity and travel patterns are generated for each individual, activities in the same category (e.g., maintenance or discretionary) can be summed up to obtain the total amount of time spent on each category for each individual. Similarly, travel time can also be obtained.

An alternative to these applied approaches is a theoretical approach in which individuals are assumed to exercise a behavioral principle in decision making, subject to a number of constraints. Then, demand functions of travel time and money expenditures may be derived from such a framework. The utility maximization framework proposed in Section 6 of this report falls into the theoretical approach category. Here, individuals are assumed to maximize their utilities, which are functions of travel time expenditures. In maximizing their utilities, individuals are subject to a number of constraints. Once the demand functions are assumed, some of the same applied modeling techniques may be used to estimate the parameters[5]. The rest of this section discusses pros and cons of these different modeling techniques, and of the utility maximization approach itself.

The advantage of the single linear equation approach is its simplicity to estimate and interpret. The disadvantage is its inability to handle the potential association between activity duration and travel time expenditure correctly. Suppose one intends to regress a single linear equation with travel time expenditure as the dependent variable, but activity duration is endogenous to the process (i.e., a function of travel time). If OLS estimation were used, either including or excluding activity duration on the right hand side results in inconsistent and biased estimates. Due to the endogeneity with activity duration, one may use 2SLS in which one first regresses

[5] E.g., for the demand functions derived from the utility maximization framework in Section 6, either single linear equations, seemingly unrelated equations, or structural equations systems may be used.

activity duration against a number of exogenous variables and obtains the predicted values of activity duration and then regresses travel time expenditure using the predicted values of activity duration.

Due to the potential association between travel time expenditure and activity duration, it may be more insightful to examine both ends of the relationship. One may hypothesize that not only does activity duration affect travel time expenditure, but also vice versa. The single linear equation is incapable of examining this two-way relationship, even with 2SLS. Additionally, because 2SLS still estimates two single linear equations separately, the information contained in both equations is not fully utilized. To remedy this problem, one may regress equations for activity duration and equations for travel time expenditure simultaneously and this is where seemingly unrelated regression equations and structural equations modeling come into the picture. A seemingly unrelated equations system assumes exogeneity of the explanatory variables but allows correlation of error terms across equations. The structural equations system goes one step further; it allows endogeneity of explanatory variables.

All approaches discussed so far are linear regression models. With these models, one regresses one or more dependent variables against a set of variables (which can be both endogenous and exogenous). These models assume that the dependent variable is unrestricted and continuous. Sometimes, the dependent variable of interest has limited response categories: for example, the observed measure on relative desired mobility discussed earlier has only five response categories. In this case, it is best to use ordinal multinomial models. None of these models account for the dependence of the choice of whether to terminate travel on the duration of the endogenous variables themselves. This sometimes becomes undesirable because one may hypothesize that the longer a person has traveled, the more likely he is to want to terminate that travel. In this case, the likelihood that a trip will be terminated (and hence affect total travel time expenditure) depends upon how long the trip has lasted. Duration models are designed to account for such a dependence.

Despite their promising aspects, duration models are not without problems. In application, it is often assumed that not only the duration of the dependent variable itself, but also other explanatory variables affect the likelihood that the spell (dependent variable) will terminate. In order to obtain consistent estimates, the set of explanatory variables entered must be exogenous. This can hardly be the case if duration models were applied in our context to model travel time expenditure, due to the endogeneity of activity duration. Like the single equation approach, duration models are incapable of examining the two-way relationship between activity time expenditure and travel time expenditure, as can be done with a structural equations system. Another issue that arises if duration models were applied in our context is that the estimated travel time expenditure is the total duration of multiple spells of travel during the study period (e.g., a week). In other words, the travel time expenditure of interest is not continuous. This may well complicate the shape of the hazard rate curve.

Finally, the utility maximization framework itself is not without limitations. One common criticism is that the assumption of utility maximization is unrealistic. Indeed, due to limited information and information processing capabilities, individuals quite often do not maximize, but rather satisfy, their preferences. Within the utility maximization framework, the derivation of closed-form demand functions is sometimes not possible. This sometimes limits the way explanatory variables are entered into the utility function. In our context, we used a flexible functional form for the utility function and closed-form demand functions are then derived through approximation of the cost function.

In conclusion, then, no single approach to modeling travel time and money expenditures is dominant. Each has its pros and cons, and selection of a specific approach is at the discretion of the analyst. It would be interesting to compare different approaches using the same data. Flood (1985) used single linear equations, structural equations modeling, and utility maximization to analyze data on household allocation. He concluded that results (in terms of signs of coefficients) were similar for these approaches. It would be desirable to conduct additional comparative studies of this nature, but Flood's results suggest a certain amount of robustness with respect to the modeling approach taken.

