D-STOP

Technical Report 119

A Comprehensive Dwelling Unit Choice Model Accommodating Psychological Constructs within a Search Strategy for Consideration Set Formation

Chandra Bhat
Center for Transportation Research

December 2015

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
D-STOP/2016/119		

4. Title and Subtitle	5. Report Date
A Comprehensive Dwelling Unit Choice Model Accommodating Psychological Constructs within a Search Strategy for Consideration Set Formation	December 2015
	6. Performing Organization Code
7. Author(s)	**8. Performing Organization Report No.**
Chandra R. Bhat	Report 119
9. Performing Organization Name and Address	**10. Work Unit No. (TRAIS)**
Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, Texas 78701	11. Contract or Grant No. DTRT13-G-UTC58
12. Sponsoring Agency Name and Address	**13. Type of Report and Period Covered**
Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, Texas 78701	14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.

16. Abstract

This study adopts a dwelling unit level of analysis and considers a probabilistic choice set generation approach for residential choice modeling. In doing so, we accommodate the fact that housing choices involve both characteristics of the dwelling unit and its location, while also mimicking the search process that underlies housing decisions. In particular, we model a complete range of dwelling unit choices that include tenure type (rent or own), housing type (single family detached, single family attached, or apartment complex), number of bedrooms, number of bathrooms, number of storeys (one or multiple), square footage of the house, lot size, housing costs, density of residential neighborhood, and commute distance. Bhat's (2014) generalized heterogeneous data model (GHDM) system is used to accommodate the different types of dependent outcomes associated with housing choices, while capturing jointness caused by unobserved factors. The proposed analytic framework is applied to study housing choices using data derived from the 2009 American Housing Survey (AHS), sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. The results confirm the jointness in housing choices, and indicate the superiority of a choice set formation model relative to a model that assumes the availability of all dwelling unit alternatives in the choice set.

17. Key Words	18. Distribution Statement
Latent psychological constructs, MACML estimation approach, mixed dependent variables, structural equations models, integrated land use-transportation modeling, housing choices	No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	60	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. The author would like to acknowledge support from a Humboldt Research Award from the Alexander von Humboldt Foundation, Germany. The author is grateful to Lisa Macias for her help in formatting this document, and to Subodh Dubey for help with coding and running specifications. Three anonymous reviewers provided useful comments on an earlier version of this paper.

.

Table of Contents

List of Illustrations

List of Tables

Chapter 1. Introduction

The home is usually considered as the base location for individuals, the place that most people start their activities from each day and the place that most people come back to at the end of the day. Thus, it has been well established for some time now that the home location can act as a facilitator or a suppressor of out-of-home activity pursuits of individuals living in the home, based on the relative spatial location of the home vis-à-vis activity opportunity locations (see, for example, Bhat and Koppelman, 1993). In turn, the residential location choices of households, at an aggregate level, impact the built environment as transport, land use, and urban form change in response to where people live. This bidirectional and dynamic interaction between where people choose to live and how the built environment evolves is at the heart of integrated land-use and transportation modeling (see, for example, Bhat and Guo, 2004, Sener *et al*., 2011, Pagliara and Wilson, 2010, and Zolfaghari *et al*., 2012). More broadly, the decision of residential location fundamentally determines the connection between the household and the rest of the urban environment, and can have a profound impact on the overall quality of life of individuals in the household. As a consequence, the study of residential location choice has attracted considerable attention in a wide variety of disciplines well beyond transportation, including real estate science, ecology, actuarial science, psychology, and urban and regional economics.

Many different approaches have been considered in the literature to model residential location choice. One approach is based on a gravity-type formulation, which uses an aggregate-level relationship to characterize a distance-decay specification for the residence location-workplace interchange (see Lowry, 1964 and Wilson, 1970). Another approach is based on Alonso's (1960) bid-rent model that assumes that households compete for land and locate in concentric circles, with the density of households fading with distance from a monocentric employment center. Households' location decision is considered to be based on a trade-off between commuting time and land prices. The bid-rent model has been extended to consider other observed factors (such as the location of good schools, accessibility to activity opportunities, and crime rates) and unobserved factors (see Ellickson, 1981, Martinez, 2008, and Hurtubia and Bierlaire, 2011). However, the dominant approach to model residential location is based on the discrete choice formulation originally proposed by Lerman (1976) and McFadden (1978). Such a formulation has an appealing underlying microeconomic basis, and enables the analysis of trade-offs among a wide range of factors affecting the decision of where to locate. It also allows sensitivity variations (across socio-demographic segments of the population) to location attributes (see Bhat and Guo, 2004, Bhat and Guo, 2007, Prashker *et al.* (2008), Pinjari *et al.* 2009, and Zolfaghari *et al*., 2012).

In the context of the discrete choice formulation, many important issues become relevant and need to be addressed. Two such very important (and as we will discuss later, inter-related) considerations deal with the analysis unit used for the alternatives in the choice set and the choice set construction.

1.1 The Analysis Unit

A majority of residential location choice models consider an aggregate spatial region (for example, census tracts or traffic analysis zones or neighborhoods) as the analysis unit. Typically referred to as zone-based models, these models focus on households' choice of the spatial zone to reside in, as a function of a suite of zone characteristics (such as zone-based accessibility measures to pursue out-of-home activities, crime rates, quality of schools in the zone, commute times of workers in the household, zonal race and income distributions relative to household's race and income, respectively) and interactions of household characteristics with the zonal characteristics (see, for example, Chen *et al*., 2008, Pinjari *et al*., 2011, de Palma *et al*., 2007, and Bhat and Guo, 2007). The advantage of zone-based models is that data preparation is rather easy for estimation and forecasting, because the zone-based attributes are constructed anyway for use in other components of a travel demand model. However, there are many limitations of using a zone-based model. First, one has to define a spatial resolution for the "zone" and develop a configuration of the zonal structure. Unfortunately, this can imply that two points in space very proximal with one another may end up in different zones (because of the discretization of space), with potentially different zone-level attributes attached to the two points. The end-result is that the model estimates can be different based on how space is compartmentalized into zones. This problem, referred to in the literature as the Modifiable Areal Unit Problem (MAUP), is a long standing one with no clear way out to resolve the problem (see Openshaw, 1978, Guo and Bhat, 2004). Second, any dwelling unit related variables that one may want to use to explain residential location (such as owned single family unit versus a rented unit in a multi-family apartment complex) also has to be aggregated up to the zone level. This results in a situation where all points in space within a zone take on the same dwelling unit attribute values, even though there will be variation in these attributes within a zone. In essence, heterogeneities in space within a zone get ignored, and could lead to an aggregation bias in model estimates (see Heckman and Sedlacek, 1985). Besides, the decision of housing involves more than just location; households do consider dwelling unit attributes (such as number of bedrooms, number of bathrooms, square footage, and housing costs) too, and this is completely ignored by zone-based models. Third, micro-scale land-use policies cannot be analyzed using zone-based models because of the coarseness of the definition of space.

A second unit of analysis used in the literature corresponds to that of a parcel or a building. Doing so has many advantages over the zone-based models, because there is no ambiguity in the definition of the spatial unit of analysis, and thus the problems arising from MAUP in the zone context are non-existent in the parcel context. The use of a parcel can also help improve the specification of accessibility attributes, and provide the fine resolution needed for the analysis of micro-scale land-use policies. This approach has been adopted by Lee *et al*. (2010) and Lee and Waddell (2010), who developed a parcel-based residential location choice model for the Puget Sound region, though many of the locational explanatory variables they used were only defined at the zone level. The problem with parcel-based models is that, like the zone-based models, they do not consider dwelling unit attributes that are made jointly with the physical location of the residence. Thus, these models do not distinguish between two parcels with very different types of dwelling units. Besides, such models have not been used much because of the need for high spatial resolution data on parcels (which can be difficult to put together) and the computational issues associated with the high number of parcels as alternatives. In particular, unless the restrictive multinomial logit model form is used, estimation with a sample of parcel alternatives requires the introduction of appropriate correction terms (see further discussion in Section 1.3).

A third unit of analysis is that of the dwelling unit. One can then use a zonal-level spatial resolution or a parcel-level spatial resolution for the physical location unit. The advantage of this approach is that it can accommodate dwelling unit attributes. But a zone-level spatial representation brings the same disadvantages as those listed earlier for zone-based models in terms of spatial aggregation, while a parcel-based spatial representation has important advantages. Further, the combination of the dwelling unit and a parcel provides a unique "point" identification of residential location. Examples of the use of a zone-based dwelling unit approach include Guevara (2010) and Bayer *et al*. (2005), while examples of the parcel-based dwelling unit approach include Habib and Miller (2009) and Eliasson (2010). In both cases, considerations related to choice set formation are very important because the universal choice set explodes in size very quickly (particularly with a parcel-based spatial resolution), as we discuss in the next section.

1.2 Choice Set Generation

An important issue in the discrete choice modeling of residential choice is the alternatives from which a family chooses its location. This is non-trivial because the number of spatial location alternatives can range from a few hundreds to thousands in the case of zone-level models to hundreds of thousands in the case of parcel-level or dwelling unit level models. There is a relatively vast body of literature in the social-behavioral science literature now that suggests that decision-makers, when confronted with a vast array of possible options, whittle down those options to just a few using heuristic non-compensatory screening rules, followed by a much more careful and possibly compensatory process to make the final decision (see Tversky, 1972 and Manrai and Andrews, 1998). Further, many earlier studies of choice set formation have demonstrated how ignoring the choice set formation process, and assuming that all alternatives are available and evaluated using a utility-maximizing structure, can lead to biased parameter estimates, leading to incorrect sensitivity to variable changes and poor forecasting performance (see, for example, Shocker *et al.*, 1991, Williams and Ortuzar, 1982, Swait, 2001, Basar and Bhat, 2004, and Bell, 2007). The fundamental reason for this is that considering the full choice set is tantamount to assuming that the choice of one alternative implies an underlying preference ordering in which the chosen alternative is the highest ranked over all other alternatives. This may not be the case because individuals may not be aware of some alternatives and/or may use heuristics to simplify the choice process to reduce cognitive, emotional, and time/money search costs that come with choice option overloading. In particular, the cognitive costs are associated with the mental energy expenditure to collect information on each alternative in the choice set and make a "rational" choice (see Shugan, 1980 and Botti and Hsee, 2010); the emotional costs relate to the psychological distress that accrues from a consumer not being entirely sure about her/his preference ordering in the presence of a large number of options and/or experiencing a high degree of loss aversion due to having to reject a large number of options in the process of selecting one option (see Carmon *et al*., 2003); the time costs refer to the opportunity cost of collecting and processing information on a large number of alternatives, and the money costs refer to fiscal investments in the information gathering process. Indeed, Simon's pioneering work emphasized the use of heuristics and short-cuts to quickly circumscribe the set of possibilities to choose from, due to humans being "cognitive misers" and having "bounded rationality" (Simon, 1986).

The typical approach in the literature on residential choice modeling (whether at the zone-level, parcel-level, or dwelling level) has been to use a random sampling approach to sample a subset of the universal choice set of alternatives (for example, Bhat and Guo, 2007, Habib and Miller, 2009, Lee and Waddell, 2010, Guevara, 2010, Eliasson, 2010, Chen *et al*., 2008, and de Palma *et al*., 2007). This reduces the computational burden of estimating a discrete choice model with a large number of alternatives. With the introduction of appropriate correction terms, the approach can also provide consistent and asymptotically normal estimates for most discrete choice models, assuming that households consider all possible alternatives (see McFadden, 1978, and Guevara and Ben-Akiva, 2013a,b). However, sampling of alternatives is simply a statistical device to reduce the computational burden of considering all alternatives during estimation. At a fundamental level, it does implicitly assume that households do choose from the universal choice set, and it ignores the presence of any search behavior process heuristics and pruning tactics in residential choice decisions. As already discussed, this can lead to inappropriate forecasts and inaccurate policy sensitivity in response to changes in exogenous variables. Another substantial problem with the sampling approach is in forecasting. As discussed by Wegener (2011), even if it is true that households actually consider the universal choice set, a very large number of alternatives leads to a situation where there is little difference in the predicted choice probabilities across alternatives, which can lead to instability and poor predictions when such a residential location choice model is used in forecasting or in evaluating the effects of a change in a policy variable.

Another approach in the literature is to acknowledge the presence of a dynamic spatial choice process in which households get exposed to an evolving set of alternatives over a period of time (with potentially changing attributes of alternatives such as housing costs), search and construct what they believe to be a set of credible and feasible alternatives during each evaluation occasion in a first stage decision process, and then make a final choice from the alternatives remaining at the end of the first stage in a second stage decision process at some point in time (Habib and Miller, 2007). In the absence of direct observation on the first stage search process, analysts attempt to mimic this underlying choice set formation process assuming (1) a search strategy and (2) a specific approach to implement the search strategy to form choice sets. Zolfaghari *et al*. (2013) provide a good review of search strategies and their implementations. Briefly, the search strategies, as originally identified by Huff (1986) (and supported by his empirical analysis based on direct observation on the search process of households looking to purchase homes in the San Fernando Valley of Los Angeles), are likely to be a combination of three underlying cognitive psychology approaches: supply constraint-based, area-based, or anchor points-based. The supply constraint-based approach assumes that households will concentrate their search on areas where their housing needs (in terms of dwelling and parcel attributes) are most likely to be met. As well, the approach recognizes that different households may consider different alternatives because financial or access constraints or the social capital available at the disposal of the household may modulate and/or meter information flow (see also Bell, 2009, who emphasizes these considerations in a parental school choice context). The area-based search approach states that once a specific geographic market (and/or area type) has been identified for housing search, households will concentrate and persist their search within that market because of start-up and information-processing costs involved in shifting attention to another area. The anchor-points based approach is based on the notion that households will circumscribe their searches around specific anchor points and consider only those alternatives that are within a specific threshold distance of the anchor points. Most studies in the literature that consider search processes in a spatial context (such as residential choice or activity location choice) focus on the anchor-based approach (see, for example, Thill and Horowitz, 1997, Bhat, 1999, Bhat and Zhao, 2002, Elgar *et al*., 2009, and Rashidi *et al*., 2012).

The implementation of a search strategy itself is generally accomplished using a deterministic set-up or a probabilistic set-up. The deterministic set-up is based on specifying a fixed threshold for each household based on the predicted distribution of distances from one or more anchor points. For example, in Zolfaghari *et al*.'s (2012) zone-level residential choice model, they develop a Weibull-distributed model for average commute time (that is, the commute time averaged across all workers in a household) as a function of the number of vehicles, number of workers. Then, the 90th percentile commute time is declared as the threshold commute distance for zones to be considered in the choice set of the household (that is, 90% of households with given values of number of vehicles, number of workers, and household income have an average commute time less than the threshold).[1] The probabilistic set-up acknowledges the lack of precise information about the search process, and accommodates the uncertainty inherent in the choice set formation process. Thus, it is more representative of the true behavioral process underlying choice set formation relative to the deterministic set-up. In this probabilistic set-up, which typically uses Manski's (1977) two-stage modeling paradigm, the overall probability of choice of an alternative is developed as the sum (across all possible non-empty choice subsets of the universal choice set) of the product of probability of a choice set (formed through a non-compensatory conjunctive heuristic process) and the probability of the alternative given the choice set (typically formulated as a conventional compensatory utility maximization process). In Manski's approach, the two stages are considered as separate and independent mental processes, even though the second-stage choice is made from the retained (but latent to the analyst) choice set in the first stage (see also Bovy, 2009). Swait and Ben-Akiva's (1987) random constraint-based approach or its variants are typically used to form the probabilities of the non-empty choice subsets in the first step in a practical manner. Kaplan *et al.* (2011, 2012a,b) adopt a Manski-like approach for rental apartment choice modeling, but with the important difference that they overtly "observe" the choice set of respondents (rather than the

[1] Another approach to choice set formation is to attribute a probability weight to each possible alternative (based on, for example, the desired commute time distribution), and then adopt an importance sampling scheme to populate the choice set of a size smaller than the cardinality of the universal choice set. But, as discussed in Zolfaghari *et al.* (2012), if sampling correction terms are used to effectively undo the importance sampling, this is no different from using the universal choice set and the procedure effectively completely ignores any behavioral search element in the analysis. If no correction terms are used, the behavior assumed is equivalent to that of the deterministic set-up.

choice set being latent as in the Manski formulation). Specifically, they use information on three search criteria for developing the choice set in the first step. The search criteria included whether respondents (631 students in a University in Israel) were willing to share a rental apartment, location preference between two neighborhoods in the vicinity of the University, and the maximal rent price. That is, Kaplan *et al.* undertook a web survey-based experimental design to elicit information on both the search process (used to form the choice set) as well as the choice process from the constituted choice set. By doing so, the likelihood function can effectively be developed separately for the two stages and does not require the development, during estimation, of the probabilities of all possible choice subsets of the universal choice set in the first step.[2,3]

1.3 The Current Study in Context

In the current study, we adopt a dwelling unit level of analysis and consider a probabilistic choice set generation approach for residential choice modeling. The use of a dwelling unit has many advantages, as already discussed earlier. Most importantly, the decision of housing necessarily involves physical dwelling unit attributes (that is, amenities within the home) as well as location attributes (that is, accessibility to activity opportunities outside the home). That is, residential models that consider only physical location-related attributes miss out on important behavioral elements that drive housing choices. At the same time, we assume a two-stage modeling paradigm to accommodate the process by which households decide on a dwelling unit. But the key innovation we introduce here is that rather than motivate the first stage from an elimination-by-aspects kind of a principle in the first stage (as in Kaplan *et al*., 2012b), we consider the first stage as a high-level (non-compensatory) decision process regarding housing attributes

[2] Zolfaghari *et al*. (2013) criticize the Kaplan *et al.* model as being deterministic in that only the probability of the choice set formed from the observed search process is included in the first consideration part of the Manski model. The situation is in fact much more nuanced. Essentially, an issue is that Kaplan *et al.* assume that only alternatives that <u>simultaneously</u> meet all the search criteria are considered by respondents, while Zolfaghari *et al*. suggest that all feasible combinations (subsets) of the multiple search criteria should also be considered in forming the possible choice sets. However, Kaplan *et al*. were consistent in that only those dwelling unit alternatives that met <u>all</u> the selected thresholds on the search criteria were presented to respondents in the choice stage. Thus, based on their experimental design, they take the position that they actually "observed" the choice set (in the estimation stage). So, while it is true that different respondents may consider different search criteria (such as only the pricing-based search criterion or only the number of bedrooms criterion), allowing combinations of search criteria leads to an explosion in the number of dwelling unit alternatives for at least some choice sets. This itself is not behaviorally reasonable from a cognitive standpoint in the context of residential location choice, where there can be scores of dwelling units if, for example, only a one-dimensional search criterion is used. In our study, we exploit the fact that, even as they are forming choice sets, households already have formed a general preference for a range of dwelling unit attributes they seek. However, at this high-level preference development stage, they are not undertaking any detailed comparative (and compensatory) evaluation of actual dwelling units, but rather develop a multi-dimensional (and non-compensatory) set of preferences for dwelling unit attributes.

