A Discrete-Continuous Choice Model with Perfect and Imperfect Substitutes

Abdul Rawoof Pinjari
Associate Professor
University of South Florida
Department of Civil and Environmental Engineering
4202 E. Fowler Ave, ENB 118, Tampa, Florida 33620
Tel: (813) 974-9671; Fax: (813) 974-2957
Email: apinjari@usf.edu

Vijayaraghavan (Vijay) Sivaraman
University of South Florida
Department of Civil and Environmental Engineering
4202 E. Fowler Ave, CUT 230, Tampa, Florida 33620
Tel: (813) 974-7165; Fax: (813) 974-2957
Email: sivaraman@cutr.usf.edu

Note: This paper is currently in review with *Transportation Research Part B.*

ABSTRACT

This study formulates a random utility maximization (RUM)-based framework to analyze discrete-continuous choices from a mix of perfect and imperfect substitutes in the choice set. The framework recognizes that some alternatives may be perfect substitutes in that not more than a single alternative is consumed from those alternatives while other alternatives may be imperfect substitutes in that potentially multiple alternatives can be consumed from those alternatives. The key to this formulation is a utility form that is linear with respect to consumption across perfectly substitutable alternatives and non-linear with respect to consumption across imperfectly substitutable alternatives. The linear utility form ensures perfect substitution while the non-linear form allows imperfect substitution. In addition to the RUM formulation, the paper presents a procedure to apply the proposed framework for forecasting purposes. Further, the formulation is extended to accommodate multiple linear budget constraints, as opposed to a single budget constraint.

As an empirical demonstration, the proposed framework is applied to develop a joint model of annual, long-distance vacation destination and mode choices to simultaneously analyze the vacation destinations that a household visits over an entire year, along with the time allocations and the travel mode to each of the visited destinations. The framework recognizes that the vacation destinations are imperfect substitutes in that a household can potentially visit multiple destinations over a year, while the travel mode alternatives to a destination are perfect substitutes in that only one primary mode is chosen to travel to a destination. In addition, the framework recognizes that households operate under both time and money budget constraints by simultaneously accommodating both the constraints in the model. The empirical model is estimated on the 1995 American Travel Survey (ATS) data, with the United States divided into 210 alternative vacation destinations.

Keywords: *discrete-continuous choice, perfect substitutes, imperfect substitutes, time and money budgets, multiple constraints, long-distance travel, destination and mode choice*

1 INTRODUCTION

Many consumer choice situations involve decisions of “what to choose” from a set of discrete goods (or alternatives) along with the decisions of “how much to consume” of the chosen good(s). Such discrete-continuous choices are rather common in consumer decisions and of interest in a variety of social sciences, including transportation, economics, and marketing. Examples include households’ choice of vacation destinations and corresponding time allocation, and grocery shoppers’ choice of brand choice and purchase quantity.

A special case of discrete-continuous choices is the single discrete-continuous (SDC) choice, where consumers choose a single discrete alternative along with the corresponding continuous quantity decision. In such situations, the choice alternatives can be viewed as *perfect substitutes* where the choice of one alternative precludes the choice of other alternatives. Another case is when the choice alternatives are *imperfect substitutes* where the choice of one alternative does not necessarily preclude the choice of other alternatives. In such situations, consumers can potentially choose multiple discrete alternatives, along with the corresponding continuous quantity choice decisions. For example, a household might choose to visit multiple vacation destinations over a given time frame. Similarly, a grocery shopper might choose a variety of brands of a product, as opposed to a single brand. Such multiple discrete-continuous (MDC) choices are being increasingly recognized and modeled in the recent literature.

A more general case of discrete-continuous choices includes both SDC and MDC choices, where consumers choose at most a single alternative from a subset of alternatives and potentially multiple alternatives from the remaining alternatives. Such situations arise when the choice set comprises a mix of both *perfect* substitutes (from which no more than a single alternative could be consumed) and *imperfect* substitutes (from which potentially multiple alternatives could be consumed). As reviewed in the next section, the vast majority of choice modeling literature has been focused on analyzing SDC choices, while there recently has been growing interest in analyzing MDC choices. Not much exists in the literature on modeling consumer behavior involving both SDC and MDC choices from choice sets that comprise a mix of perfect and imperfect substitutes. To fill this gap, the current paper formulates a unified random utility maximization (RUM) framework that can be used as a joint MDC-SDC modeling framework to analyze discrete-continuous choices from a combination of perfect and imperfect substitutable choice alternatives. In addition to the RUM formulation, the paper presents a procedure to apply the proposed framework for forecasting purposes. As importantly, the formulation can be easily extended to accommodate multiple linear budget constraints, as opposed to a single budget constraint.

As an empirical demonstration, the proposed framework is applied to develop joint models of annual, long-distance vacation destination and mode choices to simultaneously analyze the vacation destinations that a household visits over an entire year, along with the time allocation and the travel mode to each of the visited destinations. The empirical framework recognizes that the vacation destinations are *imperfect substitutes* in that a household can potentially choose to visit multiple destinations over a year, while the travel mode alternatives to a destination are *perfect substitutes* in that only one primary mode is chosen to travel to a destination. Based on this framework, three different empirical models are estimated using household-level long-distance travel data from the 1995 American Travel Survey (ATS): (1) A model that considers only the time constraint, (2) A model that considers only the money constraint, and (3) A model that considers both time and money constraints in the formulation.

The next section presents an overview of the RUM approach for modeling discrete-continuous choices. Section 3 develops the utility specification and econometric structure of the proposed framework, along with a forecasting procedure for the framework with a single constraint. In addition, this section extends the framework to consider multiple linear budget constraints. Section 4 presents and discusses the empirical models. Section 5 concludes the paper.

2 RAMDOM UTILITY MAXIMIZATION (RUM)-BASED DISCRETE-CONTINUOUS CHOICE MODELS

A variety of approaches have been used in the literature to model discrete-continuous choices (see Bhat, 2008 for a review). Among these, a particularly attractive approach is based on the classical microeconomic consumer theory of random utility maximization (RUM).[1] Specifically,

[1] Another approach is to employ a multivariate statistical system, where separate equations are used to model each component of choice – with statistical correlations between the random components of different equations. The approach has been widely used to model SDC choices by *statistically* joining a discrete choice equation with a continuous regression equation in a bivariate statistical system or in a switching regression formulation (Amemiya, 1974; Heckman, 1974; Heckman, 1979; Lee, 1983; Maddala, 1983). However, such a reduced form multivariate statistical system is typically not based on an underlying economic theory. Besides, although the approach has been used to model multiple discrete/discrete-continuous choices (Manchanda et al., 1999; Edwards and Allenby, 2003;

consumers are assumed to optimize a direct utility function $U(\mathbf{x})$ over a bundle of non-negative consumption quantities $\mathbf{x}=(x_1,...,x_k,...,x_K)$ subject to a budget constraint, as:

$$Max\ U(\mathbf{x}) \text{ such that } \mathbf{x}.\mathbf{p}=E \text{ and } x_k \geq 0\ \forall k=1,2,...,K \quad (1)$$

In the above Equation, $U(\mathbf{x})$ is a quasi-concave, increasing and continuously differentiable utility function with respect to the consumption quantity vector $\mathbf{x}$, $\mathbf{p}$ is the vector of unit prices for all goods, and E is a budget for total expenditure. Geometrically, the optimal consumption bundle for this utility maximization problem is the point in consumption space where the budget line meets (tangentially) the indifference curves corresponding to the utility function. The form of the utility function governs the characteristics of the optimal consumption bundle. If $U(\mathbf{x})$ is such that the marginal utility of consumption at the point of zero consumption (i.e., baseline marginal utility) is either infinity or indeterminate for each and every good, it results in *interior solution* (positive consumption) for all goods. In this case, the indifference curves are asymptotic to the consumption axes precluding the possibility of *corner solutions* (zero consumption). On the other hand, goods with a finite value of baseline marginal utility result in indifference curves that meet the consumption axes with a finite slope and allow *corner solutions*.

Within the context of corner solutions, again the functional form of $U(\mathbf{x})$ determines whether the formulation corresponds to a single discrete-continuous (SDC) case or a multiple discrete-continuous (MDC) case. A linear utility function with respect to consumption results in the SDC case where the choice alternatives are *perfect substitutes* (Deaton and Muellbauer, 1980; page 262) in that a utility maximizing consumer chooses only one alternative. Linear utility forms result in linear indifference curves that intersect the budget line only on a consumption axis – hence the optimal consumption bundle is a corner solution in which only one alternative is chosen. On the other hand, a non-linear utility form that allows for diminishing marginal utility and corner solutions results in the MDC case where the choice alternatives are treated as *imperfect substitutes* allowing for the possibility of "multiple discreteness".

A notable contribution to the analysis of SDC choices is the work by Hanemann (1984), who proposed a general class of utility functions that assures perfect substitutability among choice alternatives. An example of his utility forms is a bivariate function given by:

Srinivasan and Bhat, 2006; Fang, 2008), it becomes cumbersome to statistically tie all equations for discrete and continuous components into a multivariate system for more than a modest number of choice alternatives.

$U(\mathbf{x}) = u\left(x_1, \sum_{k=2}^{K} \psi_k x_k \right)$, where the first good x_1 is as an outside good[2] and the other goods k (k = 2,…, K) are inside goods, each with its own quality index ψ_k (k = 2,…, K). As can be observed, this functional form is linear with respect to the consumption of different inside goods. As such, the functional form ensures that, in addition to the outside good, only one of the inside goods (k = 2,3,…,K) is consumed – hence perfect substitutability among inside goods. To derive the demand functions implied by this utility function, Hanemann used an *indirect utility* approach which involves solving the dual of the optimization problem in (1). Specifically, the analysis starts with the specification of a conditional indirect utility function and makes use of "virtual prices" to determine the discrete choice and Roy's identity to derive the Marashallian demand functions (also see Lee and Pitt, 1986). Among other studies that adopt this approach are Chiang (1991) and Chintagunta (1993), who extended Hanemann's formulation to include the possibility of no inside goods being selected.[3]

While most literature in the area of discrete-continuous choice analysis has been geared toward SDC choices, the past decade has seen a surge of interest in analyzing MDC choices. The indirect utility approach, however, is difficult to use for analyzing MDC choices. An alternative approach, due to Hanemann (1978) and Wales and Woodland (1983), is to employ the Karush-Kuhn-Tucker (KKT) conditions of optimality to solve the utility maximization problem in (1) (also see Bockstael et al., 1987). Specifically, a randomly distributed (over the population) utility function results in randomly distributed KKT conditions, which in-turn form the basis for deriving the likelihood expressions for consumption patterns. Recent years have witnessed significant developments in the use of this KKT approach for analyzing MDC choices in the fields of environmental economics (Phaneuf et al., 2000; von Haefen et al., 2004; Phaneuf and Smith, 2005; Kuriyama et al., 2011), marketing (Kim et al., 2002; Satomura et al., 2011) and transportation (Bhat 2005; Bhat 2008). Notable among these is Bhat's (2005, 2008) multiple discrete-continuous extreme value (MDCEV) model. The MDCEV model is based on a Box-Cox transformation of a translated constant elasticity of substitution (CES) utility function that

[2] The outside good represents a composite of all goods other than the *K-1 inside* goods of interest to the analyst. The presence of the outside good helps in ensuring that the budget constraint is binding. A typical assumption is that the prices and characteristics of the goods grouped into the outside category do not influence the choice and resource allocation among the inside goods (see Deaton and Muellbauer, 1980).

[3] Also see Dubin and McFadden (1984) for a slightly different approach for SDC choice analysis, which begins with the Marshallian demand functions and utilizes Roy's identity to solve for the implied indirect utility function.

subsumes many other non-linear utility forms proposed in the literature as special cases, and enables a clear interpretation of the structural parameters. Further the specification of type-1 extreme value distributed random utility components in the MDCEV model leads to simple closed form likelihood expressions making it easy to estimate the structural parameters. Thanks to these advances, KKT-based models are being increasingly used to analyze a variety of MDC choices relevant to transportation planning, including individuals' activity participation and time-use (Bhat, 2005; Habib and Miller, 2008) and vehicle ownership and usage (Ahn et al., 2008; Bhat et al., 2009; Jaggi et al., 2011). On the methodological front, recent literature in this area has started to enhance the basic formulation in Equation (1) along different directions, including: (a) toward more flexible, non-additively separable functional forms for the utility specification (vasquez-Lavin and Hanemann, 2009; Lee et al., 2010; Bhat et al., 2013a), (b) toward greater flexibility in the specification of the constraints faced by the consumer (such as the consideration of multiple linear budget constraints; Satomura et al., 2011; Castro et al., 2012; Parizat and Sachar, 2010) and (c) toward more flexible stochastic specifications for the random utility functions (Pinjari and Bhat, 2010; Pinjari, 2011; Bhat et al., 2013b).

Despite the above advances, it is worth noting that most literature in this area has focused on analyzing choice situations that fall into either the SDC case or the MDC case. However, as discussed earlier, several choice situations can potentially include both SDC and MDC choices, as a result of choice from a combination of perfect and imperfect substitutable alternatives in the choice set. To the authors' knowledge, only a few recent studies by Bhat and colleagues (Bhat et al., 2006; Bhat et al., 2009; Eluru et al., 2010) use joint MDC-SDC modeling frameworks for such choice situations. However, the model formulations in these studies are critically hinged on the assumption that prices per unit consumption do not vary across the perfect substitutes. Specifically, their model formulations remain consistent with utility maximization only when there is no price variation across the perfect substitutes. In this paper, we formulate a unified random utility maximization framework that can accommodate perfect and imperfect substitutes regardless of the presence or absence of price variation. As discussed in the next section, the key to this approach is a utility form that is non-linear with respect to consumption across different imperfectly substitutable alternatives but linear with respect to consumption across perfectly substitutable alternatives. In addition to the RUM formulation, we present a procedure to apply the proposed framework for forecasting purposes. Furthermore, building on recent literature

(Satomura et al., 2011) we demonstrate how the formulation can be extended to accommodate multiple linear budget constraints, as opposed to a single budget constraint.

