An Integrated Model of Residential Location, Work Location, Vehicle Ownership, and Commute Tour Characteristics

Rajesh Paleti
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: rajeshp@mail.utexas.edu

Chandra R. Bhat *(corresponding author)*
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: bhat@mail.utexas.edu

Ram M. Pendyala
Arizona State University
School of Sustainable Engineering and the Built Environment
Room ECG252, Tempe, AZ 85287-5306
Phone: 480-727-9164; Fax: 480-965-0557
Email: ram.pendyala@asu.edu

November 15, 2012

ABSTRACT
This paper offers an econometric model system that simultaneously considers six different activity-travel choice dimensions in a unifying framework. The six dimensions include residential location choice, work location choice, auto ownership, commuting distance, commute mode, and number of stops on commute tours. The paper presents the modeling methodology in detail as well as estimation results for a joint model system estimated on a data set extracted from the 2009 National Household Travel Survey. Estimation results show substantial presence of correlated unobserved effects (self-selection) across choice dimensions, underscoring the value offered by joint equations model systems in the travel modeling field.

INTRODUCTION

There is a growing and important body of evidence that supports the notion that people make a multitude of choices as a "bundle", choosing a series of location and activity-travel attributes that define their lifestyle jointly. This simultaneous selection of a number of choice dimensions across the varied temporal scales calls for the development and deployment of model systems wherein a number of choice behaviors are captured jointly while accounting for both observed and unobserved effects that affect the behaviors of interest. This paper is aimed at formulating and estimating a multi-dimensional integrated choice model system that connects a multitude of choices across disparate temporal scales, *i.e.*, the long term, the medium term, and the short term.

The evidence in favor of attempting to model a multitude of choice dimensions in a joint modeling framework is quite irrefutable and growing (Abraham and Hunt, 1997). Notably, the body of work examining the impact of land use measures on travel behavior suggests that there are considerable self-selection effects wherein households tend to locate in neighborhoods that have attributes consistent with their lifestyle and mobility preferences (Bhat and Guo, 2007; Cao *et al.*, 2008a). For example, households that are not auto-oriented choose to locate in transit and pedestrian friendly neighborhoods that are characterized by mixed and high land use density, and then the good transit service may also further structurally influence mode choice behaviors. If that is the case, then it is likely that the choices of residential location, vehicle ownership, and commute mode choice (for example) are being made jointly as a bundle. That is, residential location may structurally affect vehicle ownership and commute mode choice, but underlying propensities for vehicle ownership and commute mode may themselves affect residential location in the first place to create a bundled choice. This is distinct from a sequential decision process in which residential location choice is chosen first (with no effects whatsoever of underlying propensities for vehicle ownership and commute mode on residential choice), then residential location affects vehicle ownership (which is chosen second, and in which the underlying propensity for commute mode does not matter), and finally vehicle ownership affects commute mode choice (which is chosen third). The sequential model is likely to over-estimate the impacts of residential location (land use) attributes on activity-travel behavior because it ignores self-selection effects wherein people who locate themselves in such neighborhoods were auto-disoriented to begin with. These lifestyle preferences and attitudes constitute unobserved factors that simultaneously impact long term location choices, medium term vehicle ownership choices, and short term activity-travel choices; the only way to accurately reflect their impacts and capture the "bundling" of choices is to model the choice dimensions together in a joint equations modeling framework that accounts for correlated unobserved lifestyle (and other) effects as well as possible structural effects.[1]

[1] In joint limited-dependent variable systems in which one or more dependent variables are not observed on a continuous scale, such as the joint system considered in the current paper that has several discrete dependent variables, the structural effects of one discrete variable on another can only be in a single direction. That is, it is not possible to have correlated unobserved effects underlying the propensities determining two observed discrete dependent variables, as well as have the observed discrete variables themselves structurally affect each other in a bi-directional fashion. This creates a logical inconsistency problem. For example, in the example provided earlier, the underlying propensity for vehicle ownership can impact the propensity to reside in a certain type of location (due to observed factors such as income levels and unobserved factors such as auto-orientation), and residential location itself can have a structural impact on vehicle ownership propensity. But then it is not possible to have vehicle ownership level also structurally impact the propensity to reside in a certain type of location. Doing so would lead to a situation where the probabilities of all the possible combinations of discrete observations will not sum to one (see Maddala, 1983, page 119 for a good discussion). Intuitively, the propensities are the precursors to the actual observed variables, and, when both the decisions are co-determined, it is impossible to have both observed variables

There is a large body of work on joint equations modeling in location and activity-travel choices with a view to better understand the bundling of choice behaviors while addressing the challenges associated with estimating such econometric model systems. The formulation, specification, and estimation of multi-dimensional choice model systems in which there are a variety of dependent variable types (continuous, ordinal, multinomial, count) has proven to be a challenging task because of the need to evaluate large multi-dimensional integrals of mixtures of distributions in such model systems. As a result, a number of papers in this domain have limited the number of choice dimensions considered to two or have adopted alternative approaches (such as structural equations modeling methods which cannot adequately handle multinomial choice variables) to estimate models with more than two dependent variables.

This paper attempts to overcome the limitations associated with previous work in the specification and estimation of multi-dimensional model systems of location and activity-travel choices. In this study, six choice dimensions are tied together in a joint modeling framework. Residential location and workplace location choices are long term multinomial choice variables, commute distance (which is an outcome of residential location and workplace location choices) is a long term continuous variable, household vehicle ownership is a medium term ordinal dependent variable, commute mode choice is a short-term multinomial travel choice variable, and finally, number of stops made during commute tour is an ordinal dependent variable. These six variables are tied together in a temporal framework as shown in Figure 1a while recognizing the bundling of these choice dimensions associated with the jointness or simultaneity in decision-making. The model system is estimated on a San Francisco Bay Area subsample of the 2009 National Household Travel Survey (NHTS) using the Maximum Approximate Composite Marginal Likelihood (MACML) approach (Bhat, 2011) that provides both computational tractability and numerical accuracy in the estimation of such multi-dimensional econometric model systems with mixtures of dependent variables.

The remainder of this paper is organized as follows. The next section provides a brief review of the literature on simultaneous equations modeling in activity-travel behavior. The third section offers a description of the data, while the fourth section presents the methodology in detail. The fifth section presents model estimation results, while the sixth and final section offers concluding thoughts.

MULTI-DIMENSIONAL ACTIVITY-TRAVEL CHOICE MODELING

The recognition of simultaneity in choice making behaviors has its roots in microeconomic consumer choice theory as evidenced by the partial or general equilibrium class of models developed by LeRoy and Sonstelie (1983) who investigated relationships between residential choice, income, and mode choice, Brown (1986) who postulated that residential location and commute travel mode are goods that consumed simultaneously, and DeSalvo and Huq (1996, 2005) who jointly model residential location, income, and commute mode choice.

In the transportation domain, examples of simultaneous equations models of location and activity-travel choice behaviors abound. Bagley and Mokhtarian (2002) specify and estimate a nine-equation structural equations model system to explore relationships across residential

---

structurally affect one another. In the current paper, we estimate models with each possible structural direction impact, and choose the one that provides a better data fit (which also turns out to one the one that is conceptually intuitive). However, it is critical to note that, regardless of which directionality of structural effects comes out to be better (or even if both directions are not statistically significant), the system is a joint bundled system because of the correlation in unobserved factors impacting the underlying propensities.

location, travel choices, work location, and attitudinal variables. Choo and Mokhtarian (2004) also explore the influence of attitudinal variables on traveler choices by focusing on vehicle type choice. Attitudinal variables, that are often unobserved, play an important role in shaping a multitude of choices, thus calling for the bundling of choices in a simultaneous equations framework where such correlated unobserved factors can be adequately reflected. Van Acker and Witlox (2010a, 2010b) also use structural equations modeling approaches to explore relationships between built environment attributes and vehicle use in a simultaneous equations modeling framework. Vance and Hedel (2007) model the choice of driver status and vehicle use (distance traveled) simultaneously using an instrumental variables approach. Vega and Reynolds-Feighan (2009) employ a cross-nested logit model to study the simultaneous choices of residential location and travel mode under two scenarios of employment (central city versus suburb). Ye *et al.* (2007) use a bivariate probit modeling framework to examine the relationship between trip chaining and mode choice, while Konduri *et al.* (2011) employed a probit-based joint discrete-continuous model to tie vehicle type choice and tour length (distance) together. The latter study was further extended in Paleti *et al.* (2011) who jointly modeled four key dimensions of tours – namely, tour complexity, passenger accompaniment, vehicle type choice, and tour length. Brownstone and Golob (2009) used Bayesian estimation approaches to jointly analyze residential location choice in the context of vehicle type choice and usage and find significant presence of endogeneity in the choice dimensions examined. A similar study was undertaken by Eluru *et al.* (2009), except that they employed Copula-based estimation approaches. Krizek (2003) introduces a tour-based framework to analyze relationships jointly among neighborhood access, number of tours, tour type, and tour distance, while Waddell *et al* (2007) jointly modeled residential location and work place location by assuming strict sequentiality between the two decisions, but allowing the sequentiality structure to vary across households using an endogenous discrete mixture approach.