[3] The literature has also seen applications of a single stage search and choice decision process that we do not discuss in detail in this paper (the reader is referred to Cascetta and Papola, 2001, Elrod *et al*., 2004, Martinez *et al*., 2009 and Bierlaire *et al*., 2010). These single stage models tie the search and choice components very closely together within a single compensatory process. That is, an assumption implicit in these models is that the availability of an alternative in the choice set is a direct function of its utility. On the other hand, the two stage models allow the possibility that alternatives that may have a high utility in the second choice stage may not even enter into the picture because of locational or cost or other constraints in the non-compensatory first stage search process.

that includes some aggregate spatial representation of residential location. Importantly, and different from earlier two-stage Manski type choice models, the first and second stages in our system do not even have the same dependent variables and the same unit of analysis for the dependent variable(s). Thus, the choice in the first stage focuses on specific physical characteristics of the dwelling unit (such as, say a single-family detached one storey “to-be-owned” three bedroom home with two bathrooms), along with ranges of some other dwelling unit physical attributes (such as, say a 1,500-2,000 square foot home with a good yard size of 5,000-7,500 square feet costing between $200,000-$250,000 and within 2 miles of the workplace) and a preference to live in some aggregated representation of space (say, a certain neighborhood or area of a city). The choice in the second stage focuses on the precise dwelling unit physical characteristics for those attributes that are chosen in ranges in the first stage, along with the parcel-level spatial location of the dwelling unit (given the high-level dwelling unit attribute choices made in the first stage). This representation is consistent with a combination of supply constraint-based, area-based, or anchor points-based search strategies. Specifically, consistent with the supply constraint-based approach, we represent the first stage search not as a screening mechanism, but as an integral part of trying to maximize search efficiency (and minimize cognitive burden) by increasing the chances of a “hit” for the desired housing attributes and parcel locational attributes combination. Next, not inconsistent with the area-based search process, we use revealed preference data on actual housing choices rather than use experimental or web-based search data that may be onerous to collect. That is, we use the observed housing choices and an aggregate spatial representation of the observed residential location choice as the observed outcomes for the dependent variables in the first stage. Effectively, the premise is that the dwelling unit chosen was preceded by the choice of the dwelling unit attributes at a relatively aggregate level, including the geographic market (or aggregate spatial representation) in which the chosen dwelling unit is actually located (due to the persistence of search within the initially preferred geographic submarket). Finally, consistent with the anchor-based approach, the aggregate spatial preference of dwelling unit location in the first stage is based on a grouped (coarse) representation of the average commute distance (across workers in the household) from the actual chosen dwelling unit location to the work places of employed individuals. The underlying notion again is that the actual dwelling unit location is preceded by the choice of an aggregate geographic submarket to reside in based on commute distance.

In summary, our study accommodates the fact that housing choices involve both characteristics of the dwelling unit and its location (rather than many studies that divorce the analysis of these two completely and examine only dwelling unit choices or only physical residential location choice), while also mimicking the search process that underlies housing decisions. In particular, and unlike all earlier studies, we model a complete range of dwelling unit choices that include the following dimensions in the first stage: tenure type (rent or own), housing type (single family detached, single family attached, or apartment complex), number of bedrooms, number of bathrooms, number of storeys (one or more), square footage of the house (in a ranger), lot size (in a range), housing costs (in a range), density of residential neighborhood, and household average commute distance (in a range).[4] Among the dimensions examined in this paper, tenure choice and number of stories are binary outcomes, housing type and density of residential neighborhood are nominal (unordered multinomial) outcomes, the number of bedrooms and bathrooms are count outcomes, and square footage of the house and the land, housing costs, and household average commute distance are treated as grouped outcomes with underlying continuous variables (grouped outcomes are similar to ordinal outcomes, with the difference that the thresholds that demarcate various groups are observed and do not need to be estimated; see Bhat, 1994). The reason for the treatment as grouped outcomes stems from the notion that households, in the first consideration stage, make choices of what they desire in terms of general ranges of housing attributes, and then follow through only in the second evaluative stage with a rigorous comparison of actual dwelling units to make a final choice. To acknowledge this, and also to estimate comprehensive dwelling unit choice models from the revealed preference choice of dwelling unit, we exploit the idea that the final observed dwelling unit choice provides an indication of the broader preferences at the first consideration stage.

In this paper, we focus exclusively on the first consideration stage. The extension to the estimation of dwelling unit given the choice set is relatively straightforward using traditional random utility choice models, because of the winnowing down of the dwelling unit alternatives and the separation of the mental processes between the first consideration and second cross-evaluative (among desirable dwelling units) phases (because of which estimation of the second choice stage can be undertaken easily from the dwelling units that fall within the multidimensional profile actually chosen by each household in the estimation sample). Note also that this second stage will include many location-related and accessibility attributes (not included in the first stage) because of the fine resolution for the unit of analysis of space (because the dwelling unit alternatives in this second stage are defined at the parcel-level). Of course, the model structures from the first stage and second stage are also very different. In the first stage, there are multiple

[4] There is a vast theoretical and empirical literature that focuses exclusively on the tenure decision, including those based on a portfolio analysis-based framework, a utility-based discrete choice approach, and a risk-based evaluation perspective (see, for example, Kain and Quigley, 1972, Li, 1977, Henderson and Ioannides, 1983, Sinai and Souleles, 2005, Davidoff, 2006, and Flavin and Nakagawa, 2008). There is also substantial literature on the mobility decision (that is, whether to move or not, given the current dwelling unit choice), which we do not consider in the current paper (though we appreciate the importance of including this dimension in a future effort; the emphasis on the analysis of a comprehensive set of dwelling unit attributes considerably narrowed down the possible data sets available, and the data used in the current analysis did not have adequate information on the mobility decision). But there has been very limited literature on considering a comprehensive set of dimensions characterizing housing stock, as is the focus of the current study. But, for the analysis of two or three dimensions of the housing stock, see Quigley (1976), Lerman (1977), Cho (1997), Quigley (1985), Rapaport (1997), Boheim and Taylor (1999), Skaburskis (1999), Yates and Mackay (2006), Frenkel and Kaplan (2014). There also has been a substantial literature on housing tenure/mobility (discrete choice) and quantity of housing demand (continuous choice), but this vast literature uses hedonic relationships to estimate the quantity of housing demand as the market value of a dwelling unit divided by a constructed price of a standardized unit of the flow of housing services. However, the demand for housing services in such studies is rather abstract and does not correspond to individual dimensions of the dwelling unit. Examples of this literature include Lee and Trost (1978), Rosen (1979), Dubin and McFadden (1984), Ermisch *et al*. (1996), Rapaport (1997), Rouwendal and Meijer (2001), Goodman (2002), Barrios-García and Rodríguez-Hernández (2008), and Chen and Jin (2014). Besides, previous housing stock studies have not motivated their analysis from a search theory perspective to winnow down the choice set for the dwelling unit in a parcel, as is the primary motivation for the current paper.

dependent variables, each corresponding to a physical dwelling unit attribute or an aggregate representation of space. In the second stage, there is a single nominal variable corresponding to dwelling unit choice, with the alternatives corresponding to all the dwelling units consistent with the first stage choice. Thus, this second stage takes the more familiar "unlabeled" multinomial model form used in the literature (see, for example, Newman and Bernardin Jr, 2010).

For forecasting, one can form different choice sets that exhaust the combinations of the housing attributes, next form the probability of each combination from the estimated parameters and the probability of each dwelling unit choice given the units within each choice set, and then sum the product of the two probabilities across all combinations to get the probability of choice of each dwelling unit. But this process is much easier to implement in a microsimulation framework where the probabilities of each choice set get translated into a deterministic choice in a first step. Then, all those dwelling units that are available in the market and that fit the desires of the household (as deterministically obtained after the first step) can be evaluated using the estimated choice model to assign the household to a dwelling unit.[5]

There are several salient aspects of the current paper. First, unlike earlier studies in the housing modeling literature that focus either only on physical location or only on a very limited set of housing choices, the current study examines a whole suite of housing choices at the dwelling unit level. As indicated in a comprehensive review of earlier studies on housing choices, Coulombel (2010) states "there is a wall separating the issues of location and dwellings characteristics in academic research, and the interplay between the two is still not fully understood (Hilber, 2005). This might represent the most important lack as for now." Second, the entire set of dimensions characterizing housing choices is modeled accommodating a non-compensatory search process combined with a compensatory choice model for the specific dwelling unit. Third, the use of a large set of multidimensional housing attributes at the consideration stage in our study leads to a small set of desirable options to make the final selection from, as is likely to be the case behaviorally (see Rashidi *et al*., 2012). This immediately obviates the need for sampling during estimation. At the same time, the number of possible choice sets to be formed in the forecasting stage is still kept to a manageable number while also retaining behavioral realism. Fourth, in the past, a high-dimensional model at the consideration stage has not been estimated because of the computational complexity in doing so when there are mixed dependent outcomes (*i.e.*, variables). In the current study, we address this issue by using Bhat's (2014) generalized heterogeneous data model (GHDM) system that is capable of accommodating a range of different types of dependent outcomes, while capturing jointness caused by unobserved factors. We do so by introducing latent psychological constructs that influence multiple housing choice outcomes in a measurement equation, engendering a parsimonious covariance structure across the many outcomes because the latent constructs themselves are specified in a structural system to be a function of exogenous variables and correlated random error terms. The approach

[5] Technically, as discussed earlier, one has to identify whether or not a household is looking for a new residence (either because of a move from within a metropolitan region or because of a new household migrating into the region), along with the choice of housing attributes (see, for example, Eluru *et al*., 2009, Lee and Waddell, 2010, and Kortum *et al*., 2012). We leave this extension for the future.

represents a powerful dimension-reduction technique, and accommodates non-nominal as well as nominal variables. For example, we specify that an environmentally conscious household will locate itself in a multi-family household setting with fewer bathrooms and bedrooms, in a one storey home for energy efficiency, in a high density neighborhood that supports non-auto modes of transportation, in a dwelling unit with a smaller square footage and smaller land acreage, and closer to the work place. Besides, including the latent psychological constructs is consistent with the perspective that dwelling unit choice involves an affective dimension in that it constitutes a lifestyle choice. Fifth, we propose the use of Bhat's Maximum Approximate Composite Marginal Likelihood (MACML) approach for estimation of the resulting two stage model. The proposed MACML procedure requires the maximization of a function that has no more than bivariate normal cumulative distribution functions to be evaluated.

The rest of this paper is structured as follows. Section 2 presents the model formulation, while Section 3 presents the empirical application of the model. Section 4 concludes the paper by highlighting the important results and identifying future research directions.

Chapter 2. The GHDM Model Formulation

Let q be the index for households $(q = 1, 2, ..., Q)$, which we will suppress in parts of the presentation below. Assume that all error terms in the GHDM model for a household are independent of other household error terms.

2.1 Structural Equation Model

Consider the latent variable *(i.e.*, an unobserved lifestyle variable or a latent psychological construct) z_l^* for a specific household, with l being the index for latent variables ($l = 1, 2, \ldots, L$). Write z_l^* as a linear function of covariates:

$$z_l^* = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{1}$$

where $\boldsymbol{w}$ is a $(\tilde{D} \times 1)$ vector of observed covariates (excluding constant), $\boldsymbol{\alpha}_l$ is a corresponding $(\tilde{D} \times 1)$ vector of coefficients, and η_l is a random error term assumed to be standard normally distributed for identification purposes (see Stapleton, 1978). Next, define the $(L \times \tilde{D})$ matrix $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_L)'$, and the $(L \times 1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*, ..., z_L^*)'$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, ..., \eta_L)'$. Let $\boldsymbol{\eta} \sim MVN_L[\boldsymbol{0}_L, \boldsymbol{\Gamma}]$, where $\boldsymbol{0}_L$ is an $(L \times 1)$ column vector of zeros, and $\boldsymbol{\Gamma}$ is an $(L \times L)$ correlation matrix. In matrix form, we may write Equation (1) as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha w} + \boldsymbol{\eta}. \tag{2}$$

$\boldsymbol{z}^*$ is a vector of latent psychological constructs (variables).

2.2 Measurement Equation Model Components

We will consider a combination of grouped, count, and nominal outcomes (indicators) of the underlying latent variable vector $\boldsymbol{z}^*$. In the past, SEM-type models have primarily used binary, ordinal, and continuous outcomes, but we use grouped, count, and nominal variables in the current paper.

Consider N grouped outcomes (indicator variables) for the individual, and let n be the index for the grouped outcomes $(n = 1, 2, ..., N)$. Also, let J_n be the number of categories for the n^{th} grouped outcome $(J_n \geq 2)$ and let the corresponding index be j_n $(j_n = 1, 2, ..., J_n)$. Let $\tilde{y}_n^*$ be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the n^{th} grouped variable. Assume that the household under consideration chooses the a_n^{th} grouped category. Then, in the grouped response formulation (see Bhat, 1994) for the household, we may write:

$$\tilde{y}_n^* = \tilde{\boldsymbol{\gamma}}_n' \boldsymbol{x} + \tilde{\boldsymbol{d}}_n' \boldsymbol{z}^* + \tilde{\varepsilon}_n, \text{ and } \tilde{\psi}_{n,a_n-1} < \tilde{y}_n^* < \tilde{\psi}_{n,a_n}, \tag{3}$$

where $\boldsymbol{x}$ is a fixed $(A\times 1)$ vector of exogenous variables (including a constant) as well as possibly the observed values of other endogenous grouped, count variables, and nominal variables (introduced as dummy variables).[6] $\tilde{\boldsymbol{\gamma}}_{\boldsymbol{n}}$ is a corresponding $(A\times 1)$ vector of coefficients to be estimated, $\tilde{\boldsymbol{d}}_{\boldsymbol{n}}$ is an $(L\times 1)$ vector of latent variable loadings on the n^{th} grouped outcome, the $\tilde{\psi}$ terms represent thresholds, and $\tilde{\varepsilon}_n$ is a normal random error term for the n^{th} grouped outcome; $\tilde{\varepsilon}_n \sim N(0,\sigma_n^2)$. For each grouped outcome, $\tilde{\psi}_{n,0} < \tilde{\psi}_{n,1} < \tilde{\psi}_{n,2} ... < \tilde{\psi}_{n,J_n-1} < \tilde{\psi}_{n,J_n}$; $\tilde{\psi}_{n,0} = -\infty$ and $\tilde{\psi}_{n,J_n} = +\infty$, which are observed thresholds that do not need to be estimated. For later use, let $\tilde{\boldsymbol{\psi}}_{\boldsymbol{n}} = (\tilde{\psi}_{n,1}, \tilde{\psi}_{n,2}, \tilde{\psi}_{n,3} ..., \tilde{\psi}_{n,J_n-1})'$ and $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\psi}}_1', \tilde{\boldsymbol{\psi}}_2', ..., \tilde{\boldsymbol{\psi}}_{\boldsymbol{N}})'$. Stack the N underlying continuous variables $\tilde{y}_n^*$ into an $(N\times 1)$ vector $\tilde{\boldsymbol{y}}^*$, and the N error terms $\tilde{\varepsilon}_n$ into another $(N\times 1)$ vector $\tilde{\boldsymbol{\varepsilon}}$. Let $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \tilde{\gamma}_2 ..., \tilde{\gamma}_N)'$ $[(N\times A)$ matrix] and $\tilde{\boldsymbol{d}} = \left(\tilde{\boldsymbol{d}}_{\boldsymbol{1}}, \tilde{\boldsymbol{d}}_{\boldsymbol{2}}, ..., \tilde{\boldsymbol{d}}_{\boldsymbol{N}}\right)$ $[(N\times L)$ matrix], and let $\boldsymbol{\Xi}$ be the diagonal matrix of dimension N representing the covariance matrix of $\tilde{\boldsymbol{\varepsilon}}$. That is,

$$\boldsymbol{\Xi} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \sigma_N^2 \end{bmatrix} \tag{4}$$

The diagonal specification for the matrix above helps identification (see Bhat, 2014 for a detailed discussion), though the presence of the unobserved $\boldsymbol{z}^*$ vector will, in general, generate covariance among the grouped outcomes. Finally, stack the lower observed thresholds for the decision-maker $\tilde{\psi}_{n,a_n-1}\,(n = 1, 2, ..., N)$ into an $(N\times 1)$ vector $\tilde{\boldsymbol{\psi}}_{\boldsymbol{low}}$ and the upper thresholds $\tilde{\psi}_{n,a_n}\,(n = 1, 2, ..., N)$ into another vector $\tilde{\boldsymbol{\psi}}_{up}$. Then, in matrix form, the measurement equation for the grouped outcomes (indicators) for the household may be written as:

$$\tilde{\boldsymbol{y}}^* = \tilde{\boldsymbol{\gamma}}\boldsymbol{x} + \tilde{\boldsymbol{d}}\boldsymbol{z}^* + \tilde{\boldsymbol{\varepsilon}}, \quad \tilde{\boldsymbol{\psi}}_{\boldsymbol{low}} < \tilde{\boldsymbol{y}}^* < \tilde{\boldsymbol{\psi}}_{\boldsymbol{up}}. \tag{5}$$

[6] 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 grouped, count, and nominal variables (which we will more generally refer to as limited-dependent variables), the structural effects of one limited-dependent 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 limited-dependent variables, as well as have the observed limited-dependent variables themselves structurally affect each other in a bi-directional fashion. This creates a logical inconsistency problem (see Maddala, 1983, page 119 for a good discussion). It is critical to note that, regardless of which directionality of structural effects among the endogenous variables is specified (or even if no relationships are specified), the system is a joint bundled system because of the correlation in unobserved factors impacting the underlying propensities. The reader is referred to Bhat (2014) for additional details.