3 MODEL FORMULATIONS

Sections 3.1 through 3.7 develop the utility specification and econometric structure of the proposed framework for modeling households' annual vacation destination and mode choices and corresponding time allocations. The same framework can be employed or easily extended to model many other choice situations involving perfect and imperfect substitutes. Section 3.8 outlines a procedure/algorithm to apply the framework for prediction purposes when the model parameters are available. Section 3.9 extends the framework to consider multiple linear budget constraints.

3.1 Choice Alternatives

Let j (=*1,2,3,…,J*) be the index to represent the vacation destination alternatives available to households, l (=*1,2,…,L*) be the index to represent the travel mode alternatives, and jl be the index to represent a vacation destination and travel mode combination. Let $\mathbf{t} = (t_1, t_2, ..., t_j, ..., t_J)$ be the vector of vacation time allocations by a household to each of the destination alternatives. Considering that one can travel to a destination by any of the available modes, one can expand each element t_j of $\mathbf{t}$ as a sub-vector $(t_{j1}, t_{j2}, ..., t_{jL})$ representing the vacation time allocation to destination j reached by each of the available travel modes. In the subsequent discourse, we will assume that $L = 2$ (i.e., only two modes available for traveling to any destination). Thus, $\mathbf{t}$ can be expressed as $\mathbf{t} = \left((t_{11}, t_{12}), (t_{21}, t_{22}), ..., (t_{j1}, t_{j2}), .., (t_{J1}, t_{J2})\right)$. However, the model can be generalized for $L > 2$ in a straight forward fashion.

Over the time frame of a year, a household may choose to visit none, one, or more destinations (although not necessarily all destinations). Thus, one can expect the data to exhibit *imperfect substitution* (hence, *multiple discreteness*) among destination choice alternatives. For the chosen destinations, however, households are observed to travel by a single mode of travel regardless of the number of times they visited the destination. Thus, if a destination j is visited, the entire time t_j allocated for the destination would be allocated to only one element in the time-allocation sub-vector (t_{j1}, t_{j2}) for that destination while the time allocation to the other element would be zero, exhibiting *perfect substitution* (hence, *single discreteness*) among mode choice alternatives. In this formulation, for ease in notation but without losing generality, we assume that the first modal alternative is the chosen alternative for any chosen destination (i.e., $t_{j1} = t_j$ and $t_{j2} = 0$, if $t_j > 0$).

3.2 Utility Form

To model a household's vacation destination and mode choices over an annum, consider the following utility function (the subscript for the household is suppressed for simplicity):

$$U(\mathbf{t}, e_0) = \sum_{j=1}^{J} \gamma_j \ln \left\{ \frac{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2})}{\gamma_j} + 1 \right\} + \psi_0 \ln e_0 ; \; \psi_{jl} > 0, \psi_0 > 0, \gamma_j > 0 \;\; \forall j = 1, 2, ..., J \text{ and } l = 1, 2 \tag{2}$$

In the above utility function, the first term represents the utility accrued due to vacation. Specifically, the term $\gamma_j \ln\left\{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2}) / \gamma_j + 1\right\}$ is a sub-utility function representing the utility U_j accrued due to consuming $t_j (= t_{j1} + t_{j2})$ amount of time at a vacation destination *j*. Note that utility is assumed to be additively separable across destinations in that the total utility from vacation over the time frame of a year is the sum of utility accrued from the time spent at all the vacation destinations *j* (= *1, 2, …, J*) over the year. For each destination *j*, however, the functional from is not separable in the time allocations, t_{j1} and t_{j2}, by the different modes of travel.

The second term in Equation (2), $\psi_0 \ln e_0$ completes the utility function to form an incomplete demand system. Specifically, e_0 is a numeraire Hicksian composite outside good representing annual expenditure for all purposes other than vacation (i.e., income – annual expenditure on vacation). This outside good is assumed to be "essential" with some positive consumption by all households. The presence of this term recognizes that only a part of the available annual budget (e.g., annual income) is spent on vacation.

Households are assumed to allocate the annual income (*E*) available to them to maximize the utility in Equation (2) subject to the following constraint:

$$\sum_{j=1}^{J} \left(p_{j1} t_{j1} + p_{j2} t_{j2} \right) + e_0 = E \tag{3}$$

In the above equation, p_{jl} represents the money-price of consuming unit time of a vacation destination *j* traveled by mode *l* (*l = 1, 2*). These prices accommodate two components – (a) destination costs that do not depend on the mode of travel (e.g., lodging, dining, and entertainment costs) and (b) travel costs that depend on the mode of travel – with the latter leading to the difference between the prices by different modes of travel (*l*) for a same destination *j*. The Hicksian composite outside good is assumed to be a numeraire with unit money-price.

3.3 Interpretation of Parameters

To interpret the role of ψ_{jl} (*j =1,2,…,J; l =1,2*), consider the marginal utility of the utility function in (2) with respect to t_{jl}, $\frac{\partial U(\mathbf{t},e_0)}{\partial t_{jl}} = \frac{\psi_{jl}}{\left\{\frac{(\psi_{j1}t_{j1}+\psi_{j2}t_{j2})}{\gamma_j}+1\right\}}$. It can be observed that $\frac{\partial U(\mathbf{t},e_0)}{\partial t_{jl}} = \psi_{jl}$ when $t_{j1}=0$ and $t_{j2}=0$. Thus, ψ_{jl} can be interpreted as the marginal utility for destination-mode combination alternative *jl* at the point of zero time allocation to all destination-mode combination alternatives that share the destination *j* (in short, baseline marginal utility for destination-mode combination alternative *jl*).[4] It will be shown later in the paper that an alternative with a greater price-normalized baseline marginal utility ($i.e., \psi_{jl}/p_{jl}$) is more likely to be chosen than other alternatives sharing the same destination. In other words, a greater value of ψ_{jl}/p_{jl} implies a lower likelihood of corner solution for the alternative. In addition, ψ_{jl} also plays a role in determining the continuous consumption quantity t_{jl} for alternative *jl,* if that alternative is chosen. Specifically, between two different chosen alternatives with equal values of γ_j, the alternative with a greater value of baseline marginal utility will have a greater amount of consumption.

ψ_{jl} can be expressed as a function of the destination (*j*) attributes and the modal attributes (*l*) and interactions of these attributes with household socio-demographic characteristics, collectively represented as $\mathbf{z}_{jl}$ (more details in Section 3.6). At this point, it is reasonable to assume

---

[4] Note, however, that ψ_0 cannot be interpreted as the baseline marginal utility for the outside good, since $\left[\partial U(\mathbf{t},e_0)/\partial e_0\right]\Big|_{e_0=0}$ is not ψ_0.

that the utility form exhibits the property of weak complementarity (Maler, 1974; Herriges et al., 2004; von Haefen, 2007) that the attributes ($\mathbf{z}_{jl}$) of an alternative do not matter unless that alternative is consumed (i.e., $\frac{\partial U(\mathbf{t}, e_0)}{\partial \mathbf{z}_{jl}} = 0$ if $t_{jl} = 0$).

The primary role of γ_j, as discussed in Kim et al. (2002) and Bhat (2008), is to translate the utility function so that the indifference curves become asymptotic to the consumption axes at $(-\gamma_1, -\gamma_2, ..., -\gamma_J)$. Consequently, the indifference curves strike the consumption axes (in the positive orthant as long as $\gamma_j > 0$) with a finite slope and result in a possibility of corner solutions (i.e., zero consumption). In addition to allowing corner solutions, as explained in Bhat (2008), differences in the γ_j terms allow differential rates of satiation (i.e., diminishing marginal utility) across different destinations. Specifically, all else being the same, a destination alternative with a greater γ_j value exhibits a slower rate of satiation (hence greater amount of consumption) than those with smaller values of γ_j.

3.4 Perfect and Imperfect Substitutes

The utility form in Equation (2) is non-linear with respect to time allocations (t_{jl}, t_{kl}) across different destinations (j, k), regardless of the mode (l) used to travel to the destinations. As discussed in many previous studies (e.g., Kim et al., 2002; Bhat, 2005), the non-linear form allows diminishing marginal utility to accommodate the possibility of multiple destinations being chosen (i.e., imperfect substitution across destination choice alternatives). However, the utility form is linear with respect to time allocation across different modes of travel available for a same destination (t_{j1}, t_{j2}). The linear form ensures perfect substitution across mode choice alternatives for a destination in that a utility maximizing consumer chooses only one mode of travel to any chosen destination. To verify this, one can examine the marginal rate of substitution (MRS) between time allocations to different choice alternatives. The MRS between time allocations for destination-mode alternatives across different destinations $(t_{jl}, t_{kl};\ j \neq k)$ is given by:

$$MRS(t_{jl}, t_{kl}) = \frac{\partial U(\mathbf{t}, e_0)}{\partial t_{jl}} \Bigg/ \frac{\partial U(\mathbf{t}, e_0)}{\partial t_{kl}} = \frac{\psi_{jl}}{\left\{ \frac{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2})}{\gamma_j} + 1 \right\}} \Bigg/ \frac{\psi_{kl}}{\left\{ \frac{(\psi_{k1} t_{k1} + \psi_{k2} t_{k2})}{\gamma_k} + 1 \right\}} \qquad (4)$$

As can be observed, the MRS between destination-mode alternatives that do not share a same destination is not a constant and dependent on the time allocations to the two destination alternatives. This leads to non-linear indifference curves, implying imperfect substitution (hence multiple discreteness) across different destination choice alternatives. On the other hand, The MRS between time allocations for destination-mode alternatives that share the same destination (t_{j1}, t_{j2}) is a constant and independent of the time allocations to the two alternatives, as given below:

$$MRS(t_{j1}, t_{j2}) = \frac{\partial U(\mathbf{t}, e_0)}{\partial t_{j1}} \bigg/ \frac{\partial U(\mathbf{t}, e_0)}{\partial t_{j2}} = \psi_{j1} \big/ \psi_{j2} \quad (5)$$

Therefore, the indifference curves between such alternatives are straight lines with a slope of $-\psi_{j1} \big/ \psi_{j2}$. Such goods whose indifference curves are straight lines are called *perfect substitutes* in the microeconomic literature (see Deaton and Muellbauer, 1980; page 262) since the choice of one good by a utility maximizing consumer precludes the choice of others by the same consumer.

In summary, while the utility form allows imperfect substitution across destination choice alternatives (a household can potentially choose multiple destinations over an annum), it also ensures perfect substitution across different mode choice alternatives for a destination (only one mode is chosen to travel to a destination). This feature arises automatically from the utility form regardless of the presence or absence of price variation, eliminating the need for an explicit constraint to impose perfect substitutability among mode choice alternatives.

3.5 KKT Conditions of Optimal utility

To solve the household's utility maximization problem subject to the budget constraint, one can form the Lagrangian as below, and derive Karush-Kuhn-Tucker (KKT) conditions of optimality:

$$L = \psi_0 \ln e_0 + \sum_{j=1}^{J} \gamma_j \ln \left\{ \frac{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2})}{\gamma_j} + 1 \right\} - \lambda \left[e_0 + \sum_{j=1}^{J} (p_{j1} t_{j1} + p_{j2} t_{j2}) - E \right] \quad (6)$$

The KKT condition for the numeraire outside good is below:

$$\frac{\partial L}{\partial e_0} = 0 \text{ since } e_0 > 0 \text{, or } \lambda = \frac{\psi_0}{e_0} . \quad (7)$$

The KKT conditions of optimality for the inside goods are given next:

$$\frac{\partial L}{\partial t_{jl}} = \lambda p_{jl}, \text{ if } t_{jl} > 0; \forall j = 1, 2, ..., J; l = 1, 2$$

$$\frac{\partial L}{\partial t_{jl}} < \lambda p_{jl}, \text{ if } t_{jl} = 0; \forall j = 1, 2, ..., J; l = 1, 2$$

or,

$$\begin{aligned} &\frac{\left(\psi_{jl} / p_{jl}\right)}{\left\{\frac{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2})}{\gamma_j} + 1\right\}} = \lambda \text{ if } t_{jl} > 0; \forall j = 1, 2, ..., J; l = 1, 2 \\ &\frac{\left(\psi_{jl} / p_{jl}\right)}{\left\{\frac{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2})}{\gamma_j} + 1\right\}} < \lambda \text{ if } t_{jl} = 0; \forall j = 1, 2, ..., J; l = 1, 2 \end{aligned} \qquad (8)$$

The expression in the left side of the above two conditions represents the price-normalized marginal utility of time consumption for destination-mode combination *jl* (*l=1,2*). We now expand on these KKT conditions for two different cases: (1) for chosen destinations, and (2) for non-chosen destinations.