More recently, Eluru *et al.* (2010) and Pinjari *et al.* (2011) constitute key efforts to build integrated multi-dimensional choice models that tie longer term location choices and shorter term activity-travel choices together. Both of these studies showed strong evidence of the bundling of choices with correlated unobserved effects. Many of the studies cited in this section have noted the computational challenges associated with estimating multi-dimensional choice models, particularly in the presence of a mixture of dependent variable types. However, recent advances in estimation methods, and in particular, the emergence of the Maximum Approximate Composite Marginal Likelihood (MACML) approach (Bhat, 2011), have provided the much needed computational breakthroughs needed to estimate multi-dimensional choice model systems and bring them closer to modeling practice.

DATA

The data for this study is derived from the 2009 National Household Travel Survey (NHTS) which is conducted by the US Department of Transportation on a periodic basis to obtain information about the travel characteristics of the population for a 24 hour travel diary period. For the current study, the survey subsample from the San Francisco Bay Area is extracted for analysis and model estimation purposes. This was done to limit the scope of the geographic region, deal with manageable sample sizes, and take advantage of secondary census data for the region (available from a previous study) that can be merged to the records of the NHTS. As the paper involves the modeling of work location (among other dimensions), the subsample extracted for this study includes only employed individuals who have a fixed work location outside home and who have provided complete travel diary data that includes information on commute tours, mode choice, and stop-making behavior.

Census tract data for the San Francisco Bay Area was merged with the NHTS data records to help characterize household and workplace locations. Instead of using the classic definition of spatial unit choice (identified by census tract or traffic analysis zone), this paper employs categories of land use density to characterize location choices. This helps make the definition of choice alternatives clear and manageable and more effectively captures the notion that people are looking for a built environment (land use density) that suits their mobility and lifestyle preferences. In other words, people are not choosing between tract A or B, but rather between a unit that offers a built environment of certain attributes versus another unit that offers a different built environment. Residence and workplace locations are categorized into four possible alternatives based on housing unit density (housing units per square mile).

After extensive data cleaning, the final estimation sample includes 1,480 employed individuals. Besides residence and work locations, a number of other dependent variables were constructed for this sample. The commute distance is simply a measure of separation between the residence and work locations as reported in the travel diary. Vehicle ownership is reported by respondents as well. For commute tour mode, the mode that was used in the work-to-home (half) tour was designated as the chosen alternative. If transit was used for any leg of the journey, then the commute tour mode was designated as transit. Four modal alternatives – drive alone, shared ride, transit, and walk/bike – characterized the mode choice for more than 99 percent of the tours. The few people whose commute tours did not fall within one of these four modal alternatives were omitted from the final estimation sample. Finally, the total number of stops made during the home-to-work and work-to-home tours constituted the last dependent variable of the study.

The sample of 1,480 employed individuals exhibited socio-economic and demographic characteristics suitable for undertaking a model estimation effort such as that undertaken in this paper. The distribution of individuals in the four residential location alternatives is as follows:

- 0-499 housing units per square mile: 22.6%
- 500-1999 housing units per square mile: 30.9%
- 2000-3999 housing units per square mile: 29.9%
- ≥ 4000 housing units per square mile: 16.6%

The distribution of individuals with respect to work locations is somewhat similar except that higher percent of individuals (32.4%) work in low density (0-499) tracts while a smaller percent (20.5%) of individuals work in higher density (2000-3999) tracts. With respect to vehicle ownership, 1.8 percent of the employed individuals indicate residing in households with no vehicle. This fraction is lower than that for the general population, but such differences are expected when considering a pure worker sample. About 47 percent of individuals reside in two-vehicle households, 23.2 percent reside in three-vehicle households, and 15 percent reside in households with four or more vehicles.

An examination of commute mode share shows that 72.6 percent of individuals commute by drive alone, 16.1 percent by shared ride, 8 percent by transit, and 3.2 percent by bicycle/walk. The average commute distance is 13.5 miles with a standard deviation of 14.4 miles. The distribution of stop-making shows that 47 percent of commuters make zero (non-work) stops within the commute tours. This is in contrast to 17.4 percent of commuters who make one stop, 16.7 percent who make two stops, 8.8 percent reporting three stops, 5.5 percent reporting four stops, and 4.5 percent reporting five or more stops.

In summary, the data set offered a rich source of information and appropriate variation in dependent variables suitable for estimating a multi-dimensional choice model system with a mixture of dependent variable types. The model specification included a range of individual, household, and employment characteristics.

MODELING METHODOLOGY

This section presents a detailed description of the modeling methodology developed for estimating a multi-dimensional choice model system involving a mixture of dependent variable types. Figure 1a shows the various interdependencies that might exist in the choice continuum that this study intends to explore. The solid lines represent possible relationships within single time bands while the hollow lines represent relationships across temporal bands (scales). There can be joint decisions within a single temporal band as well as decisions that are interlinked across different temporal bands. The remainder of this section presents the formulation.

Model Framework

Let there be G nominal (unordered-response) variables for an individual, and let g be the index for the nominal variables (g = 1, 2, 3, …, G). In the empirical context of the current paper, G=3 (the nominal variables are residential location, work location, and commute mode choice). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 3$) and let i_g be the corresponding index (i_g = 1, 2, 3, …, I_g). Note that I_g may vary across individuals, but index for individuals is suppressed at this time for ease of presentation. Also, it is possible that some nominal variables do not apply for some individuals, in which case G itself is a function of the individual q. However, the model is developed at the individual level, and so this notational nuance does not appear in the presentation here.

Consider the g^{th} nominal variable and assume that the individual under consideration chooses the alternative m_g. Also, assume the usual random utility structure for each alternative i_g.

$$U_{gi_g} = \boldsymbol{\beta}'_g \boldsymbol{x}_{gi_g} + \varepsilon_{gi_g}, \tag{1}$$

where $\boldsymbol{x}_{gi_g}$ is a (K_g×1)-column vector of exogenous attributes, $\boldsymbol{\beta}_g$ is a column vector of corresponding coefficients, and ε_{gi_g} is a normal error term. Let the variance-covariance matrix of the vertically stacked vector of errors $\varepsilon_g[=(\varepsilon_{g1}, \varepsilon_{g2}, ..., \varepsilon_{gI_g})']$ be $\boldsymbol{\Omega}_g$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Omega}_g$ for identification. Under the utility maximization paradigm, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$, since the individual chose alternative m_g. Let $u^*_{gi_g m_g} = U_{gi_g} - U_{gm_g} \ (i_g \neq m_g)$, and stack the latent utility differentials into a vector $\boldsymbol{u}^*_{\boldsymbol{g}} = \left[\left(u^*_{g1m_g}, u^*_{g2m_g}, ..., u^*_{gI_g m_g}\right)'; i_g \neq m_g\right]$. $\boldsymbol{u}^*_{\boldsymbol{g}}$ has a mean vector of $\boldsymbol{b}_{\boldsymbol{g}}\,(\boldsymbol{\beta}'_{\boldsymbol{1}}\boldsymbol{z}_{\boldsymbol{g1m}_g}, \boldsymbol{\beta}'_{\boldsymbol{1}}\boldsymbol{z}_{\boldsymbol{g2m}_g}, ...\boldsymbol{\beta}'_{\boldsymbol{1}}\boldsymbol{z}_{\boldsymbol{gI}_g\boldsymbol{m}_g})'$, where $\boldsymbol{z}_{\boldsymbol{gi}_g\boldsymbol{m}_g} = \boldsymbol{x}_{\boldsymbol{gi}_g} - \boldsymbol{x}_{\boldsymbol{gm}_g}, i_g = 1, 2, ..., I_g; i_g \neq m_g$. To obtain the covariance matrix of $\boldsymbol{u}^*_{\boldsymbol{g}}$, define $\boldsymbol{M}_{\text{g}}$ as an $(I_g - 1) \times I_g$ matrix that corresponds to an $(I_g - 1)$ identity matrix with an extra column of –1's added as the m_g^{th} column. Then, one may write:

