NISS

Accounting for Uncertain Travel Time: A Logit Model Assuming a Weighted Utility Function

Pia M. Koskenoja

Technical Report Number 71
October, 1997

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Accounting for Uncertain Travel Time:
a Logit Model Assuming a Weighted Utility Function

Pia M. Koskenoja
National Institute of Statistical Sciences
and
Institute of Statistics and Decision Sciences
at Duke University
(919) 685-9329
piak@niss.org

October 29, 1997

ABSTRACT

Weighted utility logit model assumes a preference structure that is compatible with weighted utility, a class of generalizations of expected utility. This weighted utility logit model distinguishes between two types of uncertainty: uncertainty due to modeler's imperfect information and measurement errors, which are accounted by the traditional error term of the logit model, and the inherent uncertainty in the choice situation, when the decision maker knows only the distribution of travel times.

The traditional mean value utility model, where travel time is presented as a sure attribute of an alternative, is a special case of both expected utility and weighted utility model, with a risk parameter $\alpha=0$. It is shown that one cannot derive the general form of weighted utility model from the general form of expected utility.

The weighted utility logit model is estimated for a continuous, normally distributed uncertain variable, and an uncertain variable with a discrete distribution. The parameter restrictions for risk aversion and first order dominance are specified for the uncertain variable with a discrete distribution. Both weighted utility logit models and the traditional mean value utility logit model are evaluated in a Monte Carlo simulation. The weighted utility models accurately identify correct parameter values in a wide range of plausible risk parameter (α) values, including the neighborhood of zero. If a true weighted utility model is misspecified and estimated as a traditional mean value model, the results are biased.

ACKNOWLEDGMENTS

Most of this work was conducted at the University of California at Irvine, Department of Economics. I am grateful to David Brownstone, Soo Hong Chew, Duncan Luce, Kenneth Small, and Scott Richman for their valuable comments. The research was partially supported by the Yrjö Jahnsson Foundation and by National Science Foundation grants NSF-DMS 9313013 and NSF-DMS 9208758.

Introduction

Uncertainty affects choice. This fact has been acknowledged by transportation researchers[1], but the explicit incorporation of that fact in demand models has been incidental and ad hoc. In this paper I suggest applying weighted utility theory to incorporate the uncertainty of travel time explicitly in a logit choice model.

I start by reviewing earlier approaches to unreliable travel time in transportation demand models. After discussing some of the shortcomings of those approaches I review the weighted utility theory and the added flexibility it brings to behavioral assumptions. I operationalize two weighted utility models and derive the implications of monotonicity and risk aversion on parameter values. In the end I test the weighted utility models and a benchmark linear utility model in a Monte Carlo simulation with a range of plausible parameter values for the weight parameter.

Earlier approaches to unreliable travel time

Travel time reliability is often characterized by the variance of travel time distribution. However, due to the expense of data collection or other reasons, the practice has been to measure the travel time only a few times and estimate the mean travel time from these measurements. The estimate ('engineering time') has then been used as if it were a sure attribute

[1] A British research team found that value of reliability in public transportation, measured as keeping to the scheduled services, was consistently high. They also conducted a survey on private transportation, which indicated 40% higher values for time savings when driving in congested driving conditions. Congested conditions was taken as a proxy for unreliable arrival time (The MVA consultancy et al., 1987).

of the mode in the choice model. Another practice has been to ask the traveler how long the trip took and use this estimate as a sure attribute of the used transportation mode. It has also been a standard treatment to use a utility function which is linear in parameters, if not in characteristics.

It is possible to justify these practices by assuming that the linear function is a local approximation on a non-linear relationship and that as long as the model is used to predict outcomes of only minor adjustments of independent variables, this approximation would give reasonably accurate estimates. This line of thinking may be invalid. McCord and Villoria (1986) test linearity of utility functions of 12 individuals from Stated Preference data. They present the subjects' comparisons of unspecified travel modes with only two attributes: travel time and monetary costs. The stated preferences are used to form an individual three dimensional utility surface for each participant. None of these individuals can be classified as exhibiting a linear utility function even in small changes of time and cost and six are classified as exhibiting systematic deviations from linearity (p. 24). Moreover, across-subject consistency of the form of utility function is absent: all the twelve utility surfaces exhibit monotonicity, but not a clear pattern of convexity or concavity.

