Analysis of the Relationship between the Timing and Duration of Maintenance Activities Using Alternative Joint Discrete-Continuous Modeling Systems

Xin Ye (corresponding author)
Ph.D. Candidate
Department of Civil and Environmental Engineering
University of South Florida, ENB118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: (813) 974-8726
Fax: (813) 974-2957
Email: yehengxin@gmail.com

Ram M. Pendyala, Ph.D.
Professor
Department of Civil & Environmental Engineering
PO Box 875306
Arizona State University
Tempe, AZ 85287-5306
Phone: (480) 727-9164
Fax: (480) 965-0557
Email: ram.pendyala@asu.edu

Initial Submission: Aug. 6, 2006
Revision and Resubmission: Nov. 15, 2006

Word Count: 8,784 + 6 × 250 = 10,284

Submitted for Presentation Only
86th Annual Meeting of the Transportation Research Board
January 21-25, 2007, Washington, D.C.

ABSTRACT

This paper investigates the causal relationship between time-of-day choice and duration of maintenance activity episodes. An understanding of the causal relationship between these two choice behaviors may aid in the development of activity-based travel demand modeling systems that attempt to incorporate these two key aspects of activity behavior. The relationship between these two aspects of travel behavior is represented in this paper by considering two different causal structures: one structure in which activity duration is determined first and influences time-of-day choice, another structure in which time-of-day choice is determined first and influences activity duration. These two causal structures are estimated within two different modeling frameworks (Lee model vs. mixed model) that accommodate random error covariance in different ways. The analysis and model estimation are performed separately for non-commuter and commuter samples drawn from the 2000 Swiss Microcensus travel survey. The non-nested statistical test for discrete choice model is extended for rigorously comparing the alternative causal structures in joint discrete-continuous model. The Lee model estimation results indicate that the causal structure in which activity duration precedes time-of-day choice performs significantly better for both non-commuters and commuters. However, the mixed model estimation results indicate that the causal structure in which time-of-day choice precedes activity duration performs significantly better for non-commuters. For commuters, mixed model fails to give a conclusive result because the two causal structures are almost equally supported by the data.

Keywords: activity duration, activity timing, mixed logit model, causal structure, simultaneous equations, time of day, endogenous variable

1. INTRODUCTION

Activity-based model is able to explicitly address the importance of time in forming activity and travel pattern (Axhausen and Garling 1992). A critical advantage of activity-based model is that it is capable of integrating temporal dimension into travel modeling process (Pas and Harvey 1997). In the current context of transportation planning, travel demand management (TDM) and transportation control measures (TCM) are inherently linked to the temporal dimension of activity and travel pattern. Thus, activity-based approach that recognizes the temporal dimension can offer a more behaviorally sound modeling framework for conducting policy analysis and impact studies (Bhat and Koppelman 1999).

There are two critical aspects of temporal dimension that plays important role in activity-travel demand modeling: timing and duration of activity episode. Activity-based analysis allows one to answer the two critical questions: when does an activity episode start and for how long does the activity episode last? In recent years, activity-based research has focused on the analysis of individual activity episodes so that both of these aspects may be studied in detail (Bhat 1998; Bhat and Misra 1999).

Studies that focused on daily time allocations to various activity types were not able to address the time-of-day choice in activity engagement (Kasturirangan et al. 2002). Thus, conducting activity-based analysis at the individual activity episode level is crucial to gaining an understanding of the relationships between activity timing and duration (Steed and Bhat 2000). The causal relationship between activity timing and duration is an important component of activity-based travel demand modeling systems that aim to explicitly capture the temporal dimension (Pendyala et al 2002). A practical activity-based model system may incorporate submodels for activity timing and activity duration. It is important to identify the specification and application sequence of these submodels for developing an activity-based model system that can truly represent travelers' decision making process. However, the relationship between activity timing and duration are ambiguous. On the one hand, one may hypothesize that the timing of an activity affects its duration. Perhaps activity episodes pursued during peak periods are of short duration while those pursued in off-peak periods are longer in duration. On the other hand, the duration of an activity may affect its timing. Perhaps activities of longer duration are scheduled during the off-peak periods while activities of shorter duration are scheduled during peak periods.

This paper attempts to shed light on this relationship by exploring alternative causal structures in two different joint discrete-continuous modeling systems. The next section formulates these two joint modeling systems. The third section introduces the dataset and presents description of data for modeling analysis. The modeling estimation results from alternative causal structure in different modeling systems are presented and compared in the fourth section. The fifth section compares the performance of models under alternative causal structures in different modeling systems. Conclusions are drawn and discussions are made in the last section.

2. MODELING METHODOLOGY

2.1 Background

In this study, activity timing is modeled as discrete choice when activity beginning time is classified into a number of discrete categories (e.g. AM peak, PM peak, Midday, Off-peak), while the activity duration for each episode can be treated as continuous variable. Travelers may jointly make decisions on activity timing and activity duration but not all the influential factors regarding activity timing and activity duration can be observed. Thus, a modeling methodology is required to accommodate the random error covariance between models for timing and duration. Analogous to Seemingly-Unrelated Regression (SUR) model and Structural Equations Model (SEM) for continuous dependent variables (Greene, 2003), the covariance between random errors can be introduced into a joint model system. However, discrete choice is usually modeled in a logit-based modeling framework, where the random error terms are gumbel distributed. Unlike normal distribution, correlation cannot be accommodated into a joint distribution linking two gumbel distributions or linking one gumbel distribution and one normal distribution. From the perspective of multivariate statistics, there are infinite number of possible joint distributions given one gumbel marginal distribution, one normal marginal distribution and a constant correlation between them. As bivariate normal distribution can allow a constant correlation between its two marginal univariate normal distributions, Lee (1983) proposes a transformation that converts gumbel error terms into normal error terms so as to establish a bivariate normal distribution between discrete choices and continuous variable (called as Lee model hereinafter for brevity). Bhat (1998) applied Lee model for joint activity/travel behavior analysis.

2.2 Joint Discrete-Continuous Modeling System Based on Lee Transformation (Lee Model)

Pendyala and Bhat (2004) extended this modeling framework by specifying endogenous unordered discrete variables and endogenous continuous variables as explanatory variables in mutual model functions for modeling the causal relationship between these two types of endogenous variables. Similar to SEM, the joint estimation is necessary in this modeling system for consistently estimating the coefficient of the endogenous variables. The following is modeling formulation and estimation method based on Lee transformation adopted in Pendyala and Bhat (2004).

Let i be an index for alternatives in discrete choice set (i = 1, 2,…, I) and let q be an index for observations (q = 1, 2,…, Q). Consider the following equation system:

$$\begin{cases} u_{qi}{}^{*} = \beta_i' z_{qi} + \gamma_i a_q + \varepsilon_{qi} \\ a_q \;\; = \theta' x_q + \delta' D_q + \omega_q \end{cases} \tag{1}$$

ε_{qi} ~ i.i.d. Gumbel(0,1), $\omega_q \sim N(0,\sigma^2)$, where $u_{qi}{}^{*}$ is the latent utility associated with the i^{th} choice for the q^{th} observation, D_q is a vector of dummy variables of length I representing discrete choice, δ is a column vector of coefficients representing the effects of different discrete choice on activity duration, ε_{qi} is a standard gumbel distributed error term assumed to be independently and identically distributed (i.i.d.) across alternatives and observations, a_q is a continuous variable and γ_i is its coefficient. The error term ω_q is assumed to be i.i.d. normally distributed across observations with a mean of zero and variance of σ^2. In Equation system (1), the alternative i will be chosen (i.e., D_{qi} =1) if the utility of that alternative is the maximum of I alternatives. Define

$$v_{qi} = \max_{j=1,2,\ldots,I,\; j\neq i} u_{qj}^{*} - \varepsilon_{qi} \tag{2}$$

and $v_{qi}^{*} = \Phi^{-1}[F_i(v_{qi})]$, (3)

where $$F_i(y) = \frac{\exp(y)}{\exp(y) + \sum_{j\neq i} \exp(\beta_j' z_{qj})} \tag{4}$$

Then, Equation system (1) may now be rewritten as:

$$\begin{cases} D_{qi}^{*} = \Phi^{-1}[F_i(\beta_i' z_{qi} + \gamma_i a_q)] - v_{qi}^{*},\; D_q = 0 \text{ if } D_{qi}^{*} < 0,\; D_q = 1 \text{ if } D_{qi}^{*} > 0 \\ a_q = \theta' x_q + \delta' D_q + \omega_q \end{cases} \tag{5}$$

The detailed mathematical derivation can be found in Pendyala and Ye (2004). A correlation r_i between the error terms v_{qi}^* and ω_q is allowed to accommodate common unobserved factors influencing the discrete choice and the continuous variable. Since a_q is partially determined by ω_q and v_{qi}^* is correlated with ω_q if r_i is unequal to zero, a_q is correlated with random error term v_{qi}^* in the first equation. Similarly, D_q is also correlated with random error term ω_q in the second equation. The endogenous nature of dependent variables D_q and a_q entails the full-information maximum likelihood method to jointly estimate their coefficients γ and δ. Limited-information maximum likelihood estimation (recursive estimation) does not provide consistent estimators for the coefficients of endogenous variables.

