# Generalized Random Utility Model

by

Joan Walker[♦] and Moshe Ben-Akiva[◂]

for publication in the

*RU2000 Special Issue of Mathematical Social Sciences*

February 2001
Revised February 2002

## Abstract

Researchers have long been focused on enriching Random Utility Models (RUMs) for a variety of reasons, including to better understand behavior, to improve the accuracy of forecasts, and to test the validity of simpler model structures. While numerous useful enhancements exist, they tend to be discussed and applied independently from one another. This paper presents a practical, generalized model that integrates many enhancements that have been made to RUM. In the generalized model, RUM forms the core, and then extensions are added that relax simplifying assumptions and enrich the capabilities of the basic model. The extensions that are included are:

- *Flexible Disturbances* in order to allow for a rich covariance structure and enabling estimation of unobserved heterogeneity through, for example, random parameters;
- *Latent Variables* in order to provide a richer explanation of behavior by explicitly representing the formation and effects of latent constructs such as attitudes and perceptions;
- *Latent Classes* in order to capture latent segmentation in terms of, for example, taste parameters, choice sets, and decision protocols; and
- *Combining Revealed Preferences and Stated Preferences* in order to draw on the advantages of the two types of data, thereby reducing bias and improving efficiency of the parameter estimates.

The paper presents a unified framework that encompasses all models, describes each enhancement, and shows relationships between models including how they can be integrated. These models often result in functional forms composed of complex multidimensional integrals. Therefore, an estimation method consisting of Simulated Maximum Likelihood Estimation with a kernel smooth simulator is reviewed, which provides for practical estimation. Finally, the practicality and usefulness of the generalized model and estimation technique is demonstrated by applying it to a case study.

*Keywords:* random utility model, choice model, choice behavior, discrete choice analysis, logit kernel, mixed logit, flexible disturbances, latent variables, latent classes, combining datasets

[♦] Massachusetts Institute of Technology and Caliper Corporation, 1172 Beacon Street, Newton, MA 02461, jwalker@alum.mit.edu

[◂] Massachusetts Institute of Technology, Room 1-181, 77 Massachusetts Avenue, Cambridge, MA 02139, mba@mit.edu

## *Introduction*

There has been much research aimed at improving the specification of the random utility (or discrete choice) model, which models a decision-maker's choice among a set of mutually exclusive alternatives. A guiding philosophy in these developments is that such enhancements lead to a more behaviorally realistic representation of the choice process, and consequently a better understanding of behavior, improvements in forecasts, and valuable information regarding the validity of simpler model structures. A number of important extensions have been presented in the literature, including:

- *Flexible Disturbances*
- *Latent Variables*
- *Latent Classes*
- *Combining Revealed Preferences and Stated Preferences*

This paper develops a generalized random utility model that integrates these extensions. In the course of the presentation of the generalized model, a brief review of each of the individual enhancements is also provided. The generalized models often result in functional forms composed of complex multidimensional integrals. Therefore, an estimation method consisting of Maximum Simulated Likelihood Estimation and a kernel probability simulator is reviewed, which provides for practical estimation. Finally, the generalized random utility model is used in a case study to demonstrate its practicality and effectiveness.

## *The Random Utility Model*

The framework for the standard random utility model is shown in Figure 1. The model is based on the notion that an individual derives utility by choosing an alternative. The *utilities* $U$ are latent variables, and the observable *preference indicators* $y$ are manifestations of the underlying utilities. The utilities are assumed to be a function of a set of *explanatory variables* $X$, which describe the *decision-maker* $n$ and the *alternative* $i$. The resulting utility equation can be written as:

$$U_{in} = V(X_{in};\beta) + \varepsilon_{in} \ , \tag{1}$$

where:

$U_{in}$ is the utility of alternative $i$ $[i = 1,..., J_n]$ for decision-maker $n$ $[n = 1,..., N]$ ($U_n$ is a vector of utilities for decision-maker $n$),

$X_{in}$ is a vector of explanatory variables describing alternative $i$ and decision-maker $n$ ($X_n$ is a matrix of explanatory variables describing all alternatives and decision-maker $n$),

$\beta$ is a vector of unknown parameters,

$V$ (called the *systematic utility*) is a function of the explanatory variables and unknown parameters $\beta$, and

$\varepsilon_{in}$ is a random disturbance for $i$ and $n$ ($\varepsilon_n$ is the vector of random disturbances, which is distributed $\varepsilon_n \sim D(\theta_\varepsilon)$, where $\theta_\varepsilon$ are unknown parameters).

RUM assumes utility maximization:

$$\text{Decision-maker } n \text{ chooses } i \text{ if and only if } U_{in} \geq U_{jn} \text{ for all } j \in C_n, \quad (2)$$

where $C_n$ is the set of $J_n$ alternatives faced by $n$.

The choice probability equation is then:

$$P(i \mid X_n; \beta, \theta_\varepsilon) = \text{Prob}[U_{in} \geq U_{jn}, \forall j \in C_n]. \quad (3)$$

Employing classical techniques, estimation involves maximizing the likelihood of the preference indicators ( $y_n$ ). The likelihood is derived from the structural equation, in this case the utility equation (1) , and the measurement equation, which defines $y_n$ as a function of the utilities via the utility maximization equation (2). For example, the measurement equation for choice data is: $y_{in} = 1$ for the chosen alternative and 0 otherwise for all $i \in C_n$, which leads to the following likelihood:

$$P(y_n \mid X_n; \beta, \theta_\varepsilon) = \prod_{i \in C_n} P(i \mid X_n; \beta, \theta_\varepsilon)^{y_{in}}. \quad (4)$$

The formulation of the RUM is grounded in classic microeconomic consumer theory; brings in the random utility paradigm pioneered by Thurstone (1927), Marshak (1960), and Luce (1959); and incorporates the manner of specifying utilities developed by Lancaster (1966) and McFadden (1974). Starting from this general formulation, assumptions on the distributions of the disturbances lead to various choice models, most commonly the probit or General Extreme Value[1] (GEV) forms such as logit and nested logit.

## *The Generalized Random Utility Framework*

The generalized random utility model is an integration of various enhancements that have been made to the RUM paradigm. In contrast to the basic RUM framework shown in Figure 1, the framework for the generalized model is provided in Figure 2. First note that the basic RUM (highlighted) forms the core (or kernel) of the generalized model. Extensions are then added to the kernel that relax simplifying assumptions and enrich the capabilities of the model. The extensions include the addition of *flexible disturbances* to be able to reproduce any desirable disturbance structure (for example, relax the IIA condition[2] of logit or specify random parameters), the explicit modeling of latent psychological explanatory variables such as attitudes and perceptions (*latent variables*), and the inclusion of latent segmentation of the population (*latent classes*). Because it would be difficult to estimate such a model with *revealed preferences* (i.e., market behavior) alone, other indicators are incorporated into the framework to aid in estimation of the behavioral relationships. These additional indicators include both *stated preferences* as well as psychometric measurements (or *indicators*) to provide information on the latent constructs (for example, survey questions regarding attitudes and perceptions). As will become

[1] A large class of models derived by McFadden (1978).

[2] See Ben-Akiva and Lerman (1985) for a discussion of IIA.

apparent, this formulation of the generalized model offers complete flexibility, enables straightforward combination of models, and is computationally feasible.

This generalized model is the integration of a number of extensions that have been made to the random utility model. It draws ideas from a great number of researchers, including, among others, McFadden (1984), Bolduc and Ben-Akiva (1991), and McFadden and Train (2000) who introduced flexible disturbances to the logit model; Cambridge Systematics (1986), McFadden (1986), Ben-Akiva and Boccara (1987) and Morikawa, Ben-Akiva, and McFadden (2002) who laid out the original ideas for incorporating latent variables and psychometric data into choice models; Kamakura and Russell (1989) who pioneered latent class models; Gopinath (1995) who developed rigorous and flexible models for capturing latent class segmentation in choice models; and Ben-Akiva and Morikawa (1990) who developed the models for combining revealed and stated preferences.

## *The Components of the Generalized Random Utility Model*

Before discussing further the generalized model, each of the components that make up the model is highlighted below. The objective is to cover the motivation and basic equations for each extension independent of the others, which will aid in presenting the formulation of the generalized model later. Only brief overviews are provided here; references are provided throughout and the case study presented later provides concrete examples of the equations.

### *Component 1: The Kernel*

As described above, RUM forms the kernel of the generalized model. The idea is that the kernel represents the straightforward, raw-bones formulation of the choice model. The kernel could be of any number of forms. Whatever the chosen form, it will have a utility equations of the form presented in Equations (1), a measurement equation relating the preference indicators ( $y_n$ ) to the utilities via a utility maximization equation, choice probability functions as in Equation (3), and a likelihood function.

A key to the selection of the kernel for a generalized RUM is that it be relatively easy to compute. Therefore a GEV kernel (for example, logit or nested logit) is the most likely candidate. For the description that follows, logit kernel is often employed. Taking this as an example, the disturbances ( $\varepsilon_n$ ) are assumed to be i.i.d. Extreme Value[3] with scale parameter $\mu$ [4], which leads to the logit probability equation:

$$P(i \mid X_n; \beta, \mu) = \frac{e^{\mu V(X_{in};\beta)}}{\sum_{j \in C_n} e^{\mu V(X_{jn};\beta)}}. \qquad (5)$$

[3] See Ben-Akiva and Lerman (1985) for the form and properties of the distribution.

[4] Per convention, the special notation for the scale parameter of logit is used rather than the generic disturbance parameter $\theta$ as defined in Equation (1).

However, note that the form of the kernel depends on the data at hand. In the case study, the preference indicators are ordinal data, and so the kernel is an ordinal logit model. Additionally, it may be computationally advantageous to use a more complex kernel such as nested logit. There is a trade-off between building the complexity in the kernel, which can have computational advantages but makes the likelihood more complicated and less flexible, or building the complexity through the flexible disturbances as described in the next section.

### *Component 2: Flexible Disturbances and Simulated Probabilities*

The limitations of the commonly used and easy to estimate GEV models derive from the rigidity of the disturbance structure. Conversely, RUMs with more flexible disturbance structures (for example, probit) can be computationally burdensome. One remedy to these issues is to develop a model that combines the advantages of both. That is, a random utility model that has both a highly flexible disturbance component (for example, probit-like disturbances) as well as an additive Extreme Value disturbance. The former allows for a rich covariance structure and the latter aids in computation. The elements of the generalized model related to this extension are highlighted in Figure 3: simply the RUM kernel plus flexible disturbances. The RUM equations in (1) and (2) are extended by assuming that $\varepsilon_n$ has a simple structure, say as i.i.d. Extreme Value, and introducing a flexible disturbance term, denoted $\xi_n$. Like $\varepsilon_n$, $\xi_n$ has a distribution (different than $\varepsilon_n$) and set of unknown parameters: $\xi_n \sim D(\theta_\xi)$. The utility is then written as:

$$U_n = V(X_n, \xi_n; \beta) + \varepsilon_n . \tag{6}$$

If the flexible disturbances $\xi_n$ are given and $\varepsilon_n$ are i.i.d. Extreme Value, the probability conditional on $\xi_n$ is a logit formulation as in Equation (5):

$$P(i \mid X_n, \xi_n; \beta, \mu) = \frac{e^{\mu V(X_{in}, \xi_{in}; \beta)}}{\sum_{j \in C_n} e^{\mu V\left(X_{jn}, \xi_{jn}; \beta\right)}} , \tag{7}$$

or, more generally for any assumption on $\varepsilon_n$:

$$P(i \mid X_n, \xi_n; \beta, \theta_\varepsilon) . \tag{8}$$

This must be integrated over $\xi_n$ to obtain the unconditional choice probability of interest:[5]

$$P(i \mid X_n; \beta, \theta_\varepsilon, \theta_\xi) = \int P(i \mid X_n, \xi; \beta, \theta_\varepsilon) f(\xi \mid \theta_\xi) d\xi , \tag{9}$$

where: $f(\xi \mid \theta_\xi)$ is the joint density function of $\xi_n$.

[5] Throughout the paper, the subscript $n$ is used to denote a particular realization of a random variable (as in $\xi_n$), and this subscript is removed when the variable is a variable of integration (as in $\xi$ of Equation (9)).

The advantage of this formulation is that $P(i \mid X_n; \beta, \theta_\varepsilon, \theta_\xi)$ can be naturally estimated with a tractable, unbiased and smooth simulator (McFadden and Train, 2000), which is computed as:

$$\hat{P}(i \mid X_n; \beta, \theta_\varepsilon, \theta_\xi) = \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} P(i \mid X_n, \xi_n^d; \beta, \theta_\varepsilon), \qquad (10)$$

where $\xi_n^d$ denotes draw $d$ from the distribution of $\xi_n$, given parameters $\theta_\xi$.

This probability simulator enables the estimation of high dimensional integrals with relative ease, and serves a key role in maintaining computability of the Generalized RUM. The advantage to this formulation is that nearly any disturbance can be specified via $\xi_n$. For example, normally distributed random parameters can be specified by replacing any $\beta$ with $(\beta + \xi_n)$ where $\xi_n \sim Normal(0, \theta_\xi)$. The case study provides an example of specifying lognormally distributed random parameters, and Hensher and Greene (2001) and Train (2001) describe other types of distributions. Ben-Akiva, Bolduc, and Walker (2001) propose a factor analytic form for $\xi_n$, and show how this general representation can be trivially used to specify heteroscedasticity, nested structures, cross-nested structures, random parameters, and auto-regressive processes. Furthermore, McFadden and Train (2000) proved that any RUM can be approximated by this formulation.

The literature currently only includes discussion of such models with a logit kernel (also known as mixed logit). The earliest presentations of estimated logit kernel models were in random parameter logit specifications, which appeared 20 years ago in the papers by Boyd and Mellman (1980) and Cardell and Dunbar (1980). The more general form of the model came about through the quest for smooth probability simulators for use in estimating probit models (McFadden, 1989), and developed into its logit kernel form by Bolduc and Ben-Akiva (1991) and the powerful mixed logit approximation theorem of McFadden and Train (2000). The model has recently grown in popularity, including appearing in a popular econometrics textbook (Greene, 2000). Most of the of estimated logit kernel models presented in the literature are in the area of random parameters (for example, Brownstone and Train, 1999, Goett, Hudson, and Train, 2000, Mehndiratta and Hansen, 1997, and Revelt and Train, 1998), but there have also been logit kernel models incorporating heteroscedasticity (Ben-Akiva and Bolduc, 1996, and Greene, 2000), nesting (Ben-Akiva and Bolduc, 1996), cross-nesting (Bhat, 1997), dynamics (Srinivasan and Mahmassani, 2000), and auto-regressive processes (Bolduc, Fortin and Fournier, 1996).

### *Component 3: Latent Variables*

Often in behavioral sciences, there are concepts of interest that are not well defined and cannot be directly measured, for example knowledge, ambition, or personality. These concepts are referred to as latent constructs. While there exists no operational models to directly measure these constructs, latent variable modeling techniques (advanced by a number of researchers including Keesling (1972), Jöreskog (1973), Wiley (1973), and Bentler (1980)) are often applied to infer information about latent variables. These techniques are based on the hypothesis that although the construct itself cannot be observed, its effects on measurable variables (called *indicators*) are observable and such relationships provide information on the underlying latent variable.

The objective for this component of the generalized model is to explicitly incorporate the psychological factors, such as attitudes and perceptions, affecting the utility by modeling them as latent variables. Various approaches have been used to achieve this objective (see Ben-Akiva, Walker, et al., 1999, for a discussion), and there have been numerous applications in specific behavioral models (as referenced below). The general formulation described here was presented in Ben-Akiva, Walker, et al. (1999).