Let there be C count outcomes for a household, and let c be the index for the count outcomes $(c = 1, 2, ..., C)$. Let the count index be k_c $(k_c = 0, 1, 2, ..., \infty)$ and let r_c be the actual observed count value for the household. Then, following the recasting of a count model in a generalized ordered-response probit (GORP) formulation (see Castro, Paleti, and Bhat, or CPB, 2012 and Bhat *et al*., 2015), a generalized version of the negative binomial count model may be written as:

$$\breve{y}_c^* = \breve{\boldsymbol{d}}_c' \boldsymbol{z}^* + \breve{\varepsilon}_c,\ \breve{\psi}_{c,r_c-1} < \breve{y}_c^* < \breve{\psi}_{c,r_c}, \tag{6}$$

$$\breve{\psi}_{c,r_c} = \Phi^{-1}\left[\frac{(1-\upsilon_c)^{\theta_l}}{\Gamma(\theta_c)} \sum_{t=0}^{r_c} \left(\frac{\Gamma(\theta_c + t)}{t!} (\upsilon_c)^t\right)\right] + \varphi_{c,r_c},\ \upsilon_c = \frac{\lambda_c}{\lambda_c + \theta_c},\ \text{and } \lambda_c = e^{\breve{\boldsymbol{\gamma}}_c' \boldsymbol{x}}. \tag{7}$$

In the above equation, $\breve{y}_c^*$ is a latent continuous stochastic propensity variable associated with the count variable c that maps into the observed count r_c through the $\breve{\boldsymbol{\psi}}_c$ vector (which is a vertically stacked column vector of thresholds $(\breve{\psi}_{c,-1}, \breve{\psi}_{c,0}, \breve{\psi}_{c,1}, \breve{\psi}_{c,2}, ...)'$. $\breve{\boldsymbol{d}}_c$ is an $(L \times 1)$ vector of latent variable loadings on the c^{th} count outcome, and $\breve{\varepsilon}_c$ is a standard normal random error term. $\breve{\boldsymbol{\gamma}}_c$ is an $(A \times 1)$ column vector corresponding to the vector $\boldsymbol{x}$. Φ^{-1} in the threshold function of Equation (7) is the inverse function of the univariate cumulative standard normal. θ_c is a parameter that provides flexibility to the count formulation, and is related to the dispersion parameter in a traditional negative binomial model $(\theta_c > 0 \,\forall\, c)$. $\Gamma(\theta_c)$ is the traditional gamma function; $\Gamma(\theta_c) = \int_{\tilde{t}=0}^{\infty} \tilde{t}^{\,\theta_c - 1} e^{-\tilde{t}}\, d\tilde{t}$. The threshold terms in the $\breve{\boldsymbol{\psi}}_c$ vector satisfy the ordering condition (*i.e.*, $\breve{\psi}_{c,-1} < \breve{\psi}_{c,0} < \breve{\psi}_{c,1} < \breve{\psi}_{c,2} .... < \infty \ \forall\ c$) as long as $\varphi_{c,-1} < \varphi_{c,0} < \varphi_{c,1} < \varphi_{c,2} .... < \infty$. The presence of the φ_c terms in the thresholds provides flexibility to accommodate high or low probability masses for specific count outcomes without the need for cumbersome traditional treatments using zero-inflated or related mechanisms in multi-dimensional model systems (see CPB, 2012 for a detailed discussion). For identification, we set $\varphi_{c,-1} = -\infty$ and $\varphi_{c,0} = 0$ for all count variables c (so $\breve{\psi}_{c,-1} = -\infty\ \forall\ c$). In addition, we identify a count value e_c^* $(e_c^* \in \{0,1,2,......\})$ above which φ_{c,k_c} $(k_c \in \{1,2,......\})$ is held fixed at φ_{k,e_c^*}; that is, $\varphi_{c,k_c} = \varphi_{c,e_c^*}$ if $k_c > e_c^*$, where the value of e_c^* can be based on empirical testing. Doing so is the key to allowing the count model to predict beyond the range available in the estimation sample. For later use, let $\boldsymbol{\varphi}_c = (\varphi_{c,1}, \varphi_{c,2}, ..., \varphi_{c,e_c^*})'$ $[(e_c^* \times 1)$ vector] (assuming $e_c^* > 0$), $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1', \boldsymbol{\varphi}_2', \dots \boldsymbol{\varphi}_C')'$ $\left[\left(\sum_c e_c^*\right) \times 1 \text{ vector}\right]$, $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_C)'$ $[(C \times 1)$ vector$]$, and $\breve{\boldsymbol{\psi}} = (\breve{\boldsymbol{\psi}}_1', \breve{\boldsymbol{\psi}}_2', ..., \breve{\boldsymbol{\psi}}_C')'$. Also, stack the C latent variables $\breve{y}_c^*$ into a $(C \times 1)$ vector $\breve{\boldsymbol{y}}^*$, the C error terms $\breve{\varepsilon}_c$ into another $(C \times 1)$ vector $\breve{\boldsymbol{\varepsilon}}$. Let $\breve{\boldsymbol{\varepsilon}} \sim MVN_C(\boldsymbol{0}_C, \mathbf{IDEN_C})$ from identification considerations, and stack the lower thresholds of the individual $\breve{\psi}_{c,r_c-1}\left(c=1,2,...,C\right)$ into a $(C\times 1)$ vector $\breve{\boldsymbol{\psi}}_{\boldsymbol{low}}$, and the upper thresholds $\breve{\psi}_{c,r_c}\left(c=1,2,...,C\right)$ into another $(C\times 1)$ vector $\breve{\boldsymbol{\psi}}_{\boldsymbol{up}}$. Define $\breve{\boldsymbol{\gamma}}=(\breve{\boldsymbol{\gamma}}_{\mathbf{1}},\breve{\boldsymbol{\gamma}}_{\mathbf{2}},...,\breve{\boldsymbol{\gamma}}_{\mathbf{C}})'$ $[(C\times A)$ matrix] and $\breve{\boldsymbol{d}}=\left(\breve{\boldsymbol{d}}_1,\breve{\boldsymbol{d}}_{2,},...,\breve{\boldsymbol{d}}_{\boldsymbol{C}}\right)'$ $[(C\times L)$ matrix]. With these definitions, the latent propensity underlying the count outcomes may be written in matrix form as:

$$\breve{\boldsymbol{y}}^*=\breve{\boldsymbol{d}}\boldsymbol{z}^*+\breve{\boldsymbol{\varepsilon}},\quad \breve{\boldsymbol{\psi}}_{\boldsymbol{low}}<\breve{\boldsymbol{y}}^*<\breve{\boldsymbol{\psi}}_{\boldsymbol{up}} \tag{8}$$

Finally, let there be G nominal (unordered-response) variables for an individual $(g=1,2,3,...,G)$. Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g\geq 2$) and let i_g be the corresponding index $(i_g=1,2,3,...,I_g)$. Consider the g^{th} nominal variable and assume that the individual under consideration chooses the alternative m_g. Also, assume the usual utility structure (see, for example, Bhat, 2000) for each alternative i_g.

$$\boldsymbol{U}_{\boldsymbol{gi_g}}=\boldsymbol{b}'_{gi_g}\boldsymbol{x}+\boldsymbol{\vartheta}'_{gi_g}(\boldsymbol{\beta}_{gi_g}\boldsymbol{z}^*)+\varsigma_{gi_g}, \tag{9}$$

where $\boldsymbol{x}$ is the same fixed vector as earlier, $\boldsymbol{b}_{gi_g}$ is an $(A\times 1)$ column vector of corresponding coefficients, and ς_{gi_g} is a normal error term. $\boldsymbol{\beta}_{gi_g}$ is a $(N_{gi_g}\times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i_g, and $\boldsymbol{\vartheta}'_{gi_g}$ is an $(N_{gi_g}\times 1)$-column vector of coefficients capturing the effects of latent variables and its interaction effects with other exogenous variables (see Bhat and Dubey, 2014). Let $\boldsymbol{\varsigma}_{\boldsymbol{g}}=(\varsigma_{g1},\varsigma_{g2},...,\varsigma_{gI_g})'$ ($I_g\times 1$ vector), and $\boldsymbol{\varsigma}_{\boldsymbol{g}}\sim MVN_{I_g}(0,\boldsymbol{\Lambda}_{\mathbf{g}})$. Taking the difference with respect to the first alternative, only the elements of the covariance matrix $\breve{\boldsymbol{\Lambda}}_g$ of the covariance matrix of the error differences, $\breve{\boldsymbol{\varsigma}}_{\boldsymbol{g}}=(\breve{\varsigma}_{g2},\breve{\varsigma}_{g3},...,\breve{\varsigma}_{gI_g})$ (where $\breve{\varsigma}_{gi}=\breve{\varsigma}_{gi}-\breve{\varsigma}_{g1}$, $i\neq 1$), are estimable. Further, the variance term at the top left diagonal of $\breve{\boldsymbol{\Lambda}}_g$ $(g=1,2,...,G)$ is set to one to account for scale invariance. $\boldsymbol{\Lambda}_{\mathbf{g}}$ is constructed from $\breve{\boldsymbol{\Lambda}}_{\mathbf{g}}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and column are filled with values of zeros. In addition, the usual identification restriction is imposed that one of the alternatives serves as the base when introducing alternative-specific constants and variables that do not vary across alternatives (that is, whenever an element of $\boldsymbol{x}$ is individual-specific and not alternative-specific, the corresponding element in $\boldsymbol{b}_{gi_g}$ is set to zero for at least one alternative i_g). To proceed forward, define $\boldsymbol{U}_g=(U_{g1},U_{g2},...,U_{gI_g})'$ ($I_g\times 1$ vector), $\boldsymbol{b}_g=(\boldsymbol{b}_{\boldsymbol{g}1},\boldsymbol{b}_{\boldsymbol{g}2},\boldsymbol{b}_{\boldsymbol{g}3},...,\boldsymbol{b}_{gI_g})'$ ($I_g\times A$ matrix), and $\boldsymbol{\beta}_g=(\boldsymbol{\beta}'_{g1},\boldsymbol{\beta}'_{g2},...,\boldsymbol{\beta}'_{gI_g})'$ $\left(\sum_{i_g=1}^{I_g}N_{gi_g}\times L\right)$ matrix. Also, define the $\left(I_g\times\sum_{i_g=1}^{I_g}N_{gi_g}\right)$ matrix $\boldsymbol{\vartheta}_g$, which is initially filled with all zero values. Then, position the $(1\times N_{g1})$ row vector $\boldsymbol{\vartheta}'_{g1}$ in the first row to occupy columns 1 to N_{g1} , position the $(1\times N_{g2})$ row vector $\boldsymbol{\vartheta}'_{g2}$ in the second row to occupy columns N_{g1} +1 to $N_{g1}+N_{g2}$, and so on until the $(1\times N_{gI_g})$ row vector $\boldsymbol{\vartheta}'_{gI_g}$ is appropriately positioned. Further, define $\boldsymbol{\varpi}_g=(\boldsymbol{\vartheta}_g\boldsymbol{\beta}_g)$ ($I_g\times L$ matrix), $\vec{G}=\sum_{g=1}^{G} I_g$, $\tilde{G}=\sum_{g=1}^{G}(I_g-1)$, $\tilde{T}=\sum_{g=1}^{G}T_g$, $\boldsymbol{U}=\left(\boldsymbol{U}'_1,\boldsymbol{U}'_2,\ldots,\boldsymbol{U}'_G\right)'$ ($\vec{G}\times 1$ vector), $\boldsymbol{\varsigma}=(\boldsymbol{\varsigma}_1,\boldsymbol{\varsigma}_2,\ldots,\boldsymbol{\varsigma}_G)'$ ($\vec{G}\times 1$ vector), $\boldsymbol{b}=(\boldsymbol{b}'_1,\boldsymbol{b}'_2,\ldots,\boldsymbol{b}'_G)'$ ($\vec{G}\times A$ matrix), $\boldsymbol{\varpi}=(\boldsymbol{\varpi}'_1,\boldsymbol{\varpi}'_2,\ldots,\boldsymbol{\varpi}'_G)'$ ($\vec{G}\times L$ matrix), and $\boldsymbol{\vartheta}_{vec}=\text{Vech}(\boldsymbol{\vartheta}_1,\boldsymbol{\vartheta}_2,\ldots,\boldsymbol{\vartheta}_G)$ (that is, $\boldsymbol{\vartheta}_{vec}$ is a column vector that includes all elements of the matrices $\boldsymbol{\vartheta}_1,\boldsymbol{\vartheta}_2,\ldots,\boldsymbol{\vartheta}_G$). Then, in matrix form, we may write Equation (9) as:

$$\boldsymbol{U}=\boldsymbol{bx}+\boldsymbol{\varpi}\,\boldsymbol{z}^{*}+\boldsymbol{\varsigma}, \tag{10}$$

where $\boldsymbol{\varsigma}\sim MVN_{\vec{G}}(\mathbf{0}_{\vec{G}},\boldsymbol{\Lambda})$. As earlier, due to identification considerations, we specify $\boldsymbol{\Lambda}$ as follows:

$$\boldsymbol{\Lambda}=\begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 & 0 & 0.\cdots 0 \\ 0 & \boldsymbol{\Lambda}_2 & 0 & 0\cdots 0 \\ 0 & 0 & \boldsymbol{\Lambda}_3 & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots\ \ldots \vdots \\ 0 & 0 & 0 & 0\ \ldots\ \boldsymbol{\Lambda}_G \end{bmatrix} (\vec{G}\times\vec{G}\ \text{matrix}), \tag{11}$$

In the general case, this allows the estimation of $\sum_{g=1}^{G}\left(\frac{I_g*(I_g-1)}{2}-1\right)$ terms across all the G nominal variables (originating from $\left(\frac{I_g*(I_g-1)}{2}-1\right)$ terms embedded in each $\breve{\boldsymbol{\Lambda}}_g$ matrix; g=1,2,…,G) .

Let $\boldsymbol{\delta}$ be the collection of parameters to be estimated:

$$\boldsymbol{\delta}=[\text{Vech}(\boldsymbol{\alpha}),\text{Vech}(\boldsymbol{\Gamma}),\text{Vech}\ (\boldsymbol{\Xi}),\text{Vech}(\tilde{\boldsymbol{\gamma}}),\ \text{Vech}(\vec{\boldsymbol{d}}),\boldsymbol{\varphi},\boldsymbol{\theta},\ \text{Vech}(\boldsymbol{b}),\text{Vech}(\breve{\boldsymbol{\gamma}}),\boldsymbol{\vartheta}_{vec},\text{Vech}(\boldsymbol{\Lambda})],$$

where the operator "Vech(.)" vectorizes all the non-zero elements of the matrix/vector on which it operates. We will assume that the error vectors $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\varepsilon}}$, and $\breve{\boldsymbol{\varepsilon}}$ are independent of each other. The identification and estimation of the system is provided in an online supplementary note to streamline the presentation here[7]. Figure 1 provides a diagrammatic representation of the entire model system.