The full-information likelihood function for estimating parameters in this case is equal to:

$$L = \prod_{q=1}^{Q}\left\{\prod_{i=1}^{I}\left[\frac{1}{\sigma}\varphi(l_q)\Phi(b_{qi})\right]^{D_{qi}}\right\}, \tag{6}$$

where $\phi(.)$ is the standard normal density function, and l_q and b_{qi} are defined as follows:

$$l_q = \left(\frac{a_q - \theta' x_q - \delta' D_q}{\sigma}\right),\ b_{qi} = \left(\frac{\Phi^{-1}F_i(\beta_i' z_{qi} + \gamma a_q) - r_i l_q}{\sqrt{1-r_i^2}}\right). \tag{7}$$

In addition, the condition of logical consistency only allows two alternative recursive structures (Pendyala and Ye, 2004). The first is the case where $\gamma \neq 0$ and $\delta = 0$: continuous variable → discrete variable, where continuous endogenous variable a_q is predetermined from the linear model and appear in utility functions u_{qi}^* as an explanatory variable for discrete choices. The second case is when $\gamma = 0$ and $\delta \neq 0$: discrete variable → continuous variable, where the vector of discrete variable D_q is predetermined by the utility functions u_{qi}^* and then serves as explanatory variables in the linear model for continuous variable a_q.

Please notice the underlying problem of discrete-continuous model system based on Lee's transformation. By plugging Equation (2) into Equation (3), one may have $v_{qi}^* = \Phi^{-1}[F_i(\max_{j=1,2,...,I,\ j\neq i} u_{qj}^* - \varepsilon_{qi})]$. (8)

Since u_{qj}^* incorporates ε_j, r_i is not exactly equal to corr(ε_i, ω) (i.e. the correlation between ε_i and ω) but a non-linear function with respect to both corr(ε_i, ω) and corr(ε_j, ω). With the presence of corr(ε_j, ω), positive (negative) r_i does not necessarily indicate a negative (positive) correlation between ε_i and ω. Thus, r_i does not have a straightforward behavioral interpretation. In addition, Schmertmann (1994) shows that the Lee model places substantial restrictions on the covariance between the continuous variable and discrete choice models. Lee model is not quite robust because its distributional assumption can be easily violated. One reason for this weakness is that the distributional assumption in Lee model is not only made on random error terms but also on systematic component of utility functions consisting of explanatory variables. For more robustness, we prefer to make distributional assumption only on random error terms rather than on explanatory variables. In the following subsection, we propose an alternative joint modeling system, called mixed discrete-continuous model, which is able to directly accommodate the correlation between random error term in each utility function and random error term in linear regression model. Since distributional assumptions are made only on the random components in the model, we believe the proposed mixed discrete-continuous model should be more robust than Lee model.

2.3 Mixed Joint Discrete-Continuous Model System (Mixed Model)

Model Formulation

The gumbel random error term adopted in the utility function for discrete choice model does not allow the correlation with the random error term in continuous model or in other utility functions for discrete choice. One approach to accommodate such correlations between discrete choices and continuous variable is to employ multinomial probit model for discrete choice, where the error terms are multivariate normally distributed instead of being gumbel distributed. However, logit-based discrete choice is being applied much more widely than multinomial probit model due to its convenience, thus logit model is still adopted for modeling discrete choice in this study.

Similar to the well-known nested logit model, one can assume that the random error term in utility functions consists of two independent random components: one represents a heterogeneity which may be normally distributed and the other is standard gumbel distributed as usual. Such modeling methodology for discrete choices is known as mixed logit model (Train, 2002). Similar to mixed logit model, we may formulate the following modeling system:

$$\begin{cases} u_{qi}^{*} = \beta_i' z_{qi} + \gamma_i a_q + f_i n_{qi} + \varepsilon_{qi} \\ a_q = \theta' x_q + \delta' D_q + k m_q \end{cases}, \tag{9}$$

where u_{qi}^{*} represents random latent utility function for alternative i of decision-maker q, z_{qi} and x_q are vector of exogenous variables and β_i and θ are respective coefficients. D_q is a vector of dummy variables indicating discrete choices and a_q is the continuous dependent variable. γ_i and δ are the coefficients of D_q and a_q in mutual models. $\varepsilon_{qi} \sim$ i.i.d. Gumbel(0,1), m_q is a random error term in the continuous model, which is standard normally distributed. For accommodating the error correlations, n_{qi} and m_q are assumed to be multivariate normally distributed with zero expectations and unit standard deviations. Correlations among n_{qi} are zero and correlations between n_{qi} and m_q are ρ_i. f_i and k represent the standard deviation of normal random components in utility functions and linear regression model. In this study, we emphasize the correlation between discrete choices and continuous variable but ignore correlations among discrete choices. Under the multivariate normality assumption, one may rewrite $m_q = \sum_{j=1}^{I}(\rho_j n_{qj}) + \sqrt{1-\sum_{j=1}^{I}\rho_j^2}\,\xi_q$, where ξ_q is a new random variable which is standard normally distributed and independent of n_{qi} and ε_{qi}. The model system (9) can be rewritten as:

$$\begin{cases} u_{qi}^{*} = \beta_i' z_{qi} + \gamma_i a_q + f_i n_{qi} + \varepsilon_{qi} \\ a_q = \theta' x_q + \delta' D_q + \sum_{j=1}^{I} k\rho_j n_{qj} + k\sqrt{1-\sum_{j=1}^{I}\rho_j^2}\,\xi_q \end{cases}. \tag{10}$$

By replacing $k\rho_j$ with g_j and $k\sqrt{1-\sum_{j=1}^{I}\rho_j^2}$ with σ, the mixed joint modeling system can be reduced to

$$\begin{cases} u_{qi}^* = \beta_i' z_{qi} + \gamma_i a_q + f_i n_{qi} + \varepsilon_{qi} \\ a_q = \theta' x_q + \delta' D_q + \sum_{j=1}^{I} g_j n_{qj} + \sigma\xi_q \end{cases}, \quad (11)$$

where n_{qi} ~ i.i.d. Normal(0,1) and ξ_q ~ Normal(0,1) and ε_{qi} ~ i.i.d. Gumbel(0,1). It implies that one univariate normal heterogeneity simultaneously appearing in both latent utility function and continuous model with unequal standard deviations performs as well as multivariate normal heterogeneities for consistently estimating the coefficient γ or δ of endogenous variables. Similar to subsection 2.2, either γ or δ needs to be zero, which leads to two alternative causal structures: 1) $\gamma = 0$ and $\delta \neq 0$, discrete choice → continuous variable and 2) $\gamma \neq 0$ and $\delta = 0$, continuous variable → discrete choice.

In this mixed joint modeling system, the correlation between latent utility function and random error term in continuous model can be calculated as

$$\mathrm{Corr}(u_{qi}^*, a_q) = \frac{f_i g_i}{\sqrt{(f_i^2 + \frac{\pi^2}{6})(\sum_{j=1}^{I} g_j^2 + \sigma^2)}} \quad (12)$$

As f_i and g_i approach positive or negative infinity, $\lim[\mathrm{Corr}(u_{qi}^*, a_q)]$ is equal to 1; meanwhile, as f_i approaches positive (or negative) infinity and g_i approaches negative (or positive) infinity, $\lim[\mathrm{Corr}(u_{qi}^*, a_q)]$ is equal to -1. Thus, theoretically speaking, this specification of heterogeneity can accommodate any degree of correlation between latent utility function and continuous model. And the correlation has reasonable behavioral interpretation that positive or negative correlation can explicitly indicate the same or the opposite impact of unobserved or unspecified common variables on latent utility function and continuous dependent variable. On this aspect, mixed model is more appealing than Lee model. In addition, the mixed discrete-continuous model specifies a heteroskedastic logit model for discrete choice, which can avoid the potential IIA (Independence of Irrelevant Alternatives) problem in multinomial logit model (Bhat, 1995).

Model Estimation

Based on the mixed joint modeling system as Equation (11), we can derive the probability function for each pair of discrete and continuous observation and then use maximum likelihood method to estimate the parameters. Conditional on the heterogeneity n_{qi}, the probability of each observation is equal to the product of probability for discrete choice observation and probability density for continuous observation, noted as:

$$\mathrm{Prob}(D_q, a_q|n_{qi}) = \left\{\prod_{i=1}^{I}\left[L_i(n_{q1}, n_{q2,...,} n_{qI})^{D_{qi}}\right]\right\} F(n_{q1}, n_{q2,...,} n_{qI}), \quad (13)$$

where $L_i(n_{q1,}n_{q2,...,}n_{qI}) = \dfrac{\exp(\beta_i z_{qi} + \gamma_i a_q + f_i n_{qi})}{\sum_{j=1}^{I}\exp(\beta_j z_{qj} + \gamma_j a_q + f_i n_{qj})}$ and (14)

$$F(n_{q1}, n_{q2},\ldots, n_{qI}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(a_q - \theta' x_q - \delta' D_q - \sum_{i=1}^{I} g_i n_{qi})^2\right]. \qquad (15)$$

To obtain unconditional probability, one needs to integrate n_{qi} over its distributional domain and then has

$$\text{Prob}(D_q,a_q) = \int_{n_{qI}=-\infty}^{+\infty} \ldots \int_{n_{q2}=-\infty}^{+\infty} \int_{n_{q1}=-\infty}^{+\infty} \left[\text{Prob}(D_q, a_q \mid n_{qi})\right] d\,\Phi(n_{q1})\, d\,\Phi(n_{q2})\ldots d\Phi(n_{qI}), \qquad (16)$$

where $\Phi(.)$ represents cumulative distribution function of standard normal distribution. The likelihood function can be formulated as

$$L = \prod_{q=1}^{Q} \Pr ob(D_q, a_q). \qquad (17)$$

Because the likelihood function does not have a closed form, Maximum Simulated Likelihood Estimation Method (MSLE) is applied to estimate the model parameters. The idea is to draw a random seeds from the known distribution, input these random seeds into probability function and approximate the integral value by averaging the calculated probability values. To save computational time, Halton quasi-random sequence (Bhat, 2001) is adopted for generating these random seeds. In this paper, the number of random draws is selected as 100 for all the model estimations.

