UC Berkeley
UC Berkeley Previously Published Works

Title
Residential mobility and location choice: a nested logit model with sampling of alternatives

Permalink
https://escholarship.org/uc/item/0z29s2mt

Journal
Transportation: Planning - Policy - Research - Practice, 37(4)

ISSN
1572-9435

Authors
Lee, Brian H.
Waddell, Paul

Publication Date
2010-07-01

DOI
10.1007/s11116-010-9270-4

Peer reviewed

Residential mobility and location choice: a nested logit model with sampling of alternatives

Brian H. Y. Lee · Paul Waddell

Abstract Residential mobility and relocation choice are essential parts of integrated transportation and land use models. These decision processes have been examined and modeled individually to a great extent but there remain gaps in the literature on the underlying behaviors that connect them. Households may partly base their decision to move from or stay at a current location on the price and quality of the available alternatives. Conversely, households that are on the market for a new location may evaluate housing choices relative to their previous residence. How and the degree to which these decisions relate to each other are, however, not completely understood. This research is intended to contribute to the body of empirical evidence that will help answer these questions. It is hypothesized that residential mobility and location choice are related household decisions that can be modeled together using a two-tier hierarchical structure. This paper presents a novel nested logit (NL) model with sampling of alternatives and a proposed procedure for sampling bias correction. The model was estimated using full-information maximum likelihood estimation methods. The results confirm the applicability of this NL model and support similar findings from other empirical studies in the residential mobility and location choice literature.

Keywords Residential mobility · Location choice · Nested logit · Sampling of alternatives

B. H. Y. Lee (✉)
School of Engineering and Transportation Research Center, University of Vermont,
210 Colchester Avenue, Farrell Hall 114, Burlington, VT 05405-0303, USA
e-mail: bhylee@uvm.edu

P. Waddell
Department of City and Regional Planning, University of California,
Berkeley, 228 Wurster Hall #1850, Berkeley, CA 94720-1850, USA
e-mail: waddell@berkeley.edu

Introduction

Transportation planning has traditionally focused on travel demand forecast models that assume land use as exogenous inputs and ignore the feedback effects between changes in the transportation system and urban development (Waddell et al. 2007). Research advancements in this arena have been shifting away from these narrow views toward more holistic approaches (Miller et al. 1998). The developments of large-scale integrated transportation and land use microsimulation systems—such as ILUMASS (Strauch et al. 2005; Wagner and Wegener 2007), ILUTE (Miller and Salvini 2001; Salvini and Miller 2005), and UrbanSim (Waddell 2002; Waddell et al. 2003)—have generated increasing attention on the research of a wide assortment of travel-related econometric models. Two essential household behaviors that have been separately examined and modeled to a great extent in recent years are residential mobility and location choice and they are the subjects of this empirical study.

Residential mobility and location choice are critical components in systems of integrated transportation and land use models. They are significant long-term household decisions that intimately connect with the daily activity and travel aspects of individual lives (Clark and Dieleman 1996; Dieleman 2001). Trade-offs between housing qualities and rents, activity opportunities, and travel costs have long been recognized as fundamental considerations in both the decision to move (Brown and Moore 1970; Rossi 1955) and the selection of a residence (Alonso 1964; Muth 1969).

In the transportation-land use modeling context, there are a few recent contributions on residential mobility (e.g., Eluru et al. 2009; Habib and Miller 2008) and a considerable list of current research on residential location (e.g., Chen et al. 2008; Guevara and Ben-Akiva 2006; Guo and Bhat 2007; Lee et al. 2010; Waddell forthcoming). Some studies have linked short- and medium-term travel-related behaviors with residential mobility (e.g., Clark and Withers 1999; Clark et al. 2003; Prillwitz et al. 2006; Prillwitz et al. 2007) while numerous others made similar connections with residential location (e.g., Ben-Akiva and Bowman 1998; Eliasson and Mattsson 2000; Krizek 2003; Krizek 2006; Ng 2008; Pinjari et al. 2008a; Pinjari et al. 2008b; Sermons and Koppelman 2001; Waddell et al. 2008). Few studies have, however, explicitly modeled these closely related decisions together; one example used a stated preference approach (Kim et al. 2005). None, to the best of the authors' knowledge, has modeled residential mobility and location choice jointly in a hierarchical structure using revealed preference data, even though these two parts of the housing decision process are most likely interdependent (Wong 2002). Households may partly base their decision to move from or stay at a current location on the price and quality of the available alternatives (Clark and Onaka 1983). Conversely, households that are on the market for a new home may use their most recent residence as a point of reference to help evaluate the location choices (Lee et al. 2010; Zondag and Pieters 2005). How and the degree to which these decisions relate to each other are not completely understood. Further, there remain opportunities for improvements in the modeling of this process. The state of the art integrated microsimulation systems mentioned above all currently rely on separately and sequentially executed sub-modules for simulating residential mobility and location choice; typically, the first sub-module uses estimates of mobility rates to identify proportions of households to move and the latter sub-module distributes them over space with models of location choice.