References

Becker, Gary (1965) A Theory of the Allocation of Time. *The Economic Journal*, 493-517.

Bhat, Chandra (1996a) A Hazard-based Duration Model of Shopping Activity with Nonparametric Baseline Specification and Nonparametric Control for Unobserved Heterogeneity. *Transportation Research* **30B(3)**, 189-207.

Bhat, Chandra (1996b) A Generalized Multiple Durations Proportional Hazard Model with an Application to Activity Behavior During the Evening Work-to-Home Commute. *Transportation Research* **30B(6)**, 465-480.

Bollen, K. A. (1989) *Structural Equations with Latent Variables*. John Wiley, New York.

Cardell, N. S. (1997) Variance Components Structure for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity. *Economic Theory* **13**, 185-213.

Chen, C. and Mokhtarian, P. (1999) *A Review and Discussion of the Literature on Travel Time and Money Expenditures.* Research Report, Institute of Transportation Studies, University of California, Davis.

Christensen, L.; Jorgenson, D. W.; Lau, L. J. (1975) Transcendental Logarithmic Utility Functions. *The American Economic Review* **65(3)**, 367-383.

Deaton, A. and Muellbauer, J. (1980) An Almost Ideal Demand System. *The American Economic Review* **70(3)**, 312-326.

Ettema, Dick; Borgers, Aloys; and Timmermans, Harry (1995) A Competing Risk Hazard Model of Activity Choice, Timing, Sequencing and Duration. *Transportation Research Record* **1493**, 101-109.

Evans, A. W. (1972) On the Theory of the Valuation and Allocation of Time. *Scottish Journal of Political Economy*, 1-17.

Flood, Lennart (1985) On the Application of Time-use and Expenditures Allocation Models. Department of Economics, University of Gothenburg, Sweden.

Fujii, S.; Kitamura, R.; and Monma, T. (1997) A Study of Commuters' Activity Patterns for the Estimation of Induced Trips. *Journal of Infrastructure Planning and Management (Japan Society of Civil Engineers)* **562(IV-35)**, 109-120 (In Japanese).

Greene, William (1993) *Econometric Analysis*. Macmillan Publishing Company, New York.

Golob, Tom (1990) The Dynamics of Household Travel Time Expenditures and Car Ownership Decisions. *Transportation Research A* **24(6)**, 443-463.

Golob, Tom and McNally, Michael (1997) A Model of Activity Participation and Travel Interactions Between Household Heads. *Transportation Research B* **31(3)**, 177-194.

Hamed, Mohammad M. and Mannering, Fred L. (1993) Modeling Travelers' Postwork Activity Involvement: Toward a New Methodology. *Transportation Science* **27(4)**, 381-394.

Kiefer, N. M. (1988) Economic Duration Data and Hazard Functions. *Journal of Economic Literature* **XXVI**, 646-679.

Kitamura, Ryuichi (1984) A Model of Daily Time Allocation to Discretionary Out-of-Home Activities and Trips. *Transportation Research* **18B(3)**, 255-266.

Kitamura, Ryuichi (1990) Panel Analysis in Transportation Planning: an Overview. *Transportation Research* **24A(6)**, 401-415.

Kitamura, Ryuichi; Robinson, John; Golob, Thomas F.; Bradley, Mark; Leonard, John; and van der Hoorn, Toon (1992) *A Comparative Analysis of Time Use Data in the Netherlands and California.* Research Report Number UCD-ITS-RR-92-9, Institute of Transportation Studies, University of California, Davis, June.

Kitamura, Ryuichi; Fujii, Satoshi; and Pas, E. (1997) Time-use Data, Analysis and Modeling: toward the Next Generation of Transportation Planning Methodologies. *Transport Policy* **4(4)**, 225-235.

Kraan, Mariëtte (1996) Time for Travel. A Model for the Allocation of Time and Money. Unpublished Ph.D. Dissertation. University of Twente, Netherlands.

Levinson, David M. (1999) Space, Money, Life-stage, and the Allocation of Time. *Transportation* **26**, 141-171.

Lu, Xuedong and Pas, Eric I. (1999) Socio-demographics, Activity Participation and Travel Behavior. *Transportation Research* **33A**, 1-18.

Ma, Jun and Goulias, K. G. (1998) Forecasting Home Departure Time, Daily Time Budget, Activity Duration, and Travel Time Using Panel Data. *Paper Presented at the 77th Annual Conference of Transportation Research Board.* Washington, D. C., U.S.A.

Mueller, R. O. (1996) *Basic Principles of Structural Equation Modeling. An Introduction to LISREL and EQS*. Springer-Verlag New York, Inc., New York.

Zahavi, Yacov and Talvitie, Antti (1980) Regularities in Travel Time and Money Expenditures. *Transportation Research Record* **750**, 13-19.