Model Identification and Normalization

As stated by Walker (2001), a small number of quasi-random draws will mask the under-identification issue and yield erroneous estimators for mixed logit model therefore one has to carefully specify the heterogeneity in the mixed joint modeling system as Equation (11). Due to the slight difference between normal and gumbel distribution, standard deviation f_i of normal heterogeneity can be identified according to the differences between each pair of latent utility functions. However, in continuous model, the random error term ξ_q and heterogeneity n_{qi} are both normally distributed without any slight difference. The linear combination of normal random variables is still normally distributed, whose expectation and variance are respectively equal to the sum of expectations and the sum of variances associated with these normal random variables. Thus, the estimation of g_i should depend on an identification of f_i in latent utility function. Without identification of f_i, n_{qi} will be absorbed into ξ_q and g_i turns to be unidentifiable. If f_i is identifiable, then g_i is identifiable as well. The reason is that the identification of f_i only depends on latent utility functions for observed discrete choice without using any information from the observed continuous dependent variable. In the joint modeling system, covariance between each pair of latent utility function and continuous model can provide additional information for estimating g_i. In addition, continuous model itself can yield estimator for standard deviation of its whole random error term.

By examining the variance-covariance of utility differences, Walker (2001) established criteria for identification and normalization in a mixed logit model with flexible error structure. She found that a mixed heteroskedastic logit model with M (M > 2) alternatives at most allows (M-1) heterogeneities to be identifiable and the valid normalization is to impose zero on the smallest variance of heterogeneity. Practically, one may use a small number of quasi-random draws to estimate an unidentified mixed logit model and obtain preliminary estimation results. Then a zero restriction needs to be imposed on the smallest variance among all the estimated variances of heterogeneities. In the current mixed discrete-continuous modeling system, we initially impose zeros on all the g_i, keep all the f_i flexible and estimate an unidentified mixed logit model for discrete choices. Then, f_j with seemingly smallest absolute value is fixed at zero for normalization and identification. Since the corresponding g_j turns to be unidentifiable, it needs to be fixed at zero as well, but all the other f_i and g_i are set flexible in the final model estimation.

Simulation-based Hypothesis Test for Error Covariance

In the procedure of Maximum Simulated Likelihood Estimation (MSLE), t-test can be obtained for the estimator of each single parameter based on estimator itself and its variance from the estimated covariance-variance matrix. However, in this study, we are more concerned about the significance of the product of f_i and g_i instead of single parameter f_i and g_i since $f_i g_i$ represents covariance between two random components, which indicates sign (+ or -) and magnitude of the error correlation. There is a need to test the following null hypothesis (H_0) and alternative hypothesis (H_1):

H_0: $Cov(u_i, a) = f_i\, g_i > 0$ (positive covariance) ;
H_1: $Cov(u_i, a) = f_i\, g_i \leq 0$ (negative covariance).

As MSLE estimator $\hat{f}_i$ and $\hat{g}_i$ are essentially maximum likelihood estimator (MLE), they should be asymptotically normally distributed. The correlation between $\hat{f}_i$ and $\hat{g}_i$ can be calculated based on the corresponding off-diagonal element in the estimated covariance-variance matrix. As a result, f_i and g_i are bivariate normally distributed. One approach to calculate the probability to make type-I error (i.e. a correct null hypothesis is rejected) is to first derive the cumulative distribution function of $f_i g_i$ and then calculate the probability. However, it is rather challenging to derive a tractable cumulative distribution function for calculating this probability. In this study, a simulation approach is adopted to approximate this probability, which actually represents the significance level of error covariance.

Since the expectation and variance of the estimator for f_i and g_i and the correlation between them have been estimated in the procedure of MSLE, Monte Carlo method can be applied to generate a large number of two random seeds (denoted as f and g), which are bivariate normally distributed with estimated expectation, variance and correlation. Calculate the product of each pair of elements in f and g and then count the frequency of positive products, denoted as N_+. The probability to make type-I error (significance level) can be approximated by $(1 - N_+/N)$, where N is the total number of random seeds.

Similarly, if $\hat{f}_i \hat{g}_i$ is initially negative, null hypothesis that Cov(u_i, a) is negative needs to be tested. The significance level can be approximated by (1 – N_-/N), where N_- represents the count of negative products between f and g. In this paper, 5,000,000 pseudo random seeds (i.e. N = 5,000,000) are used for accurately estimating the significance level of error covariance estimator represented by $\hat{f}_i \hat{g}_i$.

3. DATA PREPARATION AND DESCRIPTION

The data set is derived from Swiss Travel Microcensus 2000. The trip file was used to create an out-of-home activity file where individual activity records were created from the trip records. This activity file includes information about activity type, activity timing, activity duration, and other variables pertinent to each activity episode. This section focuses on the relationship between activity timing and duration for maintenance activities. Maintenance activities include the following two activity (trip) types: shopping and service (passenger or child). These activity records were extracted from the original file to create two maintenance activity record files, one for commuters and the other for non-commuters. Commuters were defined as individuals who commuted to a work place on the travel diary day, while non-commuters were defined as those who did not commute to a work place (made zero work trips) on the travel diary day. Note that a worker (employed person) who did not commute on the travel diary day would still be classified as a non-commuter for the purpose of this study. The commuter and non-commuter maintenance activity episode data sets included complete socio-economic and activity information for the respective samples.

Prior to commencing the model development effort, descriptive analysis of the potential relationship between activity duration and timing is undertaken by non-commuter sample and commuter sample. The results are presented in Table 1. Based on a time of day distribution of all trips in the data set, four distinct time periods in which activity begins are identified. They are:

- AM peak: 6:00 AM – 8:59 AM
- Midday: 9:00 AM – 3:59 PM
- PM peak: 4:00 PM – 6:59 PM
- Off peak: 7:00 PM – 5:59 AM

Table 1 compares mean value and standard deviation of activity duration across time-of-day allocation of activity within non-commuter sample and commuter sample. To alleviate the variance of dependent variable in linear regression model, the logarithm of activity duration in minutes is specified as dependent variable in the joint modeling system, noted as “LN_DUR”. According to the mean value of “LN_DUR”, four time-of-day choices of non-commuters can be ranked into the following sequence: MIDDAY > AMPEAK > PMPEAK > OFFPEAK. However, the corresponding sequence for commuter is shifted as: MIDDAY > PMPEAK > AMPEAK > OFFPEAK. It may imply that commuters are less likely to pursue long maintenance activities in AM peak period than non-commuters due to the work schedule constraint. Generally, commuters’ maintenance activities are of shorter duration than non-commuters according to descriptive analysis.

4. MODEL ESTIMATION RESULTS

4.1 Estimation Results for Non-Commuters

Table 2 offers a description of explanatory variables used in all the causal models for time-of-day choice and duration of maintenance activities. Among these variables, "LN_DUR" is an endogenous continuous variable and "AMPEAK", "PMPEAK" and "MIDDAY" are three endogenous dummy variable indicating time-of-day choices.

Table 3 provides the estimation results of non-commuter model. The left-half table shows the model estimation results under the causal structure: duration → time-of-day, whereas the right-half table shows the results under the alternative causal structure: time-of-day → duration. For comparison, each half table is composed of three blocks. The first block offers the results from recursive estimation, i.e. a multinomial logit model for time-of-day choices among four time periods and a linear regression model for the logarithm of activity duration. The second block offers the estimation results of mixed discrete-continuous models and the third block offers the estimation results of Lee models. In the second block, "g_i/r_i" represents "g_i" in mixed discrete-continuous model, while in the third block, "g_i/r_i" represents the correlations in Lee model,

All the models provide similar estimators for exogenous variables. Particularly, the estimators in Lee models are almost identical to those in recursive models. That's because joint estimation in Lee model merely improves the efficiency of estimators for exogenous variables. However, greater difference is found between recursive model and mixed model probably because mixed logit model is used instead of multinomial logit model. All the coefficients of exogenous variables have reasonable behavioral interpretation. In all the models, "AGE" takes greatest positive coefficient in AM peak choice utility function, which indicates that elder non-commuters are most likely to schedule maintenance activities in AM peak period and least likely in Off-peak period. Elders may undertake more responsibility of taking children to school or shopping for grocery in AM peak period. Non-commuters living with more household members tend to pursue maintenance activities in AM peak period as evidenced by the positive coefficient of "HHSIZE", presumably because they have to undertake more responsibility of serving children in AM peak period. The positive coefficient of "LOW_INC" indicates that low-income non-commuters prefer to engage in maintenance activity in AM peak period, possibly because their travel are more transit-oriented or more dependent on non-motorized mode thereby less sensitive to AM peak-period traffic congestion. The non-commuters without household car are most likely to pursue maintenance activities in Midday period and least likely in Off-peak period, as evidenced by the greatest positive coefficient in Midday utility function and less positive coefficients in AM peak and PM peak utility functions. The dependency on transit might be a plausible explanation. Public transit may be the least congested in midday period and unavailable in off-peak period. The negative coefficients of "MALE" in PM peak utility and Midday utility indicate that male non-commuters dislike scheduling maintenance activity in PM peak and Midday, as compared to female, possibly because female non-commuters tend to stay at home in the morning and at night for household obligations. High-income non-commuters prefer to schedule maintenance activities in PM peak period as evidenced by the positive coefficient.