[7] (See http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/ResidentialChoice/Online_supplementary_note.pdf)

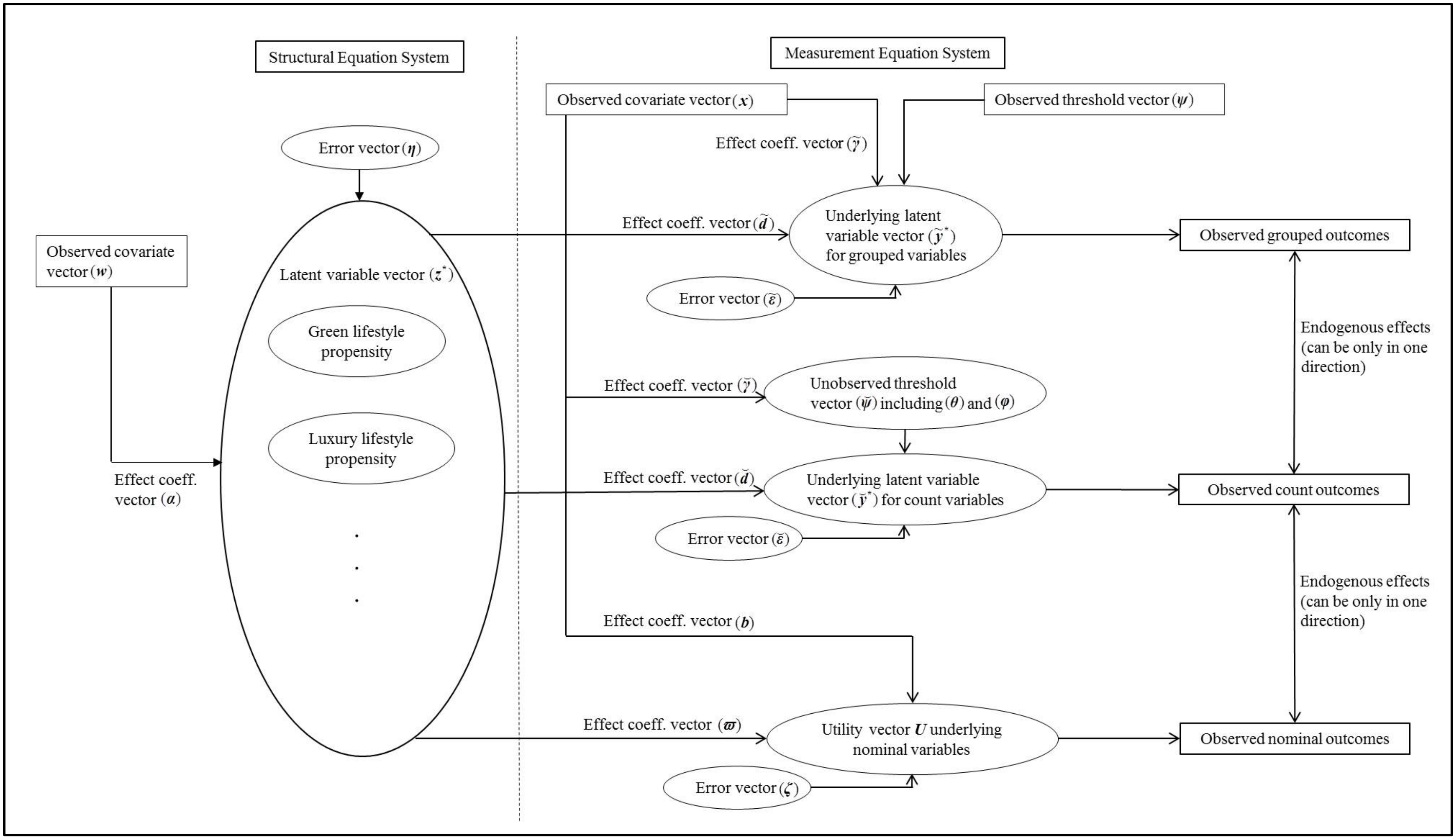

Figure 1: Diagrammatic Representation of the Model System

Chapter 3. Empirical Application

3.1 Data

Data for this study is derived from the 2009 American Housing Survey (AHS), sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. The AHS identified about 59,800 dwelling (or housing) units in the country, sampled to be representative of dwelling units in the entire country. These dwelling units were contacted by Census enumerators, either through telephone interviews or in-person interviews, and a notebook-based survey was administered to a knowledgeable household member 16 years of age or over in the dwelling unit. While detailed socio-economic information was collected regarding each member residing in the dwelling unit, the relationships among residents was based on soliciting information regarding the relationship of each resident with a single “reference person.” The “reference person” (also referred to as the “householder” in the 2009 AHS survey) was identified as the first resident member listed as owning or renting a dwelling unit in the official housing records.

Of the 59,800 eligible sample units, 6,450 could not be interviewed either because of inability to contact any interview-eligible individual in the corresponding household or because of refusal of the unit despite repeated visits. All the remaining dwelling units were interviewed between April and September 2009. The sampling units are located across 878 counties and independent cities with representation from all 50 states and the District of Columbia. Details of the sample design and survey administration procedures are available in Appendix B of U.S. Census Bureau (2011). Each sampled housing unit represents about 2000 other units in the country. A majority of the sample units (about 89%) were occupied at the time of the survey, while 11% were vacant. In addition to this national level survey, a concurrent metropolitan survey of housing was undertaken in New Orleans and Seattle. In this paper, we focus only on occupied dwelling units because of better data quality for such units than for vacant units. Further, to maintain focus on one metropolitan region, we selected the Seattle region in the State of Washington where the occupancy rates of the dwelling units was quite high. Another reason for choosing the Seattle region over the New Orleans region was the lingering effects of Hurricane Katrina in 2005 on housing markets there.

The initial sample size of occupied dwelling units in the AHS (including the national survey and the concurrent metropolitan survey) in the Seattle region was 1491. There were very few mobile homes in this sample and, after dropping these observations, the sample size reduced to 1445 fixed and occupied dwelling units. After some additional screening to remove inconsistent/unusual records (such as dwellings with zero square footage and/or zero monthly rent for rented houses), the final sample size used was 1421 with complete information on dwelling attributes as well as the socio-economic characteristics of individuals in the dwelling (interestingly, all dwelling units in our sample had at least one resident worker, presumably a reflection of the vibrant technology-driven labor market in the Seattle region).

The survey collected detailed information on the characteristics of each dwelling unit, including (but not limited to) housing type (single family-attached, single family-detached, and apartment complex with two or more apartment units), number of stories, number of bedrooms and bathrooms, house square footage, lot area, density of dwelling unit location (primary central city, secondary central city or suburb), whether the unit was currently being rented or owned (*i.e.*, tenure type), and housing costs.[8,9] In addition, information on demographics was also collected, including, for each individual in the household, the following characteristics: age, education level, sex, ethnicity/race, relationship to reference person, marital status, country of birth, U.S. citizenship status, employment status, commuting characteristics if employed (whether or not the individual worked outside the home), full-time employment (35 hours of work or more per week) or part-time employment (less than 35 hours of work), telecommuting characteristics, commute mode, time, and distance to the usual work location of the individual, and annual income by source (such as wages, tips, self-employment, rental property, retirement/pension, public welfare, and alimony or child support). The individual-level information also provided many variables at the household level (some of which were sought directly in the survey for verification purposes, though they can be computed from the individual-level information). These included household income (sum of the incomes of all individuals aged 16 years or older), household size by age groupings, number of adults (individuals over 16 years of age), number of children of different age groups, number of individuals with physical challenges, number of workers, number of workers with full-time employment, number of workers with part-time employment, number of workers with the option to work from home, education status of individuals in the household, marital status of individuals (married, widowed, separated or divorced, or never married), race/ethnicity of the householder in one of the following categories: (1) Caucasian, (2) Asian, (3) African-American, and (4) Others (Hawaiian, American Indian and Alaska Native), and the immigrant status of the household in one of the following three categories: (1) all members are foreign born (labeled as immigrant households), (2) all members are born in the U.S. (labeled as non-immigrant households), and (3) some

[8] The housing cost variable includes estimates of the equivalent monthly costs of all of the applicable following items: electricity, gas, fuel oil, other fuels (*e.g.* wood, coal, kerosene, *etc*.), garbage and trash, water and sewage, real estate taxes, property insurance, condominium fees, homeowner's association fees, land or site rent, rent, mortgage payments, home equity loan payments, other charges included in mortgage payments, and routine maintenance. Indeed, the housing cost-related questions occupy all of about 140 pages in the codebook for the American Housing Survey (U.S. Department of Housing and Urban Development, 2013). It is important to note that the monthly housing cost variable used in the current paper is a constructed variable (by the U.S. Census Bureau) from a suite of questions on housing costs and mortgage/rental payment arrangements (with respondents free to provide cost and payment estimates over any time period, not necessarily a monthly period). Additional details are available in the codebook. The overall monthly cost was constructed in this paper to represent the typical costs in residing in the dwelling unit and the value of the property and the land (computed as an equivalent total rental cost as estimated by the respondents, using all the costs listed above and replacing any mortgage payments by an estimated monthly rent as provided by the respondents).

[9] The survey data itself had information on the precise address of each dwelling unit, because dwelling units constituted the sampling frame. But the publicly available information categorizes the dwelling unit as being located in one of three density categories of neighborhoods: (1) suburb of a metropolitan statistical area MSA (population less than 250,000), (2) secondary city of an MSA (population is between 250,000 and 1,000,000) and (3) primary city of an MSA (population greater than 1 million). While this categorization is technically based on population size, we use it as a proxy for density.

members born in the U.S., and others born outside the U.S. (labeled as combination households).[10] Descriptive statistics of these household-level independent variables are provided in an online supplement to this paper at http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/ResidentialChoice/Independent_Variables.pdf.

3.2 Dependent Variables

The dependent variables considered in this study represent different attributes of the physical dwelling unit characteristics, housing costs, tenure type, housing type, density of residential location, and household average commute distance associated with the dwelling unit. These dependent variables are discussed below by variable type.

3.2.1 Grouped outcomes

(1) The square footage of the dwelling unit corresponds to the square feet of all the rooms in the dwelling unit, including basement and finished attics (but excluding unfinished attics, carports, attached garages, and porches that are unprotected from weather). This variable is classified into six grouped categories: (1) less than or equal to 1,000, (2) 1,001 – 1,500, (3) 1,501 – 2,000, (4) 2,001 – 2,500, (5) 2,501 – 3,000, and (6) greater than 3,000.

(2) The dwelling lot size refers to the square footage of the lot, including all connecting land that is owned or that is rented with the rental units. The dwelling lot size applies only to units that are not in apartment complexes. The lot size is classified into eight grouped categories as (1) less than or equal to 1,500, (2) 1,501 – 2,500, (3) 2,501 – 5,000, (4) 5,001 – 7,500, (5) 7,501 – 10,000, (6) 10,001 – 12,500, and (7) 12,501 – 15,000, and (8) greater than 15,000 square footage.

(3) Housing costs (monthly) is computed as discussed in a footnote in the earlier section, and is classified into five grouped categories: (1) less than or equal to \$1,000, (2) \$1,001 – \$1,500, (3) \$1,501 – \$2,000, (4) \$2,001 – \$2,500, (5) \$2,501 – \$3,000, and (6) greater than \$3,000.

(4) Household average commute distance (miles) is the average one-way distance in miles between the home and the workplace across those individuals who work outside the home (for brevity, from here on, we will refer to this variable as household commute distance). For the consideration stage, we group the household commute distance into seven categories: (1) less than or equal to 2 miles, (2) >2 and ≤ 5 miles, (3) >5 and ≤ 10, (4) >10 and ≤ 15, (5) >15 and ≤ 20, (6) >20 and ≤ 25, and (7) >25 miles.

The top panel of Table 1 provides descriptive statistics for the grouped variables. The statistics indicate that (a) 50% of the dwelling units are less than or equal to 1500 square feet in size, while 50% are larger than 1500 square feet, (b) About 40% of the lot sizes are less than 5,000 square feet, while 60% of the lot sizes are more than 5000 square feet

[10] We use the race of the householder to represent household race because there were only about 4% of the households in which there was a difference between the race of the householder and the race of any other member in the household. That is, there are very few mixed-race households in our sample.

(but note again that these lot sizes correspond only to units that are not apartment complexes), (c) a little less than half of the dwelling units have a monthly housing cost of \$1,500, while a little more than half have a monthly housing cost of more than \$1,500, (d) about one-half of the household commute distances are 10 miles or shorter.

3.2.2 Count outcomes

(1) The number of bedrooms refers to the count of rooms in the dwelling unit used mainly for sleeping, and rooms reserved for sleeping (e.g., guest rooms), with the qualification that dwelling units with only one room (such as a studio or efficiency apartment) are designated as having zero bedrooms.

(2) The number of bathrooms refers to the count of full bathrooms in the dwelling unit for the exclusive use of the household occupying the dwelling unit, and including a flush toilet, bathtub or shower, and a sink with hot and cold piped water.

As can be noticed from the top right panel of Table 1, 40% of the dwelling units in the sample have three bedrooms, with few units with zero or more than four bedrooms. In terms of bathrooms, the vast majority of units have one or two bathrooms.

Table 1. Descriptive Statistics

Grouped Outcomes								Count Outcomes			
Square footage of the house (Sq. feet)		Lot size (Sq. feet)		Housing cost ($)		Household average commute distance (miles)		# of bedrooms		# of bathrooms	
Category	Percentage	Category	Percentage	Category	Percentage	Category	Percentage	Value	Percentage	Value	Percentage
≤1,000	27.0	≤1,500	24.0	≤1,000	26.0	≤2	13.0	0	1.0	0	0.0
1,001 – 1,500	23.0	1,501 – 2,500	7.0	1,001 – 1,500	21.0	>2 & ≤5	14.0	1	12.0	1	45.0
1,501 – 2,000	19.0	2,501 – 5,000	9.0	1,501 – 2,000	20.0	>5 & ≤10	25.0	2	22.0	2	43.0
2,001 – 2,500	15.0	5,001 – 7,500	16.0	2,001 – 2,500	13.0	>10 & ≤15	19.0	3	40.0	3	11.0
2,501 – 3,000	7.0	7,501 – 10,000	11.0	2,501 – 3,000	9.0	>15 & ≤20	13.0	4	21.0	4	1.0
>3,000	9.0	10,001 – 12,500	10.0	>3,000	11.0	>20 & ≤25	7.0	≥5	4.0		
		12,501 – 15,000	4.0			>25	9.0				
		>15,000	19.0								

Nominal/Binary Outcomes			
Housing type	**Single family detached** 69.0	**Single family attached** 3.0	**Apartment complex** 28.0
Residential location	**Suburb of the MSA*** 65.0	**Secondary central cities of the MSA*** 11.0	**Primary central city of the MSA*** 24.0
Tenure type	**Owning** 65.0	**Renting** 35.0	
Storey type	**Single-storey** 48.0	**Multi-storey** 52.0	

*MSA: Metropolitan Statistical Area

3.2.3 Nominal/Binary outcomes

(1) Housing type is distinguished in three nominal alternatives in the sample (after removing mobile homes, as indicated earlier): single-family detached units, single-family attached units, and apartment complexes, with a dominance of single-family detached units. In our analysis, due to the very small sample share of single-family attached units (only 3%), we combine single-family attached and detached units into one category and represent housing type as a binary outcome: single family unit (72%) or apartment complex (28%). In the estimation, the single family unit is considered as the base category.

(2) The density of the dwelling unit location, as indicated earlier, is identified in three nominal alternatives: suburb, secondary city, and primary central city, with about a fourth of the dwelling units located in a primary city, and 65% in a suburban location (see under "nominal/binary outcomes" in Table 1). In the estimation, the suburb location is considered as the base category (see also Brownstone and Fang, 2010 and Brownstone and Golob, 2009 for the use of density as a representation of residential location).

(3) Housing tenure type is a binary outcome, corresponding to whether the unit is rented or owned by the occupant household. Table 1 indicates that most of the households (65.0%) choose to own their dwelling units. In the estimation, the "owned" alternative is considered as the base category.

(4) The number of storeys includes finished or unfinished basements, and finished (but not unfinished) attics. For split levels and bi-levels, the highest number of floors that are physically over each other is the designated number of stories. In this paper, we classify dwelling units into single storey units or multiple storey units. In the sample, there are 28 out of a total of 404 (about 7%) apartment dwelling units that have multiple stories (based on the definition above), and so we allow for this combination possibility. Table 1 indicates that the dwelling units are about equally split between single and multiple storey units. In the estimation, the multiple storey alternative is considered as the base category.

3.3 Latent Variables

Several earlier studies have considered lifestyle and attitude-related variables in residence-related choices (the reader is referred to Van Acker *et al*., 2011, Bohte *et al*., 2009, and Bhat *et al*., 2014 for reviews of this literature), but these studies have focused only on the location dimension of residence and not on other dwelling unit attributes as we do here. Further, rather than use extraneous indicators to characterize attitudes and lifestyles, as in most earlier studies that use an integrated choice and latent variables (ICLV) approach, we use the endogeneous outcomes themselves as indicators of the latent variables in the GHDM modeling strategy. At the same time, we use earlier descriptive studies investigating (directly or indirectly) general lifestyle-related characteristics that affect residential choice decisions as the basis to select our latent variables (or constructs). For example, Fleischer (2007) reinforce the notion that "to choose a house means to choose a lifestyle" in their investigations based on qualitative data from ethnographic fieldwork. These authors suggest that (1) the desire for green space and environmental consciousness, as well as (2) the preference for privacy, spaciousness, and exclusivity, substantially impact the types of houses and the locations of the houses people live in. Indeed, they show that these lifestyle orientations, while related to economic and education-related socio-demographics, have their own signaling considerations that are relevant to housing choice. Similarly, Schwanen and Mokhtarian (2007), in their principal components analysis of 18 attitudinal statements, identify two overarching attitudinal factors—(1) pro-high density environment factor and (2) pro-suburban housing—as determinants of housing choices. Interestingly, both Flesicher and Schwanen and Mokhtarian identify very similar latent constructs affecting housing choices, with the first factor (second factor) in Flesicher's study related closely to the first factor (second factor) in Schwanen and Mokhtarian. In our study, we use latent variables without having any explicit indicators to undertake a factor analysis on (except for the housing choice outcomes themselves). Thus, we base our identification of the latent variables on the studies above. Further, to the extent that lifestyle choices typically are precursors to the attitudinal preferences used by Schwanen and Mokhtarian, we use two latent constructs that coincide with Fleischer's lifestyle designations: (1) *Green lifestyle propensity* and (2) *luxury lifestyle propensity*. The first latent variable is a measure of the overall attitude and concern toward the environment, while the second reflects a penchant for consuming more, marked by a desire for privacy, spaciousness, and exclusivity.