3.5.1 KKT conditions for chosen destinations

If a destination *j* is chosen, as discussed before, only one of the modal alternatives will be chosen to travel to that destination. Without loss of generality, assume that the first mode is chosen (i.e., $t_{j1} > 0$ and $t_{j2} = 0$). The resulting KKT conditions are as below:

$$\frac{\left(\psi_{j1} / p_{j1}\right)}{\left(\frac{\psi_{j1} t_{j1}}{\gamma_j} + 1\right)} = \lambda \text{ since } t_{j1} > 0 \text{ and } t_{j2} = 0; \forall j \in \text{chosen destination alternatives,}$$

and (9)

$$\frac{\left(\psi_{j2} / p_{j2}\right)}{\left(\frac{\psi_{j1} t_{j1}}{\gamma_j} + 1\right)} < \lambda \text{ since } t_{j1} > 0 \text{ and } t_{j2} = 0; \forall j \in \text{chosen destination alternatives.}$$

The first of the above KKT conditions can be substituted into the second condition to result in the following condition:

$$\frac{\psi_{j2}}{p_{j2}} < \frac{\psi_{j1}}{p_{j1}} \text{ since } t_{j1} > 0 \text{ and } t_{j2} = 0; \forall j \in \text{chosen destination alternatives.} \qquad (10)$$

The above condition implies that the price-normalized baseline utility of a non-chosen mode to a chosen destination is always less than the price-normalized baseline utility of the chosen mode. Therefore, for a chosen destination, the mode choice alternative with the greatest price-normalized baseline utility would be the chosen alternative. Given this, the KKT conditions in (9) can be re-written as:

$$\frac{\left(\psi_{j1}/p_{j1}\right)}{\left\{\dfrac{\psi_{j1}t_{j1}}{\gamma_j}+1\right\}}=\lambda \text{ since } t_{j1}>0 \text{ and } t_{j2}=0; \forall j \in \text{chosen destination alternatives,}$$

and (11)

$$\frac{\psi_{j2}}{p_{j2}}<\frac{\psi_{j1}}{p_{j1}} \text{ since } t_{j1}>0 \text{ and } t_{j2}=0; \forall j \in \text{chosen destination alternatives.}$$

Substituting ψ_0/e_0 for λ, making algebraic rearrangements, and taking logarithms on both sides, one can further rewrite the KKT conditions for chosen destination alternatives as:

$$\ln(\psi_{j1})=-\ln\left(\frac{e_0}{\psi_0 p_{j1}}-\frac{t_{j1}}{\gamma_j}\right); \forall j \in \text{chosen destination alternatives,}$$

and (12)

$$\ln(\psi_{j2})-\ln(p_{j2})<\ln(\psi_{j1})-\ln(p_{j1}); \forall j \in \text{chosen destination alternatives.}$$

Note from the first of the above two conditions that the term $\left(\frac{e_0}{\psi_0 p_{j1}}-\frac{t_{j1}}{\gamma_j}\right)$ is inside a logarithmic function. During model estimation, the analyst must ensure that this term is always positive, lest it should lead to estimation breakdowns.

3.5.2 KKT conditions for non-chosen destinations

For a destination *k* that is not chosen, the time allocation would be zero for both the destination-mode alternatives corresponding to that destination (i.e., $t_{k1}=0$ *and* $t_{k2}=0$), resulting in the following KKT conditions for the destination:

$$\psi_{k1}-\lambda p_{k1}<0, \text{ and}$$
$$\psi_{k2}-\lambda p_{k2}<0; \text{ since } t_{k1}=0 \text{ and } t_{k2}=0; \forall k \in \text{non-chosen destination alternatives.}$$

Substituting ψ_0/e_0 for λ and taking logarithms on both sides, one can rewrite the KKT conditions for un-chosen destination alternatives as:

$$\ln \psi_{k1} < \ln \psi_0 + \ln p_{k1} - \ln e_0, \text{ and}$$
$$\ln \psi_{k2} < \ln \psi_0 + \ln p_{k2} - \ln e_0; \forall k \in \text{non-chosen destination alternatives.} \tag{13}$$

3.6 Econometric Structure

To complete the model specification, define the baseline marginal utility for destination-mode combination alternatives, ψ_{jl} as a function of observed and unobserved household characteristics, destination (j) characteristics, and modal (jl) characteristics as: $\psi_j = \exp(\boldsymbol{\Delta}'\mathbf{z}_{jl} + \varepsilon_{jl})$, where $\mathbf{z}_{jl}$ is a vector of destination and modal characteristics influencing the household's destination and mode choices (and their interactions with household characteristics); $\boldsymbol{\Delta}$ is a corresponding vector of parameters; and ε_{jl} is a destination-mode specific random term to accommodate the unobserved factors influencing the choice of destination-mode alternative *jl*.

For ψ_0, two different specifications have been used in the literature. Both the specifications arise from the consideration of identification issues arising from the linear budget constraint that the consumption amount of any one alternative can be derived from the consumptions of the remaining alternatives. One approach, adopted by a large number of studies in environmental economics (e.g., von Haefen et al., 2004) and recently in the marketing literature (Satomura et al., 2011), is to normalize ψ_0 to a value of 1. The reason for this normalization is that the KKT conditions are sufficient for estimating the expenditures on all but one good in the specification (e.g., the expenditures on all inside goods but the outside good). Given the expenditures on the inside goods, the outside good expenditure can be determined from the budget constraint identity. The second approach is to normalize only the deterministic component of ψ_0 to 1 and to specify a random component, as: $\psi_0 = \exp(\varepsilon_0)$, where ε_0 is a random error term capturing the unobserved factors influencing the total expenditure allocation for all inside goods (annual vacation, in this case). Bhat (2008) discusses several reasons why the latter specification should be preferred. A particular advantage with the latter specification is that including the stochastic term ε_0 on the outside good helps in capturing correlation among the random utilities of the inside goods. Such correlation helps in inducing greater competition among the consumptions of the inside goods, when compared to the competition between the inside goods and the outside good. However, the former specification is not theoretically inappropriate, especially on a Hicksian composite outside good;

except that it does not recognize greater correlations among the utility terms of inside goods. Further, recall from the discussion after Equation (12) that, to avoid estimation breakdowns, the analyst must ensure that the term $\left(\frac{e_0}{\psi_0 p_{j1}} - \frac{t_{j1}}{\gamma_j}\right)$ is always positive. Specifying ψ_0 to be a random variable with an unbounded distribution makes it difficult to impose this constraint while maintaining stability in estimation. Therefore, for convenience in model estimation, we specify ψ_0 as 1. It is relatively easier to ensure that $\left(\frac{e_0}{p_{j1}} - \frac{t_{j1}}{\gamma_j}\right)$ is positive for all individuals in the estimation data by appropriately adjusting the γ_j parameter.

To accommodate differences in satiation rates across destinations, the translation parameter γ_j can be specified as a function of destination (*j*) characteristics as: $\gamma_j = \exp(v'\mathbf{v}_j)$, where $\mathbf{v}_j$ is a vector of destination characteristics and v is a corresponding vector of parameters.

Specifying the joint cumulative distribution F of the random error terms $\left((\varepsilon_{11},\varepsilon_{12}),(\varepsilon_{21},\varepsilon_{22}),...,(\varepsilon_{j1},\varepsilon_{j2}),...,(\varepsilon_{J1},\varepsilon_{J2})\right)$ completes the random utility specification. In this paper, we assume that the random error terms have a nested extreme value structure with the following joint cumulative distribution:

$$F\left((\varepsilon_{11},\varepsilon_{12}),...,(\varepsilon_{j1},\varepsilon_{j2}),...,(\varepsilon_{J1},.\varepsilon_{J2})\right) = \prod_{j=1}^{J} \exp\left[-\left\{e^{-\left(\frac{\varepsilon_{j1}}{\sigma\theta}\right)} + e^{-\left(\frac{\varepsilon_{j2}}{\sigma\theta}\right)}\right\}^{\theta}\right] \quad (14)$$

In the above cumulative distribution function, the error terms of the modal alternatives for a specific destination *j*, $(\varepsilon_{j1},\varepsilon_{j2})$ are grouped into a nest, with a (dis)similarity parameter θ introduced to capture correlations among the random utility contributions of all destination-mode combination alternatives *jl* sharing a destination *j*. σ is a scale parameter that can be estimated due to variation in prices across the choice alternatives.

3.7 Consumption Probability Expression

Given the above-discussed econometric structure, the KKT conditions in Equations (12) and (13) can be expressed as:

$$\varepsilon_{j1} = -\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln\left(\frac{e_0}{\psi_0 p_{j1}} - \frac{t_{j1}}{\gamma_j}\right);\ \forall j \in \text{chosen destination alternatives,}$$

and (15)

$$\varepsilon_{j2} - \varepsilon_{j1} < (\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln p_{j1}) - (\boldsymbol{\Delta}'\mathbf{z}_{j2} - \ln p_{j2});\ \forall j \in \text{chosen destination alternatives.}$$

and

$$\varepsilon_{k1} < -\boldsymbol{\Delta}'\mathbf{z}_{j1} + \ln\psi_0 + \ln p_{k1} - \ln e_0,\quad \forall k \in \text{non-chosen destination alternatives,}$$

and (16)

$$\varepsilon_{k2} < -\boldsymbol{\Delta}'\mathbf{z}_{j2} + \ln\psi_0 + \ln p_{k2} - \ln e_0,\quad \forall k \in \text{non-chosen destination alternatives.}$$

The above stochastic KKT conditions can be used to derive the probability expression for the household's annual destination and mode choices and corresponding time allocations. Specifically, conditional on ψ_0, the probability that a household allocates e_0 amount of money to the outside good and the rest for vacation such that the first M of J vacation destination alternatives are chosen and the first available mode is chosen to travel to each of these destinations, with a time allocation pattern $\mathbf{t} = \big((t_{11},0),(t_{21},0),\ldots,(t_{M1},0),(0,0),\ldots,(0,0)\big)$ is given by:

$$\begin{aligned}
&P\left\{e_0,(t_{11},0),(t_{21},0),\ldots,(t_{M1},0),(0,0),\ldots,(0,0)\,|\,\psi_0\right\} \\
&= |J/\psi_0| \times \prod_{j=1}^{M} P\left\{\varepsilon_{j1} = -\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln\left(\frac{e_0}{\psi_0 p_{j1}} - \frac{t_{j1}}{\gamma_j}\right)\right\} \\
&\quad \times \prod_{j=1}^{M} P\left\{(\varepsilon_{j2} - \varepsilon_{j1}) < (\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln p_{j1}) - (\boldsymbol{\Delta}'\mathbf{z}_{j2} - \ln p_{j2})\right\} \\
&\quad \times \prod_{k=M+1}^{K} P\left\{(\varepsilon_{k1} < -\boldsymbol{\Delta}'\mathbf{z}_{k1} + \ln\psi_0 + \ln p_{k1} - \ln e_0), (\varepsilon_{k2} < -\boldsymbol{\Delta}'\mathbf{z}_{k2} + \ln\psi_0 + \ln p_{k2} - \ln e_0)\right\} \\
&= |J/\psi_0| \times \prod_{j=1}^{M} g_{\varepsilon_{j1}}\left\{-\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln\left(\frac{e_0}{\psi_0 p_{j1}} - \frac{t_{j1}}{\gamma_j}\right)\right\} \\
&\quad \times \prod_{j=1}^{M} G_{\varepsilon_{j2}-\varepsilon_{j1}}\left\{(\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln p_{j1}) - (\boldsymbol{\Delta}'\mathbf{z}_{j2} - \ln p_{j2})\right\} \qquad (17) \\
&\quad \times \prod_{k=M+1}^{K} G_{\varepsilon_{k1},\varepsilon_{k2}}\left\{(-\boldsymbol{\Delta}'\mathbf{z}_{k1} + \ln\psi_0 + \ln p_{k1} - \ln e_0), (-\boldsymbol{\Delta}'\mathbf{z}_{k2} + \ln\psi_0 + \ln p_{k2} - \ln e_0)\right\}
\end{aligned}$$

In the above expression, the Jacobian $|J/\psi_0|$ and the next two terms correspond to all chosen destinations, while the last term corresponds to all non-chosen destinations. Specifically, the terms $g_{\varepsilon_{j1}}\{.\}$ and $G_{\varepsilon_{j2}-\varepsilon_{j1}}\{.\}$ together correspond to the probability for the choice of destination j

(*j=1,2,3,..,M*) along with the corresponding time allocation and mode of travel. The term $G_{\varepsilon_{k1},\varepsilon_{k2}}\{.,.\}$ corresponds to the probability that the remaining alternatives are not chosen. $g_{\varepsilon_{j1}}\{.\}$ represents the probability density function of the random term ε_{j1}; $G_{\varepsilon_{j2}-\varepsilon_{j1}}\{.\}$ represents the cumulative density function of the difference $(\varepsilon_{j2}-\varepsilon_{j1})$ between the random terms ε_{j2} and ε_{j1}; and $G_{\varepsilon_{k1},\varepsilon_{k2}}\{.,.\}$ represents the joint cumulative density function of the random terms $(\varepsilon_{k1},\varepsilon_{k2})$.