$$\boldsymbol{u}^*_{\boldsymbol{g}} \sim N\left(\boldsymbol{b}_{\boldsymbol{g}}, \boldsymbol{\Sigma}^*_{\text{g}}\right), \text{ where } \boldsymbol{\Sigma}^*_{\text{g}} = \boldsymbol{M}_{\boldsymbol{g}} \boldsymbol{\Omega}_{\text{g}} \boldsymbol{M}'_{\boldsymbol{g}}. \tag{2}$$

The discussion above focuses on a single nominal variable g. When there are G nominal variables, consider the stacked $\left[\sum_{g=1}^{G}(I_g-1)\right]\times 1-$vector $\boldsymbol{u}^* = \left[\left(\boldsymbol{u}_1^{*\prime}, \boldsymbol{u}_2^{*\prime}, \ldots, \boldsymbol{u}_G^{*\prime}\right)^{\prime}\right]$, each of whose element vectors is formed by differencing utilities of alternatives from the chosen alternative m_g for the g^{th} nominal variable. Next, one may write:

$\boldsymbol{u}^* \sim N(\boldsymbol{b}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{b} = \left(\boldsymbol{b}_1^{\prime}, \boldsymbol{b}_2^{\prime}, \ldots, \boldsymbol{b}_G^{\prime}\right)^{\prime}$ and $\boldsymbol{\Sigma}^*$ is a $\left[\sum_{g=1}^{G}(I_g-1)\right] * \left[\sum_{g=1}^{G}(I_g-1)\right]$ matrix as follows:

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_1^* & \boldsymbol{\Sigma}_{12}^* & \cdot\ \cdot\ \cdot & \boldsymbol{\Sigma}_{1G}^* \\ \boldsymbol{\Sigma}_{21}^* & \boldsymbol{\Sigma}_2^* & \cdot\ \cdot\ \cdot & \boldsymbol{\Sigma}_{2G}^* \\ \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \boldsymbol{\Sigma}_{G1}^* & \boldsymbol{\Sigma}_{G2}^* & \cdot\ \cdot\ \cdot & \boldsymbol{\Sigma}_G^* \end{bmatrix} \tag{3}$$

The off-diagonal elements in $\boldsymbol{\Sigma}^*$ capture the dependencies across the utility differentials of different nominal variables, the differential being taken with respect to the chosen alternative for each nominal variable.

Let there be L ordinal variables for an individual, and l be the index for the ordinal variables $(l = 1, 2, \ldots, L)$. In the empirical context of the current paper, L=2 (the ordinal variables are vehicle ownership and number of stops in the commute). Also, let J_l be the number of outcome categories for the l^{th} ordinal variable $(J_l \geq 2)$ and let the corresponding index be j_l $(j_l = 1, 2, \ldots, J_l)$. Let y_l^* be the latent underlying variable whose horizontal partitioning leads to the observed choices for the l^{th} ordinal variable. Assume that the individual under consideration chooses the ${n_l}^{th}$ ordinal category. Then, in the usual ordered response formulation:

$$y_l^* = \boldsymbol{\delta}_l^{\prime} \boldsymbol{w}_l + \xi_l,\ j_l = n_l \text{ if } \psi_{n_l - 1} < y_l^* < \psi_{n_l}, \tag{4}$$

where $\boldsymbol{w}_l$ is a vector of exogenous variables relevant to the l^{th} ordinal variable, $\boldsymbol{\delta}_l$ is a corresponding vector of coefficients to be estimated, the ψ terms represent thresholds, e_l is the index for the observed outcome for the ordinal variable $(j_l = 1, 2, \ldots, J_l)$, and ξ_l is the standard normal random error for the l^{th} ordinal variable. Stack the L latent variables y_l^* into an $(L \times 1)$ vector $\boldsymbol{y}^*$, and let $\boldsymbol{y}^* \sim N\left(\boldsymbol{f}, \boldsymbol{\Sigma}_{y^*}\right)$, where $\boldsymbol{f} = \left(= (\boldsymbol{\delta}_1^{\prime}\boldsymbol{w}_l, \boldsymbol{\delta}_2^{\prime}\boldsymbol{w}_2, \ldots\ldots, \boldsymbol{\delta}_L^{\prime}\boldsymbol{w}_L\right)$ and $\boldsymbol{\Sigma}_{y^*}$ is the covariance matrix of $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_L)$. Also, stack the lower thresholds $\psi_{n_l - 1}\left(l = 1, 2, \ldots, L\right)$ into an $(L \times 1)$ vector $\boldsymbol{\psi}_{low}$ and the upper thresholds $\psi_{n_l}\left(l = 1, 2, \ldots, L\right)$ into another vector $\boldsymbol{\psi}_{up}$.

Finally, let there be H continuous variables $(y_1, y_2, \ldots, y_H)$ with an associated index h $(h = 1, 2, \ldots, H)$. In the empirical context of the current paper, H=1 (the continuous variable is natural logarithm of commute distance). Let $y_h = \boldsymbol{\gamma}_h^{\prime} \boldsymbol{s}_h + \eta_h$ in the usual linear regression fashion. Stacking the H continuous variables into a $(H \times 1)$-vector $\boldsymbol{y}$, one may write $y = N(\boldsymbol{c}, \boldsymbol{\Sigma}_y)$, where

$\boldsymbol{c} = \left(\gamma_1' s_1, \gamma_2' s_2, \ldots \gamma_H' s_H\right)'$, and $\boldsymbol{\Sigma}_y$ is the covariance matrix of $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots \eta_H)$. The variance of $\widetilde{\boldsymbol{y}} = \left(\boldsymbol{u}^*, \boldsymbol{y}^*, \boldsymbol{y}\right)$ can be written as:

$$\mathrm{Var}(\widetilde{\boldsymbol{y}}) = \widetilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Sigma}_{u^*} & \boldsymbol{\Sigma}_{u^*y^*} & \boldsymbol{\Sigma}_{u^*y} \\ \boldsymbol{\Sigma}_{u^*y^*} & \boldsymbol{\Sigma}_{y^*} & \boldsymbol{\Sigma}_{y^*y^*} \\ \boldsymbol{\Sigma}_{y^*y} & \boldsymbol{\Sigma}_{y^*y} & \boldsymbol{\Sigma}_{y} \end{bmatrix}, \tag{5}$$

where $\boldsymbol{\Sigma}_{u^*y^*}$ is a $\left[\sum_{g=1}^{G}(I_g - 1)\right] \times L$ matrix capturing covariance effects between the $\boldsymbol{u}^*$ vector and the $\boldsymbol{y}^*$ vector, $\boldsymbol{\Sigma}_{u^*y}$ is a $\left[\sum_{g=1}^{G}(I_g - 1)\right] \times H$ matrix capturing covariance effects between the $\boldsymbol{u}^*$ vector and the $\boldsymbol{y}$ vector, and $\boldsymbol{\Sigma}_{y^*y}$ is a $L \times H$ matrix capturing covariance effects between the $\boldsymbol{y}^*$ vector and the $\boldsymbol{y}$ vector. For ease in presentation, define $\widetilde{\boldsymbol{u}} = \left(\boldsymbol{u}^{*\prime}, \boldsymbol{y}^{*\prime}\right)'$, $\widetilde{g} = (b', f')'$, and $\boldsymbol{\Sigma}_{\widetilde{u}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{u}^*} & \boldsymbol{\Sigma}_{\boldsymbol{u}^*\boldsymbol{y}^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{u}^*\boldsymbol{y}^*} & \boldsymbol{\Sigma}_{\boldsymbol{y}^*} \end{bmatrix}$ and $\mathrm{Var}(\widetilde{\boldsymbol{y}}) = \widetilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\widetilde{u}} & \boldsymbol{\Sigma}_{\widetilde{u}y} \\ \boldsymbol{\Sigma}_{\widetilde{u}y} & \boldsymbol{\Sigma}_{y} \end{bmatrix}$.