The portions of the generalized model related to integrating choice and latent variable models are highlighted in Figure 4. As shown in the figure, the model has two components, a choice model and a latent variable model. The choice model is like any standard choice model, except that now some of the explanatory variables are not directly observable. The notation $X_n^*$ is used to denote these unobservable explanatory variables, and the utility equation for the choice model is then:

$$U_n = V(X_n, X_n^*; \beta) + \varepsilon_n . \quad (11)$$

If the latent variables were given, the probability of $y_n$ conditional on $X_n^*$ would be:

$$P(y_n \mid X_n, X_n^*; \beta, \theta_\varepsilon) . \quad (12)$$

This must be integrated over the distribution of the latent variables to obtain the unconditional probability of interest. This requires the latent variable structural model:

$$X_n^* = X^*(X_n; \lambda) + \omega_n , \quad (13)$$

which describes the latent variable ( $X_n^*$) as a function of observable explanatory variables ( $X_n$ ), a set of parameters ( $\lambda$ ), and a disturbance $\omega_n \sim D(\theta_\omega)$ . From this equation, the density function of the latent variables, $f(X^* \mid X_n; \lambda, \theta_\omega)$ , is obtained and the resulting unconditional probability equation is then:

$$P(y_n \mid X_n; \beta, \lambda, \theta_\varepsilon, \theta_\omega) = \int P(y_n \mid X_n, X^*; \beta, \theta_\varepsilon) f(X^* \mid X_n; \lambda, \theta_\omega) dX^* . \quad (14)$$

It is difficult to estimate this model based on the observed preference indicator alone and so psychometric data, such as responses to attitudinal and perceptual survey questions, are used as indicators of the latent psychological factors. These data ( $I_n$ ) are incorporated through the latent variable measurement equation:

$$I_n = I(X_n^*; \alpha) + \upsilon_n , \quad (15)$$

which describes the indicators ( $I_n$ ) as a function of the latent variables ( $X_n^*$), a set of parameters ( $\alpha$ ) and a disturbance $\upsilon_n \sim D(\theta_\upsilon)$ . From this equation, the density function of the indicators, $f(I_n \mid X^*; \alpha, \theta_\upsilon)$ , is obtained. Incorporating this into the likelihood leads to the final form of the integrated choice and latent variable model:

$$P(y_n, I_n \mid X_n; \beta, \alpha, \lambda, \theta_\varepsilon, \theta_\upsilon, \theta_\omega) = \tag{16}$$

$$\int P(y_n \mid X_n, X^*; \beta, \theta_\varepsilon) f(I_n \mid X^*; \alpha, \theta_\upsilon) f(X^* \mid X_n; \lambda, \theta_\omega) dX^* .$$

This technique has been applied in numerous applications. Bernardino (1996) modeled telecommuting behavior and included latent attributes such as the costs and benefits of a program. Börsch-Supan, McFadden, and Schnabel (1996) modeled the choice of living arrangements of the elderly and included a latent health characteristic. Eymann, Boersch-Supan and Euwals (2001) modeled investors' portfolio choices and included latent attitudes towards risk. Hosoda (1999) modeled shoppers' mode choices and included latent sensitivities to time, cost, comfort, and convenience. Morikawa, Ben-Akiva, and McFadden (2002) modeled intercity mode choices and included the latent attributes of comfort and convenience. Polydoropoulou (1997) modeled responses to advanced traveler information systems and included latent variables such as knowledge and satisfaction. Ramming (2002) modeled commuters' choice of route to work and included a latent characteristic that represents knowledge of the transportation system. Train, McFadden, and Goett (1987) modeled consumers' choices of public utility rate schedules and included latent characteristics such as the importance of energy consumption and the importance of finding new energy sources.

### *Component 4: Latent Classes*

As with random parameter models and latent variable models, latent class models also capture unobserved heterogeneity, but are employed when the latent variables are discrete constructs. The concept is that there may be discrete segments of decision-makers that are not immediately identifiable from the data. Furthermore, these segments (or classes) may exhibit different choice behavior as is signified by class-specific utility equations for each class $s$:

$$U_n^s = V(X_n; \beta^s) + \varepsilon_n^s . \tag{17}$$

This leads to the class-specific choice model:

$$P(i \mid X_n, s_n; \beta^s, \theta_\varepsilon^s) , \tag{18}$$

which may vary in terms of parameters, alternatives available in the choice set, or model structure.

While it cannot be deterministically identified to which class a decision-maker belongs from the observable variables, it is presumed that class membership probabilities can be estimated, which are denoted by the class membership model:

$$P(s_n \mid X_n; \gamma) . \tag{19}$$

This is the probability of decision-maker $n$ belonging to class $s$, given explanatory variables $X_n$ and parameters $\gamma$. A key issue in latent class choice models is how to specify the class membership model.

Typically these are straightforward logit equations. However, Gopinath (1995) provides details on more complex relationships that can be introduced, and why such complexity may be warranted.

Kamakura and Russell (1989) pioneered the form of the latent class choice model, which incorporates the class-specific choice model (Equation (18)) and the class membership model (Equation (19)):

$$P(i \mid X_n; \beta, \gamma, \theta_\varepsilon) = \sum_{s=1}^{S} P(i \mid X_n, s; \beta^s, \theta_\varepsilon^s) P(s \mid X_n; \gamma) , \tag{20}$$

where: $\beta = \left( \left(\beta^1\right)', ..., \left(\beta^S\right)' \right)'$ and $\theta_\varepsilon = \left( \left(\theta_\varepsilon^1\right)', ..., \left(\theta_\varepsilon^S\right)' \right)'$.

The latent class model depicted in Equation (20) is written without the inclusion of additional indicators beyond the observable choice. As displayed in the generalized framework (Figure 5), it is helpful if indicators of the latent classes are available, and the reader is referred to Ben-Akiva and Boccara (1995) and Gopinath (1995) for more information.

Latent classes have been used in many applications to improve the behavioral representation and explanatory power of choice models. Ben-Akiva and Boccara (1995) modeled commuters' mode choices allowing for different choice sets among travelers. Gopinath (1995) modeled intercity travelers' mode choices allowing for different decision protocols among classes (for example, utility maximizers versus habitual choosers). Gopinath (1995) modeled shippers' choices between train and truck allowing for different sensitivities to time and cost. Hosoda (1999) modeled shoppers' mode choices allowing for different sensitivities of time and cost, for example, distinguishing between patient and impatient travelers.

## *Component 5: Combining Stated and Revealed Preferences*

The final extension incorporated in the generalized random utility model deals with the issue of combining preference data from different sources. There are two broad classes of preference data that are used to estimate random utility models: revealed preferences (RP), which are based on actual market behavior, and stated preferences (SP), which are expressed responses to hypothetical scenarios. Each type of data has its advantages and disadvantages, namely that RP are cognitively congruent with actual behavior, but SP surveys can be collected in a tightly controlled choice environment and can provide richer information on preferences. Given these strengths and weaknesses, the two types of data are highly complementary, and combined estimators can be used to draw on the advantages of each.

The elements of the generalized model related to combining SP and RP data are highlighted in Figure 6, in which both stated preferences and revealed preferences are shown to be indicators of the unobservable utilities. The two sets of indicators lead to two sets of utility equations for the choice model:

$$U_n^{RP} = V(X_n^{RP}; \beta^{RP}) + \varepsilon_n^{RP} \text{ for RP, and} \tag{21}$$

$$U_n^{SP} = V(X_n^{SP}; \beta^{SP}) + \varepsilon_n^{SP} \text{ for SP.} \quad (22)$$

Treating the models separately (and introducing appropriate measurement equations given the form of the indicators), would lead to independent likelihood functions:

$$P(y_n^{RP} \mid X_n^{RP}; \beta^{RP}, \theta_\varepsilon^{RP}) \text{ and} \quad (23)$$

$$P(y_n^{SP} \mid X_n^{SP}; \beta^{SP}, \theta_\varepsilon^{SP}) . \quad (24)$$

A fundamental assumption in conducting SP surveys is that the trade-off relationships among major attributes (the $\beta$ 's) are common to both revealed and stated preferences. When there is such an overlap between $\beta^{RP}$ and $\beta^{SP}$, there are advantages to jointly estimating the models. The benefits include correcting bias that may exist in the SP responses, identifying the effect of new services, identifying the effects of attributes that have either limited range or are highly correlated in the RP data, and improving efficiency of the parameter estimates.

Ben-Akiva and Morikawa (1990) developed techniques for jointly estimating the RP and SP models. (See also the review in Ben-Akiva, Bradley et al., 1994, and Bhat and Castelar, 2000.) In order to combine the preference data, there are two important issues involving the disturbances that need to be considered. First, the disturbances are most likely correlated across multiple responses for a given individual. Therefore, a correlation term is often introduced to capture this effect. Using the approach suggested by Morikawa (1994), the disturbances are decomposed into the alternative- and individual-specific disturbance and the white noise: $\varepsilon_n^{RP} = \psi^{RP}\eta_n + \dot{\varepsilon}_n^{RP}$ and $\varepsilon_n^{SP} = \psi^{SP}\eta_n + \dot{\varepsilon}_n^{SP}$, where $\eta_n$ are i.i.d. normal distributed factors that are constant across responses from a single individual, and unknown parameters $\psi^{RP}$ and $\psi^{SP}$ (matrices) capture the magnitude of the correlation effect. The second issue in joint SP/RP models is that while the assumption is that the covariance structure is the same across the models, the scale is likely to vary across the two models. To address this, the scale of the RP model is still set to a fixed value for identification (for example, setting $\mu = 1$ for logit). However, the scale term of the SP model, which is denoted $\mu^{SP}$, is estimated. Incorporating these two issues into the model, the likelihood function for a joint RP/SP model with 1 RP response ( $y_n^{RP}$ ) and multiple SP responses ( $y_{n1}^{SP}, \ldots, y_{nQ_n}^{SP}$ ) per decision-maker $n$ is:

$$P(y_n^{RP}, y_n^{SP} \mid X_n; \beta, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{RP}) = \quad (25)$$

$$\int P(y_n^{RP} \mid X_n^{RP}, \eta; \beta, \psi^{RP}, \theta_\varepsilon^{RP}) \prod_{q=1}^{Q_n} P(y_{nq}^{SP} \mid X_{nq}^{SP}, \eta; \beta, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{RP}) f(\eta) d\eta ,$$

where: $y_n^{SP} = (y_{n1}^{SP}, \ldots, y_{nQ_n}^{SP})$,

$X_n$ includes all of the explanatory variables in the model, and

$\beta^{RP} = \beta^{SP} = \beta$, but this need only be the case for a subset of $\beta$ .

The first term is the RP model. The second term is the SP model. The disturbance terms of the RP model ($\theta_\varepsilon^{RP}$) must incorporate a scale constraint (for example, for logit, $\theta_\varepsilon^{RP} = \mu^{RP} = 1$). The SP model has the same disturbance structure (and $\theta_\varepsilon^{RP}$), but may vary by scale $\mu^{SP}$. The factor $\eta_n$ (and associated parameters $\psi^{RP}$ and $\psi^{SP}$) is introduced to capture correlation among multiple responses from a given individual, and must be integrated out.

These techniques are becoming fairly common in the literature and in practice. For example, joint SP/RP models have been used to model recreational site choice (Adamowicz et al., 1994), intercity mode choice (Ben-Akiva and Morikawa, 1990), congestions pricing situations (Bhat and Castelar, 2000), choices among gasoline and alternative fueled vehicles (Brownstone, Bunch, and Train, 2000), and pre-trip decisions as influenced by traveler information systems (Khattak et al., 1996).

## *Specification of the Generalized Random Utility Model*

This section integrates the extensions to RUM that were described above to build the generalized model presented in Figure 2. In specifying the generalized model, it is useful to think of two different aspects to the process. The first is specifying the behavioral model of interest, i.e., a model that explains market behavior (revealed preferences) and its underlying (assumed) relationships. The second aspect has to do with measurements, namely, incorporating stated preference and additional behavioral indicators to aid in estimating the parameters in the model of interest. These two aspects are addressed below.

### *The Generalized Choice Model*

The generalized model that explains the market behavior ( $y_n^{RP}$ ) consists of several components. The core of the model is the RUM kernel (logit or some other relatively simple specification), which is denoted:

$$P(y_n^{RP} \mid X_n^{RP}; \beta, \theta_\varepsilon^{RP}) . \qquad (26)$$

As discussed in the sections above, adding features such as flexible disturbances ( $\xi_n$, Equation (8)), latent variables ( $X_n^*$, Equation (12)), and latent classes ( $s_n$, Equation (18)) can be used to relax the limiting restrictions of the multinomial logit formulation and enrich the behavioral representation of the model. While these additional elements are all unknown factors, the probability conditional on these factors can be written as:

$$P(y_n^{RP} \mid X_n^{RP}, X_n^*, \xi_n, s_n; \beta^s, \theta_\varepsilon^{s,RP}) . \qquad (27)$$

However, because the latent variables, classes, and flexible disturbances are, in fact, unobservable, the densities are required to obtain the unconditional probability, including:

- The density function of the flexible disturbances $f(\xi_n \mid \theta_\xi)$ (as in Equation (9)),
- The density function of the latent variables $f(X_n^* \mid X_n; \lambda, \theta_\omega)$ (as in Equation (14)), and
- The class membership probability model $P(s_n \mid X_n; \gamma)$ (as in Equation (20)).

These components are used to form the generalized choice model:

$$P(y_n^{RP} \mid X_n; \beta, \gamma, \lambda, \theta_\varepsilon^{RP}, \theta_\omega, \theta_\xi) = \tag{28}$$
$$\iint \sum_{s=1}^{S} \left( P(y_n^{RP} \mid X_n^{RP}, X^*, \xi, s; \beta^s, \theta_\varepsilon^{s,RP}) P(s \mid X_n; \gamma) \right)$$
$$f(X^* \mid X_n; \lambda, \theta_\omega) f(\xi \mid \theta_\xi) dX^* d\xi \,,$$

where $\beta$ and $\theta_\varepsilon^{RP}$ (analogous to $\theta_\varepsilon$) are defined as in Equation (20), and $X_n$ includes all (observable) explanatory variables.

The conditional choice probabilities (Equation (27)) are first summed over the latent classes ($s_n$)[6] and then integrated over the unknown latent variables ($X_n^*$) and flexible disturbances ($\xi_n$). The resulting function is the probability of the revealed behavior as a function of observable explanatory variables and unknown parameters, $P(y_n^{RP} \mid X_n; \beta, \gamma, \lambda, \theta_\varepsilon^{RP}, \theta_\omega, \theta_\xi)$. While the main ideas of the generalized model are present in Equation (28), clearly even more generalization could be introduced. Further extensions include making the latent variable model class-specific, using latent variables as explanatory factors in the class membership model or for other latent variables, or incorporating flexible disturbances in the class membership model or latent variable model. It is straightforward to extend the likelihood to incorporate these cases, although issues of identification and specification become more difficult. The final model presented in the case study includes flexible disturbances in the latent variable model.

Equation (28) is the model of interest in that it explains market behavior. It also allows for a rich specification through incorporation of flexible disturbances, latent variables, and latent classes. This generalized choice model includes many types of unknown parameters, including those from the kernel ($\beta$ and $\theta_\varepsilon$), the flexible disturbances ($\theta_\xi$), the class membership model ($\gamma$), and the latent variable structural equation ($\lambda$ and $\theta_\omega$). This is a lot to estimate using only the revealed choices, and therefore other sources of data are employed. The incorporation of these data is explained next.