Age and square of age appear significant in log-linear regression model for maintenance activity duration. Negative coefficient on age and positive coefficient on age square infer a non-linear effect of age on activity duration. That's probably because middle-aged non-commuters are more responsible for household obligations at home thereby less willing to spend much time on out-of-home maintenance activities than younger and elder. Relative to female non-commuters, male non-commuters allocate less time on maintenance activities. Negative coefficient of "HHSIZE" indicates that people living with more household members spend less time in maintenance activities than those living in small family, presumably because shopping obligations can be assigned to more family members in a big household. High-income non-commuters are expected to expend less time on maintenance activities, as indicated by the negative coefficient, possibly due to more concern on time budget. Car ownership appears significant in recursive model but insignificant in the other three types of models, therefore it has been excluded from there.

The coefficients for endogenous variables are the most important outputs from the modeling estimation results. The left-half table offers non-commuter model estimation results under the causal structure that activity duration affects time-of-day choices. In recursive models, "LN_DUR", indicating logarithm of activity duration, appears positively significant in all the three utility functions for time-of-day choices. The coefficient in Midday choice utility is the greatest. The estimation results in Lee model are rather close to those in recursive models despite accommodating the error correlations possibly because the correlations are rather small, albeit statistically significant.

In the preliminary estimation of mixed logit model, "f_2" appears smallest among all the "f_i", f_2 and g_2 are therefore fixed at zero in the final mixed model. In the mixed model, the coefficients for endogenous variables differ from those in recursive model and Lee model. "LN_DUR" does not appear significant in utility functions for AM peak choice. Table 5 presents error correlations calculated according to Equation (12) and the results of simulation-based hypothesis test for error covariance estimators in all the mixed models. In the current model, the significance level of positive covariance f_1g_1 is 0.209, which is not of high level but considerable in mixed model. The positive correlation between AM peak choice utility and activity duration is rather considerable (0.231). The statistical result indicates that activity duration does not have significant impact on the utility of AM peak choice. Without accommodation of direct positive correlation between AM peak utility and activity duration choice, the coefficient of "LN_DUR" in AM peak utility function is overestimated as 0.495 in the recursive model. Table 5 shows that Corr(u_3, a) and Corr(u_4, a) are almost negligible, which offers the reason why the coefficient for endogenous variable "LN_DUR" in PM peak and Midday utility of mixed model does not substantially differ from those in recursive model. The current mixed model only supports the hypothesis that non-commuters' maintenance activities of longer duration are more likely to be pursued in PM peak period and Midday. Intuitively, non-commuters who intend to make longer maintenance activities probably prefer to start them in Midday or PM peak for avoiding peak-period congestion or institutional constraint.

In the continuous model without endogenous dummy variables, the standard deviation of normal random disturbance estimated from recursive model and Lee model are rather consistent. In the mixed continuous model, one can calculate the standard deviation of the whole normal random disturbance as 1.358 (the square root of sum of all the g_j^2 and $sigma^2$), which is almost identical to the standard deviation (1.357) in recursive model. It virtually indicates that, in the mixed continuous model, the random component has been divided into four parts: the first three parts are individually correlated to the first three utility functions and the last part is a pure random error which is uncorrelated with the utility functions. This result coincides with the a priori assumption for the mixed discrete-continuous modeling system.

The right-half table offers model estimation results of non-commuter model under the causal structure that time-of-day choices affect activity duration. Except the coefficient for endogenous dummy variable "MIDDAY" being reduced substantially (1.308 vs. 1.044), there is no considerable change for "AMPEAK" and "PMPEAK". It can be explained by the fact that "r3" is estimated as -0.307. Please notice that r_i represents $corr(v_{qi}^*, \omega)$, where v_{qi}^* can be represented by Equation (8) and ω is random component in continuous model. ε_{qi} is gumbel distributed, which is asymmetric, thus r_i is negatively related with the correlation between ε_{qi} and ω. As a result, negative "r3" partially takes account of positive correlation between ε_{q3} and ω, which explains why the positive coefficient for the third endogenous dummy variable "MIDDAY" is reduced in Lee model.

In the preliminary estimation of mixed model, "f_1" takes seemingly smallest absolute value among all the f_i, thus f_1 and g_1 are fixed at zero. The coefficients for endogenous dummy variables are close to those in recursive model and Lee model. The error correlations shown in Table 5 are too slight to substantially influence the coefficient of endogenous dummy variables, though simulation-based hypothesis test indicates that covariance f_3g_3 and f_4g_4 have high significance level (0.089 and 0.072). In the continuous model, the standard deviation (σ) of normal random disturbance estimated from recursive model and Lee model are rather consistent (1.323 vs. 1.346). In the mixed model, one can calculate the standard deviation of the whole random components as 1.328, which is close to the one from linear regression model. All the three dummy variables indicating AM peak choice, PM peak choice and Midday choice appear positively significant in all types of models. These statistical results strongly support the hypothesis that time-of-day choices of maintenance activity affect the activity duration for non-commuters. Except Lee model, both recursive model and mixed model take greatest positive coefficient on "MIDDAY", less positive coefficient on AM peak and least positive one on PM peak. This result is consistent with the descriptive analysis in Table 1, where the ranking of time-of-day categories are "MIDDAY" > "AMPEAK" >

“PMPEAK” > “OFFPEAK”, in terms of average activity duration. That is presumably because, in Midday, AM peak, and PM peak periods, non-commuters have sufficient time available for undertaking longer maintenance activities without institutional constraint such as closing time of shopping center.

4.2 Estimation Results for Commuters

The left-half part of Table 4 offers model estimation results of commuter model where activity duration affects time-of-day choices. The exogenous variables in all the models take almost identical coefficients. “AGE” takes positive coefficient in all the three utility functions, among which the one in PM peak utility is the greatest. It indicates that the elder commuters prefer to allocate maintenance activity in PM peak but does not tend to allocate it in Off-peak, presumably because the elder commuters are used to pursuing their maintenance activities in commute way from work place back home. Compared with female commuters, male commuters are more likely to undertake maintenance activity in off-peak period, as evidenced by the negative coefficients of “PMALE” in the other utility functions for the other three time periods. That’s probably because females are unwilling to go out of home at night or in early morning for security purpose. Commuters in single household are less likely to undertake maintenance activities in AM peak and midday as evidenced by the negative coefficients on “HHSIZE1” in both utility functions. That is possibly because they do not have obligation of taking children to school in AM peak period and do not have to undertake maintenance activity in midday without urgent household obligations.

Commuters with no cars in household are more likely to undertake maintenance activities in AM peak, PM peak and midday, as indicated by the positive coefficients of “CAR_0”. That’s possibly because the commuters with car are more likely to pursue activities in off-peak period since their schedule is not constrained by transit availability in that time period. The low-income commuters are more likely to undertake maintenance activities in midday, as shown by the positive coefficient which appears slightly significant. Low-income commuters may have more spare time for pursuing maintenance activities in the middle of daily work. The commute distance negatively affects AM peak engagement of maintenance activities. Uncertainty in commute time increases as distance increases, thus commuters are unwilling to undertake additional activities in AM peak period on their commute. Total daily work time negatively affects midday engagement of maintenance activities, which is consistent with expectation. The more time commuters spend on work, the less time is available for maintenance activity in midday.

In log-linear regression model for maintenance activity duration, age, gender, household size, total daily work time and car ownership are found to be significant contributing factors. The square of age is specified for capturing the non-linear impact of age on activity duration. As opposed to non-commuter model, “AGE” takes positive coefficient and “AGE_SQ” takes negative coefficient in commuter model. These results indicate that middle-aged commuters are expected to undertake longer maintenance activities than younger and elder. Middle-aged commuters may have to undertake more responsibilities outside home than younger and elder commuters. The negative coefficient of “PMALE” indicates that male commuters' activity duration is shorter than female commuters', similar to non-commuters. Commuters living with more household members are less likely to engage into longer activity duration, as evidenced by the negative coefficient of "HHSIZE", similar to non-commuters. Total daily work time negatively affects maintenance activity duration, as indicated by the negative coefficient for "WORKDUR". As expected, the more time commuters spend on work, the less time is available for maintenance activity episodes. The negative coefficient for "CAR_0" indicates that commuters without cars in household tend to allocate more time on maintenance activity than those with household cars. These commuters should heavily depend on public transit and the fixed schedule of transit service may lengthen their activity duration.

As for the estimation results for endogenous continuous variable "LN_DUR", there are no substantial difference found in all types of models. "LN_DUR" take positive coefficient in the utility functions for PM peak choice and Midday choice and take insignificant coefficient in the utility function for AM peak choice. It implies that activity duration negatively affects AM-peak or Off-peak choice of maintenance activities probably because the activities of longer duration cannot be pursued due to work schedule constraint (e.g. work starts in the morning) or institutional constraint (e.g. shopping center is closed at night and in early morning).

The minor difference in the coefficients for endogenous variable among recursive model, mixed model and Lee model is caused by slight correlations between random error terms. The unidentified mixed logit model justifies that f4 takes seemingly smallest value, thus f4 and g4 are fixed at zero in the final model. Table 5 shows, in the current mixed model, only the correlation between PM peak choice utility and activity duration is considerable (-0.129), whereas the other two correlations are negligible. In Lee model, r1 (0.285) and r4 (0.261) appear statistically significant and considerable, thus the coefficient of "LN_DUR" is somewhat less than that from recursive estimation. In the mixed discrete choice model, the standard deviation f_i for normal heterogeneity appears small and insignificant. Behaviorally, it may be a reflection that the flexibility of commuters' maintenance activities is substantially constrained by their rigid work schedule, thus there are not many unspecified factors contributing to activity time-of-day choice. In addition, Table 5 shows the significance levels of error covariance are 0.387, 0.288 and 0.429, which does not provide strong evidence for the existence of error correlations. This result infers that there are few unspecified factors simultaneously affecting commuters' maintenance activity timing and duration, which is consistent with the finding in Pendyala and Bhat (2004). The standard deviations of random error components in the continuous model are reasonable and rather consistent among all types of models. The standard deviation of the total random component in the mixed model can be calculated as 1.306, which is almost equal to the counterpart (1.307) in recursive model and is close to the counterpart (1.298) in Lee model.