Appendix

Table 2: Information on Selected US Data Sets

Data Sets	Location	Survey Year	Administration	Sample Size	Diary Period	Cost	Contact
Nationwide Personal Transportation Survey	USA	May, 1995 to July, 1996	Telephone interview	42,000 households (all hhld. members who are 5 years or older)	1 day	Free	http://www-cta.ornl.gov/npts/1995/download_table.shtml
Oregon and Southwest Washington	Portland, Oregon	Spring, 1994 to Winter 1995	Telephone and mail-back surveys	4,451 hhlds for RP data; 3,244 hhlds for SP data	2 days	Free to research organizations	Kyung-Hwa Kim at kimk@metro.dst.or.us
Bay Area Household Survey	9 countries including San Francisco, San Mateo, Santa Clara, Alameda, Contra Costa, Solano, Napa, Sonoma, and Marin	1996	Random Digit Dialing (RDD) recruitment, telephone reminder calls, mail-back surveys, and CATI data retrieval	5861 households	2 days	Free to research organizations	For additional information, contact MTC planning staff at 510-464-7700. Also see www.mtc.dst.ca.us/datamart/index.htm
Puget Sound Transportation Panel Survey 1989-1996	4 counties including King, Kitsap, Pierce, and Snohomish	1989-1996	Random Digit Dialing (RDD), and mail-back surveys	About 1700 households	1 day	Free to research organizations	For additional information, contact PSRC planning staff at 206-464-7964

Table 3: Components of Selected US Data Sets

Data Set	Types of Data Collected		Variables
Nationwide Personal Transportation Survey (NPTS)	Stated Preference		None
	Revealed Preference	Household	Household size, number of household vehicles, income, location
		Person	Age, gender, education, relationship within the household, driver status, annual miles driven if a worker
		Attitudes	Rating of potential problems in traveling, such as mobility, congestion, safety, traffic conditions, and pavement conditions
		Vehicle	Annual miles driven (based on odometer readings recorded typically two months apart), make, model, model year
		Trip level	Trip purpose, mode, length (in miles and minutes), time of day, vehicle characteristics (if a household vehicle was used), number of occupants, driver characteristics (for private vehicle trips only and if a household member was driving)
Oregon and Southwest Washington Household Activity and Travel Surveys	Stated Preference		Pricing effects (roads, congestion and parking)
			Residential location choice
			Automobile acquisition
	Revealed Preference	Household	Address, size, survey dates, structure, income, number of phone lines, number of cell or car phones, presence/absence of visitors on the survey date, tenure at the current address, zip code of previous address, own or rent, number of vehicles, shared phone lines, and transportation disability
		Person	Gender, race, English proficiency, employment status, age, household language, drivers' license, student status, employee-related information, and student-related information
		Activity	Type, location, starting and ending times, duration, accompanying young people
		Trip	Mode, starting and ending times, cost
		Vehicle	Year, make, model, type, year purchased, fuel type, ownership, purchased as a replacement or add-on, odometer reading at beginning of the 1st survey day and at the end of the 2nd survey day

Table 3: Components of Selected US Data Sets (Continued)

Data Set	**Types of Data Collected**		**Variables**
Bay Area Household Survey	Stated Preference		Pricing effects on cost and travel time
	Revealed Preference	Household	Household size, income, type of dwelling, address
		Person	Gender, age, driver's license, employment status, number of jobs, industry, occupation, length of employment, student status, student level, race, income
		Activity	Type, location, starting and ending times
		Trip	Trips regarding across the bay or not, mode, destination, starting and ending times, vehicle used, number of people accompanying, parking cost, parking location, transit route, fare, type of payment etc.
		Vehicle	Number of vehicles, make, model, year, fuel efficiency, number of bicycles
Puget Sound Transportation Panel Survey 1989-1996	Stated Preference		None
	Revealed Preference	Household	Household income, lifecycle stage, household size, number of adults, number of children in different age groups, number of household vehicles, change of residence, zip code, census tract, traffic analysis zone etc.
		Person	Age, gender, employment status, occupation, city code for work location, travel mode to/from work, number of work days per week, frequency that children are picked up, travel mode to/from school, frequency using bus per week, have transit pass or not, driver's license, parking costs, panel participation, occupation change code, workplace change code, work zip code, work census tract, work traffic analysis zone etc.
		Attitudes	Importance ratings of travel attributes (e.g., safety, on time), performance ratings of alternative travel modes (SOV, bus, carpool), agreement and disagreement statements related to features of alternative modes, importance ratings of alternative improvements in land use, transportation, and environment.
		Trip	Mode, starting and ending times, cost, vehicles used, trip origin census tract, trip origin traffic analysis zone, trip destination census tract, trip destination traffic analysis zone, travel distance etc.