### *The Likelihood Function*

While the revealed preferences are usually the only behavior of interested in terms of explaining and predicting, there also exists a host of other behavioral indicators that can provide assistance in estimating the parameters of the behavioral model presented above. These include:

- Stated preferences, $y_n^{SP}$, which aid in estimating the parameters of the choice model ($\beta$), and

[6] The sum could be moved outside of the integration, however keeping it inside facilitates the programming when the unobserved factors overlap across classes.

- Psychometric indicators, $I_n$, which aid in estimating the parameters of the latent variable structural equation ($\lambda$ and $\theta_\omega$).[7]

To make use of this information, two more elements are introduced to the model. The first is the SP model, which is analogous to the RP model as written above (Equation (27)) with the added complications of joint SP/RP estimation (the scale effect $\mu^{SP}$ and the correlation term $\eta_n$ and parameters $\psi^{SP}$ (and $\psi^{RP}$ in the RP model) as described in the SP/RP section and Equation (25) above):

$$P(y_{nq}^{SP} \mid X_{nq}^{SP}, X_n^*, \xi_n, \eta_n, s_n; \beta^s, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{s,RP}),\ q = 1,\dots,Q_n, \tag{29}$$

where: $Q_n$ is the number of SP responses per respondent.

The SP model will share some parameters with the RP model. Thus by using appropriate experimental designs for the SP experiment, the inclusion of SP data can improve the estimation of the RP choice model parameters.

The second element is the measurement model for the latent constructs (Equation (15)). Following the discussion preceding Equation (16), the resulting density function of the indicators given the latent variables is written as:

$$f(I_n \mid X_n^*; \alpha, \theta_\upsilon). \tag{30}$$

Incorporating Equation (29) and (30) into Equation (28), the likelihood function is then:

$$P(y_n^{RP}, y_n^{SP}, I_n \mid X_n; \beta, \gamma, \alpha, \lambda, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{RP}, \theta_\upsilon, \theta_\omega, \theta_\xi) = \tag{31}$$

$$\iiint \sum_{s=1}^{S} \left( P(y_n^{RP}, y_n^{SP} \mid X_n, X^*, \xi, \eta, s; \beta^s, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{s,RP}) P(s \mid X_n; \gamma) \right)$$
$$f(I_n \mid X^*; \alpha, \theta_\upsilon) f(X^* \mid X_n; \lambda, \theta_\omega) f(\xi \mid \theta_\xi) f(\eta) dX^* d\xi d\eta,$$

where: $y_n^{SP}$, $X_n$, $\beta$, and $\theta_\varepsilon^{RP}$ are defined as in Equations (25) and (28), and

$$P(y_n^{RP}, y_n^{SP} \mid X_n, X^*, \xi, \eta, s; \beta^s, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{s,RP}) =$$
$$P(y_n^{RP} \mid X_n^{RP}, X^*, \xi, \eta, s; \beta^s, \psi^{RP}, \theta_\varepsilon^{s,RP}) *$$
$$\prod_{q=1}^{Q_n} P(y_{nq}^{SP} \mid X_{nq}^{SP}, X^*, \xi, \eta, s; \beta^s, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{s,RP}).$$

[7] Psychometric indicators can also aid in estimating the class membership probability model, $P(s \mid X_n; \gamma)$. This is not explicitly included here, because there are numerous ways in which such indicators can enter the model. See Ben-Akiva and Boccara, 1995, and Gopinath, 1995, for a treatment of indicators for latent classes.

The addition of the SP model (Equation (29)) and the density function of psychometric indicators (Equation (30)) will add nuisance parameters, which are not a part of the behavioral model of interest (i.e., Equation (28)), but also must be estimated.

Note that the latent variable structural equation (13) can be substituted throughout, and the likelihood reduces to:

$$P(y_n^{RP}, y_n^{SP}, I_n \mid X_n; \beta, \gamma, \alpha, \lambda, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{RP}, \theta_\upsilon, \theta_\omega, \theta_\xi) = \quad (32)$$

$$\iiint \sum_{s=1}^{S} \left( P(y_n^{RP}, y_n^{SP} \mid X_n, \omega, \xi, \eta, s; \beta^s, \lambda, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{s,RP}) P(s \mid X_n; \gamma) \right)$$
$$f(I_n \mid \omega; \lambda, \alpha, \theta_\upsilon) f(\omega \mid \theta_\omega) f(\xi \mid \theta_\xi) f(\eta) d\omega d\xi d\eta \,.$$

## *Estimation*

Likelihood-based estimation is described here for the generalized model, because of its straightforward interpretation, implementation, performance, and availability of rigorous hypothesis testing. Clearly, there are alternative estimation methods available, including method of simulated moments or Bayesian techniques[8]. The intention here is not to review the available methods. Instead, the focus is on maximum likelihood estimation as a practical technique for estimation of the generalized model.

Using this technique, estimation amounts to maximizing the likelihood equation for the generalized model, which is represented by Equation (31). Maximum simulated likelihood (MSL) is proposed to deal with high-dimensional integrals. There are important practical issues in applying MSL to the generalized model. One key in estimation is to write the likelihood such that the distribution over which the integral is taken is independent multivariate standard normal, because this allows the use of general estimation code. This is usually trivial to accomplish through a transformation of variables. The substitution from Equation (31) to (32) is one example, and the equation could be further reduced by transforming $\xi_n$ and $\omega_n$ to functions of independent standard normal distributions.

A second key to the estimation is to keep the dimensionality of the integral down. The dimension is determined by the flexible disturbances ($\xi_n$), the RP/SP correlation terms ($\eta_n$), and the latent variables ($\omega_n$). Similarly, it is desirable to keep the number of classes small. When the dimension of the integral is above 3, simulation techniques are required in order to evaluate the integral. The idea behind simulation is to replace the multifold integral (the likelihood function) with easy to compute probability simulators, such as Equation (10). The advantage of the kernel formulation is that it provides a tractable, unbiased, and smooth simulator for the likelihood, namely Equation (31) can be simulated with:

[8] See Train (2000) for a comparison of classical versus Bayesian estimation. The software WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml) can be used to estimate complex models with Bayesian techniques.

$$\hat{P}(y_n^{RP}, y_n^{SP}, I_n \mid X_n; \beta, \gamma, \alpha, \lambda, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{RP}, \theta_\upsilon, \theta_\omega, \theta_\xi) = \tag{33}$$

$$\frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \sum_{s=1}^{S} \Big( P(y_n^{RP}, y_n^{SP} \mid X_n, \omega_n^d, \xi_n^d, \eta_n^d, s; \beta^s, \lambda, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{s,RP}) P(s \mid X_n; \gamma) \Big)$$
$$f(I_n \mid \omega_n^d; \lambda, \alpha, \theta_\upsilon),$$

where: $\xi_n^d$, $\eta_n^d$ *and* $\omega_n^d$ are particular realizations (or draws) from $f(\xi_n \mid \theta_\xi)$, $f(\eta_n)$, and $f(\omega_n \mid \theta_\omega)$.

Thus, the integral is replaced with an average of values of the function computed at discrete points. There has been a lot of research concerning how best to generate the set of discrete points. The most straightforward approach is to use pseudo-random sequences (for example, Monte Carlo). However, variance reduction techniques (for example, antithetic draws) and quasi-random approaches (for example, Halton draws, which are used in this paper) have been found to be more efficient. See Bhat (2001) for further discussion.

Using the probability simulator, the simulated log-likelihood of the sample is:

$$\hat{L} = \sum_{n=1}^{N} \ln \hat{P}(y_n^{RP}, y_n^{SP}, I_n \mid X_n; \beta, \gamma, \alpha, \lambda, \psi^{RP}, \psi^{SP}, \mu^{SP}, \theta_\varepsilon^{RP}, \theta_\upsilon, \theta_\omega, \theta_\xi). \tag{34}$$

The parameters are then estimated by maximizing Equation (34) over the unknown parameters.

A well-known issue is that the simulated log-likelihood function, although consistent, is simulated with a downward bias for finite number of draws. The issue is that while the probability simulator (Equation (33)) is unbiased, the log-simulated-likelihood (Equation (34)) is biased due to the log transformation. In order to minimize the bias in simulating the log-likelihood function, it is important to simulate the probabilities with good precision. The precision increases with the number of draws, as well as with the use of intelligent methods to generate the draws. The number of draws necessary to sufficiently remove the bias cannot be determined a priori; it depends on the type of draws, the model specification, and the data. Therefore, when estimating these models, it is necessary to verify numerical stability in the parameter estimates as the number of draws is increased.

### *Software*

There are some generalized statistical software programs being developed, such as WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml), which have potential for use in estimating models similar to the generalized model described here. For the estimation results presented in the case study, custom software written in Fortran and using the IMSL optimization libraries was used. The advantage of this approach is that once the skeletal structure is developed (meaning a base likelihood function and simulation routines), any model specification can be fairly straightforwardly introduced and estimated without being limited by the capability of a particular statistical program. The development of the likelihood function and estimation via Simulated Maximum Likelihood described in this paper provides a blueprint for developing such estimation software.

### *Identification*

Identification can be difficult, particularly as the model gets more complex. While specific identification rules have been developed for special cases of the generalized model (see, for example, Ben-Akiva and Lerman, 1985, for standard discrete choice; Bollen, 1989, for latent variable models; and Ben-Akiva, Bolduc, and Walker, 2001, for logit kernel), there are no general necessary and sufficient conditions for identification. The best that can be done is to apply the sufficient, but not necessary technique of conditionally identifying each sub-module (as in the two-step approach applied to latent variable models). However, in many cases there remains uncertainty regarding identification and, furthermore, even models that are theoretically identified often have multicollinearity issues that impede estimation of the parameters. Therefore, the use of empirical identification tests is highly recommended. There are several possible techniques in this category, including:

- Conducting Monte Carlo experiments by generating synthetic data from the specified model structure (with given parameter values), and then attempting to reproduce the parameters using the maximum likelihood estimator. If the parameters cannot be reproduced to some degree of accuracy, then this is an indication that the model is not identified.
- Verifying that the parameters converge to the same point and likelihood given different starting values. This addresses the issue of local maxima and also of a flat (unidentified) maximum.
- Verifying that the Hessian of the log-likelihood function is non-singular (a test of local identification). This is usually performed automatically in order to generate estimates of the standard errors of estimated parameters.
- Constraining one or more parameters to different values, and verifying that the log-likelihood shifts as a result. (This is particularly useful when there are one or more suspect parameters.)
- Verifying that the parameters are stable as the number of simulation draws is increased. This is critical, as unidentified models can appear identified with a small number of draws.

## *Case Study*

In this section the generalized random utility modeling approach is applied to a mode choice application. The models presented use data collected in 1987 for the Netherlands Railway, which were gathered to assess the factors that influence the choice between rail and auto for intercity travel. The purpose of the case study is to demonstrate the usefulness of the model and estimation technique, as well as to provide a set of sample equations. In this application, the data are fixed and so the objective of the generalized model is to investigate the most effective use of the data.

A series of models are presented, which begin with simple model structures and then build up to more complex forms. This procedure of gradually adding complexity to the model is important, because it aids in designing the complex models by allowing one to develop the specification of the sub-modules within a simpler choice model environment. The simpler models also provide a means for generating good starting values for the more complex models. The first models discussed are the kernels: a base RP model, and a joint RP/SP model. The latter is used as the basis of all subsequent models. Next a series of models are presented in which the extensions of flexible disturbances, latent classes, and latent variables are each applied separately. Finally a model that integrates the individual extensions is presented.

## *Data*

The sample is comprised of people who had traveled between Nijmegen and Randstad (approximately a two-hour trip) in the 3 months prior to the survey. The following data are available for each of 228 respondents:

- *Demographic data:*
  Characteristics of the respondent, for example, age and gender.

- *Psychometric data:*
  Subjective ratings of latent attributes of rail and auto, for example, relaxation and reliability.

- *Revealed Preference data (RP):*
  Characteristics of the Nijmegen to Randstad trip made by the respondent, including:
  - the chosen mode (rail or auto);
  - characteristics of the trip, such as trip purpose (business or other), number of persons traveling, and whether or not there was a fixed arrival time requirement; and
  - attributes of the alternatives, including cost, in-vehicle and out-of-vehicle travel times, number of transfers (rail only).

- *Stated Preference data 1 (SP1 – rail versus rail):*
  Responses to an SP experiment of a choice between two hypothetical rail services.

  For each experiment, the respondent was presented with two hypothetical rail alternatives for the particular intercity trip reported in the RP experiment. Each alternative was described by travel cost, travel time, number of transfers, and level of amenities. Level of amenities is a package of different aspects such as seating room and availability, quietness, smoothness of ride, heating/ventilation, and food service, but is presented at only three levels (0, 1, and 2, the lower the better). Given the two alternatives, the respondent was asked to state his or her preference on the basis of a five-point scale:
  *1* - definitely choose alternative 1,
  *2* - probably choose alternative 1,
  *3* - not sure,
  *4* - probably choose alternative 2, and
  *5* - definitely choose alternative 2.

  Each respondent was presented with multiple pairs of choices (an average of about 13 per person).

- *Stated Preference data 2 (SP2 – rail versus auto):*
  Responses to an SP experiment of a choice between hypothetical rail and auto services.

  For each experiment, the respondent was presented with a hypothetical rail alternative and a hypothetical auto alternative for the particular intercity trip reported in the RP experiment. Each alternative was described by travel cost, travel time, number of transfers (rail only), and level of amenities (rail only). Given the two alternatives, the respondent was asked to state his or her preference on the basis of a five-point scale:

*1* - definitely choose auto,
*2* - probably choose auto,
*3* - not sure,
*4* - probably choose rail, and
*5* - definitely choose rail.

Each respondent was presented with multiple pairs of choices (an average of about 7 per person).

For additional information on the data, see Bradley, Grosvenor, and Bouma (1988).

## The Kernel for the Case Study

The first step is to determine the kernel. The RP data are binary choice, and a binary logit model is used. The SP data are ordinal choice and so an ordinal logit model is used. These base equations are described in this section.

A *linear in parameters* utility equation is assumed and this is expressed as:

$$U_{in} = X_{in}\beta + \varepsilon_{in}, \qquad (35)$$

where: $X_{in}$ is a row vector and $\beta$ is a column vector.

In a binary choice situation it simplifies the equations to use the difference between the two utilities, which is denoted as:

$$\tilde{U}_n = U_{1n} - U_{2n} = \tilde{X}_n\beta + \tilde{\varepsilon}_n, \qquad (36)$$

where: $\tilde{U}_n$ is a scalar, $\tilde{X}_n$ is a row vector equal to $(X_{1n} - X_{2n})$, $\beta$ is a column vector, and $\tilde{\varepsilon}_n$ is a scalar.

Since logit is being used for the kernel, $\tilde{\varepsilon}_n$ is the difference between two independent Extreme Value distributed random variables and is therefore logistically distributed.