An important computational barrier that has limited the development of hierarchical residential mobility and location models is the large number of potential alternatives commonly assumed to be considered in the location decision. Depending on the granularity

of the location choices, the universal set of alternatives for a household looking to relocate can range from hundreds [e.g., neighborhoods or traffic analysis zones (Guo and Bhat 2007)] to hundreds of thousands [e.g., parcels or buildings (Lee et al. 2010)]. Such large numbers are computationally infeasible and behaviorally unrealistic. One practical solution to this intractability problem is to only use a sample of these alternatives for model estimation.

In a seminal work on residential location modeling, McFadden (1978) demonstrated that the consistent estimation of choice models from a subset of alternatives is possible in the multinomial logit (MNL) form, where the property of independence from irrelevant alternatives (IIA) is assumed [see Ben-Akiva and Lerman (1985, pp. 261–269) for a textbook treatment of this topic]. In more advanced, non-MNL model forms suitable for the analysis of multi-dimensional problems, the IIA restriction is relaxed to allow for correlations among alternatives. Consistent estimators of non-MNL models based on a sample of alternatives have not yet been proven in the literature (Nerella and Bhat 2004). The latest research advances on this subject, however, are promising and suggest that a mathematical proof showing the consistent estimation of non-MNL choice models from a subset of alternatives may be possible (Bierlaire et al. 2008; Frejinger et al. 2009; Koppelman and Garrow 2005).

This paper presents a two-tier nested logit (NL) model of residential mobility and location choice using household observations and building-level residences from the central Puget Sound region. It uses random sampling of alternatives and introduces a correction procedure to ensure that the NL maximum likelihood estimator provides consistent estimates of the parameters in the presence of random sampling, something that has not been done in prior research, to the authors' knowledge. This work intends to contribute to the body of empirical evidence on residential mobility and location behaviors and improve upon the understanding of the connection between them. It also offers a means for integrated transportation and land use microsimulation systems to jointly consider these parts in the housing process.

Following this introduction, sections detailing a review of the relevant literature on residential mobility and location choice are presented. They are then followed by a qualitative explanation of the NL formulation and the sampling correction procedure; a presentation of the data, model specification, and estimation results; and a discussion of the conclusions, limitations, and directions ahead.

Review of literature

The literature on residential mobility and location choice is extensive and spans across multiple related disciplines, with different expertise matched with different emphasis. Although these decisions are closely related in the household housing process, there is an expansive range of factors that contribute to each choice. Due to the overwhelming complexity of these behaviors and the assortment of circumstances, most empirical research are concerned with aspects of one decision or the other rather than with both components as a whole. In light of this division, these subjects are reviewed in the following separate, but overlapping, subsections.

Residential mobility

Researchers have explored many different facets of residential mobility. For comprehensive reviews of this subject, readers are encouraged to survey compilation works by Clark

(1982), Clark and Dieleman (1996), and Dieleman (2001). Two of the most often cited studies in this literature are by Rossi (1955) and Brown and Moore (1970). The Rossi work is significant because it changed the focal point of residential mobility research from the study of aggregate spatial patterns to one of behaviors at the individual household level. This paradigm shift has directed most of the research performed since then and spawned a wealth of papers on the reasons for adjustments in household housing consumptions. One such work is by Brown and Moore (1970), who decomposed residential mobility into two stages. The first concerns changes in household circumstances, both internal and external, that cause dissatisfaction with the existing housing choice. The second stage involves a decision to move or stay based on a survey of available alternatives in the housing stock. This conceptualization of the interdependence between mobility and location choice represents the process that the hierarchical model structure proposed in the current empirical study is attempting to capture.

In the decades since these important works, a multitude of studies on residential mobility were conducted in many different geographic, sociologic, and economic contexts (Clark and Dieleman 1996; Li and Wu 2004). Despite differences in the environments, three intersecting findings transcend across geopolitical and social boundaries (Dieleman 2001). The first is the close connection between mobility rates and the life cycle of an individual; young adults in their twenties and thirties compose the most mobile segment of the population (Abraham and Hunt 1997; Clark and Huang 2003). The second concerns the relationship between mobility and both housing size and tenure; the rates are lower for households in larger homes and for those who own (Clark and Dieleman 1996; Clark and Onaka 1983). The third emphasizes the correlations between significant life course events and household moves; the acts of household formulation and dissolution and changes in education and work opportunities very often coincide with changes in residence (Clark and Huang 2003; Li and Wu 2004; Prillwitz et al. 2007).

Most recently, two contributions in the integrated transportation and land use modeling context are of note. Habib and Miller (2008) introduced a conceptual process for residential mobility and spatial search that is to be implemented in the ILUTE microsimulation system (Miller and Salvini 2001; Salvini and Miller 2005). In this paper, they focused on the mobility component and examined different discrete choice methods and hazard-based duration models. Their findings suggest that the use of different model formulations could reveal different reasons for household moves. Similarly, Eluru et al. (2009) analyzed residential mobility behavior for motivations behind changes in housing consumption. In addition, they endogenized the length of time spent at a residence by jointly modeling reasons to move with the duration of stay. Their results support the existence of unobserved factors affecting both dimensions. These two works represent advances in disaggregate, agent-based modeling of residential mobility and contribute to the empirical understanding of this behavior. Nevertheless, they lack explicit feedback connections between mobility and location choice.