Also, supplement the threshold vectors defined earlier as follows: $\widetilde{\boldsymbol{\psi}}_{low} = \left[\left(-\infty_{\left[\sum_{g=1}^{G}(I_g-1)\right]}\right)', \boldsymbol{\psi}'_{low}\right]'$, and $\widetilde{\boldsymbol{\psi}}_{up} = \left[\left(\boldsymbol{0}_{\left[\sum_{g=1}^{G}(I_g-1)\right]}\right)', \boldsymbol{\psi}'_{up}\right]'$, where $-\infty_M$ is a $(M\times 1)$-column vector of negative infinities, and $\boldsymbol{0}_M$ is another $(M\times 1)$-column vector of zeros. The conditional distribution of $\widetilde{\boldsymbol{u}}$ given $\boldsymbol{y}$, is multivariate normal with mean $\widetilde{\widetilde{\boldsymbol{g}}} = \widetilde{\boldsymbol{g}} + \boldsymbol{\Sigma}_{\widetilde{u}y}\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y} - \boldsymbol{c})$ and variance $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{\widetilde{\mathbf{u}}} = \boldsymbol{\Sigma}_{\widetilde{\mathbf{u}}} - \boldsymbol{\Sigma}_{\widetilde{\mathbf{u}}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{u}}\mathbf{y}}'$.

Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_G;\ \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_L;\ Vech(\boldsymbol{\Sigma}_{\widetilde{u}});\ \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_H;\ Vech(\boldsymbol{\Sigma}_y);\ Vech(\boldsymbol{\Sigma}_{\widetilde{u}\mathbf{y}})]$, where $Vech(\mathbf{A})$ represents the vector of upper triangle elements of $\mathbf{A}$. Then the likelihood function for the individual may be written as:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_y) \times \Pr\left[\widetilde{\boldsymbol{\psi}}_{low} \leq \widetilde{\boldsymbol{u}} \leq \widetilde{\boldsymbol{\psi}}_{up}\right], \\ &= \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_y) \times \int_{D_{\widetilde{u}}} \phi_{\widetilde{G}+L}(\widetilde{\boldsymbol{u}} \mid \widetilde{\widetilde{\boldsymbol{g}}}, \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{\widetilde{u}})\, d\widetilde{u}, \end{aligned} \tag{6}$$

where the integration domain $D_{\widetilde{u}} = \{\widetilde{\boldsymbol{u}} : \widetilde{\boldsymbol{\psi}}_{low} \leq \widetilde{\boldsymbol{u}} \leq \widetilde{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region of the elements of the $\widetilde{\boldsymbol{u}}$ vector determined by the vector of chosen alternatives in nominal variables and observed outcomes of ordinal variables, and $\phi_{\tilde{G}+L}(.)$ is the multivariate normal density function of dimension $\tilde{G}+L,$ where $\tilde{G}=\left(\left[\sum_{g=1}^{G}(I_g-1)\right]\right).$

The above likelihood function involves the evaluation of a $\tilde{G}+L$-dimensional integral for each individual, which can be very computationally expensive if there are several nominal variables, or if each nominal variable can take a large number of values, or if there are several ordinal variables, or combinations of these. So, the Maximum Approximated Composite Marginal Likelihood (MACML) approach of Bhat (2011), in which the likelihood function only involves the computation of univariate and bivariate cumulative distributive functions, is used in this paper.

The MACML Estimation Approach

Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the G nominal variables and L ordinal variables) of the joint pairwise probability of the chosen alternatives for an individual.

$$L_{CML}(\boldsymbol{\theta})=\phi_H(\boldsymbol{y}-\boldsymbol{c}\mid\boldsymbol{\Sigma}_{\mathrm{y}})\times\left(\prod_{g=1}^{G-1}\prod_{g'=g+1}^{G}\Pr(d_{i_g}=m_g,d_{i_{g'}}=m_{g'})\right)\times\left(\prod_{l=1}^{L-1}\prod_{l'=l+1}^{L}\Pr(j_l=n_l,j_{l'}=n_{l'})\right)$$
$$\times\left(\prod_{g=1}^{G}\prod_{l'=1}^{L}\Pr(d_{i_g}=m_g,j_l=n_l)\right). \tag{7}$$

where d_{i_g} is an index for the individual's choice for the g^{th} nominal variable. The net result is that the pairwise likelihood function now only needs the evaluation of $\tilde{G}_{gg'},\tilde{G}_{ll'}$, and $\tilde{G}_{gl}$ dimensional cumulative normal distribution functions (rather than the $\tilde{G}+L$-dimensional cumulative distribution function in the maximum likelihood function), where $\tilde{G}_{gg'}=I_g+I_{g'}-2,\ \tilde{G}_{ll'}=2,$ and $\tilde{G}_{gl}=I_g$. This leads to substantial computational efficiency. However, in cases where there are several alternatives for one or more nominal variables, the dimension $\tilde{G}_{gg'}$ and $\tilde{G}_{gl}$ can still be quite high. This is where the use of an analytic approximation of the multivariate normal cumulative distribution (MVNCD) function, as shown in Bhat (2011), is convenient. The resulting maximum approximated composite marginal likelihood (MACML) of Bhat (2011), which combines the CML approach with the analytic approximation for the MVNCD function evaluation, is solely based on bivariate and univariate cumulative normal computations. The MACML approach can be applied using a simple optimization approach for likelihood estimation. It also represents a conceptually simpler alternative to simulation techniques. Also, the MACML estimator $\hat{\boldsymbol{\theta}}_{MACML}$ is asymptotically normal distributed with mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of the Godambe's (1960) sandwich information matrix $G(\boldsymbol{\theta})$.

There are important identification and positive definiteness issues that must be taken into account during model estimation. These issues and the methods to deal with them are discussed in Paleti *et al.* (2011). In addition to the identification conditions discussed in that paper, the scale of all ordinal variables must be normalized to one in the current model system to ensure identification.

MODEL ESTIMATION RESULTS

Model estimation results are described in this section. In the interest of brevity, only key findings and highlights of model estimation results are presented. In order to arrive at the final model specification, a number of model structures depicting alternative structural relationships among endogenous variables were estimated and examined with respect to statistical measures of fit. In the end, after extensive testing, plausibility checks, and goodness-of-fit assessment, the final model specification and set of structural relationships were identified. Figure 1b shows the structural relationships among dependent variables in the final model structure adopted in this study. It is found that residential location affects work location choice utility, both of which affect commute distance. All three long-term choice variables (residential location, work location, and commute distance) affect vehicle ownership propensity. In turn, long term location choices and vehicle ownership structurally influence commute mode choice utility, which structurally impacts trip chaining patterns (number of stops propensity on the commute). It should be emphasized again that these are the structural flow of relationships in the final model specification. The model system itself is a joint equations model that treats the set of dependent variables as a "bundle" with common unobserved effects affecting multiple choice dimensions.

Long Term Choice Model Components

Table 1 presents estimation results for long term choices. The residential location choice component of the model suggests that households with younger children have a greater propensity to locate in medium- to high-density neighborhoods, but households with older children shun the highest density neighborhoods, possibly in search of lower density suburban neighborhoods with good schools. Pinjari *et al.* (2008) also reported that households with children are less likely to live in high density neighborhoods. Individuals with higher education levels favor residential locations in high density neighborhoods, suggesting that they are interested in urban lifestyles that are more environmentally friendly. This result is different from the U-shaped effect of education on residential location reported in the Brownstone and Fang (2009) study, which modeled logarithm of residential block density as a function of several household demographics. Lower income individuals tend to locate in high density neighborhoods while those seeking home ownership appear to do so in lowest density neighborhoods (likely to be in the suburbs) (see Brownstone and Fang, 2009 for similar results). Immigrant households are more likely to favor higher density neighborhoods, a result also reported by Wilson and Singer (2011) in their analysis of the 2010 American Community Survey data. The relative magnitude of the constants suggests that there is a baseline preference for low-to-medium density neighborhoods.