The measurement equation for the RP choice model is as follows:

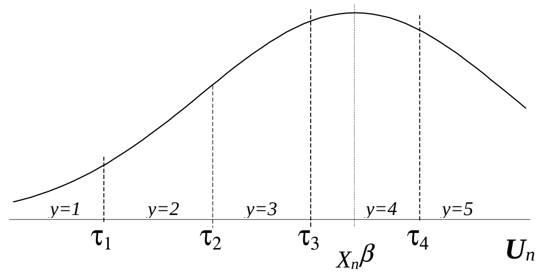
$$y_n^{RP} = \begin{cases} 1 & \textit{if } \tilde{U}_n^{RP} \geq 0 \quad \textit{(person n chose rail)} \\ -1 & \textit{if } \tilde{U}_n^{RP} < 0 \quad \textit{(person n chose auto)} \end{cases} . \qquad (37)$$

The likelihood for an RP response is then the following binary logit probability:

$$P(y_n^{RP} \mid \tilde{X}_n^{RP}; \beta, \mu^{RP}) = \frac{1}{1 + e^{-\mu^{RP}(\tilde{X}_n^{RP}\beta)y_n^{RP}}} . \qquad (38)$$

For the SP five-point preference rating, the utility is specified as in Equation (36), and threshold values $(\tau)$ are specified in the utility scale such that:

$$
\begin{aligned}
P_n(1) &= P(\tau_0 < \tilde{U}_n^{SP} \le \tau_1), \\
P_n(2) &= P(\tau_1 < \tilde{U}_n^{SP} \le \tau_2), \\
P_n(3) &= P(\tau_2 < \tilde{U}_n^{SP} \le \tau_3), \\
P_n(4) &= P(\tau_3 < \tilde{U}_n^{SP} \le \tau_4), \\
P_n(5) &= P(\tau_4 < \tilde{U}_n^{SP} \le \tau_5),
\end{aligned}
$$




where: $\tau_0 = -\infty$, $\tau_5 = \infty$, and
$\tau = (\tau_0, ..., \tau_5)'$.

The measurement equation for the SP ordinal choices are then defined as follows:

$$y_{in}^{SP} = \begin{cases} 1 & if\ \tau_{i\text{-}1} < \tilde{U}_n^{SP} \le \tau_i \\ 0 & otherwise \end{cases}, \qquad i = 1, ..., 5, \tag{39}$$

and the vector of these indicators is $y_n^{SP} = (y_{1n}^{SP}, ..., y_{5n}^{SP})'$.

The likelihood for each ordinal preference rating is then:

$$P(y_n^{SP} \mid \tilde{X}_n^{SP}; \beta, \tau, \mu^{SP}) = \prod_{i=1}^{5} \left( \frac{1}{1+e^{-\mu^{SP}(\tilde{X}_n^{SP}\beta - \tau_i)}} - \frac{1}{1+e^{-\mu^{SP}(\tilde{X}_n^{SP}\beta - \tau_{i-1})}} \right)^{y_{in}^{SP}}, \tag{40}$$

where there is a different specification (and different $\mu$ and $\tau$) for each SP dataset.

For the Rail versus Rail stated preference data, the order of the alternatives should be irrelevant[9], and therefore the threshold parameters are constrained to be symmetric, i.e., $\tau_1 = -\tau_4$ and $\tau_2 = -\tau_3$. A likelihood ratio test was used to verify that the data support this constraint. For the Rail versus Auto stated preference data, the symmetry condition is not necessary (and a likelihood ratio test verified that it does not hold). However, a constraint must be imposed to identify the model, and $\tau_2 = -\tau_3$ is used.

## *Revealed Preference Kernel*

The first model presented using the mode choice data is a binary logit model using the revealed preference data. The likelihood for this model is as written in Equation (38), and $\mu^{RP}$ is set to 1 for identification. The estimation results are shown in Table 1. Robust standard errors[10] and t-statistics are reported for all models. The check marks in the $U_{rail}$ and $U_{auto}$ columns signify whether the parameter is

---

[9] If there is not a survey response bias.

[10] Using the robust asymptotic covariance matrix estimator $H^{-1}BH^{-1}$, where $H$ is the Hessian (calculated numerically, in our case) and $B$ is the cross product of the gradient. (For more information, see Newey and McFadden, 1994)

included in the rail and/or auto utility. The signs of the parameters are as expected. And, with the exception of in-vehicle time and number of transfers, the parameters are significantly different from zero.

### *Joint Stated and Revealed Preference Model* [11]

First the joint RP/SP technique is applied, because this model forms the basis for all other models. For each respondent, the following choice indicators are available:

| *Type of Indicator* | *# Per Person* |
|---|---|
| Revealed preference | 1 |
| Stated preferences from rail versus rail hypothetical scenarios | $Q_n$ |
| Stated preferences from rail versus auto hypothetical scenarios | $R_n$ |

There is expected to be correlation among all rail versus auto choices made by an individual. Therefore a correlation term ($\eta_n$) and associated parameters ($\psi^{RP}$ and $\psi^{SP2}$) as described for Equation (25) are introduced and the utilities are then:

$$\tilde{U}_n^{RP} = \tilde{X}_n^{RP}\beta \quad + \psi^{RP}\eta_n \; + \tilde{\varepsilon}_n^{RP}, \tag{41}$$

$$\tilde{U}_{nq}^{SP1} = \tilde{X}_{nq}^{SP1}\beta \qquad\qquad + \tilde{\varepsilon}_{nq}^{SP1}, \qquad q = 1,\ldots,Q_n \textit{ (rail versus rail)}, \text{ and} \tag{42}$$

$$\tilde{U}_{nr}^{SP2} = \tilde{X}_{nr}^{SP2}\beta + \psi^{SP2}\eta_n + \tilde{\varepsilon}_{nr}^{SP2}, \qquad r = 1,\ldots,R_n \textit{ (rail versus auto)}, \tag{43}$$

where: $\eta_n$ is an unobservable scalar (distributed standard normal) and $\psi^{RP}$ and $\psi^{SP2}$ are unknown parameters (scalars) that capture the magnitude of the correlation.

This leads to a correlation of $\psi^{RP}\ \psi^{SP2}$ between RP and SP2 responses from the same individual, and a correlation of $\psi^{SP2}\ \psi^{SP2}$ for SP2 responses from the same individual. In addition, the RP choice indicator $y_n^{RP}$ is included in $\tilde{X}_{nr}^{SP2}$ ($\forall r$) in order to capture inertia in the SP responses towards the alternative actually chosen in the market environment.

The likelihood function for the joint model is:

$$\begin{aligned} P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid \tilde{X}_n; \beta, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2}) = \\ \int P(y_n^{RP} \mid \tilde{X}_n^{RP}, \eta; \beta, \psi^{RP})\ P(y_n^{SP1} \mid \tilde{X}_n^{SP1}; \beta, \tau^{SP1}, \mu^{SP1}) \\ P(y_n^{SP2} \mid \tilde{X}_n^{SP2}, \eta; \beta, \tau^{SP2}, \psi^{SP2}, \mu^{SP2})\phi(\eta)d\eta, \end{aligned} \tag{44}$$

where: $\phi(\eta)$ is the standard normal density function, and,

based on Equations (38) and (40) and using the utility specified by Equations (41)- (43),

[11] This specification is based on the models presented in Morikawa, Ben-Akiva, and McFadden (2002).

$$P(y_n^{RP} \mid \tilde{X}_n^{RP}, \eta; \beta, \psi^{RP}) = \frac{1}{1+e^{-(\tilde{X}_n^{RP}\beta + \psi^{RP}\eta_n) y_n^{RP}}} \ , \tag{45}$$

$$P(y_n^{SP1} \mid \tilde{X}_n^{SP1}; \beta, \tau^{SP1}, \mu^{SP1}) = \tag{46}$$

$$\prod_{q=1}^{Q_n} \prod_{j=1}^{5} \left( \frac{1}{1+e^{-\mu^{SP1}(\tilde{X}_{nq}^{SP1}\beta \ -\tau_j^{SP2})}} - \frac{1}{1+e^{-\mu^{SP1}(\tilde{X}_{nq}^{SP1}\beta - \tau_{j-1}^{SP1})}} \right)^{y_{jnq}^{SP1}} \quad \text{, and}$$

$$P(y_n^{SP2} \mid \tilde{X}_n^{SP2}, \eta; \beta, \tau^{SP2}, \psi^{SP2}, \mu^{SP2}) = \tag{47}$$

$$\prod_{r=1}^{R_n} \prod_{i=1}^{5} \left( \frac{1}{1+e^{-\mu^{SP2}(\tilde{X}_{nr}^{SP2}\beta \ + \psi^{SP2}\eta_n - \tau_i^{SP2})}} - \frac{1}{1+e^{-\mu^{SP2}(\tilde{X}_{nr}^{SP2}\beta \ + \psi^{SP2}\eta_n - \tau_{i-1}^{SP2})}} \right)^{y_{inr}^{SP2}} .$$

Note that the scale of the RP model ( $\mu^{RP}$ ) is constrained to 1 (for identification, as usual), but the scales for the SP1 and SP2 models ( $\mu^{SP1}$ and $\mu^{SP2}$ ) are estimated in order to allow them to be different from the RP model.

The estimation results are presented in Table 2. The joint model is presented along with models estimated individually on each of the three datasets. One clear benefit of the joint model is that the parameters for in-vehicle travel time and number of transfers are now statistically significant. Another benefit is that the concept of *amenities* is now captured in the model. Having these parameters be significant in the model is important for policy analysis.

## *Random Parameter (Flexible Disturbances) Model*

As an example of incorporating flexible disturbances, the attributes of the alternatives in the mode choice model (Table 2) are allowed to be randomly distributed. Separating out the parameters that are fixed ( $\beta_1$ ) from those that are allowed to vary across the population ( $\beta_{2,n}$ ), the model is now specified as follows:

$$\tilde{U}_n^{RP} = \tilde{X}_{1,n}^{RP}\beta_1 + \tilde{X}_{2,n}^{RP}\beta_{2,n} + \psi^{RP}\eta_n + \tilde{\varepsilon}_n^{RP} \ , \tag{48}$$

$$\tilde{U}_{nq}^{SP1} = \tilde{X}_{1,nq}^{SP1}\beta_1 + \tilde{X}_{2,nq}^{SP1}\beta_{2,n} \qquad + \tilde{\varepsilon}_{nq}^{SP1} \ , \qquad q = 1, \ldots, Q_n \ , \tag{49}$$

$$\tilde{U}_{nr}^{SP2} = \tilde{X}_{1,nr}^{SP2}\beta_1 + \tilde{X}_{2,nr}^{SP2}\beta_{2,n} + \psi^{SP2}\eta_n + \tilde{\varepsilon}_{nr}^{SP2} \ , \qquad r = 1, \ldots, R_n \ , \tag{50}$$

where: $\beta_1$ is a vector of unknown parameters, constant across individuals,

$\beta_{2,n}$ is a vector of $k = 1, \ldots, K$ unknown values ( $\beta_{2k,n}$ ), which vary across individuals,

$\tilde{X}_{1,n}^{RP}, \tilde{X}_{1,nq}^{SP1}, \tilde{X}_{1,nr}^{SP2} \equiv \tilde{X}_{1,n}$ are the explanatory variables with fixed parameters, and

$\tilde{X}_{2,n}^{RP}, \tilde{X}_{2,nq}^{SP1}, \tilde{X}_{2,nr}^{SP2} \equiv \tilde{X}_{2,n}$ are the explanatory variables with random parameters.

The model estimated includes the attributes of the alternatives (cost, in-vehicle time, out of vehicle time, transfers, and amenities) in $\tilde{X}_{2,n}$ and all other explanatory variables in $\tilde{X}_{1,n}$. Since in this case the parameters for $\tilde{X}_{2,n}$ must be negative to make economic sense, the distribution of each $\beta_{2k,n}$ is specified as independent[12] lognormal. To achieve this, each $\beta_{2k,n}$ is substituted with the lognormal relationship:

$$\beta_{2k,n} = -\exp(\delta_{k1} + \delta_{k2}\xi_{k,n}), \tag{51}$$

where:

$\xi_{k,n}$ are distributed independent standard Normal, $\xi_n = (\xi_{1,n},...,\xi_{K,n})'$,

$\delta_{k1}, \delta_{k2}$ are unknown parameters, $\delta = ((\delta_{1,1},...,\delta_{K,1}),(\delta_{1,2},...,\delta_{K,2}))'$,

$mean(\beta_{2k,n})$ is equal to $-\exp(\delta_{k1} + \delta_{k2}{}^2/2)$, and

$variance(\beta_{2k,n})$ is equal to $\exp(2\delta_{k1})(\exp(2\delta_{k2}{}^2) - \exp(\delta_{k2}{}^2))$.

The likelihood is then:

$$\begin{aligned} P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid \tilde{X}_n; \beta_1, \delta, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2}) = \\ \iint P(y_n^{RP} \mid \tilde{X}_n^{RP}, \xi, \eta; \beta_1, \delta, \psi^{RP}) P(y_n^{SP1} \mid \tilde{X}_n^{SP1}, \xi; \beta_1, \delta, \tau^{SP1}, \mu^{SP1}) \\ P(y_n^{SP2} \mid \tilde{X}_n^{SP2}, \xi, \eta; \beta_1, \delta, \tau^{SP2}, \psi^{SP2}, \mu^{SP2}) \prod_{k=1}^{K} \phi(\xi_k)\phi(\eta)\, d\xi\, d\eta. \end{aligned} \tag{52}$$

The results for the random parameter mode choice model are shown in Table 3. The first model is the joint SP/RP model shown in Table 2, and is repeated here for comparison purposes. The second model provides estimation results for a random parameter model in which the parameters are independently distributed. There is a significant improvement in fit (likelihood ratio test) over the model with fixed parameters. Capturing this taste variation will affect forecasts (Ben-Akiva, Bolduc, and Bradley, 1994).

The parameter estimates and standard errors reported in Table 3 for the distributed parameters are $\delta_{k1}$ and $\delta_{k2}$ (Equation (51)), and not of the mean and standard deviation of the random parameters. Table 4 reports the estimated mean and standard deviation (along with their t-stats) for this model. The statistics from the non-distributed model are also provided for comparison. Note that the mean value of the parameters from the distributed model are much higher than those suggested by the non-distributed model. Also, the standard deviations of these parameters are significant.

## *Latent Class Model*

For the latent class mode choice model, a model is estimated that is analogous to the random parameter model presented in Table 3. However, instead of representing the unobserved heterogeneity with random parameters, it is specified that there are two distinct classes of people, each with its own set of

[12] Multivariate lognormal specifications for this model are presented in Walker (2001).

parameters for the 5 attributes of the alternatives. Parameters other than those for the 5 attributes are common across the classes. The likelihood is as follows:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid X_n; \beta, \gamma, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2}) = \sum_{s=1}^{2} P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid \tilde{X}_n, s; \beta^s, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2}) P(s \mid X_n; \gamma) \quad (53)$$

where: $P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid \tilde{X}_n, s; \beta^s, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2})$ is as in Equation (44) except now there is a different set of parameters ( $\beta^s$ ) for each class, and $P(s \mid X_n; \gamma)$ is a binary logit model.

The estimation results are presented in Table 5. The model suggests that there are at least two classes. Class one is defined by younger travelers, recreational travelers, and people traveling in groups who are more sensitive to cost, in-vehicle time, and transfers. Class two is defined by business travelers and older travelers who are more sensitive to out-of-vehicle time and amenities. The sample is skewed towards class two as the class membership statistics show at the bottom of the table. The two latent classes do provide a significant improvement in fit (likelihood ratio test) over the base model, but the fit of the model falls well below that captured by the random parameter model. Therefore the continuous distribution provides more explanatory power.

## *Choice and Latent Variable Model* [13]

Our mode choice dataset includes information pertaining to the respondents' subjective ratings of various latent attributes. Following the RP portion of the survey, the respondents were asked to rate the following aspects for both rail and auto:

- Relaxation during the trip
- Reliability of arrival time
- Flexibility of choosing departure time
- Ease of traveling with children and/or heavy baggage
- Safety during the trip
- Overall rating of the mode

Responses for the first 5 attributes were in the form of a five-point scale (from very bad to very good), and the overall rating was on a ten-point scale (again, from very bad to very good).