In the context of dwelling unit choice, we expect that households with a green lifestyle propensity will locate themselves in high density areas with better access to public transit and better facilities for bicycling and walking (see, for example, de Abreu e Silva *et al*., 2012), will have a lower household commute distance, and will shy away from single family-detached housing and large lot sizes in preference for other less "space-guzzling" forms of housing consumption in the form of apartment complexes and small lot sizes, respectively. We also expect that green lifestyle propensity will have an effect on tenure type and the number of storeys, because the general consensus in the literature (see, for example, U.S. Department of Energy, 2011) points to (a) a higher energy consumption per square foot for rented houses relative to owned houses, and (b) multi-storey homes being more energy efficient and environmentally beneficial because of less geographic footprint (leading to higher density developments and less disturbance of natural landscape) and less overall surface area for the same volume (resulting in less thermal exchange with the environment). Thus, "green" households may prefer to own rather than rent, and prefer multi-storey units to single storey units. We also hypothesize that individuals with a luxury lifestyle propensity will consume more in terms of all the attributes of the housing stock (square footage, number of bedrooms, number of bathrooms, housing cost), and will prefer to own multi-storey dwelling units of the single family unit kind. To satisfy identification considerations, we specify household commute distance and lot size as indicators solely of a green lifestyle, while we specify housing cost and the number of bedrooms as indicators solely of a luxury lifestyle.

The latent variables are modeled as a function of a suite of household characteristics in the structural equation system, in which we also allow a covariance between the two latent variables. The expectation is that there will be a negative covariance between green lifestyle propensity and luxury lifestyle propensity.

3.4 Endogenous Effects

These effects correspond to recursive influences among the endogenous outcomes (see Section 2.2). These are parts of the $\overleftrightarrow{\boldsymbol{\gamma}}$ matrix (for the grouped outcomes), the $\breve{\boldsymbol{\gamma}}$ matrix (for the count outcomes), and the $\boldsymbol{b}$ matrix (for the nominal outcomes), and represent "cleansed" effects after accommodating unobserved covariance effects through the latent variables discussed in the previous section. The final directions of the recursive effects were obtained in the current paper after an extensive testing of various model specifications, and choosing the specification that provided the best data fit in terms of the composite marginal log-likelihood value (note, however, that regardless of the presence or absence of recursive effects, the model is a joint model because of the presence of latent variables that impact the many dependent variables). Figure 1 presents the directions of the endogenous relationships. Our results indicate that, after accommodating the jointness among the dependent variables caused by the latent variables, the type of housing (single family units or apartment complexes) affects housing cost, and both of these dwelling unit attributes impact square footage choice. The three aforementioned choices influence the density of dwelling unit location, and then the housing type and density of dwelling unit location impact housing tenure (own or rent). Finally, an array of physical dwelling unit attributes (storey type, lot size, number of bedrooms, and number of bathrooms) and commute distance bring up the trailing edge of the recursivity in decision-making.

3.5 Structural Equation Model Results

Table 2 provides the results for the two latent variables, *i.e.*, green lifestyle propensity and luxury lifestyle propensity. The two latent variables represent the entire household's attitudes and preferences, to be consistent with the notion that dwelling unit choices are made at a household level.

Table 2. Estimation Results of Structural Equation

Variables	Coefficient	T-stat
Green Lifestyle Propensity		
Education status (base: fraction of adults with less than a bachelor's degree)		
Fraction of adults with bachelor's or master's degree in the household	0.755	18.593
Fraction of adults with PhD degree in the household	1.629	16.706
Race (*base: Caucasian or Asian*)		
African-American or other	0.167	5.156
Gender (base: fraction of male adults in the household)		
Fraction of female adults in the household	0.167	7.216
Luxury Lifestyle Propensity		
Household income (base: less than $30,000)		
30,000 – 50,000	1.326	4.596
50,001 – 75,000	2.525	4.700
75,001 – 100,000	3.632	4.700
100,001 – 125,000	4.754	4.717
125,001 – 200,000	5.374	4.709
200,001 and above	6.790	4.701
Correlation coefficient between 'Green Lifestyle Propensity' and 'Luxury Lifestyle Propensity' latent constructs	-0.621	-3.735

3.5.1 Green Lifestyle Propensity

A variable that has been found to impact green lifestyle propensity is education status. From the data, we categorize the highest level of education status obtained by each individual in four groups: less than a bachelor's degree, bachelor's degree, master's degree, and PhD degree. To obtain a representation of education status at the household-level, we first select out only those individuals in each household who are 25 years or older. We do so because, by that age, most individuals would have finished their bachelor's degree if they do end up getting one, thus removing the mechanical effect of age on attaining a bachelor's degree. Next, from among the group of adults who are 25 years or older, we compute the fraction of adults who have attained each of the bachelor's degree, master's degree, and the PhD degree, and use these three fractions as household-level determinants of green lifestyle propensity (with the fraction of adults who have attained less than a bachelor's degree serving as a base category). The results in Table 2 indicate that household education status has a positive impact on the household's green lifestyle propensity. That is, households with a higher fraction of more educated individuals tend to have a higher green lifestyle propensity (the coefficients on the fractions corresponding to bachelor's and master's degrees were not statistically significantly different, and so have been constrained to be the same; the implication is that there is no difference between bachelor's degree attainment and master's degree attainment in "green" lifestyle propensity). The education effect is consistent with results in the social-psychological literature (see, for example, Stern, 2000, Sundblad *et al*., 2007, and Franzen and Vogl, 2013) that suggest that individuals with a higher education are (a) able to assimilate environmental information quickly, (b) more self-aware of the negative consequences of degrading the environment (such as the resulting health-related problems and global warming), (c) more cognizant of the actions that lead to degrading the environment (such as excessive driving) and benefiting the environment (such as using non-motorized means of travel), and (d) able to better project into the future and appreciate the trajectory of alarming environmental trends, even if these trends are very slow and do not pose an imminent danger to society.

The green lifestyle propensity is also influenced by the race of the household. We did not find a statistically significant difference in green lifestyle propensity between Caucasian and Asian households, and these two race categories together formed the base category (as in the 2010 Population Census for the Seattle region, the two dominant races in the sample are Caucasian and Asian; these two races accounted for 85-90% of the population in the 2010 Census and in our sample). Table 2 indicates that African-American households as well as other race households (Hawaiian, American Indian and Alaska Native) exhibit a higher green lifestyle propensity relative to Caucasian and Asian households (there was no statistical difference in green lifestyle propensity between African-American and other race households). There has been substantial environmental-psychological literature on race effects on environmental concern and attitudes (see a recent review of the literature on race and other demographic determinants of environmental attitudes by Gifford and Nilsson, 2014). The early literature on the subject (see, for example, Taylor, 1989, Mohai, 1990, and Jones, 1998) suggested that Caucasians are more likely to be concerned about the environment relative to other races, based on Maslow's theory of the hierarchy of human needs and the generally better social/economic status of Caucasians (Maslow's theory states that humans first focus on the survival-based instinct of meeting their basic material needs, and consider higher level needs such as the need for environmental quality only after the basic needs are satisfied). However, this view has been challenged more recently, with a majority of studies in the past decade either suggesting no statistically significant race-based variations or, as Liu *et al*. (2014) suggest, "the best available evidence appears to suggest that non-Whites/Blacks tend to be more concerned for the environment than Whites are." The explanation provided for the higher concern for the environment among non-whites/blacks is that minorities have been, for a long time, disproportionately shouldering the burden of negative environmental quality. For example, for many years over the past several decades, hazardous waste facilities were placed disproportionately in people of colored communities. This may have contributed to a higher sensitivity to the negative implications of environmental stressors in general (and neighborhood environmental issues, such as air quality, in particular).[11]

[11] Our study supports this latter reasoning. Specifically, when we removed race and added income categories as explanatory variables, the effect of income was not statistically significant, suggesting that Maslow's theory does not hold for our sample. Of course, income did not turn up statistically significant when introduced along with race.

Finally, our results indicate, consistent with the findings in the literature, that households with a higher fraction of female adults (lower fraction of male adults) tend to exhibit a higher green lifestyle propensity relative to households with a lower fraction of female adults (higher fraction of male adults). Social-psychological studies attribute this gender-based effect to socialization norms and relational value differences among men and women (see, for example, Liu *et al*., 2014 and Gifford and Nilsson, 2014). In particular, from a socialization norm perspective, these earlier studies indicate that, across cultures, women tend to be cooperative and interdependent in socialization, while men tend to be competitive and independent in socialization. From a relational value perspective, women tend to value the needs of others more so than men. In combination, the result is that women typically exhibit a more caring behavior and altruism than men, which manifests itself in the form of a higher environmental concern (that is, the notion that the environment is a shared asset whose quality has to be preserved for the benefit of all through cooperative efforts and responsibility) and appropriate "green" actions to minimize environmental harm.

Consistent with the recent study by Liu *et al*. (2014), our results did not find any statistical differences in environmental concern and expression (*i.e.*, green lifestyle living) based on other demographic factors such as household income and age of individuals in the household.

3.5.2 Luxury Lifestyle Propensity

There has been some limited research suggesting that the choice of a home is effectively the choice of a lifestyle (see, for example, Fleischer, 2007). Further, there has been quite extensive research on luxury consumption and its determinants in the context of non-housing related consumer goods (see, for example, Kastanakis and Balabanis, 2014 and Nwankwo *et al*., 2014). While the issues related to luxury consumption can vary based on the specific consumer good under study, there are some common themes associated with the socio-cultural motivations for a luxury lifestyle, such as signaling wealth, power and status, privileged access to limited resources, and/or uniqueness in the consumer space (see Chevalier and Gutsatz, 2012). This is the basis for our construction of the luxury lifestyle propensity as a latent construct in the current study.

Income has been well established as a dominant socio-economic factor that affects luxury lifestyles, simply because it provides the financial ability for individuals and households to better manage the social impression they want to project by consuming the goods (in terms of both quality and quantity) that are consistent with the intended signaling (see, for example, Husic and Cicic, 2009). The results in Table 2 are consistent with this notion, and show that income has a clear and statistically significant effect on luxury lifestyle propensity. Interestingly, after accommodating income effects, no other sociodemographic variable had any statistically significant effect on luxury lifestyle propensity.

These findings do back the notion that minority groups may have become more environmentally sensitive because of disproportionately shouldering the burden of negative environmental quality.

3.5.3 Correlation

The correlation coefficient between green lifestyle propensity and luxury lifestyle propensity is statistically significant at any reasonable level of significance, with a value of –0.621 and a t-statistic of –3.73. This negative correlation is reasonable, since a green lifestyle is associated with careful and conservative consumption of resources, while a luxury lifestyle correlates with extravagant living and indulgence beyond an indispensable minimum (Wiedmann *et al*., 2007). The correlation between green lifestyle propensity and luxury lifestyle propensity is not perfectly negative because luxury lifestyle is not only about excess consumption, but also about consumption of high quality, unique products. At the same time, green lifestyle propensity may also imply some level of unique, expensive, and high quality product consumption (for example organic food consumption or solar-paneled house choices).

3.6 Measurement Equation Model Components

Tables 3 and 4 provide the results for the latent variable measurement equation model components. The dependent outcomes are arranged row-wise in the recursive sequence presented in Section 3.4 and Figure 1. The exogenous variables and the latent constructs are arranged column-wise in Table 3, while the endogenous variables that affect other endogenous variables in a recursive fashion are arranged columnwise in Table 4.

For the binary and nominal outcomes (housing type, density of residential location, tenure type, and storey type), one of the alternatives is used as the "base alternative" (see Section 3.2 for the identification of the base alternative for each dimension, which is also listed in Tables 3 and 4). The coefficients in Table 3 represent the impacts of the column exogenous variables (corresponding to the $\boldsymbol{b}$ matrix in Equation (10)) and the column latent constructs (corresponding to the $\boldsymbol{\vartheta}_{vec}$ matrix embedded in the $\boldsymbol{\varpi}$ matrix of Equation (10)) on the utilities of the non-base alternatives relative to the base alternative. A '-' corresponding to an exogenous variable or a latent construct for a non-base alternative indicates that the corresponding column variable or latent construct has no differential effect on the utilities of the base alternative and the non-base alternative. Also, there is no intuitive interpretation of the constants in these models because of the presence of count exogenous variables in the model. The constant (t-statistic) corresponding to apartment (for housing type) is –0.269 (–6.55), for secondary and primary city (for density of residential location) are –0.796 (–15.32) and –2.447 (–2.47), respectively, rented dwelling (for tenure type) is –0.329 (–8.74), and single-storey (for storey type) is 0.861 (15.43). The coefficients in Table 4 represent the recursive endogenous impacts, also embedded in the $\boldsymbol{b}$ matrix in Equation (12).

For the grouped outcomes (housing cost, square footage, lot size, and commute distance), the coefficients in Table 3 represent the effects of the column exogenous variables (corresponding to the $\tilde{\boldsymbol{\gamma}}$ matrix in Equation (5)) and the column latent constructs (corresponding to the $\tilde{\boldsymbol{d}}$ matrix in Equation (5)) on the underlying latent continuous variable that represents the logarithm of the corresponding dimension. Thus, for example, in the case of housing cost, the underlying latent variable is the continuous natural logarithm of housing cost, with the thresholds demarcating the six grouped categories (see Section 3.2) corresponding to ln(1000), ln(1500), ln(2000), ln(2500), and ln(3000).

Again, similar to a linear regression, the constants in the grouped outcome regressions do not have any tangible interpretation, because of the presence of count explanatory variables. Similarly, the variances as such do not have any tangible meaning other than to provide a sense of the level of stochasticity in the underlying latent variable conditional on the exogenous variables. The constants (t-statistic) in the equations corresponding to the natural logarithm of housing cost, square footage, lot size, and household commute distance are –0.138 (–8.80), –0.057 (–3.69), 1.707 (138.96) and 1.345 (48.93), respectively. The variances (t-statistic) corresponding to the natural logarithm of housing cost, square footage, lot size, and household average commute distance are 0.280 (30.75), 0.171 (41.81), 0.593 (43.48), and 1.119 (38.10), respectively. The coefficients in Table 4 for these grouped outcomes capture any endogenous recursive effects, also embedded as elements of the $\tilde{\boldsymbol{\gamma}}$ matrix in Equation (5).

For the count outcomes (the number of bathrooms and the number of bedrooms), the Table 3 coefficients represent the effects of the column exogenous variables on the thresholds (corresponding to elements of the $\breve{\boldsymbol{\gamma}}$ matrix in Equation (7)) and the effects of the column latent constructs on the underlying latent propensity for the count variable (corresponding to the elements of the $\breve{\boldsymbol{d}}$ matrix in Equation (8)). The constant coefficients in the $\breve{\boldsymbol{\gamma}}$ matrix do not have any substantive interpretation. For the other variables, a positive coefficient in the $\breve{\boldsymbol{\gamma}}$ vector shifts all the thresholds toward the right of the count propensity scale, which has the effect of reducing the probability of the zero count (see CPB, 2012). On the other hand, a negative coefficient shifts all the thresholds to the left of the count propensity scale, which has the effect of increasing the probability of the zero counts. The Table 4 coefficients for the count outcomes correspond to endogenous recursive effects as embedded in the $\breve{\boldsymbol{\gamma}}$ matrix within the threshold. In addition to the effects mentioned above, for each count variable, we also have the dispersion parameter θ_c and the φ_c vectors in the thresholds that provide flexibility to accommodate high or low probability masses for specific count outcomes. In our empirical analysis, the final model specifications for the two count outcomes (number of bathrooms and bedrooms) collapsed to a Poisson generating process. In particular, the θ_c parameters for these two count variables (c=1,2) became quite large in the estimations $(\theta_c \rightarrow \infty)$, and the resulting specifications could not be distinguished from corresponding Poisson-based latent variable specifications. In terms of the φ_{c} vectors, we needed three flexibility terms for the "number of bedrooms" dependent outcome ($\varphi_{1,1} = 0.210$ with a t-statistic of 7.06, $\varphi_{1,2} = 0.463$ with a t-statistic of 12.43, and $\phi_{1,3} = 1.388$ with a t-statistic of 32.13). There was no need for any flexibility terms for the number of bathrooms. The constants (t-statistics) corresponding to the number of bedrooms and bathrooms were 0.233 (8.10) and –0.525 (–17.29), respectively.

The final model specification was obtained after a comprehensive iterative process in which different functional forms for variables (such as a continuous linear form, piecewise-linear forms, logarithmic form, and discrete categories) and different ways of including demographic variables (such as presence of children and marital status of adults separately versus a single combined family structure variable) were considered. Also, model coefficients were examined with respect to their intuitive behavioral interpretations and statistical significance levels. All of these issues, along with data fit, were considered in developing the final variable specification. In the next section, we discuss the effects of exogenous variables and latent constructs on all the dependent variables (characteristics of dwelling unit), followed by the effects of endogenous effects.