The right-half part of Table 4 offers model estimation results of commuter model where time-of-day choices affect activity duration. The exogenous variables in all the models take the coefficients with the same sign and slight variations in magnitude. The coefficient of "AGE" in PM utility function of Lee model is greatly different from the others due to the exclusion of insignificant "AGE_SQ" variable. In the recursive model, "WORKDUR" takes significantly negative coefficient as -0.042. In the mixed model, "WORKDUR" still takes negative coefficient as -0.034 in spite of insignificance. From behavioral perspective, total daily work time is expected to negatively affect commuters' maintenance activity duration. The positive coefficient of "WORKDUR" in Lee model is counterintuitive. The distributional assumption on the whole latent utility including both systematic component and random component may lead to this mistake in Lee model. Note that "WORKDUR" has been specified and appears significant in Midday utility function. In the Lee model, the highly negative correlation between Midday utility function and activity duration model, both of which include "WORKDUR" variable, may result in such a counterintuitive estimator for "WORKDUR". All the other coefficients for exogenous variables in all types of models take the same sign as those in alternative causal structure.

According to the justification of unidentified mixed logit model, f4 and g4 are fixed at zero in the final mixed discrete-continuous model. The estimation results for endogenous variables greatly differ from one another among various types of models. The coefficient of "AMPEAK" is not significant in recursive model. However, the coefficient of "AMPEAK" appears significantly negative (-1.422) in Lee model as the error correlation r1 is significantly negative (-0.526) which, to some degree, indicate a positive correlation between the corresponding random error terms. This estimation result is consistent with that in Pendyala and Bhat (2004), where they also found that "AMPEAK" takes negative coefficient in commuters' activity duration model. However, the estimation result for "AMPEAK" from mixed model appears insignificant. It explicitly indicates that the coefficient estimation for endogenous dummy variables is highly sensitive to the specification of error structure in the joint modeling system. "PMPEAK" takes significantly positive coefficient (0.867) in recursive model, but takes insignificant coefficient in mixed model and Lee model. The corresponding $Corr(u_2,a)$ is calculated as 0.243 and takes significance level of 0.187 in the mixed model, while the corresponding error correlation r2 is -0.648 in Lee model, basically inferring positive correlation between PM peak choice utility function and continuous activity duration model. At this point, both mixed model and Lee model yield consistent results: unspecified factors associated with PM peak choice positively affect duration of maintenance activities, but PM peak choice itself does not exert significant impact on the activity duration. "MIDDAY" takes significantly positive coefficient 0.960 in recursive model and 0.813 in the mixed model. The $Corr(u_3, a)$ is calculated as 0.080 and insignificant (significance level: 0.313), therefore "MIDDAY" coefficient in mixed model does not differ greatly from that in recursive model. However, in Lee model, the coefficient for "MIDDAY" takes much greater positive coefficient 1.791 while r3 is estimated as 0.726 basically indicating a negative correlation between the utility of midday choice and activity duration. The standard deviation of total random component in the mixed continuous model can be calculated as 1.267, which is close to the standard deviation 1.258 in recursive model and much less than the standard deviation 1.477 in Lee model. The greater standard deviation of Lee model is probably caused by the incorrect positive coefficient of "WORKDUR" in log-linear activity duration model.

In summary, for commuters, the relationship that midday choice positively affects activity duration is supported by both mixed model and Lee model. Due to the constraint of fixed work schedule, commuters usually do not have much time for maintenance activities. Since midday period includes lunch time, commuters may like to undertake a longer maintenance activity in this time period. On the other side, the relationship that activity duration positively affects PM peak choice and Midday choice is supported by all types of models. Intuitively, Midday includes lunch time at noon and PM peak period is flexible after work, during which longer maintenance activities can be undertaken on the way back home.

5. MODEL PERFORMANCE COMPARISONS BASED ON THE EXTENDED NON-NESTED TEST

The model estimation results presented in Table 3 and Table 4 generally offer plausible indications for alternative causal paradigms. A strict statistical test is required for comparing and selecting the models under alternative causal structures in favor of identifying dominant causal relationship of travel behavior among population. The causal models under alternative causal structures actually belong to non-nested structure, therefore the classical statistical tests, such as likelihood ratio test for nested structure, cannot be applied for this purpose. Two models are in nested structure if and only if one model can be reduced to the other model by imposing restrictions on the parameters. Cox (1961, 1962) initially proposed a statistical test for comparing the models of separate families of hypothesis (non-nested models). Horowitz (1982) simplified this test into a more applicable form for comparing non-nested discrete choice models. Ben-Akiva and Swait (1984) converted Horowitz' test into a form represented by Akaike Information Criterion (AIC) and collected it into the book (Ben-Akiva and Lerman, 1985). Pendyala and Bhat (2004) drew the conclusion on the basis of this non-nested test. After carefully reviewing the original paper (Horowitz, 1983), the authors consider it is inappropriate to directly apply this test to the non-nested joint discrete-continuous model. This section extends the non-nested test in discrete choice model for more rigorous comparisons within non-nested joint discrete-continuous models.

5.1 Review of Non-nested Test for Discrete Choice Model

It is necessary to review the original paper that proposed non-nested test for discrete choice model by Horowitz (1983). The original paper adopts the following goodness-of-fit measures: $\bar{\rho}_g^{\,2} = 1 - \frac{L_g - K_g/2}{L^*}$ and $\bar{\rho}_f^{\,2} = 1 - \frac{L_f - K_f/2}{L^*}$, (18)

where L_g and L_f are log-likelihood function value for model g and model f, both of which belong to non-nested structures, respectively; K_g and K_f are number of estimated parameters in model g and model f, respectively; L^* is log-likelihood function value of the model without any explanatory variables or any parameters. Horowitz (1983) derived that $\Pr[\bar{\rho}_f^{\,2} - \bar{\rho}_g^{\,2} > z] \le \Phi[-\sqrt{-2L^* z}\,]$. (19)

5.2 Non-nested Test Extended into Joint Discrete-Continuous Model System

Non-nested test for discrete choice model is originated from Cox' separate family of hypothesis test. Cox' test can be applied not only for discrete choice model, but also for any models estimated by maximum likelihood method. Discrete-continuous model is not an exception. Suppose we have a linear regression model as $y = \beta_0 + x'\beta + u$. A naive model with minimum number of parameters is required to provide L^* value in Equation (21). Unlike discrete choice model, the linear regression model needs to contain at least two parameters: constant β_0 and standard deviation σ of normal error term. Then one may have a naïve linear regression model as $y_i = \beta_0 + \sigma n$, $n \sim \text{Normal}(0,1)$. Ordinary Least Square (OLS) procedure yields estimators as

$$\hat{\beta}_0 = \sum_{i=1}^{N} y_i / N = \bar{y} \text{ and } \hat{\sigma} = \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2 / (N-1)}\,, \tag{20}$$

where N is sample size. For linear regression model, it is easy to show that MLE estimators are exactly equal to OLS estimators. Under normality assumption on the random error term, the probability density and log-probability density function for each continuous observation i can be expressed as

$$f_i = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp[-\frac{(y-\bar{y})^2}{2\hat{\sigma}^2}] \text{ and } \ln(f_i) = [-\frac{(y-\bar{y})^2}{2\hat{\sigma}^2}] - \ln(\sqrt{2\pi}\hat{\sigma}) \tag{21}$$

By replacing the parameters with OLS estimators and summing up log-probability density value over the sample, one may obtain L^*(continuous observations)

$$= \sum_{i=1}^{N} \ln(f_i) = -(N-1)/2 - N\ln(\sqrt{2\pi}\hat{\sigma})\,. \tag{22}$$

The log-likelihood function value for naive discrete choice model is the same as usual:
L^*(discrete observations) = N ln(1/J), (23)
where J represents the number of alternatives in choice set.
L^*(total) = L^*(continuous observations) + L^*(discrete observations)

$$= -(N-1)/2 - N\ln(\sqrt{2\pi}\hat{\sigma}J)\,. \tag{24}$$

By plugging L^*(total) into Equation (19),

we obtain $\Pr[\bar{\rho}_f^{\,2} - \bar{\rho}_g^{\,2} > z] \le \Phi\{-\sqrt{z[N-1+2N\ln(\sqrt{2\pi}\hat{\sigma}J)]}\}$, (25)

By replacing $\bar{\rho}_f^{\,2}$ and $\bar{\rho}_g^{\,2}$ with standard adjusted likelihood ratio indices, we may have

$$\Pr[\bar{\rho}_2^{\,2} - \bar{\rho}_1^{\,2} > z] \le \Phi\{-\sqrt{z[N-1+2N\ln(\sqrt{2\pi}\hat{\sigma}J)] + K_2 - K_1}\}\,, \tag{26}$$

where $\bar{\rho}_1^{\,2} = 1 - \frac{L(\beta_1) - K_1}{L(0)}$ and $\bar{\rho}_2^{\,2} = 1 - \frac{L(\beta_2) - K_2}{L(0)}$, (27)

where $\bar{\rho}_1^{\,2}$ and $\bar{\rho}_2^{\,2}$ are adjusted likelihood ratio index for model g and f; $L(\beta_1)$ and $L(\beta_2)$ are log-likelihood value at convergence in model g and f; L(0) is log-likelihood value at zero (no parameters for discrete choice model and two parameters: β_0 and σ for linear regression model); K_2 and K_1 are the number of parameters in model g and model f.