Residential location choice

The analysis of residential location choice at the household level was largely enabled by the development of discrete choice modeling methods. The early applications by Lerman (1976) and McFadden (1978) on this subject paved the way for a generation of research on identifying different contributing factors and making connections with travel-related behaviors. Much of this work is centered on the utility maximization concept where housing choice is represented as a bundle of other associated choices and households must

make trade-offs between various housing qualities, neighborhood attributes, and accessibilities to best meet their needs.

Some studies concentrated on linking residential location with different modes and other travel choice behaviors (Eliasson and Mattsson 2000; Krizek 2006; Pinjari et al. 2008a, b). Others have examined commuting factors and the relations between the locations of residence and workplace (Clark and Withers 1999; Waddell et al. 2008). Abraham and Hunt (1997), for one, have attempted to do both. The role of accessibility, to both work and non-work activities, in residential location choice has received a large share of the attention (Ben-Akiva and Bowman 1998; Chen et al. 2008; Lee et al. 2010). Various modeling issues such as self-selection (Krizek 2003; Pinjari et al. 2008a) and price endogeneity (Guevara and Ben-Akiva 2006) are also being addressed.

The multinomial logit (MNL) formulation mentioned in the introduction is only one of a family of discrete choice models but it has been, and still is, widely used in practice and research. The popularity of the MNL model in residential location choice applications is due mainly to properties that allow for efficient computations and consistent estimations with a subset of alternatives. Researchers have attempted to navigate around this sampling issue and use more advanced models with different strategies. Habib and Kockelman (2008) estimated a series of NL and joint MNL models with full enumeration of all possible alternatives. Kim et al. (2005) and Pinjari et al. (2008a) developed NL and joint binary-ordered logit models, respectively, by using categories of locations to significantly reduce the dimensionality. Others have simply used non-MNL models with sampling of alternatives and ignored sampling biases (Yagi and Mohammadian 2008; Zhou and Kockelman 2008); this strategy, as demonstrated by Nerella and Bhat (2004) in their explorations of this subject, is highly discouraged.

The sampling issue is more pressing now than ever as research in integrated transportation and land use model systems require the use of disaggregate data and representations for increasing behavioral realism. Up until recent years, residential location has been commonly operationalized as aggregate units such as neighborhoods or traffic analysis zones (Guo and Bhat 2007). Although full enumerations of hundreds of these alternatives are undesirable, it remains, in some cases, computationally feasible. With the most recent location representations all the way down to the parcel or building levels (Lee et al. 2010; Waddell forthcoming), however, the sampling of alternatives for model estimation has become essential for practical use on current computers. It is this timely need that has helped motivate the direction for this research paper.

Nested logit model formulation with sampling of alternatives

In this study, residential mobility and location choice are assumed to be two related stages of the household housing process that can be modeled jointly using a two-tier nested logit (NL) structure and full-information maximum likelihood estimation. Without any assumptions on the temporal sequence of the decisions, the nesting structure is assumed to include a binomial mobility choice of stay or move at the top level and a multinomial location choice at the bottom level. The stay nest only includes one location alternative (i.e., staying in the current location), whereas the move nest includes multiple choices for move locations. Figure 1 shows this residential mobility and location choice NL model.

The mathematical formulation of this model follows the utility maximizing NL model developed by McFadden (1978); for a detailed examination on this formulation, refer to Koppelman and Wen (1998). Since the standard NL model does not ensure consistent estimation of the model parameters with a subset of alternatives, due to its divergence from
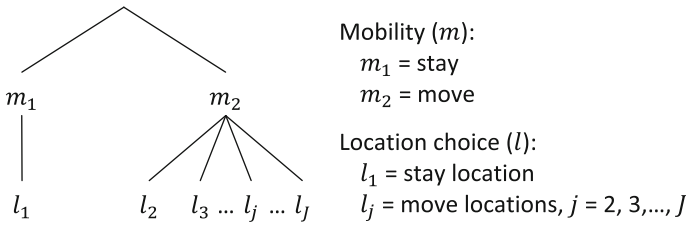
Fig. 1 Two-tier nested structure of residential mobility and location choice

the IIA assumption, a sampling correction procedure was added to the estimation procedure to address this problem. The probability of a household choosing location l is defined as

$$P(l) = P(l|m) \cdot P(m), \tag{1}$$

where $P(l|m)$ is the conditional probability of l being chosen given nest m and $P(m)$ is the marginal probability of choosing nest m.

The bottom level conditional choice probability is equivalent to the standard multinomial logit (MNL) equation and has the form

$$P(l|m) = \frac{e^{V_l \mu_l}}{\sum_{l' \in L_m} e^{V_{l'} \mu_l}}, \tag{2}$$

where V_l represents the observable components of the utility function for each elemental alternative l and μ_l is the associated scale parameter. When the simple random sampling strategy is used to draw a subset of alternatives, no correction is needed here as it has been shown that the standard MNL form produces consistent estimates of the model parameters due to the IIA property (McFadden 1978). For other sampling strategies, such as independent importance sampling or stratified importance sampling, different additional adjustment terms as described in Ben-Akiva and Lerman (1985, pp. 261–269) are needed to correct for sampling bias.