If one assumes that the true utility function is linear in characteristics, then focusing only on the mean of characteristics is appropriate. But for a risk averse decision maker a certainty equivalent of an uncertain alternative is always smaller than the expected outcome of that alternative, making the true utility function concave. If the true utility function is concave, treating the mean of a random variable as a sure variable represents a too high value for that characteristic and will lead to underestimation of the parameter assigned to that characteristic. This deficiency can be overcome by transforming the original variables and making the utility function linear in these transformed parameters. Gaudry and Wills (1978) estimate parameters of Box-Tukey and Box-Cox transformations jointly with parameters of explanatory variables. Their results give more plausible signs and elasticities for the parameters of explanatory variables, specially when contrasted with the standard additive logit model.

Winston (1981) approaches the problem of uncertain freight transit times by taking a Taylor-series expansion of an unspecified general utility function. The expansion is taken around the means of the uncertain attributes. This becomes an expected utility expression when he takes the expectation of it. The first terms of the series is the utility of the mean value of the attribute as if it were certain, and the second term disappears. The third term is second derivative of the utility function divided by two and multiplied by the variance of the uncertain attribute. He ignores the higher order terms. The expected utility expression becomes

$$EU[z] = b_0 + b_{1_1}\mu_1 + b_{1_2}\sigma_1^2 + \ldots + b_{z_1}\mu_z + b_{z_2}\sigma_z^2 + \ldots + b_{Z_1}\mu_Z + b_{Z_2}\sigma_Z^2$$

where z = 1,...,Z denotes the uncertain attributes.

Winston uses a random parameters specification, where each shipment receiver is assigned individual parameter value representing his "unobserved deviations from mean attitude toward risk or mean tastes" (p.987) for each random attribute. He uses only the means of random attributes, and aggregates the higher order terms together with the unobserved utility term into a new error term. He then divides the new error term into a random-parameter component and an iid error term so that the coefficient for the utility of the mean value of a attribute contains the effect of a mean attitude towards risk in that attribute.

In the empirical part Winston estimates transit mode choices for 12 commodity groups. The parameters for freight charges, mean and standard deviation of transit time, and reliability were treated as random. Reliability is defined as the ratio of standard deviation of transit time to the mean of transit time. This leads to a specification where he uses two reliability measures: the standard deviation and the ratio of the standard deviation to the mean. Using two measures of one effect in a regression is likely to result in estimated parameters which depend on each other and are unreliable.

The results suggest multicollinearity: in all but one industry either standard deviation of transit time or the ratio of standard deviation to mean has a positive sign. These results indicate that shippers prefer randomness in transit times. Of the 12 commodity groups 9 have opposite signs in standard deviation and the ratio, and the remaining three have statistically insignificant parameters and in addition two of those have positive signs for both unreliability measures.

Senna (1991) uses an exponential utility function. He uses an exponential utility function

$$U=\alpha t^{\beta}$$

and manipulates it into an expected utility expression of travel time t:

$$E(U) = \alpha E(t^{\beta}) = \alpha([E(t^{\frac{\beta}{2}})]^2 + [\sigma(t^{\frac{\beta}{2}})]^2) .$$

Senna uses Stated Preference data to estimate an additive utility function with the derived expression for uncertain travel time and a certain monetary cost. His estimation results in β=0.7, which implies risk prone preferences (as he indicates any β<1 value would). He also notes that the R^2 is only 0.092. He suspects that the results are poor due to subjects's inability to trade off time and money. Another experiment where the subjects trade off unreliability of travel time to mean travel time gives β=0.2. Since this result is counterintuitive, Senna proposes further study in the area.

Senna (1994) uses the same model to estimate separate models for commuters and non-commuters, with fixed and flexible arrival times. He reports that commuters with fixed arrival times have β=0.5, but all non-commuters and commuters with flexible arrival times have β=1.4. He apparently uses a sequential estimation where the time parameter (β) is estimated first and then inserted in a model trading off the time measure against monetary cost. These models still have a moderately low R^2s , 0.059-0.148.