The probability that the adjusted likelihood ratio index of model f is greater by some z > 0 than that of model g, given that model g is the true model, is asymptotically bounded by the right-hand side of Equation (26). If the model with the greater $\bar{\rho}^2$ is selected, then this bounds the probability of erroneously choosing the incorrect model over the true specification. Using this procedure, discrete-continuous models under alternative causal structures can be compared against one another.

Table 6 compares goodness-of-fit measurements of mixed and Lee models under alternative causal structure. The extended non-nested test for discrete-continuous model is adopted for identifying the dominant causal structure among the population. For non-commuters, both recursive model and mixed model indicate that the model in which time-of-day choices affects activity duration provides better goodness-of-fit in terms of adjusted likelihood index. Also, non-nested test rejects the model in which activity duration affects time-of-day choice at high significance level. However, the dominant causal structure is opposite in Lee model, where the causal structure that activity duration affecting activity timing offers better goodness-of-fit measurement. This result is consistent with that in Pendyala and Bhat (2004) who applied Lee model for identifying causal relationship between activity time-of-day choices and activity duration based on the survey data from Florida, USA.

For commuters, the causal structure remains ambiguous in recursive model and mixed model. In the recursive model, the causal structure that activity timing affects activity duration provides seemingly better goodness-of-fit measurement and non-nested test yields the bounding probability as 0.147. In the mixed model, the causal structure that activity duration affects activity timing provides seemingly better goodness-of-fit measurement, but non-nested test yields the bounding probability as 0.300. Insignificant result from non-nested test indicates that mixed model fails to identify the dominant causal relationship between activity timing and activity duration since two alternative causal structures are equally supported by commuters' sample. On the other side, Lee model supports opposite conclusion that the causal relationship of duration affecting time-of-day choices is dominant among commuters, but Pendyala and Bhat (2004) did not draw conclusive results for commuters using Lee model. It is surprising to see that not only are coefficient estimators of endogenous variables sensitive to the error structure, but also the dominant causal structure will change in response to distributional assumption on error covariance of the joint modeling system.

6. DISCUSSIONS AND CONCLUSIONS

This paper has presented an analysis of the relationship between activity timing (time-of-day choice) and activity episode duration for maintenance activities including shopping and service. The analysis involved the estimation of joint models of activity timing and duration separately for commuters and non-commuters while allowing two types of error correlations between the timing and duration model equations. Time-of-day choice was modeled as a discrete choice variable involving four alternative periods of the day while duration was modeled using a log-linear formulation.

Two different causal structures were considered:

- Activity timing (time of day choice) affects activity duration
- Activity episode duration affects activity timing (time of day choice)

Both of these causal structures were estimated on the non-commuter and commuter sample activity episodes to identify the appropriate causal structure for each sample group. The identification of such causal relationships between activity engagement phenomena is very important from several key perspectives. First, the identification of appropriate causal structures will help in the development of accurate activity-based travel demand model systems that intend to capture such relationships at the level of the individual traveler and activity episode. Second, the knowledge of the true causal relationships underlying decision processes will help in the accurate assessment and impact analysis of alternative transportation policies such as variable pricing, parking pricing, and telecommuting. The dominant causal relationship between timing and duration has not been consistently identified through two types of models. For both commuters and non-commuters, Lee model supports the causal relationship that activity duration is determined first and then influences time of day choice. However, the mixed discrete-continuous model supports the alternative causal relationship for non-commuters: time-of-day choices are first determined and then influence activity duration. For commuters, mixed discrete-continuous model fails to draw a conclusion. Since the mixed model is more behaviorally interpretable than Lee model, we considered that time-of-day choice affecting activity duration is more plausible dominant causal relationship among population.

On the methodological aspect, this paper employs two different joint discrete-continuous modeling systems: mixed model and Lee model for consistently estimating the endogenous dummy variables in continuous model and the endogenous continuous variable in multinomial discrete choice model. Both mixed model and Lee model adopt Full Information Maximum Likelihood (FIML) method based on distributional assumption of error structure in a simultaneous modeling system. The error structure in mixed model is a more behaviorally interpretable than the one in Lee model. However, the likelihood function of mixed model does not have closed form and Monte Carlo integral is required to approximate the likelihood function. Maximum Simulated Likelihood Estimation (MSLE) based on Monte Carlo integral is very time-consuming. In addition, the simulation bias cannot be avoided in MSLE. More quasi-random seeds for simulation can alleviate the simulation bias, but the accuracy is traded off with time consumption in estimation procedure. Lee model has closed form based on Lee transformation, thus the estimation procedure of Lee model takes much less time than that of mixed model (a few minutes vs. a few hours).

Relative to Lee model, mixed model makes a progress by imposing distributional assumption merely on the random error terms and releasing the distributional assumption on the whole utility functions for discrete choice. In mixed model, it is more confident to conclude whether there exists correlation between utility function and function of continuous model because the common random heterogeneities can directly take account of the common unspecified variables from the model functions. It is more difficult to explain the estimated correlation in Lee model than in mixed model. The estimated correlation r_i cannot directly reflect the correlation between utility function and continuous model. The correlation between the random error terms is rooted in common omitted variables in model functions, which needs to be linearly specified into the model if observable and quantifiable. The covariance of error structure assumed in Lee model does not reflect this mechanism very well, but the mixed discrete-continuous model can directly accommodate the correlations by linearly specifying the heterogeneity into its flexible error structure. Thus, we consider that the proposed mixed joint model is more behaviorally sound and statistically robust than Lee model.

However, it does not necessarily mean the mixed model is a perfect method in dealing with such problems. The dependency on distributional assumption is a common disadvantage of mixed model and Lee model. Maximum likelihood estimation is always consistent and efficient as long as the distributional assumption is valid and all the parameters are identifiable. However, the distributional assumption is vulnerable in many cases, particularly when it is made to take account of unobserved heterogeneities. Since the coefficient of endogenous variable is highly sensitive to distributional assumption, a robust specification of error structure is important. One alternative is to introduce non-parametric heterogeneities into the joint model system. In econometric literature, Mroz (1987 and 1999) applied discrete factor approximation to estimate endogenous dummy variable in a continuous model, where heterogeneity is assumed to be non-parametrically and discretely distributed in place of being parametrically and continuously distributed (e.g. normal distribution or log-normal distribution). However, it is not easy to apply this method in practice since the log-likelihood function is not globally concave and has multiple peaks in its multidimensional domain. A large number of starting values need to be explored to avoid the pitfall of local maxima. Another alternative is to use semi-parametric approach for more robust estimation results. Econometricians have made considerable advances in semi-parametric method for more robust estimation for discrete choice model (e.g. Lewbel, 2000). These approaches are potentially applicable to travel behavior analysis for more robust estimation of endogenous variables in discrete choice model. A wide space is remained for further exploration in the future.

ACKNOWLEDGEMENTS

The authors thank Prof. Chandra Bhat and his graduate research assistant, Abdul Rawoof Pinjari, of the Department of Civil, Architectural & Environmental Engineering at the University of Texas at Austin for their assistance and suggestions. In addition, the authors benefited from discussions with Pan Liu, graduate research assistant in the Department of Civil and Environmental Engineering at the University of South Florida in the final stage of this study. Finally, the authors thank Dr. Giovanni Gottardi of Jenni + Gottardi, AG, Zurich and the Swiss Ministry of Transport for providing the data for this study.

REFERENCES

Axhausen, K. and Garling, T. (1992) Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews* **12** 493-517.

Ben-Akiva, M.E. and Swait, J.D. (1984) The Akaike Likelihood Ratio Index. *Transportation Science* **20(2)** 133-136.

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

Bhat, C.R. (1995) A Heteroscedastic Extreme Value Model of Intercity Mode Choice. *Transportation Research B* **29(6)** 471-483.

Bhat, C.R. (1998) A model of post-home arrival activity participation behavior. *Transportation Research B* **32(6)** 387-400.

Bhat, C.R. and Koppelman, F.S. (1999) A Retrospective and Prospective Survey of Time-Use Research. *Transportation* **26(2)** 119-139.

Bhat, C.R. and Misra, R. (1999) Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. Transportation 26(2) 193-229.

Bhat, C.R. (2001) Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research B* **35** 677-693.

Cox, D. (1961) Tests of Separate Families of Hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, Berkeley: University of California Press, 1961.

Cox, D. (1962) Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society B* **24** 406-424.

Greene, W.H. (2003) *Econometric Analysis*, Fifth Edition, Pearson Education, Inc., NJ.
Horowitz, J.L. (1983) Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models. *Transportation Science* **17(3)** 319-350.

Kasturirangan, K, Pendyala, R.M. and Koppelman, F.S. (2002) Role of History in Modeling Daily Activity Frequency and Duration for Commuters. *Transportation Research Record* **1807** 129-136.

Lee, L.F. (1983) Generalized econometric models with selectivity. *Econometrica* **51(2)** 507-512.

Lewbel, A. (2000) Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables. *Journal of Econometrics* **97** 145-177.

Mroz, T.A. (1987) The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica* **55** 765-799.