$$g_{\varepsilon_{j1}}\left\{-\mathbf{\Delta'z}_{j1}-\ln\left(\frac{e_0}{\psi_0 p_{j1}}-\frac{t_{j1}}{\gamma_j}\right)\right\}=$$
$$\frac{1}{\sigma}\exp\left(\frac{\mathbf{\Delta'z}_{j1}}{\sigma}\right).\left(\frac{e_0}{\psi_0 p_{j1}}-\frac{t_{j1}}{\gamma_j}\right)^{\frac{1}{\sigma}}.\exp\left\{-\exp\left(\frac{\mathbf{\Delta'z}_{j1}}{\sigma}\right).\left(\frac{e_0}{\psi_0 p_{j1}}-\frac{t_{j1}}{\gamma_j}\right)^{\frac{1}{\sigma}}\right\} \quad (18)$$

$$G_{\varepsilon_{j2}-\varepsilon_{j1}}\left\{(\mathbf{\Delta'z}_{j1}-\ln p_{j1})-(\mathbf{\Delta'z}_{j2}-\ln p_{j2})\right\}=$$
$$\frac{\exp\left(\dfrac{\mathbf{\Delta'z}_{j1}-\ln p_{j1}}{\sigma\theta}\right)}{\exp\left(\dfrac{\mathbf{\Delta'z}_{j1}-\ln p_{j1}}{\sigma\theta}\right)+\exp\left(\dfrac{\mathbf{\Delta'z}_{j2}-\ln p_{j2}}{\sigma\theta}\right)} \quad (19)$$

$$G_{\varepsilon_{k1},\varepsilon_{k2}}\left\{(-\mathbf{\Delta'z}_{k1}+\ln\psi_0+\ln p_{k1}-\ln e_0),(-\mathbf{\Delta'z}_{k2}+\ln\psi_0+\ln p_{k2}-\ln e_0)\right\}=$$
$$\exp-\left\{\left(\frac{e_0}{\psi_0}\right)^{1/\sigma}\left\langle\exp\left(\frac{\mathbf{\Delta'z}_{k1}-\ln p_{k1}}{\sigma\theta}\right)+\exp\left(\frac{\mathbf{\Delta'z}_{k2}-\ln p_{k2}}{\sigma\theta}\right)\right\rangle^{\theta}\right\} \quad (20)$$

Note that the discourse so far is based on the assumption that only two modal alternatives are available for each destination and that the first mode is chosen for each visited destination. However, the probability expression in Equation (16) can be easily extended to the general case where L (>2) number of modes are available to travel to each destination. Let the first *M* destination alternatives be the chosen destinations (*j = 1,2,…,M*), the chosen mode for each alternative be indexed by l_j, and that the time allocated to a chosen destination-mode combination alternative jl_j be t_{jl_j}. Then the probability expression for the household's annual destination-mode choices and time allocations is given below:

$$P\left\{e_0,(0,..,t_{1l_1},..,0),(0,..,t_{2l_2},..,0),...,(0,..,t_{Ml_M},..,0),(0,0,...,0),...,(0,0,...,0)\,|\,\psi_0\right\}$$

$$=|J/\psi_0|\times\prod_{j=1}^{M}\frac{1}{\sigma}\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{jl_j}}{\sigma}\right).\left(\frac{e_0}{\psi_0 p_{jl_j}}-\frac{t_{jl_j}}{\gamma_j}\right)^{\frac{1}{\sigma}}.\exp\left\{-\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{jl_j}}{\sigma}\right).\left(\frac{e_0}{\psi_0 p_{jl_j}}-\frac{t_{jl_j}}{\gamma_j}\right)^{\frac{1}{\sigma}}\right\}$$

$$\times\prod_{j=1}^{M}\frac{\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{jl_j}-\ln p_{jl_j}}{\sigma\theta}\right)}{\sum_{l=1}^{L}\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{jl}-\ln p_{jl}}{\sigma\theta}\right)} \tag{21}$$

$$\times\prod_{k=M+1}^{K}\exp-\left\{\left(\frac{e_0}{\psi_0}\right)^{1/\sigma}\left\langle\sum_{l=1}^{L}\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{kl}-\ln p_{kl}}{\sigma\theta}\right)\right\rangle^{\theta}\right\}$$

The term $|J/\psi_0|$ in the above expression is the determinant of the Jacobian matrix (Conditional on ψ_0) obtained from applying change of variables calculus between the vector of stochastic terms $(\varepsilon_{1l_1},\varepsilon_{2l_2},\ldots,\varepsilon_{jl_j},\ldots,\varepsilon_{Ml_M})$ for all chosen destination-mode combination alternatives and the corresponding vector of time allocation variables $(t_{1l_1},t_{2l_2},\ldots,t_{jl_j},\ldots,t_{Ml_M})$. This determinant does not have a compact form but the ih^{th} element of the matrix can be computed as:

$$J_{ih}/\psi_0=\frac{\partial}{\partial t_{hl_h}}\left[-\boldsymbol{\Delta}'\mathbf{z}_{il_i}-\ln\left(\frac{e_0}{\psi_0 p_{il_i}}-\frac{t_{il_i}}{\gamma_i}\right)\right]=\frac{\left(\frac{p_{hl_h}}{\psi_0 p_{il_i}}+\frac{\delta_{ih}}{\gamma_i}\right)}{\left(\frac{e_0}{\psi_0 p_{il_i}}-\frac{t_{il_i}}{\gamma_i}\right)};(i,h=1,2,...,M) \tag{22}$$

where, δ_{ih} is an indicator that takes a value of 1 if $i = h$, or zero otherwise.

Note that all Equations (15) through (22), including the probability expression in Equation (21) are conditional on ψ_0. Recall from the earlier discussion that we assume ψ_0 to be equal to 1. Thus, simply substituting 1 for ψ_0 in Equations (21) and Equation (22) will provide the unconditional probability expression and the corresponding Jacobian expression.

3.8 Forecasting with the Proposed Model

Until recent past, forecasting with KKT demand models was pursued to be very difficult, especially in the presence of imperfect substitutes and corner solutions. von Haefen et al. (2004) and Pinjari and Bhat (2011) proposed computationally efficient algorithms for forecasting with

KKT demand models with only imperfect substitutes in the choice set. In this paper, we expand on the Pinjari and Bhat (2011) algorithm to the case with a mix of perfect and imperfect substitutes in the choice set.

Recall from Equation (10) that, for any chosen destination alternative *j* (*=1, 2,…, J*), the price-normalized baseline marginal utility of a chosen destination-mode alternative is always greater than that of other alternatives sharing that same destination. One can write the same condition as:

$$\frac{\psi_{jl_j}}{p_{jl_j}} = \underset{l=1,2,\dots,L}{Max}\left(\frac{\psi_{jl}}{p_{jl}}\right), \qquad (23)$$

where $\frac{\psi_{jl_j}}{p_{jl_j}}$ is the price-normalized baseline utility of chosen destination-mode alternative jl_j.

Similarly, one can show using the KKT conditions derived in Section 3.5 that the price-normalized baseline utility of any chosen destination-mode alternative is greater than that of destination-mode alternatives for any non-chosen destination alternative. Specifically, one can show that:

$$\underset{l=1,2,\dots,L}{Max}\left(\frac{\psi_{kl}}{p_{kl}}\right) < \lambda < \underset{l=1,2,\dots,L}{Max}\left(\frac{\psi_{jl}}{p_{jl}}\right); \text{ if } k \text{ is non-chosen and } j \text{ is a chosen destination.} \qquad (24)$$

This is because $\left(\frac{\psi_{kl}}{p_{kl}}\right) < \lambda \;\; \forall l = (1,2,..,L)$ or $\underset{l=1,2,\dots,L}{Max}\left(\frac{\psi_{kl}}{p_{kl}}\right) < \lambda$ for non-chosen destinations and $\frac{\left(\psi_{jl_j} / p_{jl_j}\right)}{\left(\frac{\psi_{jl_j} t_{jl_j}}{\gamma_j} + 1\right)} = \lambda$ or $\underset{l=1,2,\dots,L}{Max}\left(\frac{\psi_{jl}}{p_{jl}}\right) > \lambda$ for chosen destinations. Note from Equation (24) that the Lagrange multiplier λ is smaller than the price-normalized baseline marginal utility of any chosen destination-mode alternative but greater than that of all destination-mode alternatives belonging to non-chosen destinations.

One can rearrange the KKT condition for a chosen destination-mode alternative in Equation (11) to express the corresponding time allocation t_{jl_j} as:

$$t_{jl_j} = \left(\frac{\psi_{jl_j}}{p_{jl_j}} - \lambda\right)\frac{\gamma_j}{\lambda \psi_{jl_j}}. \qquad (25)$$

Substituting this expression for t_{jl_j} into the budget constraint and solving for λ gives:

$$\lambda = \frac{\left(\psi_0 + \sum_{j=1\,\text{to}\,M} \gamma_j\right)}{\left(E + \sum_{j=1\,\text{to}\,M} \frac{\gamma_j p_{jl_j}}{\psi_{jl_j}}\right)} \tag{26}$$

In the above expression, $j = (1,2,\ldots M)$ are the chosen alternatives and l_j is the chosen mode for a chosen destination j. It is important to note that the summation terms in the above expression include chosen destination-mode alternatives only. Given the insights from Equations (23) through (26), the forecasting algorithm for the model proposed in this paper comprises four basic steps as outlined next.

Step 0:

- Assume that only the outside good is chosen. Initialize the total number chosen destinations as: $M = 0$, and Lagrangian multiplier as: $\lambda = \frac{\psi_0}{E}$.
- Given the input data ($\mathbf{z}_{jl}$, p_{jl}), model parameters ($\mathbf{\Delta}, \gamma_j$), and the simulated error term (ε_{jl}) draws, compute the price-normalized baseline utility values $\left(\psi_{jl} / p_{jl}\right)$ for all destination-mode alternatives.

Step 1:

- For each destination j $(=1,2,\ldots,J)$, pick the modal alternative with the maximum price-normalized baseline utility value, $\underset{l=1,2,\ldots,L}{Max}\left(\psi_{jl} / p_{jl}\right)$. Label the corresponding destination-mode alternative as jl_j. Then $\underset{l=1,2,\ldots,L}{Max}\left(\psi_{jl} / p_{jl}\right)$ can be denoted by $\left(\psi_{jl_j} / p_{jl_j}\right)$.
- Arrange all the J destination alternatives available to the consumer in the descending order of $\underset{l=1,2,\ldots,L}{Max}\left(\psi_{jl} / p_{jl}\right)$.

Step 2:

- If $M = 0$ and $\lambda > \underset{l=1,2,\ldots,L}{Max}\left(\frac{\psi_{1l}}{p_{1l}}\right)$(i.e., the maximum price-normalized baseline utility among all modes available to the first destination alternative in the above arrangement),

- Set the optimal consumption of the outside good as $e_0^* = E$ and that of all destination-mode alternatives as zero (i.e., no destinations are visited in this case) and stop.

- Else, if $M = 0$ and $\lambda < \underset{l=1,2,\ldots,L}{Max}\left(\frac{\psi_{1l}}{p_{1l}}\right)$,
  - Set $M = M+1$ (because the next destination alternative must be a chosen alternative; see Equation (24)).
  - Update the Lagrangian multiplier as $\lambda = \left(\psi_0 + \sum_{j=1\,\text{to}\,M}\gamma_j\right)\Big/\left(E + \sum_{j=1\,\text{to}\,M}\frac{\gamma_j p_{jl_j}}{\psi_{jl_j}}\right)$.

Step 3:

- If $M = J$,
  - Compute the optimal consumption of the outside good as $e_0^* = (\psi_0 / \lambda)$
  - Compute the optimal consumptions of the first M destination alternatives (and corresponding chosen mode alternatives) as $t_{jl_j}^* = \left(\frac{\psi_{jl_j}}{p_{jl_j}} - \lambda\right)\frac{\gamma_j}{\lambda\psi_{jl_j}}$, and stop.
- Else, if $M < J$, go to step 4.

Step 4:

- If $\lambda < \underset{l=1,2,\ldots,L}{Max}\left(\frac{\psi_{(M+1)l}}{p_{(M+1)l}}\right)$ (i.e., the maximum price-normalized baseline utility among all modes available to the destination alternative in position $M+1$),
  - Set $M = M+1$ and update the Lagrangian multiplier as:

$$\lambda = \left(\psi_0 + \sum_{j=1\,\text{to}\,M}\gamma_j\right)\Big/\left(E + \sum_{j=1\,\text{to}\,M}\frac{\gamma_j p_{jl_j}}{\psi_{jl_j}}\right).$$

  - Go to step 3.
- Else, if $\lambda > \underset{l=1,2,\ldots,L}{Max}\left(\frac{\psi_{(M+1)l}}{p_{(M+1)l}}\right)$ (i.e., the maximum price-normalized baseline utility among all modes available to the destination alternative in position $M+1$),
  - Compute the optimal consumption of the outside good as $e_0^* = (\psi_0 / \lambda)$

- Compute the optimal consumptions of the first M destination alternatives in the above arrangement (and corresponding chosen mode alternatives) as $t^*_{jl_j} = \left(\frac{\psi_{jl_j}}{p_{jl_j}} - \lambda \right) \frac{\gamma_j}{\lambda \psi_{jl_j}}$.
- Set the consumptions of all other destination-mode alternatives as zero and stop.

The above outlined forecasting procedure can be applied repeatedly over a large number of simulated error term (ε_{jl}) draws to obtain distributions for the consumer choice predictions. The procedure can be used to apply empirical models for forecasting and policy prediction purposes as well as to simulate data that exhibits a mix of both perfect and imperfect substitution among choice alternatives.