Another way to approximate the expected non-linear utility is to take a polynomial approximation of an unspecified function. The approximation is:

$$U(t) = b_0 - b_1 t - b_2 t^2 - R$$

where R denotes the residual higher order terms. Ignoring the higher order terms this transforms to expected utility:

$$EU[t] = b_0 - b_1 E[t] - b_2 E[t^2].$$

Since

$$E[t] = \mu_t, \qquad E[t^2] = \sigma_t^2 - \mu_t^2$$

the utility expression can be written

$$EU[t] = b_0 - b_1\mu_t - b_2(\sigma_t^2 - \mu_t^2).$$

These expected utility expressions have unpleasant characteristics. In Senna's form both parameters (α,β) affect the squared mean and variance proportionally. This is unwelcome, because one can easily come up with example situations where the importance of mean and variance do not vary proportionally. Actually, it would be an interesting empirical result, but it is better to estimate it than assume it. Proportionality also requires that the mean and variance have to have the same sign. This is acceptable in transportation, where both the travel time and its variance are seen as costs (or negative outcomes), but the utility expression is not flexible enough for positive outcomes. The interpretation of estimated parameters is unclear.

The polynomial approximation has the fault that all approximations have: one cannot be sure about the magnitude of the omitted higher order terms. The parameters are more separated, but not fully: both b_1 and b_2 determine the response to changes in the mean, but b_2 doubles as a risk attitude parameter. This creates ambiguity about the interpretation of the estimated parameters. In the following chapters I present an application of weighted utility that has totally separated parameters.

Weighted Utility

Another way to approach unreliability is to use a more general form of utility function than expected utility. One generalization of expected utility is weighted utility, which will be used in this paper. Decision theory literature cites the Allais paradox as an example of behavior which violates expected utility: a decision maker prefers an alternative T where he gets a smaller but sure prize to an alternative A where he would either get a larger prize, same prize as in T, or - with a small probability - will get nothing. However, if both prizes are made less likely by combining the alternatives T and A with equal positive probability to get nothing, the decision makers quite often prefer the mixture of A and the worst outcome. This behavior violates the independence axiom of expected utility.

One analogy in a transportation context is travel time in mode choice: A decision maker considers a reliable but slow mode T (eg. train) to be certain to take him to his destination in 20 minutes. He compares this to the alternative mode A (eg. auto), which takes normally 15 minutes, but which has a positive probability of a traffic jam to delay the arrival time to either 20 minutes or 50 minutes. When both car and train alternatives are worsened to a probability mixture of the worst outcome (50 minutes) and the original alternatives, expected utility requires the decision maker to prefer the mixture with the previously preferred alternative. Weighted utility theory allows him to prefer either one and behave as if he would scale the probabilities differently in these two situations.

The situation can be represented in a convex hull of three probability measures p, q, and r.

$$H(\{p,q,r\}) = \{\lambda_1 p + \lambda_2 q + \lambda_3 r : \lambda_i \geq 0, \ \ \Sigma\lambda_i = 1\}$$