These responses may provide useful information on the behavior. The question is how to use it. Frequently, such data are directly inserted as explanatory variables in the choice model, resulting in highly significant parameter estimates and large improvements in model fit. However, there are several issues with such an approach. First, the data are not available for forecasting, so if forecasting is desired then such a specification is problematic. Second, the multicollinearity inherent in responses to such a

[13] This specification is based on the models presented in Morikawa, Ben-Akiva, and McFadden (2002).

string of questions often makes it difficult to estimate the full set of parameters. The third and most fundamental issue is that it is not clear that such data are causal. For these reasons, the latent variable modeling approach is used, which assumes that these responses are indicators for a smaller number of underlying latent attributes. Furthermore, these latent attributes can be explained by observable attributes of the alternatives and characteristics of the respondent.

The equations for the RP/SP choice and latent variable model follow. First, some notes on the model:

- All variables, including the latent variables and their indicators, are measured in terms of the difference between rail and auto. This was done to simplify the specification: it reduces the dimensionality of the integral by two, it cuts down on the number of parameters, and it lowers the potential for multicollinearity among the latent variable structural equations.
- The indicators in differenced form have a 9-point scale for the first 5 attribute ratings, and a 19-point scale for the *overall* attribute rating, and therefore are treated as continuous variables.
- A combination of exploratory and confirmatory analysis was used to arrive at the final structure of the latent variable model, which consists of two latent variables labeled *comfort* and *convenience*.
- The indicators pertain to the RP choice, and therefore the latent variables are specified using only RP data in the structural equation. However, it is hypothesized that these latent perceptions also impact the stated preference rail versus auto experiment (SP2), and so the latent variables are included in the SP2 model, but are allowed to have different weights (i.e., $\beta$ 's).

To specify the joint choice and latent variable model, the structural and measurement equations for both the latent variable component and the choice component are needed. Linear in the parameters forms are assumed in all cases, and the equations are as follows:

*Latent variable structural equations:*

$$\tilde{X}^*_{ln} = \tilde{X}_n \lambda + \omega_{ln}, \quad l = 1,2, \text{ and } \quad \omega_n \sim \text{i.i.d. standard Normal.} \tag{54}$$

The variances of the disturbance $\omega_n$ are set equal to 1 to set the scale of the latent variables (necessary for identification).[14]

*Choice model structural equations are as in Equations (41) to (43) but with the addition of the latent variable:*

$$\tilde{U}_n^{RP} = \tilde{X}_n^{RP}\beta + \tilde{X}_n^*\beta_{X^*}^{RP} + \psi^{RP}\eta_n + \tilde{\varepsilon}_n^{RP} \ , \tag{55}$$

$$\tilde{U}_{nq}^{SP1} = \tilde{X}_{nq}^{SP1}\beta \qquad\qquad + \tilde{\varepsilon}_{nq}^{SP1} \ , \qquad q = 1,\dots,Q_n \ , \tag{56}$$

---

[14] Models were estimated that allowed a covariance term (i.e., non-orthogonal latent variables), but the covariance term was consistently insignificant.

$$\tilde{U}_{nr}^{SP2} = \tilde{X}_{nr}^{SP2}\beta + \tilde{X}_n^*\beta_{X^*}^{SP2} + \psi^{SP2}\eta_n + \tilde{\varepsilon}_{nr}^{SP2}\ , \qquad r = 1,\ldots,R_n\ . \tag{57}$$

*Latent variable measurement equations:*

$$\tilde{I}_{bn} = \tilde{X}_n^*\alpha_b + \upsilon_{bn}\ , \quad b = 1,\ldots,6\ \text{, and} \tag{58}$$

$\upsilon_{bn} \sim$ independent Normal with mean 0 and standard deviation $\theta_{\upsilon_b}$.

*Choice model measurement equations are as in Equations(37) and (39).*

The likelihood function for the joint model is:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2}, \tilde{I}_n \mid \tilde{X}_n; \beta, \beta_{X^*}^{RP}, \beta_{X^*}^{SP2}, \alpha, \lambda, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2}, \theta_\upsilon) = \tag{59}$$

$$\begin{aligned}\iint & P(y_n^{RP} \mid \tilde{X}_n^{RP}, \tilde{X}^*, \eta; \beta, \beta_{X^*}^{RP}, \psi^{RP})\ P(y_n^{SP1} \mid \tilde{X}_n^{SP1}, \tilde{X}^*; \beta, \tau^{SP1}, \mu^{SP1}) \\ & P(y_n^{SP2} \mid \tilde{X}_n^{SP2}, \tilde{X}^*, \eta; \beta, \beta_{X^*}^{SP2}, \tau^{SP2}, \psi^{SP2}, \mu^{SP2}) \\ & f(\tilde{I}_n \mid \tilde{X}^*; \alpha, \theta_\upsilon)\, f(\tilde{X}^* \mid \tilde{X}_n^{RP}; \lambda)\phi(\eta)d\tilde{X}^* d\eta\ ,\end{aligned}$$

where: $P(y_n^{RP} \mid \ldots)$, $P(y_n^{SP1} \mid \ldots)$, and $P(y_n^{SP2} \mid \ldots)$ are as in Equations (45) to (47), but with the latent explanatory variables (i.e., the utilities as in Equations (55) to (57)),

$$f(\tilde{I}_n \mid \tilde{X}^*; \alpha, \theta_\upsilon) = \prod_{b=1}^{6} \frac{1}{\theta_{\upsilon_b}} \phi\left(\frac{\tilde{I}_{bn} - \tilde{X}^*\alpha_b}{\theta_{\upsilon_b}}\right), \text{ and}$$

$$f(\tilde{X}^* \mid \tilde{X}_n^{RP}; \lambda) = \prod_{l=1}^{2} \phi\left(\tilde{X}_l^* - \tilde{X}_n\lambda_l\right).$$

The likelihood is a three-dimensional integral: one for the correlation term and one for each latent variable. To estimate the model, the structural equation is substituted throughout, and the likelihood function is then an integral over three independent standard normal distributions.

The results of the model are provided in Table 6, and again the base RP/SP model (shaded) is provided for comparison.[15] In this case, the latent variables of comfort and convenience are borderline significant in the choice model (t-stats of 1.0 to 2.3). The latent variable model appears to reasonably capture the latent constructs, and it does add richness to the behavioral process represented by the model. However, the impact is certainly not overwhelming. The log-likelihood for just the choice model portion of the joint model is also reported. Note that there are various ways to calculate this log-likelihood. What is reported is the case in which the latent variable score and distribution are extracted using the structural equation only (a partial information extraction), and then the log-likelihood of the choice model is calculated given this information. This method is representative of the forecasting process, in which the

[15] There are 3 additional explanatory variables in the choice and latent variable model (age dummy, first class rail rider, and free parking), which enter the latent variable structural equations. These variables were tested in the base RP/SP model and are not significant (t-stats of 0.9, 0.4, and 0.2, respectively).

measurement equation is not used (since the indicators are not known). The log-likelihood actually decreases slightly compared with the base choice model. The decrease in fit for the choice model portion does not necessarily mean that the joint model is inferior. As long as the parameters for the latent variables in the choice model are significant, then the latent variable portion is bringing some explanation to the model. The best method to determine the magnitude of the benefits of the joint choice and latent variable model is to perform forecasting tests using either a hold out sample or real data.

## *Generalized Models*

The estimation results thus far have provided examples of integrating joint RP/SP models with random parameters, latent variable, and latent class, individually. Here the information provided by these specifications is used to estimate a model that integrates the different models. The model presented is a combination of the random parameter specification (Table 3) and the choice and latent variable specification (Table 6). The latent class specification (Table 5) is not included because it aimed to capture the same concept of the random parameter specification, but was not as successful.

In addition to combining the choice and latent variable specification with the random parameter specification, some of the parameters in the latent variable structural equation are also allowed to be randomly distributed. To keep the dimension of the integral down and to avoid potential multicollinearity issues, it was important to be selective in terms of the parameters that are distributed. Four parameters in the choice model (those with the most significant distributions from the random parameter model presented in Table 3) and three parameters in the structural equations (those with highest significance in the fixed parameter model presented in Table 6) are selected. The likelihood for this model is:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2}, \tilde{I}_n \mid \tilde{X}_n; \beta, \beta_{X^*}^{RP}, \beta_{X^*}^{SP2}, \alpha, \lambda, \delta, \tau^{SP1}, \tau^{SP2}, \psi^{RP}, \psi^{SP2}, \mu^{SP1}, \mu^{SP2}, \theta_\upsilon) =$$

$$\iint P(y_n^{RP} \mid \tilde{X}_n^{RP}, \tilde{X}^*, \xi, \eta; \beta, \beta_{X^*}^{RP}, \delta, \psi^{RP}) \; P(y_n^{SP1} \mid \tilde{X}_n^{SP1}, \tilde{X}^*, \xi; \beta, \delta, \tau^{SP1}, \mu^{SP1})$$

$$P(y_n^{SP2} \mid \tilde{X}_n^{SP2}, \tilde{X}^*, \xi, \eta; \beta, \beta_{X^*}^{SP2}, \delta, \tau^{SP2}, \psi^{SP2}, \mu^{SP2})$$

$$f(\tilde{I}_n \mid \tilde{X}^*; \alpha, \theta_\upsilon)\, f(\tilde{X}^* \mid \tilde{X}_n^{RP}, \xi; \lambda, \delta) \prod_{k=1}^{K} \phi(\xi_k) \phi(\eta) d\tilde{X}^* d\xi d\eta \;,$$

where negative lognormally distributed parameters are specified as: $-\exp(\delta_{k1} + \delta_{k2}\xi_{k,n})$, and normally distributed parameters are specified as: $(\delta_{k1} + \delta_{k2}\xi_{k,n})$.

The results are presented in Table 7. Not surprisingly, the combined model incorporates the advantages of each of the previously individual specifications: the random parameters significantly (likelihood ratio test) increase the fit of the model (and suggest that the parameters are, indeed, not constant in the population) and the latent variables add behavioral richness to the specification.

Several additional model combinations were estimated with the objective of both improving the overall fit and behavioral representation of the model, as well as strengthening the relationship between the latent variables and the choice model. The full set of results is available in Walker (2001).

### *Conclusions from the Case Study*

The experimental results for the mode choice dataset explored various specifications and demonstrated the practicality of the generalized model. The models are summarized in Table 8. The generalized models introducing stated preferences and random taste variation greatly improves the fit of the model, whereas latent variables and latent classes had less significant impacts. The generalized model incorporated the better fit of the random parameter specification with the behavioral richness of the latent variable specification. It is important to note that no general conclusions can be drawn on the various models from the series of estimation results presented in this paper. The results will vary based on the application and data. The generality that should be drawn is the process of developing the subcomponents of the model and then using this information to develop generalized specifications.

## *Conclusion*

This paper presented a flexible and powerful generalized random utility model that integrates key extensions to RUM, and demonstrated its practicality and effectiveness by applying the approach to a case study. The generalized random utility model makes use of different types of data that provide insight into the choice process, allows for any desirable disturbance structure (including random parameters and nesting structures) through the factor analytic disturbances, and provides means for capturing latent heterogeneity and behavioral constructs through the latent variable and latent class modeling structures. Through these extensions, the choice model can capture more behaviorally realistic choice processes and enable the validity of more parsimonious structures to be tested. Furthermore, the generalized model can be practically estimated via use of a smooth simulator and a maximum simulated likelihood estimator. The added complexity raises a number of interesting challenges for further research, including issues of identification, interpretation, the trade-offs between different estimation techniques, where and how to build the complexity into the model, and the development of generalized estimation software.

## *Acknowledgements*

The authors gratefully acknowledge Denis Bolduc, Dinesh Gopinath, Tony Marley, Taka Morikawa, three anonymous Reviewers, and the participants of the Hybrid Choice Model workshop at the 5th tri-annual Invitational Choice Symposium for their valuable input on this work. This research was partially supported by grants from the National Science Foundation and the University Transportation Centers Program.

## *References*

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities. Journal of Environmental Economics and Management 26, 271-292.

Ben-Akiva, M., Boccara, B., 1987. Integrated Framework for Travel Behavior Analysis. Presented at the International Association of Travel Behavior Research (IATBR) Conference, Aix-en-Provence, France.

Ben-Akiva, M., Boccara, B., 1995. Discrete Choice Models with Latent Choice Sets. International Journal of Research in Marketing 12, 9-24.

Ben-Akiva, M., Bolduc, D., 1996. Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure. Working paper, Massachusetts Institute of Technology.

Ben-Akiva, M., Bolduc, D., Bradley, M., 1994. Estimation of Travel Choice Models with a Distributed Value of Time. Transportation Research Record 1413.

Ben-Akiva, M., Bolduc, D., Walker, J., 2001. Specification, Estimation, & Identification of the Logit Kernel (or Continuous Mixed Logit) Model. Working paper, Massachusetts Institute of Technology.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., Rao, V., 1994. Combining Revealed and Stated Preferences Data. Marketing Letters 5(4), 335-350.

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

Ben-Akiva, M., Morikawa, T., 1990. Estimation of Travel Demand Models from Multiple Data Sources. Transportation and Traffic Theory, 461-476.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 1999. Integration of Choice and Latent Variable Models. Forthcoming, International Association of Traveler Behavior Research (IATBR) book from the 1997 Conference in Austin, Texas.

Bentler, P.M., 1980. Multivariate Analysis with Latent Variables. Annual Review of Psychology 31, 419-456.

Bernardino, A.T., 1996. Telecommuting: Modeling the Employer’s and the Employee’s Decision-Making Process, Garland Publishing, New York, New York.

Bhat, C.R., 1997. Accommodating Flexible Substitution Patterns in Multi-dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice. Transportation Research B 32(7), 455-466.

Bhat, C.R., 2001. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. Forthcoming, Transportation Research.

Bhat, C.R., Castelar, S., 2000. A Unified Mixed Logit Framework for Modeling Revealed and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in the San Francisco Bay Area. Forthcoming, Transportation Research.

Bolduc, D., Ben-Akiva, M., 1991. A Multinominal Probit Formulation for Large Choice Sets. Proceedings of the 6th International Conference on Travel Behaviour 2, 243-258.

Bolduc, D., Fortin, B., Fournier, M.A., 1996. The Impact of Incentive Policies to Influence Practice Location of General Practitioners: A Multinomial Probit Analysis. Journal of Labor Economics 14, 703-732.

Bollen, K. A., 1989. Structural Equations with Latent Variables, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

Börsch-Supan, A., McFadden, D.L., Schnabel, R., 1996. Living Arrangements: Health and Wealth Effects. In: Wise, D.A., (Ed.), Advances in the Economics of Aging. The University of Chicago Press.

Boyd, J.H., Mellman, R.E., 1980. The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis. Transportation Research A 14, 367-378.

Bradley, M., Grosvenor, T., Bouma, A., 1988. An Application of Computer-based Stated Preference to Study Mode-switching in the Netherlands. Planning and Transport Research and Computation, 307-319.

Brownstone, D., Bunch, D.S., Train, K., 2000. Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles. Transportation Research B 34, 315-338.

Brownstone, D., Train, K., 1999. Forecasting New Product Penetration with Flexible Substitution Patterns. Journal of Econometrics 89, 109-129.

Cambridge Systematics, Inc., 1986. Customer Preference and Behavior Project Report. Prepared for the Electric Power Research Institute.

Cardell, N.S., Dunbar, F.C., 1980. Measuring the Societal Impacts of Automobile Downsizing. Transportation Research A 14, 423-434.

Eymann, A., Boersch-Supan, A., Euwals, R., 2001. Risk Attitude, Impatience, and Portfolio Choice. Working paper, University of Mannheim, Germany.

Goett, A., Hudson, K., Train, K., 2000. Customers' Choice Among Retail Energy Suppliers: The Willingness-to-Pay for Service Attributes. Energy Journal 21(4), 1-28.