Table 3. Measurement Equation Estimates

Dependent Variables	Family structure of the household (base: married couple with no children)								Number of children (base: no children)			
	Single person never married (NM) household		Single person separated/divorced (S/D) household		Nuclear family household		Other households		# of children ≤ 10 years old		# of children 11-16 years old	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Housing type (base: single family)												
Apartment	1.103	32.288	0.627	15.900	-0.500	-8.600	0.210	5.575	-	-	-	-
Housing cost (dollars)	-	-	-	-	-	-	-	-	-	-	-	-
Square footage of the dwelling unit	-	-	-	-	-	-	-	-	0.026	3.037	0.056	7.595
Residential location (base: suburb)												
Secondary city	0.371	7.166	-	-	-	-	-	-	-	-	-	-
Primary city	1.001	2.814	-	-	-	-	-	-	-	-	-	-
Tenure type (base: owned)												
Rented	0.119	3.145	0.362	8.779	-	-	0.625	14.574	-	-	-	-
Storey type (base: multi-storey)												
Single-storey	0.194	4.780	-	-	-	-	0.261	5.995	-	-	-	-
Lot size (sq. feet)	-	-	-	-	-	-	-	-	0.043	3.531	0.043	3.531
# of bedrooms	-	-	-	-	-	-	-	-	0.090	7.717	0.140	14.221
# of bathrooms	-	-	-	-	-	-	-	-	0.048	2.973	0.048	2.973
Household average commute distance (miles)	-	-	-	-	-	-	-	-	-	-	-	-

Table 3 (Cont.) Measurement Equation Estimates

Dependent Variables	Number of adults (17 and above)		Presence of seniors (65 and above) (Yes=1, No=0)		Presence of a disabled person (Yes=1, No=0)		Number of full time workers		Number of workers with the option to work from home	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Housing type (base: single family)										
Apartment	-	-	-	-	-	-	-0.118	-6.018	-0.132	-4.485
Housing cost (dollars)	-	-	-	-	-	-	-	-	-	-
Square footage of the dwelling unit	0.059	12.251	0.101	6.508	-	-	-	-	0.021	2.084
Residential location (base: suburb)										
Secondary city	-	-	0.139	2.255	-0.112	-2.172	-	-	-	-
Primary city	-	-	0.626	2.507	-	-	0.247	2.378	0.121	4.130
Tenure type (base: owned)										
Rented	-	-	-0.815	-14.960	0.163	3.632	-	-	-	-
Storey type (base: multi-storey)										
Single-storey	-	-	-0.201	-3.776	-	-	-	-	-	-
Lot size (sq. feet)	-	-	0.120	3.566	-	-	-	-	0.051	2.514
# of bedrooms	0.179	33.872	-	-	-	-	-	-	-	-
# of bathrooms	-	-	-	-	-	-	-	-	-	-
Household average commute distance (miles)	-	-	-	-	-	-	-	-	-	-

Table 3 (Cont.) Measurement Equation Estimates

Dependent Variables	Immigration status (base: native household)		Household monthly income (base: $30,000 or less)						Latent Constructs			
	Immigrant and combined households		30,000 – 50,000		50,001 – 75,000		75,001 and above		Green lifestyle propensity		Luxury lifestyle propensity	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Housing type (base: single family)												
Apartment	0.415	14.892	-	-	-	-	-	-	-	-	-0.222	-4.408
Housing cost (dollars)	-	-	-	-	-	-	-	-	-	-	0.133	4.613
Square footage of the dwelling unit	-	-	-	-	-	-	-	-	-	-	0.064	4.583
Residential location (base: suburb)												
Secondary city	-	-	-	-	-	-	-	-	0.305	3.979	-0.092	-4.090
Primary city	-	-	-	-	-	-	-	-	1.736	2.679	-	-
Tenure type (base: owned)												
Rented	0.248	8.068	-	-	-	-	-	-	-0.265	-7.205	-0.213	-4.445
Storey type (base: multi-storey)												
Single-storey	-	-	-	-	-	-	-	-	-0.261	-7.274	-0.041	-3.279
Lot size (sq. feet)	-	-	-	-	-	-	-	-	-0.264	-16.742	-	-
# of bedrooms	-	-	-	-	-	-	-	-	-	-	0.110	4.440
# of bathrooms	-	-	-	-	-	-	-	-	-	-	0.138	4.358
Household average commute distance (miles)	-	-	0.184	5.581	0.381	10.716	0.462	14.011	-0.364	-12.419	-	-

Table 4. Endogenous Effects

Dependent Variables	Housing type (base: detached/ attached)		Housing cost (base: $1,000 dollars or less)									
	Apartment		1001 – 1500		1501 – 2000		2001 – 2500		2501 – 3000		3001 and above	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Housing cost	-0.410	-26.475	-	-	-	-	-	-	-	-	-	-
Square footage of the dwelling unit	-0.640	-52.861	-	-	-	-	0.131	11.124	0.182	12.856	0.330	20.277
Residential location (base: suburb)												
Secondary city	-	-	-	-	-	-	-	-	-	-	-	-
Primary city	0.443	2.416	-	-	0.302	2.166	0.653	2.379	0.653	2.379	0.653	2.379
Tenure type (base: owned)												
Rented	1.416	38.329	-	-	-	-	-	-	-	-	-	-
Storey type (base: multi-storey)												
Single-storey	2.069	41.994	-	-	-0.101	-2.937	-0.101	-2.937	-0.406	-8.414	-0.406	-8.414
Lot size (sq. feet)	-	-	-	-	-	-	-	-	-	-	0.188	6.309
# of bedrooms	-	-	0.124	8.569	0.233	16.725	0.312	19.263	0.312	19.263	0.312	19.263
# of bathrooms	-	-	-	-	-	-	0.099	3.213	0.099	3.213	0.166	4.615
Household average commute distance	-	-	-	-	-	-	-	-	-	-	-	-

Table 4 (Cont.) Endogenous Effects

Dependent Variables	Square footage of house (base: 1,000 sq. feet or less)										Residential location (base: suburb)				Tenure type (base: rented)	
	1001 – 1500		1501 – 2000		2001 – 2500		2501 – 3000		3001 and above		Primary city		Secondary city		Owned	
	Coeff	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**	**Coeff**	**T-stat**
Housing cost	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Square footage of the dwelling unit	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Residential location (base: suburb)																
Secondary city	-0.140	-2.912	-0.140	-2.912	-0.140	-2.912	-0.140	-2.912	-0.140	-2.912	-	-	-	-	-	-
Primary city	-0.847	-2.684	-0.847	-2.684	-0.847	-2.684	-1.656	-2.607	-1.656	-2.607	-	-	-	-	-	-
Tenure type (base: owned)																
Rented	-	-	-	-	-	-	-	-	-	-	0.238	6.956	0.160	3.746	-	-
Storey type (base: multi-storey)																
Single-storey	-0.384	-7.898	-1.454	-27.291	-1.454	-27.291	-2.205	-30.361	-2.205	-30.361	-0.663	-18.785	-	-	0.263	6.893
Lot size (sq. feet)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.273	14.590
# of bedrooms	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.427	38.894
# of bathrooms	0.379	12.364	0.379	12.364	0.379	12.364	0.379	12.364	0.379	12.364	-	-	-	-	-	-
Household average commute distance	-	-	-	-	-	-	-	-	-	-	-0.421	-15.430	-	-	0.230	9.103

3.6.1 Effects of Explanatory Variables

The family structure of the household is the first variable in Table 3, and is introduced in the form of a set of categories (with married couples with no children, or simply couple family, forming the base category): (a) single person (never married) household (from here on referred to as a single person NM household), (b) single person (separated/divorced) household (or single person S/D household), (c) nuclear family household (married couple with one or more children), and (d) other households (such as single parent households, presence of a widow with other adults in the household, roommates, and joint families with two or more adults). The last category of "other" households comprised less than 9.5% of all households, with very few households in any single sub-category in this broad "other" category to meaningfully differentiate between. The results in Table 3 indicate that single person NM households and single person S/D households are more likely (than couple and nuclear families) to live in rented apartments rather than owned single family homes, perhaps as a way of increasing social interaction and activity participation opportunities. This is reinforced, in particular, for single person NM households who are more likely to live in high density secondary and primary cities rather than in the suburbs. Our finding is consistent with that of Boehm and Schlottmann (2014) who note that "In particular, single males, single females, and those homeowners recently experiencing a change in marital status (*e.g.*, divorce, death of a spouse, *etc.*) would be more likely to transition back to rental tenure." Single person NM households also appear to prefer single-storey units to multi-storey units. In contrast, nuclear families have the highest preference of all family types to live in a single family dwelling unit, perhaps because of a perception that such dwelling units provide more security and a better quality of living space for the children (see Mulder and Lauster, 2010). Finally, in the category of family structure variables, relative to couple family households, "other" households prefer rented single-storey apartments.

Interestingly, the family structure variables discussed above that represent the make-up of a household (the relationship among individuals residing within a dwelling unit) directly influence the non-size attributes of dwelling unit choice, but do not directly influence the size-related attributes of dwelling unit choice (note, however, that because of the structural (endogenous) effect of housing type on other housing dimensions, there is an indirect influence of family structure on size-related attributes; for example, single person households prefer apartment living, which, in a recursive manner, implies a smaller square footage of the dwelling unit). On the other hand, the number of children and number of adults variables in Table 3 represent the effect of household size in non-single person households and directly impact the size-related attributes. In particular, and as expected, as the number of children or the number of adults increase, so does the preference for a higher square footage of the dwelling unit and the number of bedrooms in the unit, with this positive effect being stronger for older children relative to younger children and for adults relative to children (see Frenkel and Kaplan, 2014 for a similar result). But, while the number of children also has a positive impact on lot size and number of bathrooms (with this effect not being different across age groups within children), the number of adults does not have an impact on lot size and only an indirect effect on the number of bathrooms (this latter effect is because of a structural or endogenous effect of square footage on the number of bathrooms, as discussed in the next section). The implication is that lot size and number of bathrooms are much more tied with the number of children than with the number of adults. This is reasonable, because dwelling units with a large lot size provide outdoor play space for children, while more bathrooms may be preferred to accommodate the biological needs of children. Another explanation related to the effect of children on lot size and number of bathrooms (as well on square footage and number of bedrooms) may be derived from Mulder and Lauster's (2010) dramaturgical framework of housing and motherhood. In this framework, Mulder and Lauster propose that housing serves as both staging matter used by actors (the main actor here being the mother, along with other close family members) to set up the stage performance of motherhood/proper family life presented to the wider social world, as well as staging matter that partitions "the performance space within the family into adequate frontstage and backstage performance space." In particular, Mulder and Lauster test and show that a higher number of bathrooms and a roomy dwelling unit are representative stage elements to set up as well partition performance space to present what society has come to expect and appreciate as a good performance of privileged motherhood. Mulder and Lauster (2010) also discuss this two-way interaction between housing and motherhood, and conclude that "family is as much a context for understanding housing needs and residential outcomes as housing is a context for understanding family events." Ström (2010) similarly finds a strong inter-relationship between dwelling unit size attributes and childbearing.

Households with senior members (65 years or over) seem to favor dwelling units with large square footage located in primary and secondary cities (rather than located in the suburbs), and are predisposed to owning multi-storey units with large lot sizes. The preference for large square footage and large lot sizes may be attributed to older members spending more of their time within their dwelling units and their lots (see Andersson and Abramsson, 2012), and so desiring to increase that activity space. The preference for locating in primary and secondary cities (rather than in the suburbs) is presumably at least a partial reflection of the "empty-nesters flocking back to the downtown areas" effect that provides better accessibility to activity centers such as parks, supermarkets, entertainment centers, restaurants, and hospitals (see, for example, Litman, 2009). The preference for residing in owned multi-storey units among households with older individuals has also been found in Carter (2011).

Households that have an individual with some form of physical disability are more likely to reside in either the primary city or the suburbs (relative to the moderate density locations in the secondary city), and rent rather than own. The higher likelihood to reside in the primary city for such households may be attributed to the need for easier access to destinations and opportunities (including medical care). A similar result was found in Bhat *et al*. (2014) for households with individuals with a prolonged medical condition. The higher likelihood to rent rather than own among households with disabled individuals may reflect a disadvantage such households continue to have in the owned housing market, despite the institution of policies such as the 1990 Americans with Disabilities Act (ADA) and the 1991 Fair Housing Act Accessibility Guidelines.

The number of full time workers and the number of workers with the option to work from home both increase the propensity to reside in single family dwelling units in the primary city. The preference to live in the primary city may be a reflection of the benefits of knowledge spillovers (through formal and informal personal interactions) that occur in dense urban regions, and that provide and allow workers to retain (and enhance) their competitive edge in the market place (see Autant-Bernard and LeSage, 2011). As indicated by Boterman and Sleutjes (2014), this is true even for workers who have the option to work from home because "E-mail and mobile telecommunication complement, rather than replace, face-to-face contacts." Additionally, the results reveal that the number of workers with the option to work from home has a positive impact on dwelling unit square footage and lot size. These effects are intuitive because of the need for additional office space. Overall, it appears that households with many workers (especially those working from home) prefer single family, large-sized units in the primary city that affords them social interaction opportunities and high accessibility (see also Kunzmann, 2009, Yigitcanlar *et al*., 2008, Yigitcanlar, 2010, and Frenkel *et al*., 2013 for similar results). Interestingly, while primary cities are typically associated on the supply side with rental apartment units, in the last decade, there has been a dramatic increase in the supply of the number of single family homes in dense primary cities. In particular, between 2000 and 2010, a total of over 4.5 million single family units were newly built in the large metropolitan areas, while the corresponding figure for apartment units was only 590,000 (see Cox, 2011). For the Seattle region, the distribution of single family versus apartment units in the primary city was about 60% to 40% according to the 2010 American Community Survey, a distribution also reflected in the sample used in the current paper. This supply of single family units should facilitate the fulfillment of the demand for such units in the primary city (of course, it is still true that more single family units are located in the suburbs; for the Seattle region, the overall distribution of single family units in the primary city, secondary city, and suburb locations is 20%, 10%, and 70%, respectively).

The effect of immigrant status indicates that immigrant households and combination households are more likely to reside in rented apartment complexes than non-immigrant households (there was no difference between immigrant and combination households for each of the housing type and tenure type effects). This is not inconsistent with the notion of the "American dream" being tied to owning a single-family home (see, for example, Carter, 2011). Also, immigrants may prefer rented apartment living because of lower access to mainstream sources of financing because of the absence of a credit history, or because of past living behaviors with multiple families living in a single housing complex in their originating countries, or due to self-preservation and identity considerations that encourage them to share multi-unit quarters with other people of their own culture to feel a sense of closeness/belonging, as well as retaining a sense of security and place in a foreign land. Interestingly, our results show that, while housing unit and location decisions are made jointly (due to the presence of latent variables affecting multiple housing dimensions), the structural effect of immigrant status on the density of residential location manifests itself indirectly through housing type choice rather than a direct structural effect (note that there is no effect of immigrant/combination households on the density of residential location in Table 3). Many earlier studies, on the other hand, consider the immigrant status effect directly on the density of residential location choice (see, for example, Wilson and Singer, 2011, Logan *et al*., 2011, and Bhat *et al*., 2014). Overall, our results suggest that the reason for the preference for high density living among immigrants may be more tied to factors such as the absence of a credit history and self-preservation considerations (that is, preference for living in apartment units, which are more abundant in supply in high density areas) than a direct preference for high density living (though we have to emphasize that this is speaking to the strength of the direction of relationship among the many housing unit choices, and not suggesting a hierarchical framework of housing unit-related choices, given that the housing unit choices are being made jointly).

Finally, the household commute distance increases with an increase in household income, a result found in many earlier studies too (see, for, example, Rashidi *et al*., 2012, Paleti *et al*., 2013, and Surprenant-Legault *et al*., 2013). An explanation for this effect is that individuals who are able to command higher wages in the market are willing to commute long distances to improve their market potential, while commuting long distances is simply not worthwhile for those who command relatively low wages in the market place (Madden, 1985). Another explanation is related to occupation segregation; that is, occupations with relatively low wages, such as clerical, sales, and services positions, tend to be more localized geographically, engendering a positive relationship between income and commute distance (see Clark and Wang, 2005). Note also that household income has an indirect impact on almost all dimensions of the dwelling, through its impact on the luxury propensity latent construct. In the next section, we briefly discuss these latent construct effects.

3.6.2 Latent Construct Effects

The latent construct effects are quite consistent with our hypotheses in Section 3.3. Specifically, according to the results in Table 3, households with a high green lifestyle propensity generally have a preference for owned, multiple-storey units on a compact lot in higher density neighborhoods with a short commute distance, while households with a high luxury lifestyle propensity are inclined toward owning high-priced large multi-storey single family homes in sparsely dense (suburb) or dense (primary city) neighborhoods (and not in moderately dense neighborhoods in the secondary city).