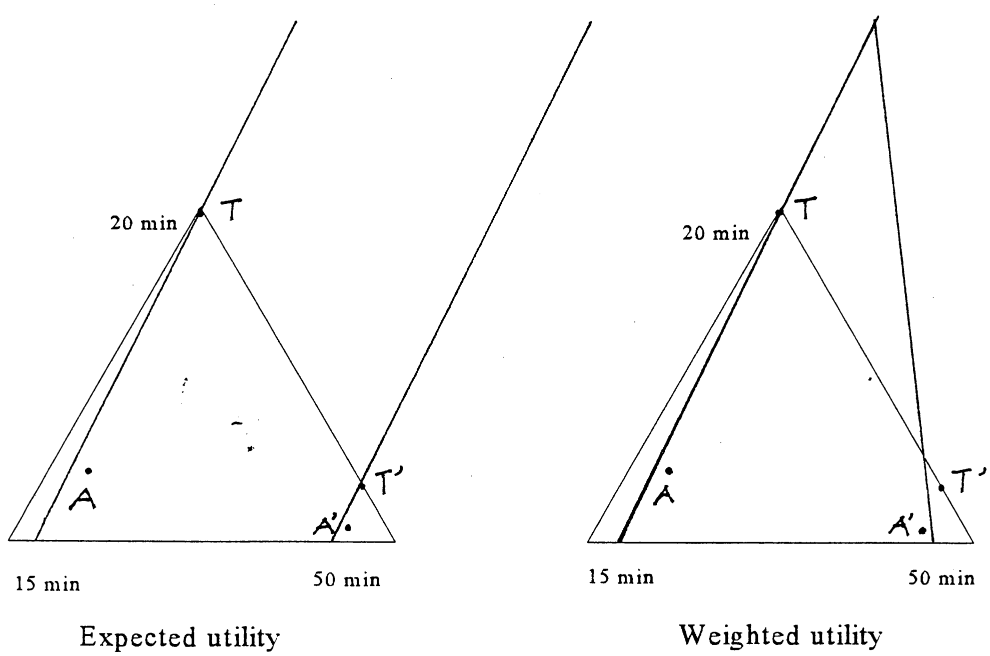
When it is illustrated barycentrically, each point in the hull corresponds to a point in a equilateral triangle with the probability measures as vertices. In this example each of the three measures is a degenerate distribution with only one sure outcome: 15 minutes, 20 minutes and 50 minutes. When the perpendicular distance from each side to its opposite vertex is 1, λ_1p+λ_2q+λ_3r is the point with perpendicular distances λ_1, λ_2, and λ_3 from sides qr, pr, and pq, respectively.

When the indifference curves are drawn in this space, weighted utility requires the indifference curves to be linear and intersect in a common point, outside the triangle. Expected utility is a special case of weighted utility where the probabilities are 'weighted' by a constant for all outcomes, and the common intersection point is at infinity, which makes the indifference lines parallel. This is demonstrated in Figure 1.

The train alternative is represented by point T, where the travel time is 20 minutes for sure. The auto alternative is represented by point A, which is close to outcome 15 minutes, but shows some probability for outcomes 20 minutes and 50 minutes. When the decision maker compares the auto and train alternatives, she prefers the train, which is shown in both expected utility and weighted utility cases by the indifference curves intersecting point T, but above point A. When both of these alternatives are worsened at same rate to a probability mixture A'= αA + (1-α)(50 minutes) and T'= αT + (1-α)(50 minutes), the decision maker behaving according to expected utility has to prefer T' to A', while the decision maker behaving according to weighted utility can prefer A' to T'.

Weighted utility behavior cannot be derived from expected utility by transforming the independent variables. The demonstration in 3-outcome space is presented in Appendix A. Chew (1982) proved the more general case.

Figure 1. Allais paradox presented in the context of travel time for Expected Utility and Weighted Utility



Axiomatically weighted utility deviates from expected utility by loosening the independence axiom. Von Neumann and Morgenstern axiomatized expected utility through preference order $\succ$ on P, where P is a nonempty set of probability measures p,q,...(which in the previous example would correspond to the travel time probability distributions of train and auto). The set of probability measures P is defined on a Boolean algebra $\mathfrak{C}$ of subsets of X, the set of outcomes. This means that for each $p \in P$, $p(A) \geq 0$ for every $A \in \mathfrak{C}$, $p(A \cup B) = p(A) + p(B)$ whenever A and B are disjoint events in $\mathfrak{C}$, and p=1 on the universal event X in $\mathfrak{C}$, and that $\mathfrak{C}$ is defined to be closed under complementation and finite unions. P is assumed convex, which guarantees that convex combinations of probability measures p and q are in P. Under these assumptions expected utility is axiomatized:

A1. Order: $\succ$ on P is asymmetric and negatively transitive.

A2. Independence: $p \succ q \Rightarrow \lambda p + (1-\lambda)r \succ \lambda q + (1-\lambda)r$, for $0<\lambda<1$.