Gopinath, D.A., 1995. Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand. Ph.D. Dissertation, Massachusetts Institute of Technology.

Greene, W.H., 2000. Econometric Analysis, Fourth Edition, Prentice Hall, Upper Saddle River, New Jersey.

Hensher, D.A., Greene, W.H., 2001. The Mixed Logit Model: The State of Practice and Warnings for the Unwary. Working paper, University of Sydney.

Hosoda, T., 1999. Incorporating Unobservable Heterogeneity in Discrete Choice Model: Mode Choice Model for Shopping Trips. Masters Thesis, Massachusetts Institute of Technology.

Jöreskog, K.G., 1973. A General Method for Estimating a Linear Structural Equation System. In: Goldberger, A.S., Duncan, O.D., (Eds.), Structural Models in the Social Sciences. Academic Press, New York.

Kamakura, W.A., Russell, G.J., 1989. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. Journal of Marketing Research 25, 379-390.

Keesling, J.W., 1972. Maximum Likelihood Approaches to Causal Analysis. Ph.D. Dissertation, University of Chicago.

Khattak, A., Polydoropoulou, A., Ben-Akiva, M., 1996. Modeling Revealed and Stated Pretrip Travel Response to Advanced Traveler Information Systems. Transportation Research Record 1537, 46-54.

Lancaster, K., 1966. A New Approach to Consumer Theory. Journal of Political Economy 74, 132-157.

Luce, R.D., 1959. Individual Choice Behavior, Wiley, New York.

Marschak, J., 1960. Binary Choice Constraints on Random Utility Indicators. In: Arrow, K., (Ed.), Stanford Symposium on Mathematical Methods in the Social Sciences. Stanford University Press.

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P., (Ed.), Frontiers of Econometrics. Academic Press.

McFadden, D., 1978. Modelling the Choice of Residential Location. In: Snickars, F., Weibull, J., (Eds.), Spatial Interaction Theory and Planning Models. North-Holland.

McFadden, D., 1984. Econometric Analysis of Qualitative Response Models. In: Griliches, Z., Intriligator, M.D., (Eds.), Handbook of Econometrics II. Elsevier Science Publishers.

McFadden, D., 1986. The Choice Theory Approach to Marketing Research. Marketing Science 5(4), 275-297.

McFadden, D., 1989. A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. Econometrica 57(5), 995-1026.

McFadden, D., Train, K., 2000. Mixed MNL Models for Discrete Response. Journal of Applied Econometrics 15(5), 446-470.

Mehndiratta, R.M., Hansen, M., 1997. Analysis of Discrete Choice Data with Repeated Observations: Comparison of Three Techniques in Intercity Travel Case. Transportation Research Record 1607, 69-73.

Morikawa, T., 1994. Correcting State Dependence and Serial Correlation in the RP/SP Combined Estimation Method. Transportation 21, 153-165.

Morikawa, T., Ben-Akiva, M., McFadden, D., 2002. Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data. Econometric Models in Marketing 16, 27-53.

Newey, W., McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing. In: Engle, R., and McFadden, D., (Eds.), Handbook of Econometrics IV, 2111-2245.

Polydoropoulou, A., 1997. Modeling User Response to Advanced Traveler Information Systems (ATIS). Ph.D. Dissertation, Massachusetts Institute of Technology.

Ramming, M.S., 2002. Network Knowledge and Route Choice. Ph.D. Dissertation, Massachusetts Institute of Technology.

Revelt, D., Train, K., 1998. Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level. Review of Economics and Statistics 80(4), 647-657.

Srinivasan, K.K., Mahmassani, H.S., 2000. Dynamic Kernel Logit Model for the Analysis of Longitudinal Discrete Choice Data: Properties and Computational Assessment. Presented at the International Association of Travel Behavior Research (IATBR) Conference, Gold Coast, Queensland, Australia.

Thurstone, L., 1927. A Law of Comparative Judgment. Psychological Review 34, 273-286.

Train, K.E., 2001. Discrete Choice Methods with Simulation, forthcoming Cambridge University Press.

Train, K.E., 2000. A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit. Working paper, University of California at Berkeley.

Train, K.E., McFadden, D.L., Goett, A.A., 1987. Consumer Attitudes and Voluntary Rate Schedules for Public Utilities. The Review of Economics and Statistics LXIX(3), 383-391.

Walker, J.L., 2001. Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables. Ph.D. Dissertation, Massachusetts Institute of Technology.

Wiley, D.E., 1973. The Identification Problem for Structural Equation Models with Unmeasured Variables. In: Goldberger, A.S., Duncan, O.D., (Eds.), Structural Models in the Social Sciences. Academic Press, New York.