The marginal choice probability of choosing nest m is

$$P(m) = \frac{e^{V'_m \mu_m}}{\sum_{m' \in M} e^{V'_{m'} \mu_m}}, \tag{3}$$

where V'_m is the logsum (or inclusive value) associated with nest m and μ_m is the top level scale parameter. The logsum represents the expected value of the maximum of the random utilities of all alternatives in nest m. In a standard NL formulation with full enumeration of all available elemental alternatives, the logsum equals the log of the denominator of the conditional probability multiplied by $1/\mu_l$, or

$$V'_m = \left(1/\mu_l\right) \ln \left(\sum_{l' \in L_m} e^{V_{l'} \mu_l} \right) \tag{4}$$

In this application, as is common in location choice models, the universal choice set is massive in size, and it is not computationally feasible to estimate the bottom level choice model using all available alternatives. If this was a MNL model of location choice only, then random sampling of alternatives provides a consistent estimate of the parameters of the choice model. The proposed model structure, however, is of the NL form with both residential mobility and location choice, and random sampling of alternatives is not

possible without introducing sampling bias. In spite of this, prior research on residential location has employed NL with random sampling of alternatives, generally without any mention of the bias introduced by using sampling in a model structure that does not rely on the IIA property. This paper introduces a correction for this sampling bias by implementing a maximum likelihood estimator for the NL model that adjusts the logsum to account for the presence of sampling of alternatives at the bottom level of the nesting structure. By the Slutsky Theorem (Ben-Akiva and Lerman 1985, p. 19), which states that a continuous function of consistent estimates is also consistent, an *expanded* logsum may be used as a consistent estimate of the logsum derived from the full set of alternatives. The idea is that the sampling of alternatives in the logit produces consistent estimates of the parameters, and then by Slutsky one can expand the logsum to the full choice set to get a consistent estimate of the logsum. This is intuitive, since it simply means that it is necessary to account for the sample proportion within each nest in order to achieve a consistent estimate of the logsum for the nest. Failing to do so would provide a biased measure of the logsum. For simple random sampling, the expanded logsum is

$$V'_m = \left(1/_{\mu_l} \right) \ln \left\{ \sum_{l' \in L_m} \left[\left(1/_{R'_l} \right) e^{V'_l \mu_l} \right] \right\}, \tag{5}$$

where R'_l is the sampling rate, $0 < R'_l < 1$, which only applies to the sampled non-chosen alternatives. The observed chosen alternative, by design, is not sampled and always included in the choice set and, thus, has an expansion factor of one. This scaling adjustment approximates the true logsum because the alternatives are sampled randomly from the full choice set and these random samples are varied across households. If other sampling strategies are used, then there may be different sampling rates for different sample groups and that should be reflected within the summation component of the expanded logsum calculations.

Note that of the two scale parameters, μ_l and μ_m, only one of them or the ratio between the two can be identified in the model estimation. It is necessary to normalize one of the parameters to equal one and estimate the other. Specification with upper normalization (i.e., μ_m set to one) is equivalent to one with lower normalization (i.e., μ_l set to one) but the former approach has the simplicity of having only one parameter to consider whereas the latter approach raises the problem of which lower level nest to use when there are multiple nests with more than one alternative (Carrasco and Ortuzar 2002). This issue, however, is not relevant here as there is only one nest with multiple alternatives, and a lower normalization is adopted following the approach described in Koppelman and Wen (1998). In this case, μ_m must be between 0 and 1 as a condition of consistency. As the coefficient of the logsum, the parameter μ_m can be interpreted as an indicator of the hierarchical nature of the nesting structure. If the estimate of this parameter approaches 0, then the decision process is considered to be strictly hierarchical. Whereas if μ_m equals one, then the two choices are considered independent and the NL model reduces to a single-stage MNL model. In the case where there is only a single alternative in a nest (e.g., nest m_1, the stay nest with only l_1, the stay location), that nest is considered degenerate and μ_m equals one. If there is only a single alternative in every nest, then the NL model also collapses to the MNL model.

Empirical application in the Central Puget Sound region

The NL model of residential mobility and location choice described above was applied in the Seattle, Washington metropolitan area using a 2006 household activity survey. This

application modeled intra-regional household movements and did not concern immigration from or emigration to locations outside of the study area. The model was estimated simultaneously using a new full-information maximum likelihood NL estimation package in the UrbanSim microsimulation system software (Waddell 2002; Waddell et al. 2003). The following subsections describe the data, explain the model specifications, and present the estimation results.

Data

The Puget Sound Regional Council (PSRC) 2006 Household Activity Survey (Cambridge Systematics Inc. 2007) was the primary source of data for modeling residential mobility and location choice. This was complemented by 2005 building and parcel assessments from county tax assessors; 2000 business establishment data from the state unemployment insurance database; 2000 census data; 2006 travel model results from the PSRC; and other regional geo-spatial data representing environmental, political, and planning features.

Executed as a 2-day activity and travel diary, the PSRC 2006 survey obtained valid data from 4,739 households, representing 10,516 persons. The data includes x–y coordinate residential location information for all household residences at the time of the survey, which enabled assignment to individual buildings for modeling. The survey also contains the length of time each household has resided at that current home and, thus, it is possible to identify households who recently changed residences. Of the full set of surveyed households, a subset of 1,677 moved in the 5-year period prior to 2006 and the data has the x–y coordinate location of each of these households' previous home. This group of households who moved is considered to be the "Movers", while the rest of the surveyed households are the "Stayers".