In terms of work location choice, it is found that there is a strong positive association between residential location density and work location density utility. It appears that people may be working in locations that are at least as dense as their residential neighborhoods, which is not surprising given that employment tends to locate such that workers can easily access jobs. Ebertz (2009) found similar results when jointly examining residential and work location choices of a household. Specifically, the study found that households have the highest baseline utility preference for living and working in metropolitan areas. Males are less likely to work in higher density locations. Individuals with higher education levels tend to find jobs in higher density areas (consistent with their residential location). Full time workers are less likely to work in high density areas, but self-employed individuals are more likely to do so. It is possible that self-employed individuals seek high density areas where business opportunities abound. Immigrants are less likely to work in high density areas (in contrast to their residential location choice), but tend to favor higher density locations (similar to non-immigrant households) as they assimilate into the country over a period of time. Asians are less likely to work in higher density neighborhoods, while African Americans are more likely to do so.

The commute distance is similarly affected by a number of socio-economic variables. Males, full-time employees, and African Americans exhibit longer commutes, while lower income individuals and those with children have lower commuting distances. The first result that men have longer commutes than women is consistent with the findings of earlier literature on commute travel patterns (see Sermons and Koppelman, 2001, and Vovsha *et al.,* 2012). Full-time workers, on the hand, might be trading off commute distance with higher wages (Ebertz, 2009). Those who own a home have longer commutes, presumably because they reside in distant suburbs to a greater degree. As residential location density or work location density increases, the commuting distance decreases; suggesting that there is an observed impact of density on commuting distance even after controlling for other factors and reflecting endogeneity through a simultaneous equations model system.

Medium Term Choice Model Component

The vehicle ownership model takes the form of an ordered response model. The results are presented in Table 2. Higher levels of auto ownership are associated with a larger number of persons in the household. Thus, as number of adults, number of children, number of full time workers, number of self-employed individuals, and number of individuals with more than one job in the household (in which the sample respondent resides) increase, so does auto ownership. On the other hand, the presence of senior adults in the household or the prevalence of a medical condition has a negative impact on auto ownership presumably because these individuals have mobility limitations. As income levels fall, so do auto ownership levels as evidenced by the trend in negative coefficients associated with income dummy variables. Higher density residential location is associated with lower levels of auto ownership, presumably because these neighborhoods are better served by alternative modes and people who locate in such neighborhoods are not necessarily auto-oriented to begin with. Home ownership and longer commutes appear to contribute to higher levels of auto ownership. All of these indications are consistent with expectations and with the now vast literature on auto ownership modeling (for example, see Potoglou and Susilo, 2008, Ma and Srinivasan, 2010, and Pinjari *et al.*, 2011)

Short Term Choice Model Component

Table 3 presents the model estimation results for the short-term choice components. There is negative baseline preference associated with the use of alternative modes of transport as evidenced by the negative constants. Older individuals are less likely to share a ride or bike/walk, possibly due to physical limitations. Males are less likely to share a ride, but more likely (than females) to use transit or bicycle and walk (see Pinjari *et al.*, 2011). Low education levels are associated with alternative mode use, possibly because these individuals are in low paying jobs, having lower income, and cannot afford to commute by car. Self-employed individuals are more likely to drive alone, possibly due to the flexibility that they need in seeking business opportunities. Those with a flexible work schedule are more likely to use alternative modes of transport. Immigrants are more likely to share a ride or use transit, but this effect dampens as the immigrants stay longer in the US and assimilate into the general population, as also noticed by Blumenberg and Norton (2010). Even after controlling for all other factors and endogeneity across choice dimensions, it is found that residential and workplace location density impact commute mode choice utility. Higher density location choices appear to contribute to greater levels of transit mode choice. Those working in high density tracts show a lower propensity to bicycle and walk, possibly because the areas are not conducive to non-motorized mode use (although conducive to transit use). As expected, and observed in earlier studies (for example, Van Acker and Witlox, 2010b), high levels of vehicle ownership negatively impact alternative mode use due to increased auto availability).

The final dependent variable is that of number of stops on the commute tours (an ordinal response variable). Consistent with expectations, higher levels of education, holding multiple jobs, and flexible work schedules are associated with higher levels of stop-making propensity. Immigrants tend to have a lower stop-making propensity, while Caucasians and individuals with children in the household tend to have a higher stop-making propensity (due to serve-child trips). As the number of adults increases, stop making responsibilities are likely shared through household interactions, and stop-making propensity at the individual level drops (see Ye *et al.*, 2007). Similar task allocation effects are seen with respect to number of workers and number of self-employed individuals in the household. Lower income individuals have a lower stop-making propensity, possibly because the lower income does not afford them the opportunity to participate in other discretionary activities (Portoghese *et al.*, 2011 also observe this result). Those residing in the highest density neighborhoods tend to engage in more stops, possibly because there are more destination opportunities that can be visited during the commute tour. In other words, higher residential density does not necessarily bring about inefficiencies in tour formation or activity engagement (where a person repeatedly returns home and starts a new tour to engage in new activities). Mode choice affects stop making behavior with those in shared ride mode likely to make more stops (to drop off and pick up passengers) and bicycle and walk commuters engaging in fewer stops, possibly in an effort to keep commuting distance and times manageable (see Cao *et al.*, 2008b for similar results).

Self-Selection Effects and Model Assessment

Table 4 presents estimation results corresponding to the covariance matrix of utility differences, latent propensities, and continuous variables considered in this study. A number of interesting observations can be made. The significant parameter of 0.8009 in the block of covariances between modal utility differences, suggests that there are common unobserved factors affecting the choice of transit (relative to drive alone) and the choice of bicycle/walk (relative to drive alone). In other words, people's attitudes about the environment and the desire to live a "green" lifestyle (which are unobserved effects) may be simultaneously (and positively) impacting preference for transit and bicycle/walk modes. There do not appear to be any significant endogeneity effects across residential and workplace location choices. The model estimation results revealed an observed impact of residential location choice on work location choice; there do not seem to be any common unobserved effects influencing these long term location choice decisions (at least in the context of this study).

It appears that there are self-selection effects across work location choice and commute mode. The negative parameter of -0.1507 suggests that unobserved factors that contribute to a person choosing a low density area as work location are correlated with unobserved factors that make a person intrinsically less likely to walk or bicycle. These may be individuals who are more auto oriented by nature. Conversely, there are positive covariances (0.0555 and 0.2883) reflecting a positive disposition across the choice to work in high density areas and the choice of transit or shared ride as a commute mode. The unobserved factors that motivate an individual to seek a high density work place (desire for transit and pedestrian friendly options) are likely the very factors that contribute to higher level of transit and shared ride mode usage. Unobserved factors that contribute to an individual owning more vehicles (such as desiring an auto-oriented lifestyle) contribute negatively to the choice of bicycle and walk as a commute mode.

Similar self-selection effects are observed across residential location choice and number of stops, where it appears that the unobserved effects contributing to a choice of a high density residential location or work location positively impact stop-making behavior. This is plausible as a fun-loving activity-seeking person who is an extrovert may choose residential and work locations that are high density (and provide such opportunities) and support their desire to engage in a variety of activities (stops) on the way to and from work.

The log-likelihood of the final model is -10508.1 and that of the model which ignores all potential correlations between the choices considered is -10520.4. The log-likelihood ratio test statistic of comparison between the two models is 24.54. This value is significantly greater than 15.51 which is the critical chi-squared value corresponding to 8 degrees of freedom at a 95 percent confidence level, thus demonstrating the superior statistical fit in the joint model.