3.9 Accommodation of Multiple Linear Budget Constraints

The discourse so far is based on the assumption that a single linear budget constraint governs consumer choices. However, several consumer choices involve the use of multiple resources such as time and money and therefore governed by multiple constraints. In the current empirical context, households' leisure travel decisions are potentially influenced by both time and money constraints. For example, some households may have the time to travel to exotic and far away destinations but not enough money to do so. On the other hand, some households may simply not have the time for long vacations even if they are able to afford the expenses. In most cases, both time and money constraints are likely to influence the choices. To accommodate both these constraints, the households' utility maximization problem discussed in Section 3.2 can be extended as below (assuming only two modes of travel are available for each destination):

$$\underset{(\mathbf{t}, e_0, t_0)}{Max}\ U(\mathbf{t}, e_0, t_0) = \sum_{j=1}^{J} \gamma_j \ln \left\{ \frac{(\psi_{j1} t_{j1} + \psi_{j2} t_{j2})}{\gamma_j} + 1 \right\} + \ln e_0 + \ln t_0 \quad (27)$$

Subject to two linear budget constraints – one for time and the other for money – as below:

$$\sum_{j=1}^{J} \left(p_{j1} t_{j1} + p_{j2} t_{j2} \right) + e_0 = E,$$

and (28)

$$\sum_{j=1}^{J} \left(q_{j1} t_{j1} + q_{j2} t_{j2} \right) + t_0 = T.$$

In the above utility function (Equation 27), all terms in the sub-utility function $\gamma_j \ln\left\{(\psi_{j1}t_{j1} + \psi_{j2}t_{j2}) / \gamma_j + 1\right\}$ are as described in Section 3.2. The second and third terms, $\ln e_0$ and $\ln t_0$, are specified following Satomura et al.'s specification for outside goods in the presence of multiple budget constraints. Specifically, e_0 is the Hicksian composite outside good for money (i.e., income – annual expenditure on vacation) and t_0 is the Hicksian composite outside good for time representing all the non-vacation time in a year (i.e., 365 days – annual number of days spent on vacation). These terms complete the utility function to form an incomplete demand system. Both the outside goods are assumed to be "essential" with some positive consumption by all households since neither the entire year nor the whole income is typically spent completely on vacation.

In the two constraints identified in Equation (28), q_{jl} and p_{jl} (*l = 1, 2*) represent the time-prices and money-prices, respectively, of consuming unit time of a vacation destination *j* traveled by mode *l.* [5] As can be observed from the two constraints, following Satomura et al., (2011), the Hicksian composite outside good for time (t_0) is assumed to have unit time-price and zero money-price (i.e., it doesn't appear in the time constraint), while the money-specific outside good (e_0) has unit money-price and zero time-price (i.e., it doesn't appear in the money constraint).

To setup the optimality conditions for the time- and money-constrained utility maximization problem without considering the third constraint, one can form a Largrangian function as below:

$$L = U(\mathbf{t}, e_0, t_0) - \lambda\left[e_0 + \sum_{j=1}^{J}(p_{j1}t_{j1} + p_{j2}t_{j2}) - E\right] - \mu\left[t_0 + \sum_{j=1}^{J}(q_{j1}t_{j1} + q_{j2}t_{j2}) - T\right] \quad (29)$$

As described in Section 3.5, applying the KKT conditions for the outside goods result in the following Lagrangian multipliers: $\lambda = 1/e_0$ and $\mu = 1/t_0$, representing the marginal utility of time and money, respectively. Further, the following KKT conditions can be derived for the inside goods (i.e., destination-mode combinations).

---

[5] The time-price (q_j), in the current empirical context, is the amount of time that needs to be expended to consume a unit amount of vacation time. If one assumes that the time spent traveling is a part of vacation time (i.e., people derive utility from traveling for vacation), then the time-price is unity. On the other hand, if the time spent traveling is viewed only as a cost without any contribution to utility, the time-price is more than unity; since a day of vacation at a destination incurs one day of vacation plus some amount of travel. In the current empirical application, we consider the time-price to be unity assuming that travel time is a part of vacation and it contributes to utility. The money-price is simply the amount of money expended to consume a unit amount of vacation time.

$$\frac{\psi_{jl}}{\left\{\frac{(\psi_{j1}t_{j1}+\psi_{j2}t_{j2})}{\gamma_j}+1\right\}}=\tau_{jl} \text{ if } t_{jl}>0; \forall j=1,2,...,J; l=1,2$$

$$\frac{\psi_{jl}}{\left\{\frac{(\psi_{j1}t_{j1}+\psi_{j2}t_{j2})}{\gamma_j}+1\right\}}<\tau_{jl} \text{ if } t_{jl}=0; \forall j=1,2,...,J; l=1,2 \qquad (30)$$

where, $\tau_{jl}=\mu q_{jl}+\lambda p_{jl}=\frac{q_{jl}}{t_0}+\frac{p_{jl}}{e_0}$.

Again following the discourse in Section 3.2, assuming that the first M destination alternatives are chosen and that the first mode of travel is chosen to travel to each of these destinations, one can derive the following separate KKT conditions for all chosen destinations:

$$\frac{\psi_{j1}}{\left\{\frac{\psi_{j1}t_{j1}}{\gamma_j}+1\right\}}=\tau_{j1} \text{ since } t_{j1}>0 \text{ and } t_{j2}=0; \forall j \in \text{chosen destinations}$$

and (31)

$$\frac{\psi_{j2}}{\tau_{j2}}<\frac{\psi_{j1}}{\tau_{j1}}, \text{ since } t_{j1}>0 \text{ and } t_{j2}=0; \forall j \in \text{chosen destinations.}$$

Similarly, one can derive the following KKT conditions for non-chosen destinations.

$$\psi_{k1}<\tau_{j1} \text{ and } \psi_{k2}<\tau_{j2}, \text{ since } t_{k1}=0 \text{ and } t_{j2}=0; \forall k \in \text{chosen destinations} \qquad (32)$$

Assuming the same stochastic distributions as in Section 3.6, the following stochastic KKT conditions can be derived:

$$\varepsilon_{j1}=-\mathbf{\Delta}'\mathbf{z}_{j1}-\ln\left(\frac{1}{\tau_{j1}}-\frac{t_{j1}}{\gamma_j}\right); \ \forall j \in \text{chosen destinations,}$$

and (33)

$$\varepsilon_{j2}-\varepsilon_{j1}<(\mathbf{\Delta}'\mathbf{z}_{j1}-\ln\tau_{j1})-(\mathbf{\Delta}'\mathbf{z}_{j2}-\ln\tau_{j2}); \ \forall j \in \text{chosen destinations.}$$

and

$$\varepsilon_{k1}<-\mathbf{\Delta}'\mathbf{z}_{j1}+\ln\tau_{k1}, \quad \forall k \in \text{non-chosen destination alternatives,}$$

and (34)

$$\varepsilon_{k2}<-\mathbf{\Delta}'\mathbf{z}_{j2}+\ln\tau_{k2}, \quad \forall k \in \text{non-chosen destination alternatives.}$$

These stochastic KKT conditions can be used to derive the following probability expression for households' annual destination and mode choices and corresponding time allocations:

$$P\left\{t_0, e_0, (t_{11},0), (t_{21},0), ..., (t_{M1},0), (0,0), ..., (0,0)\right\}$$

$$=|J| \times \prod_{j=1}^{M} \frac{1}{\sigma} \exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{j1}}{\sigma}\right) . \left(\frac{1}{\tau_{j1}} - \frac{t_{j1}}{\gamma_j}\right)^{\frac{1}{\sigma}} . \exp\left\{-\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{j1}}{\sigma}\right) . \left(\frac{1}{\tau_{j1}} - \frac{t_{j1}}{\gamma_j}\right)^{\frac{1}{\sigma}}\right\}$$

$$\times \prod_{j=1}^{M} \frac{\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln \tau_{j1}}{\sigma\theta}\right)}{\exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{j1} - \ln \tau_{j1}}{\sigma\theta}\right) + \exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{j2} - \ln \tau_{j2}}{\sigma\theta}\right)} \tag{35}$$

$$\times \prod_{k=M+1}^{J} \exp - \left\{\left\langle \exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{k1} - \ln \tau_{k1}}{\sigma\theta}\right) + \exp\left(\frac{\boldsymbol{\Delta}'\mathbf{z}_{k2} - \ln \tau_{k2}}{\sigma\theta}\right)\right\rangle^{\theta}\right\}$$

In the above expression, $|J|$ is the determinant of the Jacobian matrix due to change of variables from the vector of time allocation variables to the corresponding stochastic terms for all chosen destination-mode combination alternatives. The ih^{th} element of the matrix is given below:

$$J_{ih} = \frac{\partial}{\partial t_{h1}}\left[-\boldsymbol{\Delta}'\mathbf{z}_{i1} - \ln\left(\frac{1}{\tau_{i1}} - \frac{t_{i1}}{\gamma_i}\right)\right] = \frac{\left(\frac{\left((q_{i1}q_{h1})/t_o^2\right) + \left((p_{i1}p_{h1})/e_o^2\right)}{\tau_{i1}^2}\right) + \frac{\delta_{ih}}{\gamma_i}}{\left(\frac{1}{\tau_{i1}} - \frac{t_{i1}}{\gamma_i}\right)}; (i, h = 1, 2, ..., M) \tag{36}$$

where, $\tau_{i1} = \frac{q_{i1}}{t_0} + \frac{p_{i1}}{e_0}$, and δ_{ih} is an indicator that takes a value of 1 if $i = h$, or zero otherwise.

Note from the KKT conditions in (33) and the probability expression in (35) that the term $\left(\frac{1}{\tau_{j1}} - \frac{t_{j1}}{\gamma_j}\right)$ is inside a logarithmic function. During model estimation, the analyst must ensure that this term is always positive, lest it should lead to estimation breakdowns.

The reader will note here that the above formulation simplifies to a formulation with a single linear budget constraint when the outside good quantity corresponding to one of the constraints is infinity (i.e., when one of the constraints is relaxed). For example, when the time constraint is relaxed (i.e., when $t_0 \to \infty$), the above formulation collapses to the model in Section 3.7 with a money budget constraint. Similarly, when the money constraint is relaxed, the formulation collapses to a model with a time budget constraint.

4 EMPIRICAL APPLICATION TO LONG-DISTANCE LEISURE TRAVEL DEMAND ANALYSIS

4.1 Background

A significant portion of passenger travel miles in the United States (US) comes from long-distance travel, especially for leisure purposes such as vacation. Statistics from national travel surveys (BTS 2001) indicate that more than one half of all long-distance travel is for leisure. Besides, long-distance leisure travel garners particular attention due to its impact on the tourism and recreation industry. Due to all these reasons, long-distance leisure travel has been studied extensively in the tourism and recreation demand literatures, and is steadily gaining importance in the transport planning/modeling arena. Most work on this topic in the transport modelling arena can be categorized into: (1) Statewide travel models in the US (e.g., Horowitz 2008), (2) National travel models in Europe (e.g., Rich et al., 2009; Fosgerau, 2001), and (3) Inter-city travel demand analysis between specific city pairs (e.g., Koppelman and Sethi, 2005). An important end-goal of all these efforts is to estimate travel flows between different regions by different modes of travel for informing various policy and investment initiatives.

A drawback of most literature in both the travel demand and tourism fields is that the analysis is typically limited to smaller time frames such as a day or a few weeks. However, analysis of the 1995 American Travel Survey data indicates that on average, a household makes less than 4 vacation trips over a year. Given the infrequent nature of long-distance leisure travel, a smaller time-frame of analysis (e.g., a day) is likely to provide a distorted picture of leisure travel flows in the nation. Intuitively, vacations are planned over longer time frames, as opposed to daily travel decisions for which shorter time frames may suffice. In this context, Eugenio-Martin's (2003) theoretical framework for tourism demand analysis suggests one year as appropriate for vacation travel analysis (also see Morley 1992). The need for considering a longer time frame of analysis is also well-recognized in the recreational demand literature, where numerous studies analyze recreational choices over longer time horizons such as seasons than over a single choice instance (see Bockstael et al., 1987; Phaneuf and Smith 2005; von Haefen and Phaneuf, 2005; Phaneuf et al., 2000).

To be sure, a few studies in the transport planning arena do consider longer time-frames for analyzing leisure travel. These include the annual leisure travel framework of van Middlekoop et al. (2004), the holiday travel module in the pan-European long-distance travel model (Rich et al., 2009), and annual vacation time allocation studies by LaMondia et al. (2008) and Van Nostrand et al. (2013). Among these studies, the paper by Van Nostrand et al. is the most relevant to the current empirical research, as discussed next.

Van Nostrand et al. applied Bhat's MDCEV framework for modeling households' annual vacation destination choices and the corresponding time allocations, recognizing that households could potentially visit multiple destinations over a year (i.e., destination choices are imperfect substitutes). Specifically, they formulated an annual time allocation model assuming that an annual vacation time budget exists for each household. Building on their work, the current paper estimates an integrated model of annual vacation destination choice and mode choices to recognize that vacation destination and mode choices are made in a joint fashion (Van Nostrand et al. do not consider mode choice jointly with destination choice). The framework recognizes that, while vacation destinations are imperfect substitutes, the travel mode alternatives to a destination are perfect substitutes in that only one primary mode is chosen to travel to a destination. In addition, the current empirical work explicitly considers the role of both time and money constraints (as opposed to a single constraint) in households' long-distance leisure travel destination and mode choice decisions, while considering the price variation across destination and mode choice alternatives.

The empirical model in this paper is estimated using household-level long-distance travel data from the 1995 American Travel Survey (ATS), as assembled by Van Nostrand et al. (2013). To define the destination choice alternatives, first each of the Metropolitan Statistical Areas (MSAs) from each of the 48 contiguous states in the US was identified as a potential vacation destination. Subsequently, the remaining non-MSA area in each state was counted as a single destination (one non-MSA area for each state. All together, the U.S. was divided into 210 destinations comprising 162 MSA destinations and 48 non-MSAs. For each destination, auto and air were considered as the two primary modes of travel.

The reader will note here that the primary intent of the current empirical application is only a demonstration of the proposed methodological framework to jointly model SDC and MDC choices in the presence of a mix of imperfect and perfect substitutes in the choice set. Further empirical investigations such as policy simulations with the proposed framework are left out for further research. To this end, one must first develop forecasting algorithms to apply the proposed framework with multiple budget constraints (the algorithm outlined in Section 3.8 is applicable only in the presence of a single budget constraint).