The observed chosen alternative for both the Movers and the Stayers is the household residential location in 2006. In reference to the two-tier nested structure of the residential mobility and location choice model proposed in Fig. 1, the chosen alternative for the Stayers is the location in the Stay nest (i.e., l_1 in nest m_1) and 29 other buildings were randomly sampled from the universal choice set to populated the move alternatives in the Move nest. For the Movers, there are two location data points: one is the 2006 location and the other is the previous residence. In this case, the 2006 location is considered l_2 in the Move nest and the previous residence is l_1 in the Stay nest. The remaining 28 alternatives (i.e., $l_3, l_4,\ldots, l_{30}$) are randomly sampled.

In addition to the location information, there are other household and individual socio-demographical characteristics and activity and travel records that are relevant to the residential mobility and location choice model. This includes household income and size; the age and work/school status of each member; the nature of the residential tenure (i.e., rent or own); commuting times; and travel to non-work activities.

Model specifications

Equations 1 through 3 and 5 from above were used for this Central Puget Sound application of the NL residential mobility and location choice model. The unit of analysis for the choice set is the individual residential building and the modeled choice behavior is that of the household. The attractiveness of location l for a household n is expressed by the utility function,

$$V_l = \alpha X_l + \beta(Y_n - P_l) + \gamma H_n X_l, \tag{6}$$

where α, β, and γ are parameters to be estimated; X_l is an array of characteristics describing alternative l; Y_n is the household annual income; P_l is the housing price in annualized rents; and H_n is an array of attributes for household n. This utility expression highlights the underlying logic that when making a residential location decision, households take account of their own characteristics (e.g., household income and size, lifecycle stage) and those of the available housing options (e.g., price, accessibility). The interaction terms $(Y_n - P_l$ and $H_nX_l)$ allow the household attributes (Y_n and H_n) to enter the choice model with the housing characteristics (P_l and X_l, respectively); the household attributes cannot be specified by themselves in the utility function because they do not vary across the alternatives and there would be no way to estimate coefficients for such variables. The income and price interaction $(Y_n - P_l)$ represents household disposable income and is specified as such to allow housing price to enter into the model.

Based on a priori knowledge from the literature and previous work on a residential location choice model using the same data for this region (Lee et al. 2010), a wide assortment of explanatory variables were examined for their theoretical and statistical significance. The ones related to residential mobility include variables that were found to be some of the most important in the mobility literature (Dieleman 2001), such as individual or household life stage (i.e., age), household size and composition, and residential tenure (i.e., rent or own). The variables relating to residential location choice that were explored include housing qualities (e.g., dwelling size per unit; building type such as single family residential or multi-family condominiums), neighborhood attributes (e.g., population density, proportions of household with children or young adults), and accessibility measures (e.g., general accessibility to regional work opportunities or local shopping activities, individual-level work travel time).

Model estimation and results

Full-information maximum likelihood techniques were used to estimate the parameters of the two-tier NL model of residential mobility and location choice. The log-likelihood function, with a lower normalization and an expanded logsum, is

$$\mathcal{L} = \sum_{n}\left\{\sum_{m\in M} y_{mn}\left[V'_m\mu_m - \ln\left(\sum_{m'\in M} e^{V'_{m'}\mu_m}\right)\right] + \sum_{l\in L_m} y_{lmn}\left[V_l - \ln\left(\sum_{l'\in L_m} e^{V'_l}\right)\right]\right\} \quad (7)$$

where y_{mn} is 1 if household n chooses nest m and 0 otherwise; y_{lmn} is 1 if household n chooses alternative l in nest m and 0 otherwise;

$$V'_m = \ln\left\{\sum_{l'\in L_m}\left[(1/R'_l)e^{V'_l}\right]\right\}; \quad (8)$$

and all else as defined previously.

After extensive experimentations with different specifications, one NL model was chosen based on the theoretical and statistical significance of the estimated parameters. The explanatory variables specified in this model and brief descriptions are summarized in Table 1 and the model estimates and results are shown in Table 2.

This NL model of residential mobility and location choice has a relatively parsimonious specification but it, nevertheless, includes the important exploratory variables that are expected to be an integral part of the household housing decision process. The estimated

Table 1 Residential mobility and location choice model explanatory variables and descriptions

Variables	Descriptions
asc_stay	Alternative specific constant for the stay location (l_1 in nest m_1)
hh_avg_adult_age	Average age of all adults (18 years and older) in the household (HH)
hh_has_kids	Dummy of HH with kids (under 18 years)
hh_has_workers	Dummy of HH with full-time workers
hh_is_home_owner	Dummy of HH who are home owners
inc_less_price	Log of HH annual income less annual imputed rent per unit
kids × SFR	Dummy of HH with kids × dummy of single-family residential building
kids × kids_HH	Dummy of HH with kids × percent HH with kids within 600m
work_accessibility	Individual worker specific, zonal travel logsum for AM HBW drive alone trips to workplace location (maximum value between up to 2 workers)