CONCLUSIONS

This paper presents an integrated econometric model system that ties together residential location choice, work location choice, commuting distance, vehicle ownership, commute mode choice, and number of stops made on commute tours. Thus, the model system includes a variety of dependent variable types commonly encountered in transport modeling contexts. The model system is estimated on a San Francisco Bay Area subsample of commuters drawn from the 2009 National Household Travel Survey data set in the United States. The paper presents the model formulation and estimation procedures; recognizing that traditional estimation methods are computationally infeasible for the type of model system specified in this paper, the study employs the maximum approximate composite marginal likelihood (MACML) estimation procedure together with a numerical approximation method to evaluate multi-dimensional integrals of the cumulative normal distribution function. The methodological breakthrough presented in this paper offers the potential to bring integrated model systems of the nature estimated in this study closer to practical reality.

Model estimation results show that the choice dimensions considered in this paper are inter-related, both through direct observed structural relationships and through correlations across unobserved factors (error terms) affecting multiple choice dimensions. The significant presence of self-selection effects (endogeneity) suggests that modeling the various choice processes in an independent sequence of models is not reflective of the true relationships that exist across these choice dimensions. The study findings suggest the following:

- Residential location choice affects work location choice utilities
- Both residential and work location choices together impact commuting distance
- Residential and work location choices, together with commuting distance, impact vehicle ownership propensity
- Both location choices, and vehicle ownership, affect commute mode choice propensities
- Commute mode choice and residential location affect number of stops propensity on commute tours.

In addition to these observed structural relationships, the examination of error covariances shows that people with a propensity for non-auto oriented lifestyles (*i.e.*, greener lifestyles) tend to locate in higher density neighborhoods, adopt alternative modes of transport for their commute, and exhibit lower levels of automobile ownership. Clearly, attitudes and lifestyle preferences play an important role in shaping the multitude of choice dimensions considered and ignoring such self-selection effects can prove costly in policy forecasting and decision making processes. Future research efforts are aimed at operationalizing integrated econometric model systems (such as that presented in this paper) within activity-based travel forecasting models so that the types of endogeneity effects uncovered in this paper can be better reflected in forecasts of travel demand under a wide range of modal and land use scenarios. For instance, the residential and work location choices are modeled as density choices and not as actual spatial location choices such as traffic analysis zones (TAZs). One way to operationalize the model developed in the paper would be to first predict the density of location choice, and subsequently use another spatial location choice model that operates on a choice set comprising of travel analysis zones (TAZs) in the predicted range of location density.

The work undertaken in this paper can be extended in two important ways. *First*, the model structure in the study is a restrictive version of a modeling system that allows mixtures of structural relationships among endogenous variables. A latent segmentation model that determines the joint probability of the observed bundle of choices as a summation (over all possible structural relationships) of the product of the probability of each structural relationship among the endogenous variables and the probability of the observed bundle of choices conditional on the structural relationship may be developed in the future to accommodate different structural relationships for different population segments. This is conceptually similar to the discrete mixture segment model of Waddell *et al.* (2007), though our system already considers jointness through the error correlations for each segment (as opposed to the sequential process of decisions for each segment in Waddell *et al*., 2007). Of course, the consideration of many more variables than in Waddell *et al.* will be an interesting estimation challenge. *Second*, we consider decision making at individual level. The model system, as it is, cannot be operationalized into an activity-based model system because it becomes difficult to maintain consistency in the household level choices across different household members. The current modeling framework can be extended to consider an even larger set of multinomial/ordinal/continuous choices at different levels of decision making (some at the household level, and some at the individual level) given that the MACML estimation technique used in the paper is robust and can be used to estimate any number of choices within a unifying framework as long as adequate data is available to extract out the system relationships. Future research should work towards developing such integrated models at different decision making levels.

ACKNOWLEDGEMENTS

The authors acknowledge the helpful comments of four anonymous reviewers on an earlier version of the paper. The authors are grateful to Lisa Macias for her help in formatting this document.

REFERENCES

Abraham, J., and Hunt, J. (1997) Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Transportation Research Record: Journal of the Transportation Research Board*, 1606, 17-24.

Bagley, M.N., and Mokhtarian, P.L. (2002) The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science*, 36(2), 279-297.

Bhat, C.R. (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, 45(7), 923-939.

Bhat, C.R., and Guo, J.Y. (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B*, 41(5), 506-526.

Blumenberg, E., and A. Norton (2010) Falling immigration rates mean falling transit ridership. *Access*, 37, 10-16.

Brown, B. (1986) Modal choice, location demand, and income. *Journal of Urban Economics*, 20(2), 128-139.

Brownstone, D., and Fang, H.A. (2009) A vehicle ownership and utilization choice model with endogenous residential density. Compendium of Papers DVD, Transportation Research Board 88th Annual Meeting, Washington, D.C.

Brownstone, D., and Golob, T.F. (2009) The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1), 91-98.

Cao, X., Mokhtarian, P.L., and Handy, S.L. (2008a) Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews*, 29(3), 359-395.

Cao, X., Mokhtarian, P.L., and Handy, S.L. (2008b) Differentiating the influence of accessibility, attitudes, and demographics on stop participation and frequency during the evening commute. *Environment and Planning B: Planning and Design*, 35, 431-442.

Choo, S., and Mokhtarian, P.L. (2004) What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A*, 38(3), 201-222.

DeSalvo, J.S., and Huq, M. (1996) Income, residential location, and mode choice. *Journal of Urban Economics*, 40(1), 84-99.

DeSalvo, J.S., and Huq, M. (2005) Mode choice, commuting cost, and urban household behavior. *Journal of Regional Science*, 45(3), 493-517.

Ebertz, A. (2009) The Determinants of Joint Residential and Job Location Choices: A Mixed Logit Approach. Working Paper 82, Institute for Economic Research at University of Munich.

Eluru, N., Bhat, C.R., Pendyala, R.M., and Konduri, K.C. (2010) A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation*, 37(4), 603-626.

Eluru, N., Sener, I.N., Bhat, C.R., Pendyala, R.M., and Axhausen, K.W. (2009) Understanding residential mobility. *Transportation Research Record: Journal of the Transportation Research Board*, 2133, 64-74.

Godambe, V. (1960) An Optimum Property of Regular Maximum Likelihood Estimation. *Annals of Mathematical Statistics*, 31, 1208-1211.

Konduri, K.C., X. Ye, B. Sana, and R.M. Pendyala (2011) A Joint Tour-Based Model of Vehicle Type Choice and Tour Length. *Transportation Research Record*: *Journal of the Transportation Research Board*, 2255, 28-37.

Krizek, J.K. (2003) Neighborhood services, trip purpose, and tour-based travel. Transportation 30(4), 387-410.

LeRoy, S.F., and Sonstelie, J. (1983) Paradise lost and regained: Transportation innovation, income, and residential location. *Journal of Urban Economics*, 13(1), 67-89.

Ma, L., and Srinivasan, S. (2010) Impact of individuals' immigrant status on household auto ownership. *Transportation Research Record: Journal of the Transportation Research Board*, 2156, 36-46.

Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, UK.

Paleti, R., Pendyala, R.M., Bhat, C.R., and Konduri, K.C. (2011) A joint tour-based model of tour complexity, passenger accompaniment, vehicle type choice, and tour length. Technical paper, Arizona State University, Tempe, AZ.

Pinjari, A.R., Pendyala, R.M., Bhat, C.R., and Waddell, P.A. (2011) Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6), 933-958.

Pinjari, A.R., Eluru, N., Bhat, C.R., Pendyala, R.M., and Spissu, E. (2008) joint model of choice of residential neighborhood and bicycle ownership: accounting for self-selection and unobserved heterogeneity. *Transportation Research Record*, 2082, 17-26.

Portoghese, A., Spissu, E., Bhat, C.R., Eluru, N., and Meloni, I. (2011) A copula-based joint model of commute mode choice and number of non-work stops during the commute. *International Journal of Transport Economics*, 38(3), 337-364.

Potoglou, D., and Susilo, Y.O. (2008) Comparison of vehicle-ownership models. *Transportation Research Record,* 2076, 97–105.

Sermons, M.W., and Koppelman, F.S. (2001) Representing the differences between male and female commute behavior in residential location choice models. *Journal of Transport Geography*, 9, 101-110.

Van Acker, V., and Witlox, F. (2010a) Commuting trips within tours: how is commuting related to land use? *Transportation*, 38(3), 465-486.