Mroz, T.A. (1999) Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome, *Journal of Econometrics* **92** 233-274.

Pas, E.I. and Harvey, A.S. (1997) Time Use Research and Travel Demand Analysis and Modeling. In Stopher, P.R. and Lee-Gosselin, M. *Understanding Travel Behavior in an Era of Change*, Elsevier, Oxford 315-338.

Pendyala, R.M, Yamamoto, T. and Kitamura, R. (2002) On the Formulation of Time Space Prisms to Model Constraints on Personal Activity-Travel Engagement. *Transportation* **29(1)** 73-94.

Pendyala, R.M. and Bhat, C.R. (2004) An Exploration of the Relationship Between Timing and Duration of Maintenance Activities. *Transportation* **31(4)** 429-456.

Pendyala, R.M. and Ye, X. (2004) Contributions to Understanding Joint Relationships Among Activity and Travel Variables. International Conference on Progress in Activity-Based Analysis, Maastricht, The Netherlands, May.

Steed, J. and Bhat, C.R. (2000) On Modeling the Departure Time Choice for Home-based Social/Recreational and Shopping Trips. *Transportation Research Record* **1706** 152-159.

Train, K. (2002) *Discrete Choice Methods with Simulation*, Cambridge University Press.

Walker, J. (2001) Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Table 1. Description of Endogenous Variables in the Samples

Time-of-Day Choices	Mean of LN_DUR	Std. Dev. Of LN_DUR	Mean of Duration (minutes)	Std. Dev. of Duration	N
Non-commuter Sample					
AMPEAK (6:00-8:59)	2.90	1.38	41.24	73.13	1142
PMPEAK (16:00-18:59)	2.68	1.40	37.22	80.59	1775
MIDDAY (9:00-15:59)	3.21	1.29	50.70	90.62	7911
OFFPEAK (19:00-5:59)	1.87	1.53	22.09	48.84	465
Total	3.04	1.36	46.45	86.40	11293
Commuter Sample					
AMPEAK (6:00-8:59)	1.95	1.33	17.38	46.52	359
PMPEAK (16:00-18:59)	2.77	1.19	31.58	66.47	1222
MIDDAY (9:00-15:59)	2.89	1.25	35.09	55.93	1467
OFFPEAK (19:00-5:59)	1.84	1.53	21.38	47.43	346
Total	2.64	1.33	30.55	58.61	3394

Table 2. Description and Definition of Variables in Timing-Duration Model

		Commuters' Activity Sample		Non-Commuters' Activity Sample	
	Sample Size	3394		11293	
Variable Name	Variable Description	Mean	Std. Dev.	Mean	Std. Dev.
AGE	Age in 100 years	0.41	0.12	0.49	0.20
AGE_SQ	The square of age (100^2 years^2)	0.18	0.10	0.28	0.20
PMALE	Person is male	0.46	0.50	0.35	0.48
HHSIZE1	Single-member family	0.27	0.44	--	--
HIGH_INC	Monthly household income > Fr 10000	--	--	0.10	0.30
LOW_INC	Monthly household income < Fr 4000	0.10	0.31	0.24	0.43
CAR_0	Household does not own car	0.13	0.34	0.20	0.40
LN_DISW	Ln(1 + commute distance in kilometers)	1.91	1.02	--	--
WORKDUR	Daily total work time (100 mins)	4.15	1.66	--	--
HHSIZE	Total number of household members	2.56	1.33	2.55	1.39
LN_DUR	Ln(1 + activity duration in minutes)	2.64	1.33	3.04	1.36
AMPEAK	Activity is scheduled in AM peak	0.11	0.31	0.10	0.30
PMPEAK	Activity is scheduled in PM peak	0.36	0.48	0.16	0.36
MIDDAY	Activity is scheduled in Midday	0.43	0.50	0.70	0.46

Table 3. Non-commuter Models

	Causal Structure (Duration → Time-of-Day)						Causal Structure (Time-of-Day → Duration)					
	Recursive Model		Mixed Model		Lee Model		Recursive Model		Mixed Model		Lee Model	
Variable	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*
Activity Time-of-Day Choice Model												
AM Peak Choice Utility												
Constant	-2.502	-11.66	-2.873	-1.33	-2.226	-10.15	-1.360	-7.13	-1.267	-5.19	-1.360	-7.14
AGE	3.760	11.84	4.557	2.34	3.768	11.88	3.805	12.29	3.916	12.05	3.800	12.27
HHSIZE	0.129	4.43	0.155	1.30	0.127	4.36	0.132	4.56	0.130	4.44	0.133	4.54
LOW_INC	0.186	2.36	0.271	1.05	0.185	2.35	0.185	2.34	0.171	2.15	0.178	2.27
CAR_0	0.678	3.35	0.677	2.80	0.647	3.24	0.795	3.95	0.801	3.92	0.814	4.04
LN_DUR	**0.495**	**12.20**	0.044	0.04	**0.377**	**8.92**	--	--	--	--	--	--
f1	--	--	2.167	0.60	--	--	--	--	0.000	--	--	--
PM Peak Choice Utility												
Constant	-0.023	-0.16	-0.009	-0.02	0.144	0.94	0.826	6.74	0.826	4.32	0.831	6.77
AGE	1.206	4.35	1.206	4.13	1.192	4.30	1.237	4.59	1.313	4.63	1.220	4.53
PMALE	-0.234	-3.25	-0.260	-1.96	-0.229	-3.18	-0.239	-3.33	0.000	--	-0.238	-3.29
HIGH_INC	0.146	1.77	0.000	--	0.156	1.89	0.180	2.20	0.181	2.18	0.185	2.26
CAR_0	0.565	2.87	0.557	2.79	0.535	2.75	0.667	3.40	0.706	3.56	0.678	3.45
LN_DUR	**0.384**	**10.22**	**0.390**	**1.93**	**0.305**	**7.73**	--	--	--	--	--	--
f2	--	--	0.000	--	--	--	--	--	0.143	0.73	--	--
Midday Choice Utility												
Constant	0.003	0.02	-0.193	-0.17	0.380	2.70	1.686	15.22	1.708	8.84	1.694	15.28
AGE	2.474	9.52	2.514	6.43	2.475	9.56	2.521	10.14	2.617	9.68	2.497	10.05
PMALE	-0.199	-3.46	-0.224	-1.39	-0.205	-3.56	-0.238	-4.20	0.000	--	-0.238	-4.19
CAR_0	0.792	4.22	0.803	4.03	0.762	4.13	0.915	4.93	0.963	5.08	0.926	4.98
LN_DUR	**0.665**	**18.81**	**0.741**	**1.83**	**0.514**	**14.02**	--	--	--	--	--	--
f3	--	--	-0.440	-0.25	--	--	--	--	-0.366	-1.35	--	--
Off-Peak Choice Utility												
f4	--	--	-0.029	-0.02	--	--	--	--	0.552	1.46	--	--
Activity Duration Model												
Constant	3.404	56.48	3.439	40.96	3.524	43.38	2.313	31.03	2.331	15.09	2.357	8.82
AGE	-0.760	-3.53	-0.959	-3.08	-0.782	-2.71	-0.897	-4.17	-0.935	-3.28	-0.822	-2.88
AGE_SQ	0.821	3.75	1.038	3.27	0.764	2.61	0.785	3.59	0.853	2.96	0.772	2.68
PMALE	-0.147	-7.42	-0.146	-5.40	-0.132	-4.94	-0.121	-6.08	-0.117	-4.49	-0.134	-5.01
HHSIZE	-0.056	-6.46	-0.060	-5.30	-0.057	-5.09	-0.052	-6.06	-0.053	-4.76	-0.059	-5.33
HIGH_INC	-0.136	-4.14	-0.141	-3.02	-0.144	-3.34	-0.126	-3.84	-0.134	-3.14	-0.142	-3.31
CAR_GE2	-0.045	-1.96	0.000	--	0.000	--	-0.023	-1.01	0.000	--	0.000	--
AMPEAK	--	--	--	--	--	--	**1.016**	**18.30**	**0.840**	**5.96**	**1.099**	**2.59**
PMPEAK	--	--	--	--	--	--	**0.794**	**15.20**	**0.740**	**3.79**	**0.802**	**2.56**
MIDDAY	--	--	--	--	--	--	**1.308**	**27.19**	**1.313**	**8.71**	**1.044**	**3.96**
g1/r1	--	--	0.364	1.18	**0.088**	**4.66**	--	--	0.000	--	0.063	0.44
g2/r2	--	--	0.000	--	**0.171**	**9.53**	--	--	**-0.876**	**10.57**	0.031	0.36
g3/r3	--	--	0.240	0.29	0.000	--	--	--	**0.504**	**5.26**	**-0.307**	**-5.80**
g4/r4	--	--	-0.284	-1.08	**0.326**	**14.92**	--	--	**-0.305**	**-3.32**	0.018	0.20
sigma	1.357	--	1.254	10.18	1.348	151.10	1.323	--	0.805	10.13	1.346	109.44