Table 2 Model estimation results

Variables	Coefficient Estimates	T-values
Stay alternative only		
asc_stay	−4.14	−23.3
hh_avg_adult_age	0.0589	8.29
hh_has_kids	0.0388	3.71
hh_has_workers	0.310	3.93
hh_is_home_owner	1.61	6.23
Move alternatives only		
inc_less_price	0.00467	3.98
kids × SFR	0.00802	2.67
kids × kids_HH	0.00202	2.16
work_accessibility	0.0250	4.02
move nest logsum (μ_m)	0.0158	3.97
Observations (N)	4739	
Log-likelihood ($\mathcal{L}$)	−7058.6	
Null log-likelihood	−47642.2	
LL ratio (ρ)	0.8518	
Adjusted $\rho \cdot (\rho')$	0.8516	

logsum parameter (μ_m) for the Move nest is small at 0.0158, which strongly supports the hierarchical NL model structure. Trials of models with fuller specifications generally resulted in μ_m estimates that were out of theoretical range. This included experiments with specifying alternative specific variables for the Stay alternative that were included in the Move nest. For this reason, only household specific variables were specified for the Stay alternative.

The stay location alternative was considered the reference choice and four household specific variables plus a constant were specified for this location. All of these variables have the expected positive sign and they are all statistically significant. The first variable, hh_avg_adult_age, captures the life cycle of the household based on the average age of the individual adult members. As expected, this estimated variable parameter suggests that older households are more likely to stay than move, which corresponds to the empirical

evidence in the literature that says young adults are generally the most mobile segment of the population (Abraham and Hunt 1997; Clark and Huang 2003). The next two variables, hh_has_kids and hh_has_workers, are both indicators of the household life cycle as well as the composition. The estimated parameters for these variables imply that households with kids and workers are less likely to move than those without. These results echo other studies that found having children and a steady job to be a constraint to the intention to move (Clark and Dieleman 1996; Kim et al. 2003). This is likely due to inertial from a combination of social, professional, and educational ties associated with a current residence. The last household specific variable is hh_is_home_owner and the estimated parameter agrees with overwhelming empirical evidence from other studies, which points to home owners as being less mobile than renters (Dieleman 2001).

The next four variables are all interactions between household characteristics and location attributes and they were specified for the move alternatives, using the stay location as a reference. The first of these variables is inc_less_price, the household income and housing price interaction. Income is a powerful indicator of a household's preference for and expected ability to afford different types of housing. This variable, as described above in Eq. 6 as $Y_n - P_l$, compares the household income to an annual imputed rent and describes the amount of resources remaining for other consumption needs after paying for housing. The price variable is theoretically vital in the decision making process but price coefficients in location choice models have been reported as statistically insignificant, small, or even positive in several empirical studies (Guevara and Ben-Akiva 2006). In this and a previous analysis (Lee et al. 2010), the authors found this specification of price with income to consistently provide statistically significant and theoretically sound coefficients. The positive sign supports the logic that the more disposable income is associated with a housing choice, the more attractive it is. The next two variables, kids × SFR and kids × kids_HH, express the desires of households with children for single-family residential buildings and neighborhoods with other children, respectively. These variables reflect the different needs and preferences between households with kids and those without and relate to life course studies that correlate significant life events such as having children with household changes in residence (Clark and Huang 2003; Li 2004; Prillwitz et al. 2007). The last variable is a measure of accessibility that captures the ease of access between each residential location choice and specific work locations, for each worker with a fixed non-home workplace. Since this measure is at the individual-level while the residential mobility and location choice model is at the household-level, it was aggregated by choosing the maximum value among the workers in multi-worker households. The positive sign of the estimated parameter suggests that in general, good access to work is a positive contributing factor both to the residential mobility and location choice decisions.

Conclusions

Residential mobility and location choice are interrelated decisions that make up the household housing process. In this paper, they were modeled jointly using a two-tier NL structure with simple random sampling of location alternatives and full-information maximum likelihood estimation. A correction procedure to account for sampling bias was proposed and executed. To the best of the authors' knowledge, this is the first empirical model of its kind, with the sampling correction, in the literature.

The model results from this research align with those found in previous empirical studies. Life stage, household composition, and housing tenure were determined to be

important household characteristics that helped to explain the residential mobility and location choice process. In general, older households, ones with children or workers, and those who own their homes are less likely to move and change residences than younger renters or those with no kids or no workers. In addition, the housing price, the building type, the neighborhood composition, and accessibility were also found to be significant factors in the mobility-location choice calculus. Relative to a household's current home, location alternatives for a move are attractive if the cost is lower and the ease of access to work is higher, and, for households with children, if the building is a single-family residential home in a neighborhood with other families with children.

This NL model explicitly accounts for the multi-dimensional nature of the housing process and represents a step forward in the modeling of residential mobility and location choice in the integrated transportation and land use model system context. The interdependence between these decisions is reflected by the location alternatives in the Move nest informing the move-stay choice and the stay location being used as a reference to compare possible move locations. There are, however, some limitations to this model and they correspond to opportunities for future research in this arena. First, the housing tenure is considered as an exogenous input in this model, but tenure choice may be another interrelated part of the housing process that is better determined jointly with residential mobility and location choice. Second, the use of a cross-sectional household survey dataset with a retrospective view at housing allowed for the incorporation of previous residences in the model but it is no substitute for a panel dataset. Life course events and changes in individual and household circumstances were not observed and these factors are known to be important in the mobility literature when considering the household level of satisfaction with an existing housing situation (Clark and Huang 2003; Li 2004). Third, it should be noted that the household survey was conducted during a real estate price bubble and the behaviors observed may not reflect those in the current economic downturn. The consideration of macroeconomic conditions would help any predictive models in forecasting future behaviors. Lastly, the sampling of alternatives with corrections in a NL structure is a useful methodological enhancement, but could be improved with more sophisticated sampling and adjustment procedures, advanced location choice set screening methods, and experimentation with different sample sizes.