Van Acker, V., and Witlox, F. (2010b) Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. *Journal of Transport Geography*, 18(1), 65-74.

Vance, C., and Hedel, R. (2007) The impact of urban form on automobile travel: disentangling causation from correlation. *Transportation*, 34(5), 575-588.

Vega, A., and Reynolds-Feighan, A. (2009) A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns. *Transportation Research Part A*, 43(4), 401-419.

Vovsha, P., Gupta, S., Freedman, J., Sun, W., and Livshits, V. (2012) Workplace Choice Model: Comparison of Spatial Patterns of Commuting in Four Metropolitan Regions. Presented at the 91st Annual Meeting of the Transportation Research Board, January 2012, Washington, D.C.

Waddell, P., Bhat, C., Eluru, N., Wang, L., and Pendyala, R. (2007) Modeling interdependence in household residence and workplace choices. *Transportation Research Record: Journal of the Transportation Research Board*, 2003, 84-92.
Wilson, J.H., and Singer, A. (2011) Immigrants in 2010 Metropolitan America: A decade of change. *State of Metropolitan America series*, Metropolitan Policy Program at Brookings, Washington D.C., October 2011.
Ye, X., Pendyala, R.M., and Gottardi, G. (2007) An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B*, 41(1), 96-113.

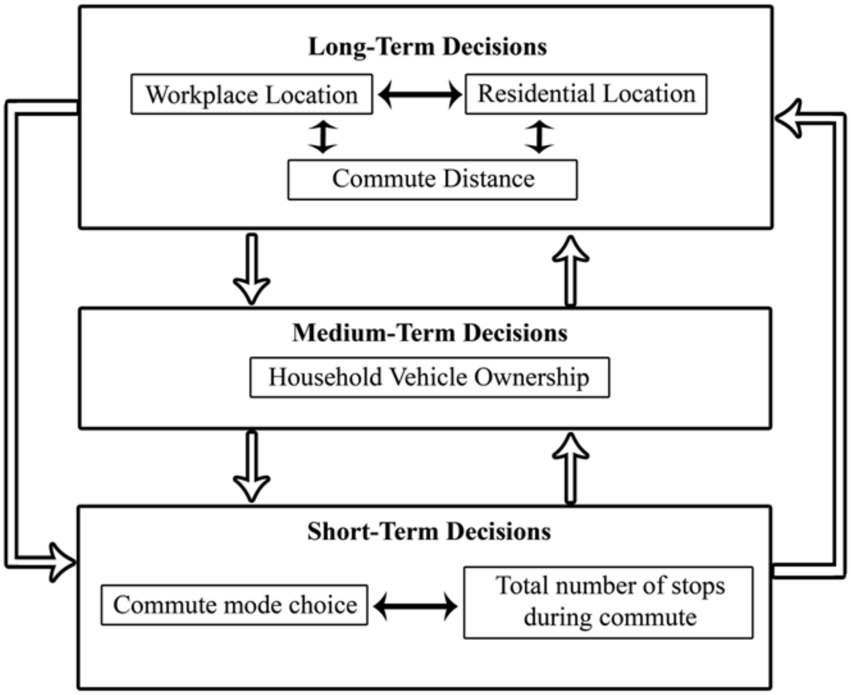

Figure 1a Possible interdependencies in the choice continuum.

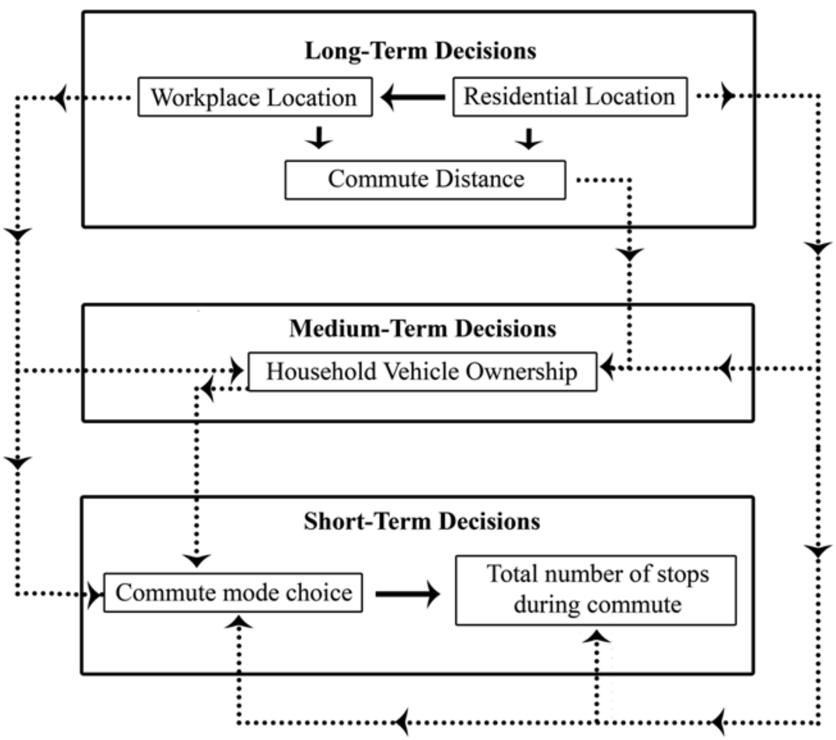

Figure 1b Relationships among endogenous variables in final model specification.

FIGURE 1 Interdependencies in the choice continuum.

TABLE 1 Integrated Model Estimation Results – Long Term Choices

Variable Description	Coef	t-stat	Coef	t-stat	Coef	t-stat
Residential Loc. Utility (Base Alt.: 0-499 housing units per square mile)	**500-1,999**		**2000-3,999**		**≥4,000**	
Constant	0.2413	2.10	0.2071	1.75	-0.0090	-0.05
Socio-economic Attributes						
Presence of children aged 6 to 10 years (Yes=1,No=0)			0.2427	2.14		
Presence of children aged 11-15 years (Yes=1,No=0)			0.2427	2.14	-0.4706	-3.96
Highest education attainment in household: College degree					0.4448	2.83
Highest education attainment in household: Post-doctoral degree					0.4807	3.05
Number of full time workers					0.1875	2.01
Number of self employed individuals			-0.1061	-1.77	-0.1061	-1.77
Number of workers with option to work from home	0.1889	2.70	0.1889	2.70	0.1889	2.70
Household income: Less than $20K (Yes=1 or No=0)			0.5132	2.71	0.5884	2.80
Housing tenure: Own house(Yes=1, No=0)	-0.1704	-1.41	-0.2501	-2.11	-0.8529	-7.12
Immigration status: Combination household			0.2153	2.67	0.2276	2.38
Immigration status: Immigrant household			0.1829	1.57	0.2378	1.87
Work Location Utility (Base Alt.: 0-499 housing units per square mile)	**500-1,999**		**2000-3,999**		**>=4,000**	
Constant	-0.2174	-2.76	-0.7019	-4.33	-0.8493	-5.20
Socio-economic Attributes						
Gender (Male = 1, Female = 0)			-0.1199	-2.00	-0.1199	-2.00
Education attainment of the worker: College degree					0.1503	1.64
Education attainment of the worker: Post doctoral degree					0.1052	1.14
Full-time employment indicator (Yes = 1, No = 0)			-0.1270	-1.51	-0.1270	-1.51
Self employed (Yes=1, No=0)			0.5575	5.08	0.3336	2.54
Immigration status (Yes=1, No=0)			-0.2598	-2.70	-0.1614	-1.61
Immigration status: Number of years since entered the US	0.0047	1.51	0.0047	1.51	0.0047	1.51
Race of household respondent: African American					0.2687	1.35
Race of household respondent: Asian	-0.2033	-2.22	-0.2033	-2.22	-0.2033	-2.22
Residential Location						
500-1,999 housing units per square mile	0.2850	2.81	0.3915	3.50	0.3749	2.94
2,000-3,999 housing units per square mile	0.3451	3.32	0.6285	5.61	0.5331	4.11
≥4,000 housing units per square mile	0.3218	2.49	0.4793	3.27	1.1748	8.03
Natural Logarithm of Commute Distance (in miles)						
Constant	1.6760	13.28				
Socio-economic Attributes						
Gender (Male = 1, Female = 0)	0.2950	5.19				
Full-time employment indicator (Yes = 1, No = 0)	0.3970	5.69				
Self-employed (Yes=1,No=0)	-0.3960	-4.64				
Flexible work schedule (Yes=1, No=0)	0.1400	2.32				
Immigration status (Yes=1,No=0)	0.2490	3.13				
Race of household respondent: African American	0.4060	1.83				
Race of household respondent: Asian	-0.0860	-0.96				
Presence of children 0-15 years (Yes=1, No-0)	-0.0960	-1.52				
Household income: Less than $20K (Yes=1 or No=0)	-0.4590	-2.73				
Household income: $20K-$45K (Yes=1 or No=0)	-0.4690	-4.27				
Household income: $45K-$60K (Yes=1 or No=0)	-0.1830	-2.05				
Household income: $60K-$75K (Yes=1 or No=0)	-0.1730	-1.74				
Housing tenure: Own house(Yes=1, No=0)	0.2930	3.72				
Residential Location						
500-1,999 housing units per square mile	-0.1250	-1.64				
2,000-3,999 housing units per square mile	-0.2710	-3.30				
≥4,000 housing units per square mile	-0.5520	-5.55				
Work Location						
500-1,999 housing units per square mile	-0.1030	-1.40				
2,000-3,999 housing units per square mile	-0.0980	-1.33				
≥4,000 housing units per square mile	-0.0980	-1.33				