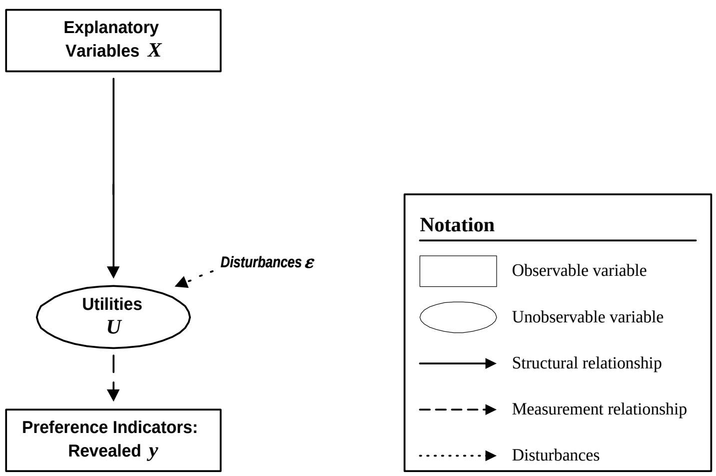



**Figure 1: Random Utility Model**

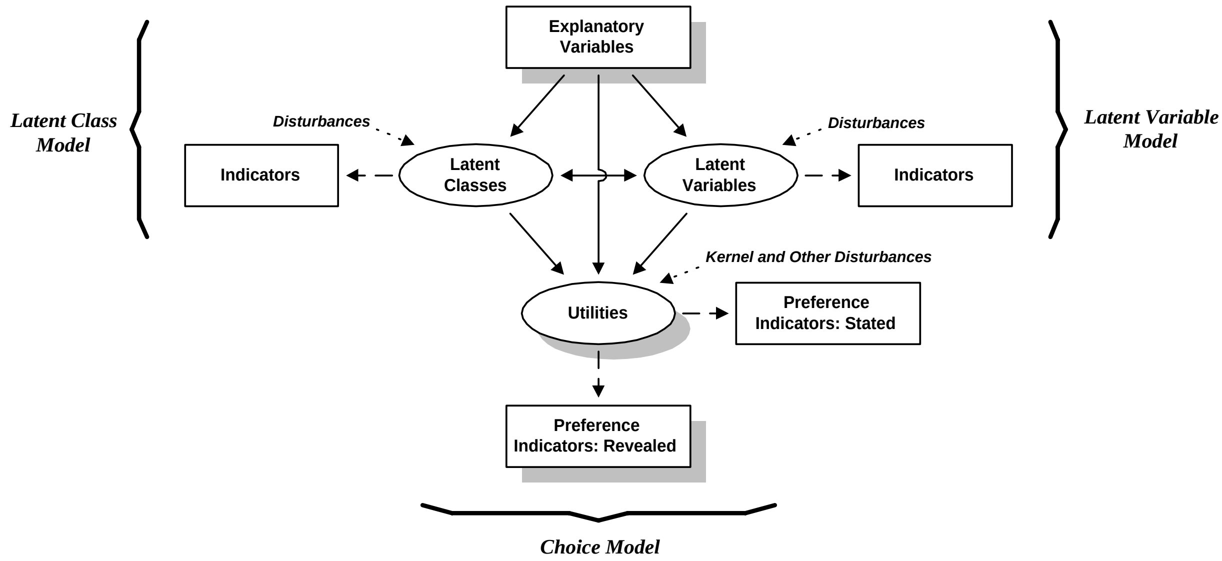



**Figure 2: Generalized Random Utility Model**

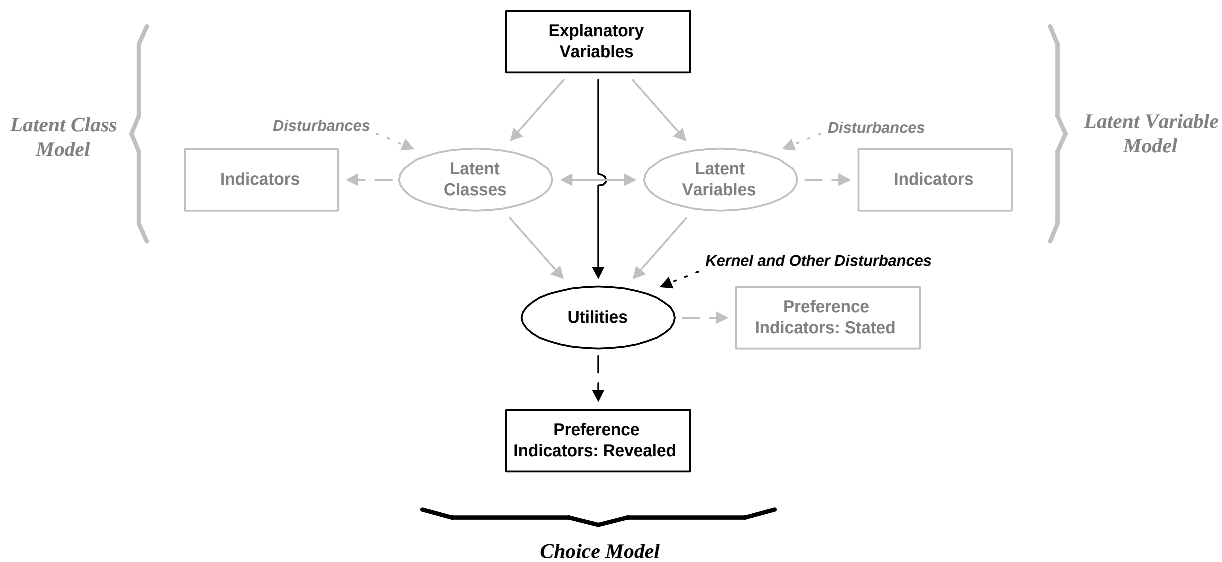



**Figure 3: Flexible Disturbances**

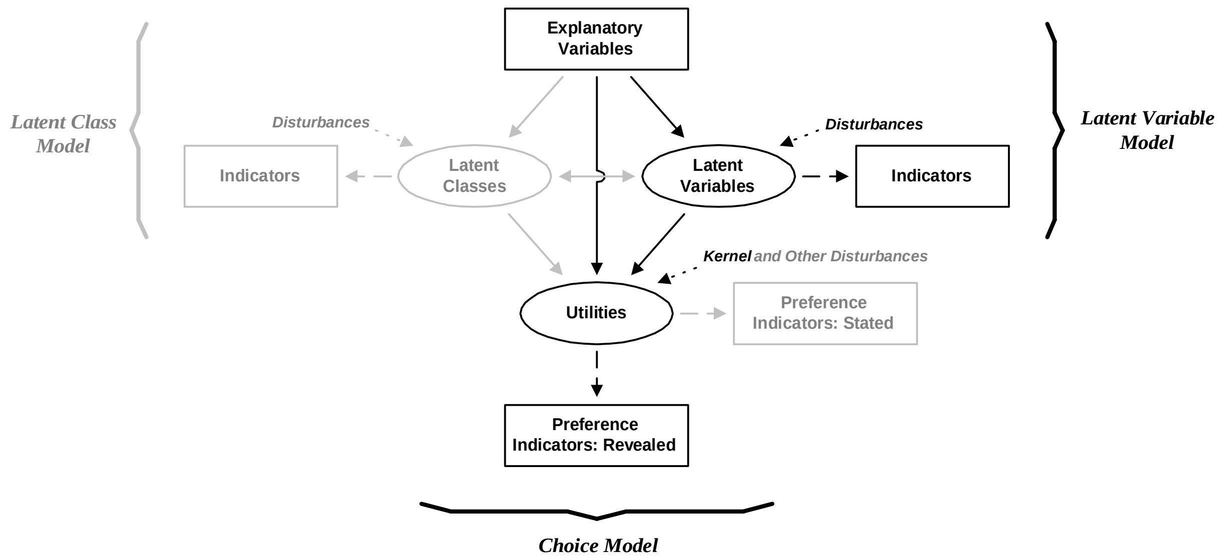



**Figure 4: Integrated Choice & Latent Variable Model**

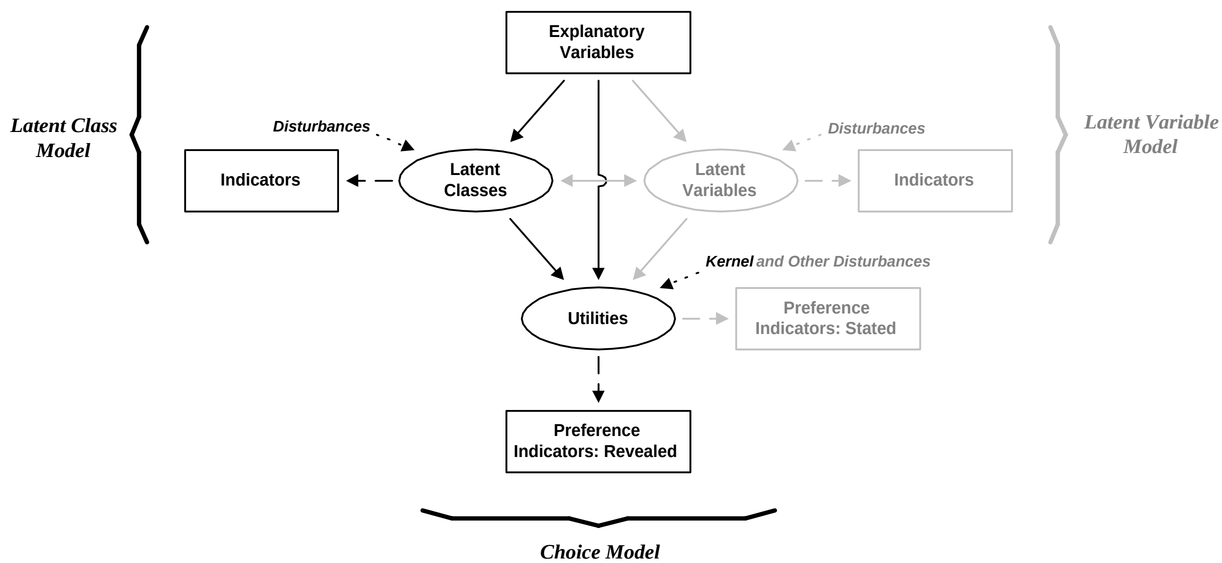



**Figure 5: Random Utility Model with Latent Classes**

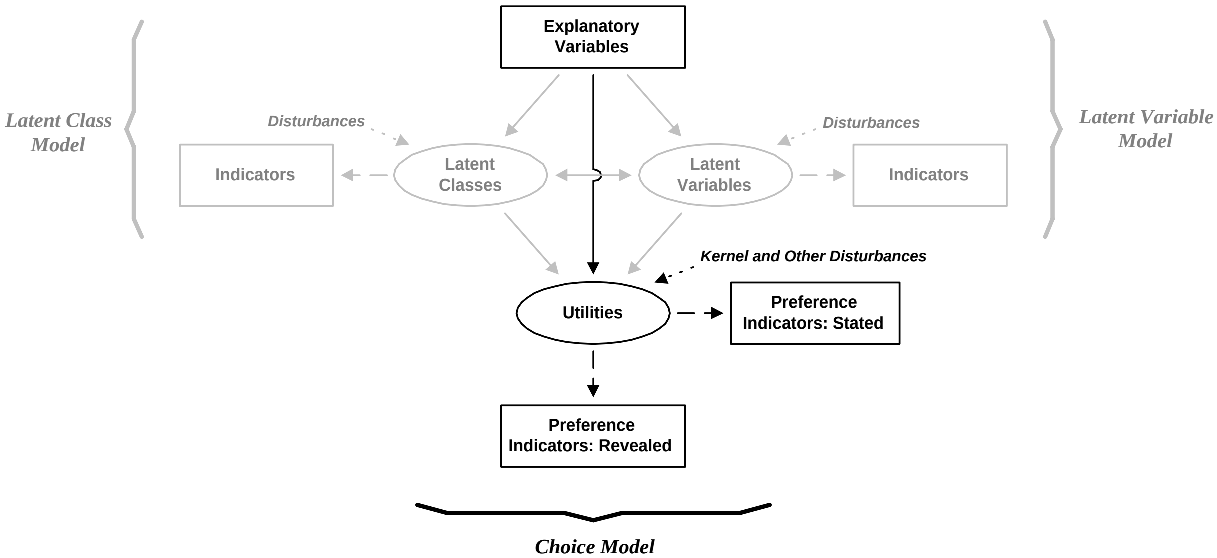



**Figure 6: Joint Revealed and Stated Preference Model**

**Table 1: Revealed Preference Binary Logit Mode Choice Model (Rail versus Auto)**

| Parameter | $U_{rail}$ | $U_{auto}$ | Est. | Std Er. | *t-stat* |
|---|---|---|---|---|---|
| Rail constant | ✓ | | 0.637 | 0.425 | *(1.5)* |
| Work trip dummy | ✓ | | 1.21 | 0.48 | *(2.5)* |
| Fixed arrival time dummy | ✓ | | 0.736 | 0.368 | *(2.0)* |
| Female dummy | ✓ | | 0.949 | 0.352 | *(2.7)* |
| Cost per person in Guilders | ✓ | ✓ | -0.0477 | 0.0122 | *(3.9)* |
| Out-of-vehicle time in hours | ✓ | ✓ | -2.90 | 0.80 | *(3.6)* |
| In-vehicle time in hours | ✓ | ✓ | -0.554 | 0.462 | *(1.2)* |
| Number of transfers | ✓ | | -0.255 | 0.255 | *(1.0)* |
| Number of observations: | | | 228 | | |
| Log-likelihood: | | | -109.89 | | |
| Rho-bar-squared: | | | 0.254 | | |

## Table 2: Joint Stated Preference & Revealed Preference Mode Choice Model

| | | | Joint RP/SP1/SP2 | | | RP Only (Rail vs. Auto) | | | SP1 Only (Rail vs. Rail) | | | SP2 Only (Rail vs. Auto) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | $U_{rail}$ | $U_{auto}$ | Est. | *Std Er.* | *t-stat* | Est. | *Std Er.* | *t-stat* | Est. | *Std Er.* | *t-stat* | Est. | *Std Er.* | *t-stat* |
| Rail constant RP | ✓ | | 0.444 | 0.493 | *(0.9)* | 0.637 | 0.425 | *(1.5)* | | | | | | |
| Rail constant SP2 | ✓ | | -0.466 | 0.777 | *(0.6)* | | | | | | | -2.10 | 0.63 | *(3.3)* |
| Work trip dummy | ✓ | | 1.17 | 0.51 | *(2.3)* | 1.21 | 0.48 | *(2.5)* | | | | | | |
| Fixed arrival time dummy | ✓ | | 0.723 | 0.381 | *(1.9)* | 0.736 | 0.368 | *(2.0)* | | | | | | |
| Female dummy | ✓ | | 0.990 | 0.381 | *(2.6)* | 0.949 | 0.352 | *(2.7)* | | | | | | |
| Cost per person in Guilders | ✓ | ✓ | -0.0608 | 0.0132 | *(4.6)* | -0.0477 | 0.0122 | *(3.9)* | -0.141 | 0.012 | *(11.8)* | -0.0703 | 0.0180 | *(3.9)* |
| Out-of-vehicle time in hours | ✓ | ✓ | -2.23 | 0.83 | *(2.7)* | -2.90 | 0.80 | *(3.6)* | | | | -0.841 | 0.935 | *(0.9)* |
| In-vehicle time in hours | ✓ | ✓ | -0.710 | 0.158 | *(4.5)* | -0.554 | 0.462 | *(1.2)* | -1.64 | 0.16 | *(10.2)* | -1.23 | 0.41 | *(3.0)* |
| Number of transfers | ✓ | | -0.100 | 0.036 | *(2.8)* | -0.255 | 0.255 | *(1.0)* | -0.238 | 0.066 | *(3.6)* | 0.0798 | 0.1995 | *(0.4)* |
| Amenities | ✓ | | -0.361 | 0.080 | *(4.5)* | | | | -0.821 | 0.073 | *(11.2)* | -0.925 | 0.237 | *(3.9)* |
| Inertia dummy (RP Choice) | ✓ | | 2.97 | 1.02 | *(2.9)* | | | | | | | 5.92 | 0.68 | *(8.7)* |
| Correlation Term RP | | | 0.686 | 0.490 | *(1.4)* | | | | | | | | | |
| Correlation Term SP2 | | | 2.44 | 0.50 | *(4.9)* | | | | | | | 3.11 | 0.29 | *(10.8)* |
| Scale (mu) SP1 | | | 2.31 | 0.50 | *(4.6)* | | | | | | | | | |
| Scale (mu) SP2 | | | 1.31 | 0.30 | *(4.4)* | | | | | | | | | |
| *Tau1 SP1 (=-Tau4 SP1)* | | | *-0.195* | ---- | ---- | | | | *-0.450* | ---- | ---- | | | |
| *Tau2 SP1 (=-Tau3 SP1)* | | | *-0.0127* | ---- | ---- | | | | *-0.0292* | ---- | ---- | | | |
| Tau3 SP1 | | | 0.0127 | 0.0036 | *(3.5)* | | | | 0.0292 | 0.0060 | *(4.9)* | | | |
| Tau4 SP1 | | | 0.195 | 0.049 | *(4.0)* | | | | 0.450 | 0.038 | *(11.7)* | | | |
| Tau1 SP2 | | | -0.986 | 0.219 | *(4.5)* | | | | | | | -1.30 | 0.13 | *(10.2)* |
| *Tau2 SP2 (=-Tau3 SP2)* | | | *-0.180* | ---- | ---- | | | | | | | *-0.238* | ---- | ---- |
| Tau3 SP2 | | | 0.180 | 0.053 | *(3.4)* | | | | | | | 0.238 | 0.055 | *(4.3)* |
| Tau4 SP2 | | | 1.32 | 0.32 | *(4.1)* | | | | | | | 1.75 | 0.18 | *(9.6)* |
| Number of observations: | | | 4680 | | | 228 | | | 2875 | | | 1577 | | |
| Number of draws (Halton): | | | 1000 | | | 1000 | | | 1000 | | | 1000 | | |
| Log-likelihood: | | | -4517.43 | | | -109.89 | | | -3131.10 | | | -1271.29 | | |
| Rho-bar-squared: | | | 0.380 | | | 0.254 | | | 0.322 | | | 0.495 | | |

# Table 3: Random Parameter Mode Choice Model

| Parameter | Base RP/SP Model (Table 2) Est. | Std. Er. | t-stat | Distributed Model Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat |
|---|---|---|---|---|---|---|---|---|---|
| Rail constant RP | 0.444 | 0.493 | *(0.9)* | 2.80 | 0.97 | *(2.9)* | | | |
| Rail constant SP2 | -0.466 | 0.777 | *(0.6)* | 4.05 | 1.20 | *(3.4)* | | | |
| Work trip dummy | 1.17 | 0.51 | *(2.3)* | 0.891 | 0.762 | *(1.2)* | | | |
| Fixed arrival time dummy | 0.723 | 0.381 | *(1.9)* | 0.513 | 0.647 | *(0.8)* | | | |
| Female dummy | 0.990 | 0.381 | *(2.6)* | 1.61 | 0.61 | *(2.7)* | | | |
| | | | | Location Parameter (δ1) | | | Dispersion Parameter (δ2) | | |
| Cost per person in Guilders | -0.0608 | 0.0132 | *(4.6)* | -2.19 | 0.26 | * | 0.993 | 0.129 | *(7.7)* |
| Out-of-vehicle time in hours | -2.23 | 0.83 | *(2.7)* | 1.56 | 0.34 | * | 0.723 | 0.166 | *(4.4)* |
| In-vehicle time in hours | -0.710 | 0.158 | *(4.5)* | 0.284 | 0.279 | * | 0.818 | 0.057 | *(14.3)* |
| Number of transfers | -0.100 | 0.036 | *(2.8)* | -2.29 | 0.33 | * | 1.96 | 0.21 | *(9.3)* |
| Amenities | -0.361 | 0.080 | *(4.5)* | -0.644 | 0.265 | * | 1.06 | 0.05 | *(20.9)* |
| Inertia dummy (RP Choice) | 2.97 | 1.02 | *(2.9)* | -0.245 | 0.680 | *(0.4)* | | | |
| Correlation Term RP | 0.686 | 0.490 | *(1.4)* | 3.19 | 1.28 | *(2.5)* | | | |
| Correlation Term SP2 | 2.44 | 0.50 | *(4.9)* | 4.14 | 1.14 | *(3.6)* | | | |
| Scale (mu) SP1 | 2.31 | 0.50 | *(4.6)* | 4.07 | 1.11 | *(3.7)* | | | |
| Scale (mu) SP2 | 1.31 | 0.30 | *(4.4)* | 1.79 | 0.48 | *(3.8)* | | | |
| *Tau1 SP1 (=-Tau4 SP1)* | *-0.195* | ---- | | *-0.241* | | | | | |
| *Tau2 SP1 (=-Tau3 SP1)* | *-0.0127* | ---- | | *-0.0159* | | | | | |
| Tau3 SP1 | 0.0127 | 0.0036 | *(3.5)* | 0.0159 | 0.0052 | *(3.0)* | | | |
| Tau4 SP1 | 0.195 | 0.049 | *(4.0)* | 0.241 | 0.081 | *(3.0)* | | | |
| Tau1 SP2 | -0.986 | 0.219 | *(4.5)* | -0.904 | 0.241 | *(3.8)* | | | |
| *Tau2 SP2 (=-Tau3 SP2)* | *-0.180* | ---- | | *-0.160* | | | | | |
| Tau3 SP2 | 0.180 | 0.053 | *(3.4)* | 0.160 | 0.055 | *(2.9)* | | | |
| Tau4 SP2 | 1.32 | 0.32 | *(4.1)* | 1.15 | 0.31 | *(3.8)* | | | |
| Number of observations: | 4680 | | | 4680 | | | | | |
| Number of draws (Halton): | 1000 | | | 20000 | | | | | |
| Log-likelihood: | -4517.43 | | | -3931.20 | | | | | |
| Rho-bar-squared: | 0.380 | | | 0.460 | | | | | |

(Cost per person through Amenities: Lognormally Distributed)

* Testing that the lognormal location parameter is different from 0 is meaningless.

# Table 4: Mean and Standard Deviations of the Distributed Parameters in Table 3

| | Base RP/SP Model (Table 2) | | | Distributed Model | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean of the parameter | | | Mean of the parameter | | | Standard Deviation | | |
| Parameter | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat |
| Cost per person in Guilders | -0.0608 | 0.0132 | *(4.6)* | -0.183 | 0.059 | *(3.1)* | 0.237 | 0.097 | *(2.4)* |
| Out-of-vehicle time in hours | -2.23 | 0.83 | *(2.7)* | -6.19 | 1.71 | *(3.6)* | 5.12 | 1.77 | *(2.9)* |
| In-vehicle time in hours | -0.710 | 0.158 | *(4.5)* | -1.86 | 0.49 | *(3.8)* | 1.81 | 0.41 | *(4.4)* |
| Number of transfers | -0.100 | 0.036 | *(2.8)* | -0.689 | 0.860 | *(0.8)* | 4.63 | 3.61 | *(1.3)* |
| Amenities | -0.361 | 0.080 | *(4.5)* | -0.922 | 0.247 | *(3.7)* | 1.33 | 0.31 | *(4.3)* |

# Table 5: Latent Class Mode Choice Model

***MODE CHOICE MODEL***

| | Base RP/SP Model (Table 2) | | | Latent Class Model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Parameters Common Across Classes | | | Parameters Unique to Class 1 | | | Parameters Unique to Class 2 | | |
| Parameter | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat |
| Rail constant RP | 0.444 | 0.493 | *(0.9)* | 1.26 | 0.756 | *(1.7)* | | | | | | |
| Rail constant SP2 | -0.466 | 0.777 | *(0.6)* | 1.42 | 0.772 | *(1.8)* | | | | | | |
| Work trip dummy | 1.17 | 0.51 | *(2.3)* | 1.10 | 0.620 | *(1.8)* | | | | | | |
| Fixed arrival time dummy | 0.723 | 0.381 | *(1.9)* | 0.641 | 0.497 | *(1.3)* | | | | | | |
| Female dummy | 0.990 | 0.381 | *(2.6)* | 1.03 | 0.432 | *(2.4)* | | | | | | |
| Cost per person in Guilders | -0.0608 | 0.0132 | *(4.6)* | | | | -0.231 | 0.063 | *(3.7)* | -0.0408 | 0.0115 | *(3.5)* |
| Out-of-vehicle time in hours | -2.23 | 0.83 | *(2.7)* | | | | -1.31 | 1.21 | *(1.1)* | -3.47 | 1.34 | *(2.6)* |
| In-vehicle time in hours | -0.710 | 0.158 | *(4.5)* | | | | -1.69 | 0.48 | *(3.5)* | -0.876 | 0.244 | *(3.6)* |
| Number of transfers | -0.100 | 0.036 | *(2.8)* | | | | -0.216 | 0.092 | *(2.3)* | -0.149 | 0.055 | *(2.7)* |
| Amenities | -0.361 | 0.080 | *(4.5)* | | | | -0.408 | 0.114 | *(3.6)* | -0.540 | 0.146 | *(3.7)* |
| Inertia dummy (RP Choice) | 2.97 | 1.02 | *(2.9)* | 0.99 | 0.696 | *(1.4)* | | | | | | |
| Correlation Term RP | 0.686 | 0.490 | *(1.4)* | 2.09 | 0.76 | *(2.8)* | | | | | | |
| Correlation Term SP2 | 2.44 | 0.50 | *(4.9)* | 2.87 | 0.73 | *(3.9)* | | | | | | |
| Scale (mu) SP1 | 2.31 | 0.50 | *(4.6)* | 2.25 | 0.59 | *(3.8)* | | | | | | |
| Scale (mu) SP2 | 1.31 | 0.30 | *(4.4)* | 1.56 | 0.35 | *(4.5)* | | | | | | |
| *Tau1 SP1 (=-Tau4 SP1)* | *-0.195* | ---- | ---- | -0.236 | ---- | ---- | | | | | | |
| *Tau2 SP1 (=-Tau3 SP1)* | *-0.0127* | ---- | ---- | -0.0154 | ---- | ---- | | | | | | |
| Tau3 SP1 | 0.0127 | 0.0036 | *(3.5)* | 0.0154 | 0.0050 | *(3.1)* | | | | | | |
| Tau4 SP1 | 0.195 | 0.049 | *(4.0)* | 0.236 | 0.070 | *(3.4)* | | | | | | |
| Tau1 SP2 | -0.986 | 0.219 | *(4.5)* | -0.895 | 0.210 | *(4.3)* | | | | | | |
| *Tau2 SP1 (=-Tau3 SP2)* | *-0.180* | ---- | ---- | -0.161 | ---- | ---- | | | | | | |
| Tau3 SP2 | 0.180 | 0.053 | *(3.4)* | 0.161 | 0.051 | *(3.1)* | | | | | | |
| Tau4 SP2 | 1.32 | 0.32 | *(4.1)* | 1.17 | 0.28 | *(4.2)* | | | | | | |
| Number of observations: | 4680 | | | 4680 | | | | | | | | |
| Number of draws (Halton): | 1000 | | | 1000 | | | | | | | | |
| Log-likelihood: | -4517.43 | | | -4283.04 | | | | | | | | |
| Rho-bar-squared: | 0.380 | | | 0.411 | | | | | | | | |