3.6.3 Endogenous Effects

As discussed in Section 3.4 and presented in Figure 2, there are a number of endogenous effects in the choice of dwelling characteristics. The reader will note that these are endogenous effects because error correlations across the many dimensions (engendered by the latent effects) are explicitly accommodated for. In contrast, many earlier studies that focus on limited dimensions of the dwelling unit assume one or more unmodeled dimensions to be exogenous. For example, it is quite typical to consider housing costs and/or density of residential neighborhood (for instance, urban versus suburban neighborhood) as an explanatory variable in tenure/housing type and housing size decisions (see, for example, Barrios-García and Rodríguez-Hernández, 2008 and Carter, 2011). But, as our results indicate, households who are luxury-oriented because of various unobserved attributes (such as a higher need to signal wealth or power than their observationally equivalent peer group) are likely to be drawn toward owning high-priced single family homes in non-moderately dense (that is, either high density or low density) neighborhoods. After accounting for the resulting correlation effects among housing cost, density of residential neighborhood, tenure, and housing type choices, our results do not show any remaining structural (causal) effect of housing cost on housing tenure decisions (see Table 4). The results also reveal that the housing type decision affects housing cost decisions, and both these decisions affect residential location and tenure decisions in a recursive fashion (though the model is a joint model because of the explicit incorporation of the stochastic latent constructs). More generally, the takeaway is that ignoring the package nature of the many housing dimensions can lead to incorrect inferences regarding structural (causal) effects among the dimensions. Table 4 provides the coefficient estimates of these endogenous effects, which are self-explanatory (and have already been diagrammatically illustrated in Figure 2, and discussed in Section 3.4).

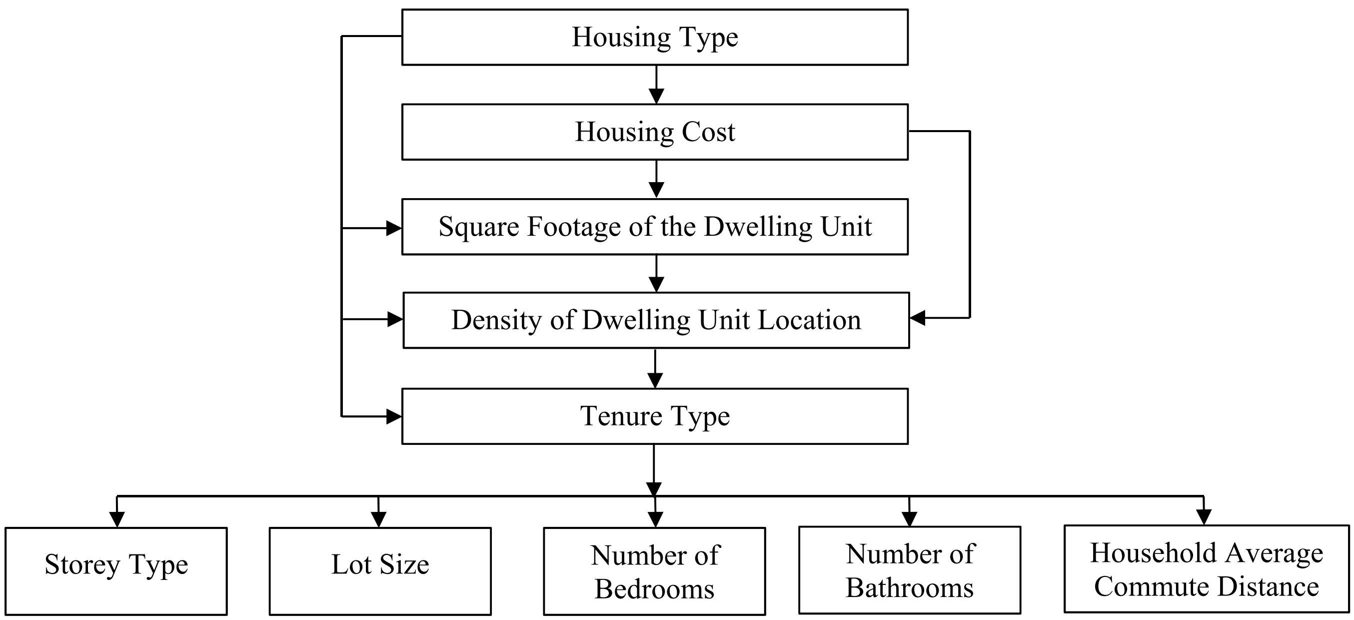

Figure 2. Recursivity in Implied Structural Effects

3.6.4 Variance-Covariance Parameters

In addition to the correlations across dimensions engendered by the latent constructs, we also allowed a general covariance structure for the utility differences (taken with respect to the base alternative of the suburb) of the three alternatives for the density of residential neighborhood. But the resulting 2×2 covariance matrix provided estimates that could not be statistically distinguished from a matrix with the value of 1.0 on the diagonal and the value of 0.5 on the off-diagonal. Thus, we fixed the 2×2 covariance matrix with 1.0 on the diagonals and 0.5 on the off-diagonal. This is equivalent, of course, to an IID error structure for the original three alternatives with a variance of 0.5 for each alternative. That is, after accommodating for the error heteroskedasticity and correlation in the utilities of the location alternatives due to the stochastic latent constructs, there is no remaining heteroskedasticity and correlation in the location utilities.

3.6.5 Data Fit

The performance of the GHDM structure used here may be compared to the one that does not consider latent constructs, maintaining the same specification otherwise as in the final GHDM model. However, this would not constitute a good "strawman" specification to test the GHDM model with. Instead, we estimate a model including the determinants of the latent constructs as explanatory variables, while maintaining the recursivity in the dimensions as obtained from our final GHDM model. Essentially, this is an independent model in that the error term correlations across the dimensions are ignored, but the best specification of the explanatory variables (including those used in the GHDM model in the structural equation system to explain the latent constructs) is considered to explain the dwelling unit dimensions. We will refer to this as the independent heterogeneous data model (or IHDM model). The GHDM and the IHDM models are not nested, but they may be compared using the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005). The CLIC takes the following form:

$$\log L^{*}_{CML}(\hat{\boldsymbol{\theta}}) = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}})^{-1}\right] \tag{12}$$

The model that provides a higher value of CLIC is preferred. Another way to examine the performance of the two models is to compute the equivalent GHDM predictive household-level likelihood value and computing the log-likelihood value across all households at convergence $\mathcal{L}(\hat{\boldsymbol{\theta}})$. The corresponding IHDM predictive log-likelihood value may also be computed. Then, one can compute the adjusted likelihood ratio index of each model with respect to the log-likelihood with only the constants for each dimension in the IHDM model:

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}) - M}{\mathcal{L}(c)}, \tag{13}$$

where $\mathcal{L}(\hat{\boldsymbol{\theta}})$ and $\mathcal{L}(c)$ are the log-likelihood functions at convergence and at constants, respectively, and M is the number of parameters (not including the constant(s) for each dimension) estimated in the model. To test the performance of the two non-nested models (*i.e.* the GHDM and IHDM models) statistically, the non-nested adjusted likelihood ratio test may be used (see Ben-Akiva and Lerman, 1985, page 172). This test determines if the adjusted likelihood ratio indices of two non-nested models are significantly different. In particular, if the difference in the indices is $(\bar{\rho}_2^2 - \bar{\rho}_1^2) = \tau$, then the probability that this difference could have occurred by chance is no larger than $\Phi\left\{-[-2\tau\mathcal{L}(c) + (M_2 - M_1)]^{0.5}\right\}$ in the asymptotic limit. A small value of the probability of chance occurrence indicates that the difference is statistically significant and that the model with the higher value of adjusted likelihood ratio index is to be preferred.

We also evaluate the data fit of the two models intuitively and informally at both the disaggregate and aggregate levels. To do so, we consider four specific dimensions (housing type, density of residential location, storey type, and tenure type) and compute marginal multivariate predictions for these four dwelling unit attributes jointly. At the <u>disaggregate level</u>, for the GHDM model, we estimate the probability of the observed marginal multivariate outcome for each household using Equation (4) in the supplementary note (essentially this entails a four dimensional rectangular integral computation after integrating out other housing dimensions), and compute an average (across households) probability of correct prediction at this four-variate level. Similar disaggregate measures are computed for the IHDM model. At the aggregate level, we design a heuristic diagnostic check of model fit by computing the predicted aggregate share of individuals for combinations of the four dimensions identified earlier that have an aggregate share of over 0.05. The predicted shares at the multivariate outcome as well as the marginal outcome levels from the GHDM and the IHDM models are compared to the actual shares, and the absolute percentage error (APE) statistic is computed.

The results of our disaggregate data fit evaluation are provided in Table 5a. The CLIC values in Table 5 clearly favor the GHDM model over the IHDM model. The same result is obtained when comparing the predictive likelihood values, the predictive adjusted likelihood ratio indices, and computing the non-nested likelihood ratio statistic; the probability that the adjusted likelihood ratio index difference between the GHDM and the IHDM models could have occurred by chance is literally zero. The average probability of correct prediction at the four-variate level of housing type, residential location density, storey type, and tenure type is 0.2235 for the GHDM model, and 0.2145 for the IHDM model. At the aggregate level, the five combinations at the four-variate level with a share of over 0.05 are identified in Table 5b. For each of these combinations, the shares predicted by the GHDM model are either superior to the IHDM model or about the same as the IHDM model. Across all five combinations, the average APE is 6.93% for the GHDM model compared to 21.17% for the IHDM. The aggregate fit measures in Table 5b reinforce the disaggregate level results in Table 5a.

In summary, the results clearly show that the GHDM model proposed here outperforms the IHDM model in the disaggregate level and aggregate level comparisons. It is also interesting to compare the model with a model that ignores the first stage search process completely, and assumes that households choose from all possible dwelling units. Such a model essentially assumes that, regardless of household demographics and attitudes, the probability of choice of each alternative along each dimension in the first stage process modeled here is equal across alternatives in that dimension. Equivalently, each multivariate combination has the same probability of being included into the choice set of the household. The log-likelihood (LL) value of the naïve model is –20,161.00, which is much worse than that of the GHDM LL value of –10,726.72, and of the IHDM LL value of –11,438.43. The average probability of correct prediction of such a naïve model at the four-variate level identified above is 0.0420, which is much worse than the corresponding GHDM value (of 0.2235) or the IHDM value (of 0.2145). Overall, the GHDM model, which has the best data fit, is the one that provides the most accurate set of alternatives for the subsequent second stage fine-level determination of housing choices at the parcel level.

Table 5a. Disaggregate Data Fit Measures

Summary Statistics	Model	
	GHDM	**IHDM**
Composite Marginal log-likelihood value at convergence	-99009.59	-99640.90
Composite Likelihood Information Criterion (CLIC)	-99257.24	-99956.41
Log-likelihood at constants	-11965.33	
Predictive log-likelihood at convergence	-10726.62	-11438.43
Number of parameters	116	146
Number of observations	1421	1421
Predictive adjusted likelihood ratio index	0.094	0.032
Non-nested adjusted likelihood ratio test between the GHDM and IHDM	$\Phi[-36.09] \ll 0.0001$	

Table 5b. Aggregate Data Fit Measures

Alternative	Observed Share	GHDM		IHDM	
		Predicted Share	**APE***	**Predicted Share**	**APE***
Owned, multi-storey, single-family unit located in primary city	0.0950	0.0912	4.00	0.0865	8.95
Owned, single-storey, single-family unit located in suburb	0.1281	0.1137	11.24	0.0888	30.68
Owned, multi-storey, single-family unit located in suburb	0.2885	0.2658	7.87	0.2720	5.72
Rented, single-storey, apartment unit located in primary city	0.0690	0.0680	1.45	0.0357	48.26
Rented, single-storey, apartment unit located in suburb	0.1119	0.1006	10.10	0.0982	12.24
Average across all combinations			**6.93**		**21.17**

*APE: Absolute Percentage Error

Chapter 4. Conclusions

A majority of residential location choice models are based on traffic analysis zones or similar aggregate spatial regions as the analysis unit. However, because of several limitations of such zone-based models, some recent studies have started using parcels as the analysis unit. However, like the zone-based models, parcel-based models also do not consider dwelling unit attribute choices that are made jointly with the physical location of the residence. Another unit of analysis is that of the dwelling unit. The advantage of this analysis unit is that it can accommodate dwelling unit attributes, but a problem is that the universal choice set explodes in size very quickly (particularly with a parcel-based spatial resolution). The typical approach to handle such a situation has been to use a sampling approach to sample a subset of the universal choice set of alternatives, and use correction terms that can provide consistent and asymptotically normal estimates for most discrete choice models. But this works only if households do choose from the universal choice set, which is not the case. It fundamentally ignores the presence of any search behavior process heuristics and pruning tactics.

In the current study, we adopt a dwelling unit level of analysis and consider a probabilistic choice set generation approach for residential choice modeling. Our representation of the search process is consistent with a combination of supply constraint-based, area-based, or anchor points-based search strategies. The study accommodates the fact that housing choices involve both characteristics of the dwelling unit and its location, while also mimicking the search process that underlies housing decisions. A complete range of dwelling unit choices are considered in the first consideration stage, including tenure type, housing type, number of bedrooms, number of bathrooms, number of storeys (one or two), square footage of the house, lot size, housing costs, density of residential neighborhood, and household commute distance. We exploit the idea that the final dwelling unit choice provides an indication of the broader preferences at the first consideration stage, and focus on this consideration stage in this paper. The many housing choices associated with the consideration stage are estimated by introducing latent psychological constructs that influence multiple housing choice outcomes in a measurement equation, engendering a parsimonious covariance structure across the many outcomes because the latent constructs themselves are specified in a structural system to be a function of exogenous variables and correlated random error terms. The model system is estimated on a Seattle subsample drawn from the 2009 American Housing Survey.

Several important conclusions may be drawn from the estimation results. First, housing choices are determined by many household socio-demographic variables, including family structure, number of children, number of adults, presence of senior adults and physically challenged individuals, number of workers, immigrant status, and household income. Second, housing choices are clearly made jointly, unlike the typical study that separates location considerations from dwelling unit characteristics. That is, residential models that consider only physical location-related attributes miss out on important behavioral elements that drive housing choices. Further, our study shows that housing studies that focus on only a limited set of housing dimensions, while considering other housing dimensions as being exogenous, run the risk of endogeneity bias because of the package nature of housing choices. Third, the gamut of housing choices are inter-related because of both common observed variables directly impacting the many dimensions of choice, as well as because of the effects of stochastic latent variables associated with green lifestyle propensity and luxury lifestyle propensity. Because these latent constructs are themselves a function of socio-demographics such as education level, income, race, and gender, the many housing choices are indirectly a function of these socio-demographic variables too. Also, the presence of unobserved factors impacting the latent constructs immediately implies that the many housing choices are correlated because of the unobserved factors. In addition, after controlling for these unobserved correlation effects, our results show a recursive pattern of endogenous effects within the housing choices. Fourth, the two latent variables provide important insights into housing choices, and reinforce the notion that dwelling unit choice involves an affective dimension and constitutes a lifestyle choice. Fifth, our results clearly show the superior data fit of the proposed GHDM model relative to a model that ignores the package nature of the housing decisions. Importantly, the proposed model is vastly superior to the commonly used residential location model that ignores all dwelling type attributes, and that assumes that households choose from all possible parcels (or zones).

In conclusion, this paper underscores the need to examine all aspects (locational and non-locational) of dwelling unit choice jointly, which, in turn, indicates a need to collect dwelling unit attributes in activity-travel surveys. One important extension of the current paper is to include the mobility decision when modeling dwelling unit attributes. While doing so is methodologically straightforward in our proposed analytic framework, the data used in the empirical analysis of this paper did not have adequate information on mobility decisions.

References

Alonso, W., 1960. A theory of the urban land market. *Papers in Regional Science*, 6(1), 149-157.

Andersson, E., Abramsson, M., 2012. Changing residential mobility rates of older people in Sweden. *Ageing and Society*, 32(06), 963-982.

Autant-Bernard, C., LeSage, J.P., 2011. Quantifying knowledge spillovers using spatial econometric models. *Journal of Regional Science*, 51(3), 471-496.

Barrios García, J.A., Rodríguez Hernández, J.E., 2008. Housing demand in Spain according to dwelling type: Microeconometric evidence. *Regional Science and Urban Economics*, 38(4), 363-377.

Basar, G., Bhat, C.R., 2004. A parameterized consideration set model for airport choice: An application to the San Francisco Bay area. *Transportation Research Part B*, 38(10), 889-904.

Bayer, P., McMillan, R., Rueben, K., 2005. Residential segregation in general equilibrium. NBER Working Paper No. 11095, National Bureau of Economic Research.

Bell, C.A., 2009. All choices created equal? The role of choice sets in the selection of schools. *Peabody Journal of Education*, 84(2), 191-208.

Bell, M.G., 2007. Mixed routing strategies for hazardous materials: Decision-making under complete uncertainty. *International Journal of Sustainable Transportation*, 1(2), 133-142.

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

Bhat, C.R., 1994. Imputing a continuous income variable from grouped and missing income observations. *Economics Letters*, 46(4), 311-319.

Bhat, C.R., 1999. An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey. *Transportation Research Part B*, 33(7), 495-510.

Bhat, C.R., 2000. A multi-level cross-classified model for discrete response variables. *Transportation Research Part B*, 34(7), 567-582.

Bhat, C.R., 2014. A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Bhat, C.R., Dubey, S.K., 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B*, 67, 68-85.

Bhat, C.R., Guo, J.Y., 2004. A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research Part B*, 38(2), 147-168.

Bhat, C.R., 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.

Bhat, C.R, Koppelman, F.S., 1993. A conceptual framework of individual activity program generation. *Transportation Research Part A*, 27(6), 433-446.

Bhat, C.R., Zhao, H., 2002. The spatial analysis of activity stop generation. *Transportation Research Part B*, 36(6), 557-575.

Bhat, C.R., Paleti, R., Castro, M., 2015. A new utility-consistent econometric approach to multivariate count data modeling. *Journal of Applied Econometrics*, forthcoming.