References

Abraham, J.E., Hunt, J.D.: Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. Transp. Res. Record **1606**, 17–24 (1997)

Alonso, W.: Location and Land Use: Toward a General Theory of Land Rent. Harvard University Press, Cambridge (1964)

Ben-Akiva, M.E., Bowman, J.L.: Integration of an activity-based model system and a residential location model. Urban Stud. **35**, 1131–1153 (1998)

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

Bierlaire, M., Bolduc, D., McFadden, D.: The estimation of generalized extreme value models from choice-based samples. Transp. Res. B **42**, 381–394 (2008)

Brown, L.A., Moore, E.G.: The intra-urban migration process: a perspective. Geogr. Ann. B **52**, 1–13 (1970)

Cambridge Systematics Inc.: Final Report: PSRC 2006 Household Activity Survey Analysis Report. Washington Department of Transportation, Sammamish (2007)

Carrasco, J.A., Ortuzar, J.D.: Review and assessment of the nested logit model. Transp. Rev. **22**, 197–218 (2002)

Chen, J., Chen, C., Timmermans, H.: Accessibility trade-offs in household residential location decisions. Transp. Res. Record **2077**, 71–79 (2008)

Clark, W.A.V.: Recent research on migration and mobility—A review and interpretation. Prog. Plann. **18**, 5–56 (1982)

Clark, W.A.V., Dieleman, F.M.: Households and Housing: Choice and Outcomes in the Housing Market. The Center for Urban Policy Research, New Brunswick (1996)

Clark, W.A.V., Huang, Y.: The life course and residential mobility in British housing markets. Environ. Plann. A **35**, 323–339 (2003)

Clark, W.A.V., Onaka, J.L.: Life-cycle and housing adjustment as explanations of residential-mobility. Urban Stud. **20**, 47–57 (1983)

Clark, W.A.V., Withers, S.D.: Changing jobs and changing houses: mobility outcomes of employment transitions. J. Regional Sci. **39**, 653–673 (1999)

Clark, W.A.V., Huang, Y., Withers, S.D.: Does commuting distance matter? Commuting tolerance and residential change. Reg. Sci. Urban Econ. **33**, 199–221 (2003)

Dieleman, F.M.: Modelling residential mobility: a review of recent trends in research. J. Hous. Built. Environ. **16**, 249–265 (2001)

Eliasson, J., Mattsson, L.: A model for integrated analysis of household location and travel choices. Transp. Res. A **34**, 375–394 (2000)

Eluru, N., Sener, I.N., Bhat, C.R., Pendyala, R.M., Axhausen, K.W.: Understanding residential mobility: a joint model of the reason for residential relocation and stay duration. In: Transportation Research Board 88th Annual Meeting, Washington, DC (2009)

Frejinger, E., Bierlaire, M., Ben-Akiva, M.: Sampling of alternatives for route choice modeling. Transp. Res. B **43**, 984–994 (2009)

Guevara, C.A., Ben-Akiva, M.E.: Endogeneity in residential location choice models. Transp. Res. Record **1977**, 60–66 (2006)

Guo, J.Y., Bhat, C.R.: Operationalizing the concept of neighborhood: application to residential location choice analysis. J. Transp. Geogr. **15**, 31–45 (2007)

Habib, K.M.N., Kockelman, K.: Modeling choice of residential location and home type: recent movers in Austin Texas. In: 87th Annual Meeting of the Transportation Research Board, Washington, DC (2008)

Habib, M.A., Miller, E.J.: Modeling residential mobility and spatial search behavior—estimation of continuous-time hazard and discrete-time panel logit models for residential mobility. In: Transportation Research Board 87th Annual Meeting, Washington, DC (2008)

Kim, J.H., Pagliara, F., Preston, J.: The intention to move and residential location choice behavior. Urban Stud. **42**, 1621–1636 (2005)

Kim, J., Pagliara, F., Preston, J.: An analysis of residential location choice behaviour in Oxfordshire UK—a combined state preference approach. Int. Rev. Public Admin. **8**, 103–114 (2003)

Koppelman, F.S., Wen, C.: Alternative nested logit models: Structure properties and estimation. Transp. Res. B **32**(5), 289–298 (1998)

Koppelman, F.S., Garrow, L.: Efficiently estimating nested logit models with choice-based samples example applications: example applications. Transp. Res. Record **1921**, 63–69 (2005)

Krizek, K.J.: Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? J. Am. Plann. Assoc. **69**, 265–281 (2003)

Krizek, K.J.: Lifestyles residential location decisions and pedestrian and transit activity. Transp. Res. Record **1981**, 171–178 (2006)

Lee, B.H.Y., Waddell, P.A., Wang, L., Pendyala, R.M.: Operationalizing time-space prism accessibility in a building-level residential choice model: empirical results from the Puget Sound region. Environ. Plann. A **42** (2010)