4.2 Data

The 1995 ATS is the source of household-level vacation destination and mode choice data used in this analysis. The survey collected information from over 60,000 American households on all long-distance trips each household made over an entire year to destinations farther than 100 miles (BTS 1995a). For each trip, the information on the purpose, mode, and destination of travel and other travel attributes such as the time spent (no. of days) on the trip and travel party size were collected. From this sample, 22,215 households reported making at least one trip over the year for one of the four leisure purposes – relaxation, sightseeing, entertainment, or outdoor recreation – by either the car mode or the air mode of travel. A random sample of 1000 households was selected for model estimation while another random sample of 500 households was selected for model validation.

In addition to the 1995 ATS, several secondary data sources were utilized to compile other required information such as: (1) the transportation level of service variables, including the travel times and costs between each origin–destination pair via air and auto modes, (2) lodging prices, and non-lodging (dining, entertainment/recreation, and other) prices at each of the 210 destinations, (3) the destination size and attraction variables for the year 1995, including land area, number of employees in different sectors (leisure and hospitality, retail, etc.), total population, and gross domestic product, and (4) the destination climate variables, including mean monthly temperatures for different months in a year, miles of coastline at the destination, and the annual number of freezing days experienced at the destination. Details on the data sources and procedures used to create the specific variables of interest can be found in Van Nostrand et al. Since the focus of this paper is on accommodating imperfect substitutability across destination choice alternatives and perfect substitutability across mode choice alternatives in the presence of price variation across the different vacation destinations and travel modes, the procedures and assumptions used to construct the price variables using the 1995 Consumer Expenditure (CEX) data are described in Appendix A.

Table 1 presents the descriptive statistics from the estimation sample used in this analysis. The households in the estimation sample have an average size of 2.86

persons/household, householder age of 45.6 years, and an average income of $50,694per annum. 38% of them have at least one child of age less than 16 years, and 13% of them have householders who are retired. About 48 % of the households in the estimation sample made multiple long-distance leisure trips over a year. Further, a significant proportion (close to 40%) of the households visited multiple destinations. Furthermore, a large percentage (81%) of households visited a destination (if they did so) only once. This suggests multiple discreteness (or variety-seeking) in households' annual destination choices. Even if households visited a destination more than once a year, a vast majority of the times (99.5% of the time in the data; not shown in the table) the same mode was used to travel across all the different trips made by a household to that same destination. This suggests perfect substitution (or single discreteness) among the mode choice alternatives to a destination. The total annual household time spent on vacation ranged from a day to about 44 days, with an average of 6 days. The estimated total annual vacation expenditure ranged from $30 to $5,660 (not shown in table), with an average of $693 per household.

The trip level characteristics suggest 84.7% of the trips in the estimation sample were by auto mode and the remaining 15.3% were by the air mode. The average round trip distance was about 990 miles. On average, the households in the sample spent about 4 days and about $437 on each trip. The last set of descriptive statistics is for the characteristics of the 210 destinations and level of service variables between the 210 x 210 possible OD pairs.

4.3 Model Estimation

The parameters of the proposed model formulations were estimated using the maximum likelihood method. Three different models were estimated: (1) A model that considers only the time budget constraint, (2) A model that considers only the money budget constraint, and (3) A model that considers both time and money budget constraints. In all three models, the destination choice alternatives are considered to be imperfect substitutes while the mode choice alternatives for each destination are considered to be perfect substitutes. The log-likelihood functions of all the three models were coded in the GAUSS matrix programming language. As discussed earlier, the term $\left(\frac{t_0}{q_{j1}} - \frac{t_{j1}}{\gamma_j}\right)$ in the time-constrained model must be always positive to avoid estimation breakdowns. Similarly, the term $\left(\frac{e_0}{p_{j1}}-\frac{t_{j1}}{\gamma_j}\right)$ must be positive during the estimation of the money-constrained model, while the term $\left(\frac{1}{\tau_{j1}}-\frac{t_{j1}}{\gamma_j}\right)$ must be positive during the estimation of the time- and money-constrained model.[6] To consider these constraints, we first attempted to use the constrained maximum likelihood (CML) module of GAUSS. But we encountered estimation instability and convergence issues with the CML module. Therefore, all the models were estimated using the maximum likelihood module of GAUSS, while ensuring positivity of the above mentioned terms at each of the iterations during estimation. To do so, if the positivity condition was not met at any iteration, the iteration-search parameters of the γ_j function were updated to ensure positivity without deviating too much from their original values.

Admittedly, the above-described approach of ensuring the positivity constraints is somewhat *ad hoc*. To verify whether the approach results in appropriate parameter estimates, we conducted simulation experiments with the single constrained model formulation. Specifically, we used the forecasting procedure described in Section 3.8 to simulate discrete-continuous choice data that exhibits both perfect and imperfect substitution patterns among choice alternatives, assuming a single linear budget constraint (i.e., the money budget constraint). Two different types of data were simulated based on two different assumptions: (1) with the satiation parameters (γ_j) assumed to be constants, and (2) with the satiation parameters (γ_j) specified as a function of alternative attributes. In both cases, the baseline utility parameters were specified as a function of alternative attributes and decision-maker characteristics. Next, maximum likelihood estimation was performed on the simulated data to retrieve the parameters used to simulate the data. As discussed earlier, at each iteration, the constraint that the term $\left(\frac{e_0}{p_{j1}}-\frac{t_{j1}}{\gamma_j}\right)$ is positive was ensured in a heuristic fashion by updating the iteration-search parameters in the γ_j function (if the constraint was not satisfied). The detailed results of this exercise are not presented here to conserve space, but the overall findings are reported briefly. The structural parameters of both

---

[6] Recall that $\tau_{j1}=\frac{q_{j1}}{t_0}+\frac{p_{j1}}{e_0}$. Therefore, $\frac{1}{\tau_{j1}}$ becomes $\frac{t_0}{q_{j1}}$ in the single(time)-constrained model and $\frac{e_0}{p_{j1}}$ in the single(money)-constrained model.

the baseline utility and satiation functions used to simulate the data could be easily retrieved through maximum likelihood estimation when the satiation parameters γ_j were assumed to be constants. Using different sets of starting values for the parameters did not generally influence either the estimability of parameters or the estimates at convergence. However, the estimation process was relatively slow with its stability and convergence depending on the starting values when the satiation parameters were specified as a function of alternative attributes. Also, the need for the above-described adjustment to the iteration-search parameters depended on the starting values. Using starting values that were closer to the true parameter values resulted in an easier estimation without having to make adjustments to the iteration-search parameters.

Based on the above insights from the simulation experiments, the empirical model estimation was carried out in a step-by-step manner, beginning with the estimation of specifications with only a constant in the satiation (γ_j) functions and using those parameter estimates as starting values for richer specifications with destination-specific variables in the γ_j functions. Different sets of starting values were explored as well. Limited explorations suggest that while not all starting values necessarily lead to convergence, the same set of parameter estimates were obtained whenever the model converged.

4.4 Model Results

For each of the three empirical models estimated – a time-constrained model, a money-constrained model, and a time- and money-constrained model – a variety of model evaluation measures, namely, log-likelihood (*LL*) at convergence, Rho-square (ρ^2), Bayesian Information Criterion (BIC), and predictive log-likelihood (*PLL*) on a validation sample of 500 households, are presented in Table 2.

As can be observed, the log-likelihood value of the time- and money-constrained model is better than that of the single, time-constrained (money-constrained) model by 572 (991) points. All other goodness of fit measures in the table (Rho-square, BIC, and PLL on a sample of 500 households) also suggest that the time- and money-constrained model performs better than the two single-constrained models. A non-nested likelihood ratio test was also conducted to compare the model fit of the time- and money-constrained model with that of the time-constrained model (which has a better log-likelihood value than the money-constrained model). To do so, a naïve time constrained model with only constants in it (with a log likelihood value of -12,483.) was considered as the base. The rho-square values for time-constrained model and the time- and money-constrained models are 0.1492 and 0.1951, respectively with respect to the naïve, time constrained model. The difference between the above adjusted rho-squared values is 0.0459. The probability that this difference could have occurred by chance is less than $\Phi(-\sqrt{-2 \times 0.0459 \times -12,483})$. This value is almost zero, suggesting that the time- and money-constrained model has a better data fit compared to the time-constrained model. All these results suggest the need to consider both the constraints.

Table 3 presents the parameter estimates from the time- and money-constrained model (or multiple-constrained model). The parameter estimates of the single constrained models are not reported here but are available from the authors. Since the time- and money-constrained model performs better than the two single-constrained models in its goodness of fit to estimation data as well as a validation sample, we use the former model to discuss the influence of different factors on households' annual destination and mode choices. While most substantive interpretations of the parameter estimates in the multiple-constrained models are not different from the single-constrained models, wherever appropriate, we discuss the differences in the interpretations from the single-constrained models. The specification of the baseline utility function (ψ_{jl}) is discussed first, followed by the specification of the translation function (γ_j).

The first set of explanatory variables in the baseline utility function have common coefficients across all destination-mode combinations (i.e., inside goods) with the outside goods as the base category (for normalization). Specifically, the constant for all destination-mode combinations is negative suggesting that households spend a smaller proportion of the year on vacation compared to the time spent on all other purposes captured in the outside goods (such as work, sleep, leisure activities pursued closer to the household). This is reasonable because the amount of annual time that a household typically spends on vacation is much less compared to the other time investments to be made in the year. The next variable, leisure employment per capita at the household location captures the influence of opportunities for leisure activities within a closer vicinity of the household (as opposed to long-distance destinations). As expected, the negative coefficient suggests that households living in places with greater leisure opportunities are likely to spend less time on long-distance vacation. This result points to higher substitution between the leisure time spent locally and the time spent on long-distance vacation for households in locations with greater leisure opportunities. While the result is intuitive with a statistically significant coefficient in the time- and money-constrained model, the corresponding coefficient was not statistically significant when only the time constraint was considered.

The second set of variables in the baseline utility function comprises destination-specific characteristics. The interpretations of these variables have reasonable substantive interpretations similar to those discussed in Van Nostrand et al. (2013) who considered only the time constraint. Specially, the logarithm of land-area variable that controls for size differences across the destination choice alternatives has a positive coefficient less than one. This can be explained based on spatial aggregation of several smaller destination alternatives into larger destinations for modeling purposes. As explained by Daly (1982), a smaller than unit coefficient suggests significant heterogeneity across the elemental destination alternatives that comprise the destination alternatives in the model. The positive coefficient on the leisure employment per capita variable reflects a greater attractiveness of destinations with higher leisure opportunities. The dummy variables for the destinations being in the same or adjacent states have positive coefficients reflecting that households are more likely to visit familiar destinations that tend to be within or adjacent to their residential state. The coefficients on the temperature variables during winter and summer suggest that destinations with moderate temperatures (65-75 degree Fahrenheit) are generally more attractive for vacation purposes.

The third set of variables is specific to the travel modes under consideration. The alternative specific constant reflects that households have a general preference to travel by car even after considering the time- and money-constraints and other mode-specific variables in the model. MSA origins and destinations are more attractive for the air mode of travel than the non-MSA origins or destinations, perhaps because of a greater access to the air travel mode in the MSAs. The last variable in this category is the round trip travel time by the alternative modes of travel, whose negative coefficient suggests that households prefer to travel by faster modes of travel. In addition to its influence on mode choice, this variable helps in accommodating that farther destinations are less attractive for vacation compared to closer destinations. Note that mode-specific travel costs are not included as explanatory variables in the time- and money-constrained model, while the travel times are included as explanatory variables.[7] This is because the travel costs are already incorporated into the money-budget constraint through the money-prices (p_{jl}) of travel to the destinations. Such money-prices help in incorporating that farther destinations are more pricy to travel to and hence less likely to be chosen because of the monetary constraint. On the other hand, as discussed earlier, the travel times were not incorporated into the time-prices (q_{jl}). This is because the time-price (q_{jl}) of allocating unit time for a destination has been set to unity assuming that traveling also contributes to the utility derived from vacation (in addition to the utility due to the time spent at the destination).

The next parameter is the scale (σ) of the error terms (ε_{jl}) in the baseline utility functions (ψ_{jl}). This parameter provides a measure of variation in the household preferences due to unobserved factors. The parameter was fixed to 1 in the time-constrained model as it could not be estimated due to the absence of price variation (Bhat, 2008). In the other two models, the parameter could very well be estimated and is significantly different from 1. Specifically the estimate is 0.748 in the money-constrained model (not reported in the table) and 0.576 in the time- and money-constrained model. These estimates suggest that the magnitude of variation in the household preferences due to unobserved factors is lower in the time- and money-constrained model than that in the two single-constrained models. This may be because accounting for both the time and money constraints together helped in capturing a greater proportion of the variation in household preferences.

The next parameter is the dissimilarity parameter (θ). The estimate for this parameter is significantly different from 1 (in all three models) suggesting the significant presence of destination-specific unobserved factors inducing correlations between the baseline utility parameters of the destination-mode combination alternatives that share the same destination. Neglecting such correlations and estimating the destination and mode choice models separately would result in significantly inferior model fit.

The last set of variables correspond to the translation parameters (γ_j), which allow for corner solutions as well as differential satiation effects across different vacation destinations. The

[7] The single, time-constrained model, on the other hand, includes mode-specific travel cost as explanatory variable with a negative coefficient. This is because the time-constrained model does not explicitly consider the money constraint.

positive coefficient on the distance variable in the γ_j function suggests that households tend to allocate greater amount of time for vacation destinations that are farther (than those that are closer). This may be because households might want to spend more time at a destination that is farther from home (if they chose to visit the destination). Besides, it generally takes greater amount of time to travel to father destinations.

Overall, the model estimation results are all reasonable and shed light on the various factors influencing households' annual vacation destination and mode choices and related time and money allocations. The results demonstrate the applicability of the proposed framework for modeling discrete-continuous choices in the presence of a mix of perfect and imperfect substitutes in the choice set. Empirically, the results highlight the need for accommodating both time and money constraints in modeling households' vacation travel choices.