Table 4. Commuter Models

	Causal Structure (Duration → Time-of-Day)						Causal Structure (Time-of-Day → Duration)					
	Recursive Model		Mixed Model		Lee Model		Recursive Model		Mixed Model		Lee Model	
Variable	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*	Coeff.	*t-test*
Activity Time-of-Day Choice Model (AM Peak Choice Utility)												
Constant	-1.050	*-3.26*	-1.098	*-2.61*	-0.900	*-2.75*	-0.973	*-3.17*	-1.161	*-0.99*	-0.954	*-3.17*
AGE	3.610	*5.56*	3.713	*5.28*	3.592	*5.50*	3.623	*5.57*	3.699	*4.65*	3.345	*5.19*
PMALE	-0.400	*-2.59*	-0.413	*-2.57*	-0.471	*-3.03*	-0.387	*-2.52*	-0.385	*-2.36*	-0.277	*-1.82*
HHSIZE1	-0.405	*-2.72*	-0.403	*-2.47*	-0.419	*-2.81*	-0.479	*-3.26*	-0.519	*-2.28*	-0.387	*-2.87*
CAR_0	0.974	*3.50*	0.987	*3.44*	0.940	*3.42*	1.013	*3.67*	1.036	*3.44*	0.944	*3.44*
LN_DISW	-0.105	*-1.83*	-0.112	*-1.81*	-0.118	*-2.06*	-0.108	*-1.91*	-0.115	*-1.68*	-0.091	*-1.70*
LN_DUR	0.035	*0.58*	-0.019	*-0.16*	0.001	*0.02*	--	--	--	--	--	--
f1	--	--	0.665	*0.85*	--	--	--	--	0.704	*0.33*	-0.954	*-3.17*
PM Peak Choice Utility												
Constant	-1.347	*-3.08*	-1.632	*-2.44*	-1.012	*-2.31*	-0.224	*-0.53*	-0.352	*-0.69*	0.404	*1.85*
AGE	5.590	*2.75*	5.352	*2.52*	5.419	*2.68*	6.036	*3.00*	6.459	*2.73*	2.396	*4.61*
AGE_SQ	-4.290	*-1.76*	-4.069	*-1.62*	-4.117	*-1.70*	-4.551	*-1.88*	-5.096	*-1.78*	0.000	--
PMALE	-0.311	*-2.44*	-0.285	*-2.07*	-0.363	*-2.84*	-0.388	*-3.12*	-0.373	*-2.82*	-0.327	*-2.70*
CAR_0	0.812	*3.41*	0.787	*3.23*	0.793	*3.38*	0.965	*4.1*	0.991	*4.05*	0.929	*3.97*
LN_DUR	**0.530**	*10.72*	**0.657**	*2.80*	**0.410**	***7.41***	--	--	--	--	--	--
f2	--	--	-0.210	*-0.56*	--	--	--	--	0.640	*1.05*	0.404	*1.85*
Midday Choice Utility												
Constant	0.483	*1.81*	0.476	*1.45*	0.916	*3.35*	1.809	*7.45*	1.893	*5.83*	1.789	*7.53*
AGE	2.522	*4.68*	2.514	*4.65*	2.515	*4.68*	2.768	*5.25*	2.791	*5.16*	2.696	*5.23*
PMALE	-0.630	*-4.94*	-0.629	*-4.91*	-0.685	*-5.37*	-0.716	*-5.76*	-0.728	*-5.39*	-0.631	*-5.21*
HHSIZE1	-0.182	*-1.98*	-0.166	*-1.66*	-0.165	*-1.79*	-0.150	*-1.64*	-0.167	*-1.62*	0.000	--
LOW_INC	0.210	*1.68*	0.213	*1.68*	0.217	*1.74*	0.217	*1.74*	0.236	*1.67*	0.184	*1.78*
CAR_0	0.640	*2.63*	0.632	*2.59*	0.619	*2.58*	0.784	*3.26*	0.784	*3.22*	0.643	*2.71*
WORKDUR	-0.274	*-12.07*	-0.280	*-10.72*	-0.278	*-12.25*	-0.280	*-12.49*	-0.301	*-5.19*	-0.282	*-12.99*
LN_DUR	**0.582**	*11.79*	**0.595**	*7.01*	**0.418**	***7.98***	--	--	--	--	--	--
f3	--	--	0.078	*0.19*	--	--	--	--	-0.403	*-0.56*	1.789	*7.53*
Off-Peak Choice Utility												
f4	--	--	0.000	--	--	--	--	--	0.000	--	--	--
Activity Duration Model												
Constant	2.713	*13.97*	2.722	*10.70*	2.834	*11.36*	1.863	*9.28*	1.997	*6.27*	1.571	*3.05*
AGE	2.601	*2.85*	2.552	*2.13*	2.587	*2.20*	2.350	*2.57*	2.658	*2.23*	2.339	*2.04*
AGE_SQ	-2.853	*-2.60*	-2.808	*-1.96*	-2.871	*-2.03*	-2.648	*-2.41*	-3.021	*-2.10*	-2.722	*-1.98*
PMALE	-0.196	*-5.64*	-0.187	*-4.13*	-0.176	*-3.92*	-0.140	*-4.02*	-0.136	*-2.68*	0.000	--
HHSIZE	-0.097	*-7.14*	-0.097	*-5.50*	-0.097	*-5.74*	-0.083	*-6.08*	-0.086	*-4.93*	-0.075	*-4.39*
WORKDUR	-0.071	*-6.77*	-0.071	*-5.21*	-0.058	*-4.22*	-0.042	*-3.86*	-0.034	*-1.49*	0.045	*2.66*
CAR_0	0.121	*2.31*	0.127	*1.83*	0.000	--	0.088	*1.67*	0.113	*1.63*	0.158	*2.11*
AMPEAK	--	--	--	--	--	--	0.071	*0.94*	-0.010	*-0.04*	**-1.422**	***-2.51***
PMPEAK	--	--	--	--	--	--	**0.867**	*14.13*	0.450	*1.50*	-0.279	*-0.56*
MIDDAY	--	--	--	--	--	--	**0.960**	*15.66*	**0.813**	*3.01*	**1.791**	***3.82***
g1/r1	--	--	0.086	*0.54*	**0.285**	***8.38***	--	--	0.094	*0.24*	**-0.526**	***-6.10***
g2/r2	--	--	1.040	*8.64*	0.094	*1.76*	--	--	0.689	*1.73*	**-0.648**	***-14.15***
g3/r3	--	--	-0.276	*-1.90*	--	--	--	--	-0.337	*-1.57*	**0.726**	***15.96***
g4/r4	--	--	0.000	--	**0.261**	***7.67***	--	--	0.000	--	0.083	*0.45*
sigma	1.307	--	0.736	*4.98*	1.298	*82.45*	1.258	--	1.004	*3.57*	1.477	*34.81*

Table 5. Error Covariance Test and Calculated Error Correlation in Mixed Discrete-Continuous Models

i	1	2	3	4
Non-commuter Model (Duration → Time-of-Day)				
Corr ($\hat{f}_i \hat{g}_i$)	0.709	--	-0.915	-0.314
Significance ($\hat{f}_i \hat{g}_i$)	0.209	--	0.128	0.549
Corr(u_i, a)	0.231	--	-0.057	0.004
Non-commuter Model (Time-of-Day → Duration)				
Corr ($\hat{f}_i \hat{g}_i$)	--	-0.086	0.196	-0.026
Significance ($\hat{f}_i \hat{g}_i$)	--	0.232	0.089	0.072
Corr(u_i, a)	--	-0.073	-0.104	-0.091
Commuter Model (Duration → Time-of-Day)				
Corr ($\hat{f}_i \hat{g}_i$)	-0.065	-0.279	-0.009	--
Significance ($\hat{f}_i \hat{g}_i$)	0.387	0.288	0.429	--
Corr(u_i, a)	0.030	-0.129	0.013	--
Commuter Model (Time-of-Day → Duration)				
Corr ($\hat{f}_i \hat{g}_i$)	-0.522	-0.548	-0.004	--
Significance ($\hat{f}_i \hat{g}_i$)	0.630	0.187	0.313	--
Corr(u_i, a)	0.036	0.243	0.080	--

Table 6. Comparison of Goodness-of-Fit of Timing-Duration Models

	Non-Commuter Models		Commuter Models	
	Duration → Time	Time → Duration	Duration → Time	Time → Duration
Sample Size	11293		3394	
LL at zero: LL(0)	-35185.0		-10476.7	
LL at constant: LL(c)	-29729.8		-9846.93	
Estimated sigma	1.3640		1.3254	
Recursive Model				
# of Parameters	25	25	29	30
LL at convergence (LL)	-29212.9	-29193.2	-9514.41	-9513.36
ρ^2 at zero	0.1697	0.1703	0.0919	0.0920
Adj. ρ^2 at zero	0.1690	0.1696	0.089082	0.089087
ρ^2 at constant	0.0174	0.0180	0.0338	0.0339
Adj.ρ^2 at constant	0.0165	0.0172	0.0308	0.0308
Non-nested Test (Prob.)		0.0006 (0.000)		0.000005(0.147)
Mixed Model				
# of Parameters	29	28	35	35
LL at convergence	-29213.1	-29196.6	-9511.55	-9511.69
ρ^2 at zero	0.1697	0.1702	0.0921	0.0921
Adj. ρ^2 at zero	0.1689	0.1694	0.088783	0.088770
ρ^2 at constant	0.0174	0.0179	0.0341	0.0340
Adj.ρ^2 at constant	0.0164	0.0170	0.0305	0.0305
Non-nested Test (Prob.)		0.0005 (0.000)	0.000013(0.300)	
Lee Model				
# of Parameters	27	28	31	31
LL at convergence	-29078.9	-29187.8	-9463.49	-9497.50
ρ^2 at zero	0.1735	0.1704	0.0967	0.0935
Adj. ρ^2 at zero	0.1728	0.1697	0.0938	0.0905
ρ^2 at constant	0.0219	0.0182	0.0389	0.0355
Adj.ρ^2 at constant	0.0210	0.0173	0.0358	0.0323
Non-nested Test (Prob.)	0.0031 (0.000)		0.0032 (0.000)	