Bhat, C.R., Astroza, S., Sidharthan, R., Jobair Bin Alam, M., Khushefati, W.H., 2014. A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. *Transportation Research Part B*, 68, 31-51.

Bierlaire, M., Hurtubia, R., Flötteröd, G., 2010. Analysis of implicit choice set generation using the constrained multinomial logit. *Transportation Research Record: Journal of the Transportation Research Board*, 2175, 92-97.

Boehm, T.P., Schlottmann, A.M., 2014. The dynamics of housing tenure choice: Lessons from Germany and the United States. *Journal of Housing Economics*, 25, 1-19.

Böheim, R., Taylor, M., 1999. Residential mobility, housing tenure and the labour market in Britain. Institute for Social and Economic Research, University of Essex.

Bohte, W., Maat, K., van Wee, B., 2009. Measuring attitudes in research on residential self-selection and travel behavior: A review of theories and empirical research. *Transport Reviews*, 29(3), 325-357.

Botti, S., Hsee, C.K., 2010. Dazed and confused by choice: How the temporal costs of choice freedom lead to undesirable outcomes. *Organizational Behavior and Human Decision Processes*, 112(2), 161-171.

Boterman, W.R., Sleutjes, B., 2014. Stated residential preferences of higher educated workers in Amsterdam and Eindhoven. HELP Report, 2.

Bovy, P.H., 2009. On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews*, 29(1), 43-68.

Brownstone, D., Fang, H.A., 2010. A vehicle ownership and utilization choice model with endogenous residential density. *University of California Transportation Center*.

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

Carmon, Z., Wertenbroch, K., Zeelenberg, M., 2003. Option attachment: When deliberating makes choosing feel like losing. *Journal of Consumer Research*, 30(1), 15-29.

Carter, S., 2011. Housing tenure choice and the dual income household. *Journal of Housing Economics*, 20(3), 159-170.

Cascetta, E. and Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C*, 9, 249-263.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B*, 46(1), 253-272.

Chen, J., Jin, M., 2014. Income elasticity of housing demand in China: micro-data evidence from Shanghai. *Journal of Contemporary China*, 23(85), 68-84.

Chen, C., Gong, H., Paaswell, R., 2008. Role of the built environment on mode choice decisions: additional evidence on the impact of density. *Transportation*, 35(3), 285-299.

Chevalier, M., Gutsatz, M., 2012. Luxury retail management: How the world's top brands provide quality product and service support. John Wiley & Sons.

Clark, W.A.V., Wang, W.F.W., 2005. Job access and commute penalties: Balancing work and residence in Los Angeles. *Urban Geography*, 26(7), 610-626.

Cho, C.J., 1997. Joint choice of tenure and dwelling type: a multinomial logit analysis for the city of Chongju. *Urban Studies*, 34(9), 1459-1473.

Coulombel, N., 2010. Residential choice and household behavior: State of the art. SustainCity Working Paper, 2.2 a, ENS Cachan.

Cox, W., 2011. More Americans move to detached houses. Accessed on Sep 16, 2014 http://www.newgeography.com/content/002506-more-americans-move-detached-houses.

Davidoff, T., 2006. Labor income, housing prices, and homeownership. *Journal of Urban Economics*, 59(2), 209-235.

de Abreu e Silva, J., Morency, C., Goulias, K.G., 2012. Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. *Transportation Research Part A* 46(8), 1252-1264.

de Palma, A., Picard, N., Waddell, P., 2007. Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics*, 62(2), 204-230.

Dubin, J.A., McFadden, D.L., 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 52(2), 345-362.

Elgar, I., Farooq, B., Miller, E.J., 2009. Modeling location decisions of office firms. *Transportation Research Record: Journal of the Transportation Research Board*, 2133, 56-63.

Eliasson, J., 2010. The influence of accessibility on residential location *Residential Location Choice* (pp. 137-164): Springer.

Ellickson, B., 1981. An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics*, 9(1), 56-79.

Elrod, T., Johnson, R.D., White, J, 2004. A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Processes*, 95(1), 1-19.

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

Ermisch, J.F., Findlay, J., Gibb, K., 1996. The price elasticity of housing demand in Britain: issues of sample selection. *Journal of Housing Economics*, 5(1), 64-86.

Flavin, M., Nakagawa, S., 2008. A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence. *The American Economic Review*, 98(1), 474-495.

Fleischer, F., 2007. To choose a house means to choose a lifestyle. The consumption of housing and class-structuration in urban China. *City & Society*, 19(2), 287-311.

Franzen, A., Vogl, D., 2013. Two decades of measuring environmental attitudes: A comparative analysis of 33 countries. *Global Environmental Change*, 23(5), 1001-1008.

Frenkel, A., Kaplan, S., 2014. The joint choice of tenure, dwelling type, size and location: the effect of home-oriented versus culture-oriented lifestyle. *Letters in Spatial and Resource Sciences*, 1-19.

Frenkel, A., Bendit, E., Kaplan, S., 2013. The linkage between the lifestyle of knowledge-workers and their intra-metropolitan residential choice: a clustering approach based on self-organizing maps. *Comput Environ Urban* 39, 151-16.

Gifford, R., Nilsson, A., 2014. Personal and social factors that influence pro-environmental concern and behaviour: A review. *International Journal of Psychology*, 49(3), 141-157.

Goodman, A.C., 2002. Estimating equilibrium housing demand for ‘stayers’. *Journal of Urban Economics*, 51(1), 1-24.

Guevara, C.A., 2010. Endogeneity and sampling of alternatives in spatial choice models. Ph.D. Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Guevara, C.A., Ben-Akiva, M.E., 2013a. Sampling of alternatives in logit mixture models. *Transportation Research Part B*, 58, 185-198.

Guevara, C.A., Ben-Akiva, M.E., 2013b. Sampling of alternatives in multivariate extreme value (MEV) models. *Transportation Research Part B*, 48, 31-52.

Guo, J.Y., Bhat, C.R., 2004. Modifiable areal units: problem or perception in modeling of residential location choice? *Transportation Research Record: Journal of the Transportation Research Board*, 1898(1), 138-147.

Habib, M., Miller, E., 2007. Modeling residential and spatial search behaviour: evidence from the greater Toronto area. Paper presented at the 6th Triennial Symposium on Transportation Analysis, Phuket Island, Thailand.

Habib, M.A., Miller, E.J., 2009. Reference-dependent residential location choice model within a relocation context. *Transportation Research Record: Journal of the Transportation Research Board*, 2133, 92-99.

Heckman, J.J., Sedlacek, G., 1985. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *The Journal of Political Economy*, 1077-1125.

Henderson, J.V., Ioannides, Y.M., 1983. A model of housing tenure choice. *The American Economic Review*, 73(1), 98-113.

Hilber, C.A., 2005. Neighborhood externality risk and the homeownership status of properties. *Journal of Urban Economics*, 57(2), 213-241.

Husic, M., Cicic, M., 2009. Luxury consumption factors. *Journal of Fashion Marketing and Management*, 13(2), 231-245.

Huff, J.O., 1986. Geographic regularities in residential search behavior. *Annals of the Association of American Geographers*, 76(2), 208-227.

Hurtubia, R., Bierlaire, M., 2011. Bid rent model for simultaneous determination of location and rent in land use microsimulations. Proceedings of the Swiss Transport Research Conference. May 11-13.

Jones, R.E., 1998. Black concern for the environment: Myth versus reality. *Society & Natural Resources*, 11(3), 209-228.

Kain, J.F., Quigley, J.M., 1972. Housing market discrimination, home-ownership, and savings behavior. *The American Economic Review*, 62(3), 263-277.

Kaplan, S., Bekhor, S., Shiftan, Y., 2011. Development and estimation of a semi-compensatory residential choice model based on explicit choice protocols. *The Annals of Regional Science*, 47(1), 51-80.

Kaplan, S., Bekhor, S., Shiftan, Y., 2012a. Hybrid Compensatory-Noncompensatory Choice Sets in Semicompensatory Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2322, 10-19.

Kaplan, S., Shiftan, Y., Bekhor, S., 2012b. Development and estimation of a semi-compensatory model with a flexible error structure. *Transportation Research Part B*, 46(2), 291-304.

Kastanakis, M.N., Balabanis, G., 2014. Explaining variation in conspicuous luxury consumption: An individual differences' perspective. *Journal of Business Research*.

Kortum, K., Paleti, R., Bhat, C.R., Pendyala, R.M., 2012. Joint model of residential relocation choice and underlying causal factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2303, 28-37.

Kunzmann, K.R., 2009. The strategic dimensions of knowledge industries in urban development. DISP – *The Planning Review*, 177 (2), 40-47.

Lee, L.F., Trost, R.P., 1978. Estimation of some limited dependent variable models with application to housing demand. *Journal of Econometrics*, 8(3), 357-382.

Lee, B.H., Waddell, P., 2010. Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation*, 37(4), 587-601.

Lee, B.H., Waddell, P., Wang, L., Pendyala, R.M., 2010. Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework. *Environment and Planning A*, 42(4), 913.

Lerman, S.R., 1976. Location, housing, automobile ownership, and mode to work: a joint choice model. *Transportation Research Record* (610).

Lerman, S.R., 1977. Location, housing, automobile, ownership and mode to work: a joint choice model. *Transportation Research A*, 13, 1-19

Li, M.M., 1977. A logit model of homeownership. *Econometrica: Journal of the Econometric Society*, 45(5), 1081-1097.

Litman, T., 2009. Where we want to be. Home location preferences and their implications for smart growth. Victoria Transport Policy Institute, Victoria, BC, 1-40.

Liu, X., Vedlitz, A., Shi, L., 2014. Examining the determinants of public environmental concern: Evidence from national public surveys. *Environmental Science & Policy*, 39, 77-94.

Logan, J.R., 2011. Separate and unequal: The neighborhood gap for blacks, hispanics and asians in metropolitan America. Project US2010 Report, 1-22.

Lowry, I.S., 1964. A model of metropolis. Memorandum RM-4035-RC, Rand Corporation, Santa Monica, CA.

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

Manrai, A.K., Andrews, R.L., 1998. Two-stage discrete choice models for scanner panel data: An assessment of process and assumptions. *European Journal of Operational Research*, 111(2), 193-215.

Manski, C.F., 1977. The structure of random utility models. *Theory and Decision*, 8(3), 229-254.

Martínez, F.J., 2008. Towards a land-use and transport interaction framework. *Handbook of Transport. Modelling*, 2nd edition, Elsevier, Amsterdam.

Martinez, F.J., Aguila, F., and Hurtubia, R., 2009. The constrained multinomial logit model: a semi-compensatory choice model. *Transportation Research Part B*, 43, 365–377.

Madden, J.F., 1985. Urban wage gradients: Empirical evidence. *Journal of Urban Economics*, 18(3), 291-301.

McFadden, D., 1978. Modelling the choice of residential location. Institute of Transportation Studies, University of California.

Mohai, P., 1990. Black environmentalism. *Social Science Quarterly*, 744-765.

Mulder, C.H., Lauster, N.T., 2010. Housing and family: An introduction. *Housing Studies*, 25(4), 433-440.

Newman, J.P., Bernardin Jr., V.L., 2010. Hierarchical ordering of nests in a joint mode and destination choice model. *Transportation*, 37(4), 677-688.

Nwankwo, S., Hamelin, N., Khaled, M., 2014. Consumer values, motivation and purchase intention for luxury goods. *Journal of Retailing and Consumer Services*, 21(5), 735-744.

Openshaw, S., 1978. An empirical study of some zone-design criteria. *Environment and Planning A*, 10(7), 781-794.

Pagliara, F., Wilson, A., 2010. The state-of-the-art in building residential location models *Residential location choice* (pp. 1-20): Springer.

Paleti, R., Bhat, C.R., Pendyala, R.M., 2013. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, 2382, 162-172.

Pinjari, A.R., Bhat, C.R., Hensher, D.A., 2009. Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B*, 43(7), 729-748.

Pinjari, A.R., Pendyala, R.M., Bhat, C.R., 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.

Prashker, J., Shiftan, Y., Hershkovitch-Sarusi, P., 2008. Residential choice location, gender and the commute trip to work in Tel Aviv. *Journal of Transport Geography*, 16(5), 332-341.

Quigley, J.M., 1976. Housing demand in the short run: An analysis of polytomous choice. *In Explorations in Economic Research*, 3(1), 76-102.

Quigley, J.M., 1985. Consumer choice of dwelling, neighborhood and public services. *Regional Science and Urban Economics*, 15(1), 41-63.

Rashidi, T.H., Auld, J., Mohammadian, A.K., 2012. A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection. *Transportation Research Part A*, 46(7), 1097-1107.

Rapaport, C., 1997. Housing demand and community choices. *Journal of Urban Economics*, 42(2), 243-260.

Rosen, H.S., 1979. Housing decisions and the US income tax: An econometric analysis. *Journal of Public Economics*, 11(1), 1-23.

Rouwendal, J., Meijer, E., 2001. Preferences for housing, jobs, and commuting: a mixed logit analysis. *Journal of Regional Science*, 41(3), 475-505.

Schwanen, T. Mokhtarian, P.L., 2007. Attitudes toward travel and land use and choice of residential neighborhood type: Evidence from the San Francisco bay area. *Housing Policy Debate*, 18(1), 171-207.

Sener, I.N., Pendyala, R.M., Bhat, C.R., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. *Journal of Transport Geography*, 19(2), 294-303.

Shocker, A.D., Ben-Akiva, M., Boccara, B., Nedungadi, P., 1991. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters*, 2(3), 181-197.

Shugan, S.M., 1980. The cost of thinking. *Journal of Consumer Research*, 99-111.

Sinai, T., Souleles, N., 2005. Owner-occupied housing as a hedge against rent risk. *Quarterly Journal of Economics*, 120, 763-789.

Simon, H.A., 1986. Rationality in psychology and economics. *Journal of Business*, S209-S224.

Skaburskis, A., 1999. Modelling the choice of tenure and building type. *Urban Studies*, 36(13), 2199-2215.

Stapleton, D.C., 1978. Analyzing political participation data with a MIMIC Model. *Sociological Methodology*, 52-74.

Stern, P.C., 2000. New environmental theories: toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*, 56(3), 407-424.

Ström, S., 2010. Housing and first births in Sweden, 1972–2005. *Housing Studies*, 25(4), 509-526.

Sundblad, E.L., Biel, A., Gärling, T., 2007. Cognitive and affective risk judgements related to climate change. *Journal of Environmental Psychology*, 27(2), 97-106.

Surprenant-Legault, J., Patterson, Z., El-Geneidy, A., 2013. Commuting trade-offs and distance reduction in two-worker households. *Transportation Research Part A*, 51, 12-28.

Swait, J., 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B*, 35(7), 643-666.

Swait, J., Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B*, 21(2), 91-102.

Taylor, D.E., 1989. Blacks and the environment toward an explanation of the concern and action gap between blacks and whites. *Environment and Behavior*, 21(2), 175-205.

Thill, J.-C., Horowitz, J.L., 1997. Modelling non-work destination choices with choice sets defined by travel-time constraints *Recent Developments in Spatial Analysis* (pp. 186-208): Springer.

Tversky, A., 1972. Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281-299.

U.S. Census Bureau, 2011. American Housing Survey for the United States: 2009. Current Housing Reports, Series H150/09, Washington, DC.

U.S. Department of Housing and Urban Development, 2013. Codebook for the American Housing Survey, Public Use File: 1997 - 2011. Office of Policy Development & Research, Washington, DC.

U.S. Department of Energy, 2011. 2010 Buildings Energy Data Book. Accessed on Sep 16, 2014. http://buildingsdatabook.eren.doe.gov/docs%5CDataBooks%5C2010_BEDB.pdf

Van Acker, V., Mokhtarian, P.L., Witlox, F., 2011. Going soft: on how subjective variables explain modal choices for leisure travel. *European Journal of Transport and Infrastructure Research* 11(2), 115-146.

Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. *Biometrika*, 92(3), 519-528.

Wegener, M., 2011. From macro to micro—how much micro is too much? *Transport Reviews*, 31(2), 161-177.

Wiedmann, K.P., Hennigs, N., Siebels, A., 2007. Measuring consumers' luxury value perception: A cross-cultural framework. *Academy of Marketing Science Review*, 7(7), 333-361.

Williams, H., Ortúzar, J. de D., 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research Part B*, 16(3), 167-219.

Wilson, A.G., 1970. Disaggregating elementary residential location models. *Papers in Regional Science*, 24(1), 103-126.

Wilson, J., Singer, A., 2011. *Immigrants in 2010 Metropolitan America: A decade of change*. Metropolitan Policy Program, Brookings Institution.

Yates, J., Mackay, D.F., 2006. Discrete choice modelling of urban housing markets: A critical review and an application. *Urban Studies*, 43(3), 559-581.

Yigitcanlar, T., 2010. Making space and place for the knowledge economy: knowledge-based development of Australian cities. *European Planning Studies*, 18(11), 1769-1786.

Yigitcanlar, T., O'Connor, K.O., Westerman, C., 2008. The making of knowledge cities: Melbourne's knowledge-based urban development experience. *Cities*, 25(1), 63-72.

Zolfaghari, A., Sivakumar, A., Polak, J.W., 2012. Choice set pruning in residential location choice modelling: a comparison of sampling and choice set generation approaches in greater London. *Transportation Planning and Technology*, 35(1), 87-106.

Zolfaghari, A., Sivakumar, A., Polak, J., 2013. Simplified probabilistic choice set formation models in a residential location choice context. *Journal of Choice Modelling*, 9, 3-13.