5 SUMMARY AND CONSLUSIONS

Discrete-continuous choices are pervasive in consumer decisions and of interest in a variety of social sciences, including transportation, economics, and marketing. A special case of discrete-continuous choices is the single discrete-continuous (SDC) choice, where consumers choose a single discrete alternative along with the corresponding continuous quantity choice. In such situations, the choice alternatives can be viewed as *perfect substitutes* in that the choice of one alternative precludes the choice of other alternatives. Another special case involves multiple discrete-continuous (MDC) choices, where the choice alternatives are *imperfect substitutes* where consumers can potentially choose multiple discrete alternatives, along with the corresponding continuous quantity choice decisions. The vast majority of choice modeling literature has been focused on analyzing SDC choices, while there has been a recently growing interest in analyzing MDC choices. A more general case of discrete-continuous choices can potentially include both SDC and MDC choices, where consumers choose not more than a single alternative from a subset of alternatives and potentially multiple alternatives from the remaining alternatives. In such situations, the choice set comprises a mix of both *perfect* substitutes (from which not more than one alternative could be consumed) and *imperfect* substitutes (from which potentially multiple alternatives could be consumed).

This paper formulates a unified random utility maximization (RUM) framework that can be used as a joint MDC-SDC modeling framework to analyze discrete-continuous choices from a combination of perfect and imperfect substitutable choice alternatives. The key to this formulation is a utility form that is linear with respect to consumption across perfectly substitutable alternatives and non-linear with respect to consumption across imperfectly substitutable alternatives. The linear utility form ensures single discreteness among the perfect substitutes (i.e., only a single choice alternative is chosen) while the non-linear form accommodates multiple discreteness among imperfect substitutes (i.e., multiple alternatives could be chosen). In addition to the RUM formulation, the paper presents a procedure to apply the proposed framework for forecasting purposes. Further, the formulation can be extended to accommodate multiple linear budget constraints, as opposed to a single budget constraint. To our knowledge, this is the first formulation in the econometric literature to account for multiple linear budget constraints and price variation to model discrete-continuous choices from a combination of perfect and imperfectly substitutable choice alternatives.

As an empirical demonstration, the proposed framework is applied to develop a joint model of annual, long-distance vacation destination and mode choices to simultaneously analyze the vacation destinations that a household visits over an entire year, along with the time allocation and the travel mode to each of the visited destinations. The formulation assumes that, over a year, households allocate a part of the total time (365 days) and money (annual income) available with them to one or more vacation destinations and make the mode choices in such a way as to maximize the utility derived from their choices. The framework recognizes that the vacation destinations are *imperfect substitutes* in that a household can potentially choose to visit multiple destinations over a year, while the travel mode alternatives to a destination are *perfect substitutes* in that only one primary mode is chosen to travel to a destination. Besides, the framework recognizes that households operate under both time and money budget constraints by simultaneously accommodating both the constraints in the model.

The proposed modeling framework is applied to the 1995 American Travel Survey (ATS) data to estimate the empirical model parameters, with the United States divided into 210 alternative long-distance vacation destinations. The ATS data provides information on the different vacation destinations visited (and the time spent on each vacation) by the surveyed households over the time-frame of an entire year. Along with this information from the ATS data, a variety of other data sources, including the Consumer Expenditure Survey (CEX) are used to synthesize information on destination attributes, and lodging costs and other costs of vacation for each of the 210 destinations.

The empirical model estimates are reasonable and shed light on the various factors influencing households' annual vacation destination and mode choices and related time and money allocations. The results demonstrate the applicability of the proposed framework for modeling discrete-continuous choices in the presence of a mix of imperfect and perfect substitutes in the choice set, while considering multiple budget constraints. In addition, the analysis demonstrates the benefit of considering both time and money budget constraints simultaneously in analyzing households' vacation travel choices. Considering the time and money constraints simultaneously lead to a significant improvement (over the single-constrained models) of the model goodness of fit in the estimation sample as well as the predictive performance (as measured by predictive log-likelihood) on a validation sample. Besides, the time- and money-constrained model demonstrated a greater capture of the variation in household preferences than the models that ignored one of the two constraints.

This study can be extended in several important directions. First, development of rigorous constrained maximum likelihood estimation techniques for the proposed formulation would help alleviate some of the estimation difficulties encountered in this study. Second, development of efficient forecasting procedures for the proposed formulation with multiple budget constraints will enable the use of the estimated model for practical forecasting and policy analysis purposes. Third, the empirical application in the current paper does not treat travel costs as *fixed* costs (in that travel costs do not vary with the no. of days spent at the destination). Instead the travel costs are combined with *variable* costs such as lodging and dining costs by assuming that the travel costs could be amortized over the number of vacation days spent at the destinations. Treating travel costs as fixed and different from variable costs makes the consumer's utility maximization problem non-linear and non-smooth with respect to the money budget constraint and makes it difficult to use KKT conditions for solving the problem. Accommodating non-linear and non-smooth budget constraints in random utility maximization-based discrete-continuous choice models is an important avenue for future research (see Parizat and Sachar, 2010 for a recent attempt at this). Fourth, on the empirical front, extending the current study to consider long-distance travel to visiting friends and family and to consider multiple visits to a same destination will be fruitful.

ACKNOWLEDGEMENTS

The first author's efforts on this paper are dedicated to his friend Siva Subramanyam Jonnavithula, a transportation engineer, who passed away in July 2013. Two anonymous reviewers provided valuable comments on a previous version of the paper. One reviewer's comments were particularly crucial in addressing key methodological issues of the formulation. Prof. Naveen Eluru provided helpful discussions on the empirical setup. Prof. Bo Zeng, an optimization expert helped confirm that the proposed utility structure ensures perfect substitutability (among mode choice alternatives available to a destination) as a result of utility maximization.

REFERENCES

Ahn, J., G. Jeong, and Y. Kim (2008). A forecast of household ownership and use of alternative fuel vehicles: a multiple discrete-continuous choice approach. *Energy Economics*, 30, 2091–2104.

Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42, 999-1012.

Bhat, C.R. (2005). A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B*, 39(8), 679-707.

Bhat, C.R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B*, 42(3), 274-303.

Bhat, C.R., S. Srinivasan, and S. Sen. (2006). A joint model for the perfect and imperfect substitute goods case: application to activity time use decisions. *Transportation Research Part B*, 40(10), 827-850.

Bhat, C.R., S. Sen, and N. Eluru (2009) The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B*, 43(1), 1-18.

Bhat, C.R., M. Castro, and A.R. Pinjari (2013a). Allowing for non-additively separable and flexible utility forms in multiple discrete-continuous models. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Bhat, C.R., M. Castro, and M. Khan (2013b). A new estimation approach for the multiple discrete-continuous probit (MDCP) choice model. Forthcoming, *Transportation Research Part B*.

Bockstaeal, N.E, Strand I.E., Jr., and M.W. Hanemann (1987). Time and the recreational demand model. *Land Economics*, 69(2), 293-302.

Bureau of Economic Analysis (1995). Gross Domestic Product by State. Retrieved January 10, 2011, from http://www.bea.gov/regional/gsp/

Bureau of Labor Statistics (1995). Employment, Hours, and Earnings—State and Metro Area. Retrieved December 2010, from http://www.bls.gov/data/#employment

Bureau of Transportation Statistics (1995a). 1995 American Travel Survey Technical Documentation. Washington, DC.

Bureau of Transportation Statistics (1995b). Airline Origin and Destination Survey (DB1B). Washington, DC.

Bureau of Transportation Statistics (1997). 1995 American Travel Survey Profile, from http://www.bts.gov/publications/1995_american_travel_survey/us_profile/entire.pdf

Bureau of Transportation Statistics (2001). NHTS – Long Distance Travel Quick Facts. http://www.bts.gov/programs/national_household_travel_survey/long_distance.html

Castro, M., C.R. Bhat, R.M. Pendyala, and S. Jara-Diaz (2012). Accommodating multiple constraints in the multiple discrete–continuous extreme value (MDCEV) choice model. *Transportation Research Part B*, 46(6), 729-743.

Chiang, J. (1991). A simultaneous approach to whether to buy, what to buy, and how much to buy. *Marketing Science*, 10(4), 297-314.

Chintagunta, P.K. (1993). Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12(2), 184-208.

Daly, A. (1982). Estimating choice models containing attraction variables. *Transportation Research Part B, 16*(1), 5-15

Deaton, A., and J. Muellbauer (1980). Economics and Consumer Behavior. Cambridge University Press, Cambridge.

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

Edwards, Y.D., and G.M. Allenby (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research*, 40, 321-334.

Eluru, N., A.R. Pinjari, R.M. Pendyala, and C.R. Bhat (2010). An Econometric Multi-Dimensional Choice Model of Activity-Travel Behavior. *Transportation Letters*, 2(4), 217-230.

Energy Information Administration (EIA) (1995). Retail Gasoline Historical Prices. Retrieved December 2010, from http://www.eia.doe.gov/oil_gas/petroleum/data_publications/wrgp/mogas_history.html

Eugenio-Martin, J. (2003). Modelling determinants of tourism demand as a five stage process: A discrete choice methodological approach. *Tourism and Hospitality Research*, 4(4), 341.

Fang, H.A. (2008). A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B,* 42(9), 736-758.

Fosgerau, M. (2001). PETRA—an activity based approach to travel demand analysis. In: Lundqvist, L., Mattsson, L.G. (eds.) National Transport Models: Recent Developments and Prospects. The Swedish Transportand Communications Research Board, Stockholm

Habib, K.M.N. and E.J. Miller (2008). Modeling daily activity program generation considering within-day and day-to-day dynamics in activity-travel behaviour.*Transportation*, 35(4), pp. 467-484.

Hanemann, M.W. (1978). A methodological and empirical study of the recreation benefits from water quality improvement. Ph.D. dissertation, Department of Economics, Harvard University.

Hanemann, M.W. (1984). Discrete/Continuous models of consumer demand. *Econometrica*, 52(3), 541-562.

Heckman, J.J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, 42(4),. 679-694

Heckman, J.J. (1979) Sample selection bias as a specification error, *Econometrica,* 47(1), 153-161.

Herriges, J.A., Kling, C.L., and D.J. Phaneuf (2004). What's the use? Welfare estimates from revealed preference models when weak complementarity does not hold. *Journal of Environmental Economics and Management*, 47, 55-70.

Horowitz, A.J. (2008). Statewide Travel Demand Forecasting – Meeting Federal Surface Transportation requirements in Statewide and Metropolitan Transportation Planning. *American Association of State Highway and Transportation Officials*.

Inter University Consortium for Political and Social Research (ICPSR), (2011): Bureau of Labor Statistics Consumer Expenditure Survey. (1995). Retrieved August 29, 2011 from http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp

Jaggi, B., C. Weis, and K.W. Axhausen (2011). Stated Response and Multiple Discrete-Continuous Choice Models: Analysis of Residuals. Technical Paper, Department of Civil and Environmental Engineering, ETH, Zuruch..

Kim, J., G.M. Allenby, and P.E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science*, 21(3), 229-250.

Koppelman, F., and V. Sethi (2005). Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior. *Transportation Research Part B: Methodological*. 39(9), 825-853.

Kuriyama, K., Y. Shoji, and T. Tsuge. (2011) Recreation demand models with spatial heterogeneity: a spatial Kuhn-Tucker model. European Association of Environmental and Resource Economists, 18th Annual Conference, Rome.

LaMondia, J., Bhat, C., and D. Hensher (2008). An annual time use model for domestic vacation travel. *Journal of Choice Modelling*, 1(1), 70-97.

Lee, S., Kim. J. and Allenby, G.M. (2010). A category-level model of asymmetric complements. Fisher College of Business Working Paper. Available at SSRN: http://ssrn.com/abstract=1596218 or http://dx.doi.org/10.2139/ssrn.1596218.

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

Lee, L-F., and M.M. Pitt (1986). Microeconometric Demand System with Binding Nonnegativity Constraints: The Dual Approach. *Econometrica* , Vol. 54, No. 5 (Sep., 1986), pp. 1237-1242

Maddala, G.S. (1983), Limited-dependent and qualitative variables in Economics, New York: Cambridge University Press

Mäler, K-G., (1974). Environmental Economics: A Theoretical Inquiry. The Johns Hopkins University Press for Resources for the Future, Baltimore, MD.

Manchanda, P., Ansari, A., and S. Gupta (1999). The “shopping basket”: A model for multicategory purchase incidence decisions. *Marketing Science*, 18, 95-114.

Morley, C.L. (1992). A microeconomic theory of international tourism demand. *Annals of Tourism Research*. 19(2), 250-267.

Parizat, S., and R. Shachar (2010).When Pavarotti meets Harry Potter at the Super Bowl?. Working paper, Tel Aviv University.

Phaneuf, D.J., C.L. Kling, and J.A. Herriges (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. *The Review of Economics and Statistics*, Vol. 82, No.1 , pp. 83-92.

Phaneuf, D.J., and V.K. Smith (2005). Recreation demand models. In Handbook of Environmental Economics, Vol. 2, K-G. Mäler and J.R. Vincent (Eds.), North Holland.

Pinjari, A.R., and C.R. Bhat (2011). computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. Technical paper, Department of Civil and Environmental Engineering, University of South Florida.

Pinjari, A.R. (2011). Generalized Extreme Value (GEV)-Based Error Structures for Multiple Discrete-Continuous Choice Models. *Transportation Research Part B*, Vol. 45, Issue. 3,, pp. 474-489.