***CLASS MEMBERSHIP MODEL***

| Parameter | Est. | Std. Er. | t-stat |
|---|---|---|---|
| Constant | -0.455 | 0.395 | *(1.2)* |
| Female dummy | -0.0832 | 0.3625 | *(0.2)* |
| Number of persons in party | 0.174 | 0.121 | *(1.4)* |
| Work trip dummy | -1.94 | 0.73 | *(2.7)* |
| Age over 40 dummy | -0.472 | 0.371 | *(1.3)* |

| *Class Membership Statistics* |
|---|
| Probability(Class 1) < 0.2 for 16% of the sample |
| 0.2 <= Probability(Class 1) < 0.4 for 19% of the sample |
| 0.4 <= Probability(Class 1) < 0.6 for 62% of the sample |
| 0.6 <= Probability(Class 1) < 0.8 for 3% of the sample |
| Probability(Class 1) >= 0.6 for 0% of the sample |

# Table 6: Choice and Latent Variable Mode Choice Model

***CHOICE MODEL***

| Parameter | Base RP/SP Choice Model Est. | Std. Er. | t-stat | Choice and Latent Variable RP/SP Model (latent variable portion below) Est. | Std. Er. | t-stat |
|---|---|---|---|---|---|---|
| Rail constant RP | 0.444 | 0.493 | *(0.9)* | -0.442 | 0.750 | *(0.6)* |
| Rail constant SP2 | -0.466 | 0.777 | *(0.6)* | -0.890 | 0.837 | *(1.1)* |
| Work trip dummy | 1.17 | 0.51 | *(2.3)* | 1.67 | 0.64 | *(2.6)* |
| Fixed arrival time dummy | 0.723 | 0.381 | *(1.9)* | 0.692 | 0.532 | *(1.3)* |
| Female dummy | 0.99 | 0.38 | *(2.6)* | 1.13 | 0.45 | *(2.5)* |
| Cost per person in Guilders | -0.0608 | 0.0132 | *(4.6)* | -0.0605 | 0.0163 | *(3.7)* |
| Out-of-vehicle time in hours | -2.23 | 0.83 | *(2.7)* | -0.983 | 0.936 | *(1.1)* |
| In-vehicle time in hours | -0.710 | 0.158 | *(4.5)* | -0.691 | 0.186 | *(3.7)* |
| Number of transfers | -0.100 | 0.036 | *(2.8)* | -0.0982 | 0.0384 | *(2.6)* |
| Amenities | -0.361 | 0.080 | *(4.5)* | -0.358 | 0.097 | *(3.7)* |
| *Latent Comfort - RP* | | | | 1.16 | 1.17 | *(1.0)* |
| *Latent Comfort - SP2* | | | | 1.16 | 0.55 | *(2.1)* |
| *Latent Convenience - RP* | | | | 1.30 | 0.76 | *(1.7)* |
| *Latent Convenience - SP2* | | | | 0.764 | 0.331 | *(2.3)* |
| Inertia dummy (RP Choice) | 2.97 | 1.02 | *(2.9)* | 2.52 | 1.24 | *(2.0)* |
| Correlation Term RP | 0.686 | 0.490 | *(1.4)* | 0.210 | 0.611 | *(0.3)* |
| Correlation Term SP2 | 2.44 | 0.50 | *(4.9)* | 2.08 | 0.64 | *(3.3)* |
| Scale (mu) SP1 | 2.31 | 0.50 | *(4.6)* | 2.32 | 0.63 | *(3.7)* |
| Scale (mu) SP2 | 1.31 | 0.30 | *(4.4)* | 1.31 | 0.42 | *(3.1)* |
| *Tau1 SP1 (=-Tau4 SP1)* | -0.195 | ---- | ---- | -0.194 | ---- | ---- |
| *Tau2 SP1 (=-Tau3 SP1)* | -0.0127 | ---- | ---- | -0.0126 | ---- | ---- |
| Tau3 SP1 | 0.0127 | 0.0036 | *(3.5)* | 0.0126 | 0.0041 | *(3.0)* |
| Tau4 SP1 | 0.195 | 0.049 | *(4.0)* | 0.194 | 0.058 | *(3.3)* |
| Tau1 SP2 | -0.986 | 0.219 | *(4.5)* | -0.988 | 0.313 | *(3.2)* |
| *Tau2 SP2 (=-Tau3 SP2)* | -0.180 | ---- | ---- | -0.181 | ---- | ---- |
| Tau3 SP2 | 0.180 | 0.053 | *(3.4)* | 0.181 | 0.065 | *(2.8)* |
| Tau4 SP2 | 1.32 | 0.32 | *(4.1)* | 1.33 | 0.44 | *(3.0)* |
| Number of observations: | 4680 | | | 4680 | | |
| Number of draws (Halton): | 1000 | | | 5000 | | |
| Log-likelihood (Choice&Latent): | | | | -6656.12 | | |
| Log-likelihood (Choice): | -4517.43 | | | -4517.97 | | |
| Rho-bar-squared (Choice): | 0.380 | | | 0.380 | | |

***LATENT VARIABLE MODEL***

*Structural Equations (2 equations, 1 per column)*

| Parameter | Comfort Equation Est. | Std. Er. | t-stat | Convenience Equation Est. | Std. Er. | t-stat |
|---|---|---|---|---|---|---|
| Constant - Comfort | 0.106 | 0.219 | *(0.5)* | | | |
| Constant - Convenience | | | | 0.489 | 0.303 | *(1.6)* |
| Age dummy - over 40 | -0.449 | 0.622 | *(0.7)* | 0.871 | 0.287 | *(3.0)* |
| First class rail rider | 0.431 | 0.567 | *(0.8)* | | | |
| In-vehicle time in hours | -0.481 | 0.331 | *(1.5)* | | | |
| Out-of-vehicle time in hours | | | | -1.18 | 0.71 | *(1.6)* |
| Number of transfers | | | | -0.122 | 0.199 | *(0.6)* |
| Free parking dummy (auto) | | | | 0.222 | 0.242 | *(0.9)* |
| Variance($\omega$) | 1.00 | ---- | ---- | 1.00 | ---- | ---- |
| Squared Multiple Correlation (SMC) | 0.092 | | | 0.230 | | |

*Measurement Equations (6 equations, 1 per row)*

| Equation | Comfort Parameters Est. | Std. Er. | t-stat | Convenience Parameters Est. | Std. Er. | t-stat | Disturbance Params. (StdDev($\upsilon$)) Est. | Std. Er. | t-stat | Fit (SMC) |
|---|---|---|---|---|---|---|---|---|---|---|
| Relaxation | 0.522 | 0.240 | *(2.2)* | 0.131 | 0.135 | *(1.0)* | 1.17 | 0.13 | *(9.3)* | 0.172 |
| Reliability | 0.331 | 0.105 | *(3.1)* | 0.446 | 0.089 | *(5.0)* | 0.899 | 0.055 | *(16.3)* | 0.263 |
| Flexibility | | | | 0.731 | 0.288 | *(2.5)* | 0.877 | 0.242 | *(3.6)* | 0.366 |
| Ease | | | | 0.571 | 0.168 | *(3.4)* | 1.15 | 0.09 | *(12.1)* | 0.188 |
| Safety | 0.381 | 0.135 | *(2.8)* | 0.132 | 0.117 | *(1.1)* | 0.803 | 0.081 | *(10.0)* | 0.197 |
| Overall Rating | 1.25 | 0.82 | *(1.5)* | 1.39 | 0.51 | *(2.7)* | 1.28 | 0.26 | *(5.0)* | 0.616 |

# Table 7: Choice and Latent Variable Mode Choice Model with Random Parameters

***CHOICE MODEL***

| | Location Parameters | | | Dispersion Parameters (δ2) | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | |
| Rail constant RP | 0.100 | 0.796 | *(0.1)* | | | | |
| Rail constant SP2 | 1.53 | 0.67 | *(2.3)* | | | | |
| Work trip dummy | 1.07 | 0.83 | *(1.3)* | | | | |
| Fixed arrival time dummy | 0.397 | 0.651 | *(0.6)* | | | | |
| Female dummy | 1.48 | 0.63 | *(2.4)* | | | | |
| Cost per person in Guilders | -2.18 | 0.29 | * | 1.02 | 0.05 | *(22.1)* | *lognormal* |
| Out-of-vehicle time in hours | 0.06 | 0.94 | *(0.1)* | | | | |
| In-vehicle time in hours | 0.228 | 0.305 | * | 0.864 | 0.040 | *(21.5)* | *lognormal* |
| Number of transfers | -2.14 | 0.38 | * | 1.76 | 0.15 | *(12.0)* | *lognormal* |
| Amenities | -0.609 | 0.271 | * | 1.13 | 0.05 | *(22.3)* | *lognormal* |
| *Latent Comfort - RP* | 2.98 | 0.84 | *(3.5)* | | | | |
| *Latent Comfort - SP2* | 3.08 | 0.87 | *(3.5)* | | | | |
| *Latent Convenience - RP* | 1.54 | 0.37 | *(4.2)* | | | | |
| *Latent Convenience - SP2* | 1.18 | 0.37 | *(3.2)* | | | | |
| Inertia dummy (RP Choice) | -1.05 | 0.57 | *(1.8)* | | | | |
| Correlation Term RP | 1.00 | ---- | ---- | | | | |
| Correlation Term SP2 | 1.84 | 0.53 | *(3.5)* | | | | |
| Scale (mu) SP1 | 4.28 | 1.24 | *(3.5)* | | | | |
| Scale (mu) SP2 | 2.03 | 0.55 | *(3.7)* | | | | |
| *Tau1 SP1 (=-Tau4 SP1)* | -0.229 | ---- | ---- | | | | |
| *Tau2 SP1 (=-Tau3 SP1)* | -0.0152 | ---- | ---- | | | | |
| Tau3 SP1 | 0.0152 | 0.0053 | *(2.9)* | | | | |
| Tau4 SP1 | 0.229 | 0.083 | *(2.8)* | | | | |
| Tau1 SP2 | -0.812 | 0.220 | *(3.7)* | | | | |
| *Tau2 SP1 (=-Tau3 SP2)* | -0.143 | ---- | ---- | | | | |
| Tau3 SP2 | 0.143 | 0.049 | *(2.9)* | | | | |
| Tau4 SP2 | 1.03 | 0.28 | *(3.7)* | | | | |
| Number of observations: | 4680 | | | | | | |
| Number of draws (Halton): | 20000 | | | | | | |
| Log-likelihood (Choice&Latent): | -6066.08 | | | | | | |
| Log-likelihood (Choice): | -3935.04 | | | | | | |
| Rho-bar-squared (Choice): | 0.458 | | | | | | |

***LATENT VARIABLE MODEL***

*Structural Equations (2 equations, 1 per column)*

| | Comfort Equation | | | | | | Convenience Equation | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Location Parameters | | | Dispersion Parameters (δ2) | | | Location Parameters | | | Dispersion Parameters (δ2) | | | |
| Parameter | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | |
| Constant - Comfort | 0.0688 | 0.1362 | *(0.5)* | | | | | | | | | | |
| Constant - Convenience | | | | | | | 0.649 | 0.239 | *(2.7)* | | | | |
| Age dummy - over 40 | -0.435 | 0.145 | *(3.0)* | | | | 0.961 | 0.286 | *(3.4)* | -0.281 | 0.072 | *(3.9)* | *normal* |
| First class rail rider | -0.434 | 0.211 | *(2.1)* | | | | | | | | | | |
| In-vehicle time in hours | -3.03 | 0.43 | * | 1.964 | 0.154 | *(12.7)* | | | | | | | *lognormal* |
| Out-of-vehicle time in hours | | | | | | | 0.246 | 0.386 | * | -0.674 | 0.133 | *(5.1)* | *lognormal* |
| Number of transfers | | | | | | | -0.294 | 0.126 | *(2.3)* | | | | |
| Free parking dummy (auto) | | | | | | | 0.147 | 0.180 | *(0.8)* | | | | |
| Variance(ω) | 1.00 | ---- | ---- | | | | 1.00 | ---- | ---- | | | | |

*Measurement Equations (6 equations, 1 per row)*

| | Comfort Parameters | | | Convenience Parameters | | | Disturbance Params. (StdDev(υ)) | | |
|---|---|---|---|---|---|---|---|---|---|
| Equation | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat | Est. | Std. Er. | t-stat |
| Relaxation | 0.408 | 0.138 | *(3.0)* | 0.136 | 0.084 | *(1.6)* | 1.20 | 0.07 | *(16.4)* |
| Reliability | 0.220 | 0.100 | *(2.2)* | 0.402 | 0.072 | *(5.6)* | 0.896 | 0.052 | *(17.1)* |
| Flexibility | | | | 0.603 | 0.109 | *(5.6)* | 0.870 | 0.087 | *(10.0)* |
| Ease | | | | 0.453 | 0.085 | *(5.3)* | 1.16 | 0.07 | *(15.8)* |
| Safety | 0.242 | 0.095 | *(2.5)* | 0.152 | 0.069 | *(2.2)* | 0.838 | 0.044 | *(19.1)* |
| Overall Rating | 1.05 | 0.13 | *(8.1)* | 1.12 | 0.12 | *(9.0)* | 1.39 | 0.14 | *(10.0)* |

## Table 8: Summary of Case Study Models

| Table | Model Description | *Model Components* Revealed Preference | Stated Preferences | Flexible Disturbances | Latent Variables | Latent Class | Number of parameters (choice model + other model) | Goodness of fit (Choice Model portion only) | Comments |
|---|---|---|---|---|---|---|---|---|---|
| 1 | RP Model | X | | | | | 8 | ---- | |
| 2 | RP/SP Model | X | X | | | | 20 | 0.380 | Key parameters became significant. New variable captured (amenities). |
| 3 | RP/SP with Random Parameters on Attributes of Auto & Rail | X | X | X | | | 25 | 0.460 | Resulted in large improvement in fit. Suggests parameters are not constant in population. |
| 5 | RP/SP with Latent Class Taste Heterogeneity | X | X | | | X | 25 + 5 | 0.411 | 2 classes: each with own set of attribute parameters. class 1: younger, recreational, group travelers more sensitive to cost, in-vehicle time, transfers class 2: older, business, travelers more sensitive to out-of-vehicle time, amenities Improvement not as strong as random parameters. |
| 6 | RP/SP with Latent Variables of 'Comfort' and 'Convenience' | X | X | | X | | 24 + 25 | 0.380 | Marginally significant latent variable parameters. Weak causal relationship. More behaviorally satisfying. |
| 7 | Generalized | X | X | X | X | | 27 + 28 | 0.458 | Combines advantages of Random Parameter and Choice and Latent Variablel models. |