Lerman, S.R.: Location housing automobile ownership and mode to work: a joint choice model. Trans. Res. Record **610**, 6–11 (1976)

Li, S.: Life course and residential mobility in Beijing China. Environ. Plann. A **36**, 27–43 (2004)

Li, S., Wu, F.: Contextualizing residential mobility and housing choice: evidence from urban China. Environ. Plann. A **36**, 1–6 (2004)

McFadden, D.: Modelling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (eds.) Spatial Interaction Theory and Planning Models, pp. 75–96. North Holland, Amsterdam (1978)

Miller, E.J., Salvini, P.A.: The integrated land use transportation environment (ILUTE) microsimulation modelling system: description and current status. In: Hensher, D.A. (ed.) Travel Behaviour Research: The Leading Edge, pp. 711–724. Pergamon Press, Amsterdam (2001)

Miller, E.J., Kriger, D.S., Hunt, J.D.: Integrated urban models for simulation of transit and land use policies. Transit Cooperative Research Program Transportation Research Board. http://wwwnapedu/catalog/9435html (1998). Accessed 30 July 2009

Muth, R.F.: Cities and Housing: The Spatial Pattern of Urban Residential Land Use. University of Chicago Press, Chicago (1969)

Nerella, S., Bhat, C.: Numerical analysis of effect of sampling of alternatives in discrete choice models. Transp. Res. Record **1894**, 11–19 (2004)

Ng, C.F.: Commuting distances in a household location choice model with amenities. J. Urban Econ. **63**, 116–129 (2008)

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

Pinjari, A.R., Pendyala, R.M., Bhat, C.R., Waddell, P.A.: Modeling the choice continuum: Integrated model of residential location automobile ownership bicycle ownership and commute tour mode choice decisions. In: Transportation Research Board 87th Annual Meeting, Washington, DC (2008b)

Prillwitz, J., Harms, S., Lanzendorf, M.: Impact of life-course events on car ownership. Transp. Res. Record **1985**, 71–77 (2006)

Prillwitz, J., Harms, S., Lanzendorf, M.: Interactions between residential relocations life course events and daily commute distances. Transp. Res. Record **2021**, 64–69 (2007)

Rossi, P.H.: Why Families Move: A Study in the Social Psychology of Urban Residential Mobility. Free Press, Glencoe (1955)

Salvini, P.A., Miller, E.J.: ILUTE: an operational prototype of a comprehensive microsimulation model of urban systems. Netw. Spatial Econ. **5**, 217–234 (2005)

Sermons, M.W., Koppelman, F.S.: Representing the differences between female and male commute behavior in residential location choice models. J. Transp. Geogr. **9**, 101–110 (2001)

Strauch, D., Moeckel, R., Wegener, M., Gräfe, J., Mühlhans, H., Rindsfüser, G.: Linking transport and land use planning: the microscopic dynamic simulation model ILUMASS. In: Atkinson, P.M., Foody, G.M., Darby, S.E., Wu, F., et al. (eds.) Geodynamics, pp. 295–311. CRC Press, Boca Raton (2005)

Waddell, P.: UrbanSim: modeling urban development for land use transportation and environmental planning. J. Am. Plann. Assoc. **68**, 297–314 (2002)

Waddell, P.: Modeling residential location in UrbanSim. In: Preston, J., Pagliara, F., Simmonds, D. (eds.) Modelling Residential Location Choice. Springer, London (forthcoming)

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., Ulfarsson, G.: Microsimulation of urban development and location choices: design and implementation of UrbanSim. Netw. Spatial Econ. **3**, 43–67 (2003)

Waddell, P., Ulfarsson, G.F., Franklin, J.P., Lobb, J.: Incorporating land use in metropolitan transportation planning. Transp. Res. A **41**, 382–410 (2007)

Waddell, P., Bhat, C.R., Eluru, N., Wang, L., Pendyala, R.M.: Modeling interdependence in household residence and workplace choice. Transp. Res. Record **2003**, 84–92 (2008)

Wagner, P., Wegener, M.: Urban land use transport and environment models: experiences with an integrated microscopic approach. disP **170**, 45–56 (2007)

Wong, G.K.M.: A conceptual model of the household's housing decision-making process: the economic perspective. Rev. Urban Reg. Dev. Stud. **14**, 217–234 (2002)

Yagi, S., Mohammadian, A.: Joint models of home-based tour mode and destination choices: applications to a developing country. Transp. Res. Record **2076**, 29–40 (2008)

Zhou, B., Kockelman, K.: Microsimulation of residential land development and household location choices: bidding for land in Austin Texas. Transp. Res. Record **2077**, 106–112 (2008)

Zondag, B., Pieters, M.: Influence of accessibility on residential location choice. Transp. Res. Record **1902**, 63–70 (2005)

Author Biographies

Brian H. Y. Lee is an Assistant Professor in the School of Engineering and the Transportation Research Center at the University of Vermont. His research agenda includes transportation and land use planning, policy evaluation, non-motorized and public transportation, and the built environment and physical activities.

Paul Waddell is a Professor in the Department of City and Regional Planning, University of California, Berkeley. His research focuses on land use, transportation and environmental modeling and planning. He leads the development of the UrbanSim planning support system, and the Open Platform for Urban Simulation (www.urbansim.org).