Rich, J., J. Brocker, C. Hanson, A. Korchenewych, O. Nielsen, and G. Vuk (2009). Report on scenario, traffic forecast and analysis of traffic on the TEN-T, taking into consideration the external dimension of the Union-Trans Tools version 2.

Satomura, T., Kim J., and G.M. Allenby (2011). Multiple-constraint choice models with corner and interior Solutions. *Marketing Science*, 30(3), 481-490.

Savageau, D., Loftus, G. (1997). Places Rated Almanac: Macmillan General Reference.

Srinivasan, S., Bhat C.R. (2006). A multiple discrete-continuous model for independent- and joint-discretionary-activity participation decision. *Transportation*, 33(5), 497-515.

U.S. Census Bureau (2000). Demographic profiles: 100-Percent and Sample Data. http://www.census.gov/census2000/demoprofiles.html

van Middelkoop, M., Borgers, A., and H.J.P. Timmermans (2004). Merlin: Microsimulation system for predicting leisure activity travel patterns. *Transportation Research Record* 1894, 20-27.

Van Nostrand, C. (2011). A Discrete-Continuous Modeling Framework for Long-Distance, Leisure Travel Demand Analysis. Master's Degree in Civil Engineering, University of South Florida, Tampa.

Van Nostrand, C., Sivaraman, V. and A.R. Pinjari (2013). Analysis of long-distance vacation travel demand in the United States: a multiple discrete-continuous choice framework. *Transportation*, 40(1), 151-171.

Vasquez-Lavin, F., and M.W. Hanemann (2009). Functional Forms in Discrete/Continuous Choice Models with General Corner Solutions. Proceedings of the 1st International Choice Modeling Conference, Harrogate..

von Haefen, R.H (2007). Empirical Strategies for Incorporating Weak Complementarity into Consumer Demand Models. *Journal of Environmental Economics and Management*, 54(1), 15-31.

von Haefen, R.H., D.J. Phaneuf, and G.R. Parsons (2004). Estimation and welfare analysis with large demand systems. *Journal of Business and Economic Statistics*, 22(2), 194-205.

von Haefen, R.H. and D.J. Phaneuf (2005), 'Kuhn-Tucker demand system approaches to nonmarket valuation', in R. Scarpa and A. Alberini (eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Dordrecht, The Netherlands: Springer, pp. 135-158.

Wales, T.J., and A.D. Woodland (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21(3), 263-285.

Table 1: Descriptive Statistics for Estimation Data

Household Socio-demographic Characteristics (in the estimation sample of 1000 households)		
Household size	Average: 2.86	Std. Dev: 1.35
Age of householder (years)	Average: 45.6	Std. Dev. : 14.08
Household yearly income	Average: $50,694	Std. Dev: $30,544
Presence of Kids	38.4%	
Householder is Retired	12.6%	
Household Leisure Travel Characteristics (in the estimation sample of 1000 households)		
Number of long distance leisure trips	Average: 2.2	Std. Dev: 2.07
1	51.6%	
2 or more	48.4%	
Number of destinations visited	Average: 1.58	Std. Dev: 0.93
1	62.7%	
2	23.8%	
3or more	13.5%	
Number of trips made to a destination	Average: 1.39	Std. Dev: 1.20
1	80.7%	
2 or more	19.3%	
Total Annual Vacation Time (Days)	Average: 6.40	Std. Dev: 5.40
Total Annual Expenditure on Vacation	Average: $693	Std. Dev: $672
Trip-level Characteristics (for 1530 leisure trips made by the 1000 households)		
Primary mode of transportation	Auto: 84.7%	Air: 15.3%
Round trip Ground Distance(miles)	Average: 990	Std. Dev: 1,116
No. of nights away from home on trip	Average: 3.84	Std. Dev: 3.19
Monetary Expenditure	Average: $437	Std. Dev: $400
Destination Characteristics (for 210 Destinations)		
Destination is an MSA	76.70%	
Ln (LandArea in square miles)	Average: 5.92	Std. Dev: 2.88
Leisure Employment Per Capita	Average: 0.12	Std. Dev: 0.14
Winter Temperature (Fahrenheit)	Average: 42.3	Std. Dev: 16.53
Summer Temperature (Fahrenheit)	Average: 82.03	Std. Dev: 8.79
Level of Service Characteristics (between 210 x 210 OD pairs)		
Highway Distance (Roundtrip)	Average: 2,622	Std. Dev: 1,749
Auto Travel Time (hours)	Average: 23	Std. Dev: 12.25
Air Travel Time (hours)	Average: 4.5	Std. Dev: 2.96
Auto Travel Cost (US dollars)	Average: $154.24	Std. Dev: $105.74
Air Travel Cost (US Dollars)	Average: $437.76	Std. Dev: $301.69

Table 2: Goodness of Fit Measure for Models Estimated in the Study

	Log-likelihood (LL) at model convergence (LL)	No. of Parameters (K)	Rho-square (ρ^2) $1-\frac{LL}{LL(C)}$	BIC $-2LL+\ln(N).K$	Predictive *LL* for 500 households (PLL)
1.Time-constrained model for destination and mode choices	-10,620	21	0.14293	21,385	-6,484
2. Money-constrained model for destination and mode choices	-11,040	21	0.16368	22,225	-6,904
3.Time and Money constrained model for destination and mode choices	-10,048	21	0.17203	20,241	-6,114

Note:
LL = Log-likelihood at model convergence
LL(C) = Log-likelihood with only constants in the model
K = No of parameters in the model
Rho-square (ρ^2) = $1-\{LL / LL(C)\}$
Bayesian Information Criterion (BIC) = $-2LL+\ln(N).K$

Table 3 : Parameter Estimates of the Time- and Money-Constrained Model

	Parameter Estimate	Std. Err.
Baseline Utility Function (ψ_{jl}) Specification		
Constant for all destination-mode combinations	**-8.614**	0.110
Leisure Employment Per Capita at household location	**-0.330**	0.171
Destination Specific Characteristics (Z_j)		
Log of Land Area	**0.170**	0.006
Leisure Employment Per Capita at Destination	**1.663**	0.088
Dummy if destination in same state as HH residence	**1.810**	0.053
Dummy if destination in adjacent state to HH residence	**1.156**	0.044
Winter (January) temperature. 65^o-75^o Fahrenheit is base		
55^o-65^o Fahrenheit	**-0.162**	0.056
45^o-55^o Fahrenheit	**-0.351**	0.058
$< 45^o$ Fahrenheit	**-0.378**	0.066
Summer (June) temperature. 65^o-75^o Fahrenheit is base		
60^o to 65^o Fahrenheit	**-0.377**	0.153
75^o to 80^o Fahrenheit	**-0.274**	0.059
80^o to 85^o Fahrenheit	**-0.251**	0.055
$> 85^o$ Fahrenheit	**-0.310**	0.059
Mode Specific Characteristics (Z_{jl})		
Alternative specific constant (Mode - auto as base)	**-0.370**	0.036
Origin is an MSA – on the air mode (auto is base)	**0.085**	0.016
Destination is an MSA – on the air mode (auto is base)	**0.176**	0.018
Round trip travel time (in days)	**-0.332**	0.023
***Scale parameter σ* of the baseline utility function**	**0.576**	0.011
Dissimilarity Parameter θ	**0.144**	0.012
Satiation Function (γ_j) Specification		
Alternative specific constant	**-3.242**	0.010
Highway distance to destination (100's miles)	**0.007**	0.001

Appendix-A: Procedures Employed to Synthesize Price Variables

There are two types of unit prices for each vacation destination and travel mode alternative – time-prices and money-prices (i.e., the q_{jl} and p_{jl} variables in the time and money constraints of equation (28)). For the current analysis, the time-prices (q_{jl}) are considered to be unity in that the amount of time needed to spend one day of vacation time is equal to one day. This makes an implicit assumption that the time spent traveling to a destination *j* is part of the vacation time t_j. That is, households derive utility not only from the time spent at a vacation destination, but also from the time spent traveling to the destination. This is reasonable because traveling for vacation might not be as onerous (it might in fact be fun) as compared to commuting. However, doing so does not account for the possibility that households tend to prefer to visit closer destinations as opposed to farther destinations. To account for such preferences, the baseline utility functions incorporate the travel time to the destination (by the corresponding mode) as an explanatory variable. One would expect a negative coefficient on this variable. The synthesis of money-prices (p_{jl}), on the other hand, required several assumptions and significant data gathering and processing, as described below.

The money-price p_{jl} is the monetary expenditure a household needs to incur to spend unit time (i.e., a day) at a vacation destination *j* traveled by mode *l* (note that the subscript for the household is suppressed for simplicity in notation). These prices comprise two components – (a) destination prices that do not depend on the mode of travel and (b) travel prices that depend on the mode of travel. The destination prices, in turn, have two components – (a1) lodging prices (i.e., lodging costs per day) and (a2) non-lodging prices (i.e., costs per day for dining, recreation, entertainment, etc.). The process used to synthesize the information on money-prices for each household to travel to each available destination by each available travel mode is described below.

First, the lodging costs and non-lodging costs per day at each destination were synthesized from the 1995 Consumer Expenditure Survey (CEX) data using a two stage process. In the first stage, the per-day lodging costs for each household was derived using a regression equation relating the per-day costs to the household's socio-demographic characteristics (income, household size, and residential Census region). This regression equation was estimated using household-level microdata on annual vacation expenditures from the 1995 Consumer

Expenditure Survey (CEX) data (see Table A1). Similarly, the per-day non-lodging costs were derived using another regression equation estimated with the CEX data on non-lodging vacation expenditures (see Table A2). Both the above mentioned regression equations recognize the variation in per-day costs by household characteristics. Thus, this approach recognizes that not every household incurs the same costs at a destination. Rather, households make the lodging choices and other expenditure choices according to their income and other characteristics. However, the regression equations do not recognize the variation in the lodging and non-lodging prices across the different destinations (because the CEX data does not provide information on which destinations were visited by the households). To accommodate the price-variation across destinations, in the second stage, the regressed per-day costs for each household were scaled by a factor capturing how pricy (or less expensive) each destination is compared to an average destination (as measured by the median per-day costs at different destinations. To implement this second state strategy, the median values of lodging costs of vacationing at each of the 210 destinations were obtained from a hotel guide database made available by VisitUSA.com (http://www.visitusa.com/state-hotels/index.htm). The lodging prices and non-lodging prices obtained in the above manner were added up to obtain the destination prices. Call such destination price as p_j, where *j* is the index for destination.

Second, using the 1995 ATS data, the number of days spent at a destination were regressed, using an ordered logit model, as a function of the household characteristics (age of householder, household size, income, presence of children), distance between origin and destination, and an indicator if the destination is an MSA (see Table A3). The resulting ordered response model estimates were used, for each household in the estimation sample, to estimate the expected number of days (n_j) that the household would spend at each destination ($j = 1,2,\ldots,J$) if the household visited that destination. Third, the money-price p_{jl} of spending a day visiting a destination *j* by mode *l* was computed as: $p_{jl} = \frac{p_j n_j}{(n_j + tt_{jl})} + \frac{tc_{jl}}{(n_j + tt_{jl})}$, where tc_{jl} and tt_{jl} are the round trip travel cost and travel time, respectively, to travel to a destination *j* (from the household's origin) by travel mode *l*. The first component of this money-price formula can be viewed as the destination price, while the second component can be viewed as the travel price. Note from the formula for p_{jl} that the money-prices are computed assuming that the travel costs

(tc_{jl}) can be amortized over the duration spent visiting a destination ($n_j + tt_{jl}$). In reality, however, travel costs are fixed costs and do not depend on the number of days spent at the destination. The current formulation cannot consider such fixed costs, a reason why we assumed that travel costs could be amortized over the no. of days spent at the destination. Enhancing the model formulation to relax this assumption and consider travel costs as *fixed* is an important avenue for further research.

Table A1: Regression Estimates for Lodging Costs per Day on Vacation (using CEX Data)

Explanatory Variables	Coefficient	t-stat
Constant	58.922	2.85
Region of Residence : Midwest (Northeast as base)	-7.473	2.77
Region of Residence : West (Northeast as base)	-5.787	2.78
HH Income < $30K (high income is base category)	-15.743	2.65
HH Income between $30K and < 75K (high income is base category)	-13.134	2.53
Household size < 3 (3 or more member household as base)	-8.473	1.91

Note: The dependent variable, lodging costs per day was derived as the total annual lodging expenditure on vacation divided by the total number of days per annum on vacation

Table A2: Regression Estimates for Non-Lodging Costs per Day on Vacation (using CEX Data)

Explanatory Variables	Coefficient	t-stat
Constant	74.464	4.61
Region of Residence : Midwest (Northeast is base category)	-10.417	4.12
Region of Residence : West (Northeast is base category)	-9.432	4.14
HH Income < $30K (high income is base category)	-19.984	3.93
HH Income between $30K and < 75K (high income is base category)	-17.121	3.76
Household size between 3 and 4 (1-2 member household is base)	11.813	3.07
Household size greater than 5 (1-2 member household is base)	20.603	4.49

Note: The dependent variable, non-lodging costs per day was derived as the total annual expenditure on food, drinks, recreation, entertainment and other local expenditures (other than lodging) during vacation divided by the total number of days per annum on vacation.

Table A3 Ordered Logit Model for No. of Days Spent at Destination (using ATS Data)

Explanatory Variables	Coefficient	t-stat
Age of householder	0.009	3.66
Household size*Low income household	-0.117	-4.22
Presence of Kids	0.187	2.74
Distance to the destination	0.044	16.02
Destination is an MSA	-0.494	-7.72

Note: Estimates of the thresholds in the ordered logit model are not reported as they do not carry significant interpretation.