TABLE 2 Integrated Model Estimation Results – Medium Term Choice (Vehicle Ownership Propensity)

Variable Description	Coef	t-stat
Thresholds		
Threshold 1 (1-2 vehicles)	-0.5866	-2.86
Threshold 2(2-3 vehicles)	0.9779	6.01
Threshold 3 (3-4 vehicles)	2.8345	17.35
Threshold 4 (4 or more vehicles)	3.8146	22.40
Socio-economic Attributes		
Number of adults in household	0.8614	20.76
Presence of children aged 11-15 years (Yes=1,No=0)	0.1481	1.71
Presence of senior adults aged 65 or over (Yes=1, No=0)	-0.2211	-2.27
Presence of an individual with prolonged medical condition (Yes=1, No=0)	-0.2293	-1.52
Highest education attainment in household: College degree	-0.2338	-2.87
Highest education attainment in household: Post-doctoral degree	-0.2997	-3.70
Number of full time workers	0.1524	2.74
Number of self employed individuals	0.1850	2.99
Number of individuals with more than one job	0.1322	1.46
Household income: Less than $20K (Yes=1 or No=0)	-0.7407	-5.13
Household income: $20K-$45K (Yes=1 or No=0)	-0.5459	-4.34
Household income: $45K-$60K (Yes=1 or No=0)	-0.3617	-3.28
Housing tenure: Own house(Yes=1, No=0)	0.7057	8.08
Residential Location		
500-1,999 housing units per square mile	-0.1078	-1.20
2,000-3,999 housing units per square mile	-0.1275	-1.39
≥4,000 housing units per square mile	-0.6695	-6.10
Work Location		
≥4,000 housing units per square mile	-0.2824	-3.16
Natural logarithm of commute distance (in miles)	0.0799	2.68

TABLE 3 Integrated Model Estimation Results – Short Term Choices

Variable Description	Coef	t-stat	Coef	t-stat	Coef	t-stat
Commute Mode Utility (Base Alternative: Drive Alone)	**Shared Ride**		**Transit**		**Walk/Bike**	
Constant	-0.6794	-3.56	-2.8180	-11.51	-2.0918	-3.85
Socio-economic Attributes						
Age (in years)	-0.0143	-3.88			-0.0110	-1.52
Gender (Male = 1, Female = 0)	-0.0878	-1.05	0.1527	1.23	0.8706	4.08
Education attainment of the worker: Less than High school	0.3578	1.35	0.3578	1.35	0.3578	1.35
Self employed (Yes=1, No=0)	-0.1851	-1.27	-0.7354	-2.59	-0.9086	-2.33
Option to work from home (Yes=1, No=0)			0.2340	1.82	0.4062	1.52
Flexible work schedule (Yes=1, No=0)	0.2906	3.37	0.3087	2.43	0.6569	2.91
Immigration status (Yes=1, No=0)	0.2612	1.74	0.3838	2.18		
Immigration status: Number of years since entered the US	-0.0053	-1.05	-0.0053	-1.05	-0.0053	-1.05
Race of household respondent: African American	-0.5177	-1.80	0.3201	1.25		
Residential Location						
500-1,999 housing units per square mile					0.8108	2.25
2,000-3,999 housing units per square mile			0.2191	1.57	0.8325	2.23
≥4,000 housing units per square mile	0.2193	1.93	0.9416	5.44	0.9486	2.43
Work Location						
500-1,999 housing units per square mile			0.2670	1.73		
2,000-3,999 housing units per square mile			0.6852	2.70	-0.6038	-2.49
≥4,000 housing units per square mile			0.6852	2.70	-0.6038	-2.49
Natural logarithm of Commute distance (in miles)			0.1555	2.17	-0.8513	-5.47
Vehicle Ownership						
Four or more vehicles	-0.1759	-3.89	-0.1759	-3.89	-0.1759	-3.89

Number of Commute Stops propensity		
Thresholds		
Threshold 1 (1-2 stops)	0.2830	1.87
Threshold 2(2-3 stops)	0.7738	5.04
Threshold 3 (3-4 stops)	1.3366	8.57
Threshold 4 (4 -5 stops)	1.7713	11.04
Threshold 4 (5 or more vehicles)	2.2254	13.60
Socio-economic Attributes		
Education attainment of the worker: College degree	0.1763	1.86
Education attainment of the worker: Post-doctoral degree	0.1654	1.57
Has more than one job (Yes=1,No=0)	0.3465	3.45
Flexible work schedule (Yes=1, No=0)	0.3431	5.08
Immigration status (Yes=1,No=0)	-0.1882	-2.23
Race of household respondent: Caucasian	0.0946	1.18
Presence of children 0-10 years (Yes=1, No-0)	0.1841	2.06
Number of adults in household	-0.1894	-4.48
Number of full time workers	0.1578	3.08
Number of self employed individuals	0.2466	3.89
Household income: Less than $20K (Yes=1 or No=0)	-0.2675	-1.46
Household income: $20K-$45K (Yes=1 or No=0)	-0.2976	-2.17
Residential Location		
≥4,000 housing units per square mile	0.1276	1.54
Commute Mode		
Shared ride	0.6481	7.58
Walk or bike	-0.5388	-2.48

TABLE 4 Covariance Matrix for the Integrated Model System

		Res	Res	Res	Work	Work	Work	Mode	Mode	Mode	# Veh	# Stops	Ln Comm Dist
		(500-1,999)	(2,000-3,999)	(≥4,000)	(500-1,999)	(2,000-3,999)	(≥4,000)	SR	TR	WB			
Res	**(500-1,999)**	1.0											
Res	**(2,000-3,999)**	0.5	1.0										
Res	**(≥4,000)**	0.5	0.5	1.0									
Work	**(500-1,999)**	0.0	0.0	0.0	1.0								
Work	**(2,000-3,999)**	0.0	0.0	0.0	0.5	1.0							
Work	**(≥4,000)**	0.0	0.0	0.0	0.5	0.5	1.0						
Mode	**SR**	0.0	0.0	0.0	0.0	0.0	**0.0555 (1.03)**	1.0					
Mode	**TR**	0.0	0.0	0.0	0.0	0.0	**0.2883 (2.12)**	0.5	1.0				
Mode	**WB**	0.0	0.0	0.0	**-0.1507 (-1.1)**	0.0	0.0	0.5	**0.8009 (1.98)**[*]	1.0			
# Veh		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**-0.3317 (-2.09)**	1.0		
# Stops		**-0.0826 (-1.35)**	**0.0973 (1.72)**	0.0	0.0	0.0	**0.1004 (1.03)**	0.0	0.0	0.0	0.0	1.0	
Ln Comm Dist		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.9804 (25.43)**

[*] T-statistic computed against 0.5 corresponding to the value in independent MNP model.