A3. Continuity: $\{p \succ q, q \succ r\} \Rightarrow (\alpha p + (1-\alpha)r \succ q$ and $q \succ \beta p + (1-\beta)r$ for some α and β in (0,1).

Weighted utility preserves axioms A1 and A3, but replaces the Independence axiom by two others: a weak independence axiom and a convexity axiom.

B1. Weak Independence: $p \sim q \Rightarrow$ for every $0 < \alpha < 1$ there is a $0 < \beta < 1$ such that for every $r \in P$, $\alpha p + (1-\alpha)r \sim \beta q + (1-\beta)r$.

B2. Convexity: For $0 < \lambda < 1$,

$\{p \succ q, p \succeq r\} \Rightarrow p \succ \lambda q + (1-\lambda)r$,

$\{q \succ p, r \succeq p\} \Rightarrow \lambda q + (1-\lambda)r \succ p$.

Weighted utility was first axiomatized by Chew and MacCrimmon in 1979, and further axiomatic work has been continued by Chew (1982, 1983), Fishburn (1981, 1983) and Nakamura (1984, 1985). The axioms of weighted utility have been tested in empirical laboratory experiments by Chew and Waller (1986), Camerer (1989) and Conlisk (1989).[2] However, I do not know of any application of weighted utility in transportation economics.

The Mode Choice Model

I present a model which maintains the additivity of the systematic utility in attributes, but introduces weighted utility to one attribute. In particular, the utility function contains a weighted utility functional of one attribute, the travel time, while it is linear in the other, cost of the trip. The approach can be generalized to more than one attribute[3].

The model is a binomial logit model which describes a mode, route, or any choice situation, where the decision maker has to decide between two mutually exclusive alternatives. In the simplest form both of the alternatives are described by only two characteristics: travel time and monetary cost of the trip. Travel time is uncertain: the decision maker has some previous experience of the same trip or he knows the travel time distribution parameters. The monetary cost is known for sure. The model contains two types of randomness: randomness due to modeler's imperfect information about the attributes affecting choice and measurement

---

[2]See Fishburn (1988) for more thorough discussion of the weighted utility.

[3]Before generalizing the model the separability of different attributes has to be solved by assuming a more general functional form. This complicates the derivation of the derivatives of the likelihood function.

error, which are captured by the traditional error term of the logit model, and the randomness inherent in the choice alternatives, modeled as the distribution of travel times. This specification is one step towards a more realistic description of a choice situation and enables a less limited estimation of the decision maker's attitude towards risk. In situations where the decision maker has some previous experience, this approach has potential use as a learning model where the traveler updates his set of observations as his experience of the mode accumulates. It provides a way to model public's response path to changes in traffic information or implementation of transportation policies.

Discrete Distribution of Travel Times

Weighted utility function for travel time can be written

$$V(\{p(t_i),\ \ i = 1,\dots, I\}) = \sum_i \frac{p(t_i)\,w(t_i)\,U(t_i)}{\sum_m p(t_m)\,w(t_m)}$$

Where

$p(t_i)$ = the probability that the trip will last t_i minutes

$w(t_i)$ = the weight the decision maker places on t_i

$u(t_i)$ = the utility of a sure t_i.

The weighted utility function reduces to expected utility expression when the weight function is constant. One can follow the tradition of logit models and use an additive utility function, which is separable in attributes. For uncertain travel time t and certain monetary cost c the utility function is:

$$V(\cdot) = b_0 - b_1 * \sum_i \frac{p(t_i)\,w(t_i)\,U(t_i)}{\sum_m p(t_m)\,w(t_m)} - b_2 * c$$

In order to derive an empirical model, one has to make an additional assumption about the form of the weight function. The weighted utility theory requires that the weight function has to be always positive to insure that the transformed probabilities are positive[4]. To simplify the estimation, a one parameter function and range of the whole positive real line is preferred. Therefore I chose an exponential: $w(t_i) = \exp(\alpha t_i)$. If $\alpha=0$, the function gets a value one throughout the domain. This would reduce the weighted utility expression to expected utility. If $\alpha>0$, the traveler emphasizes the potential of longer travel times. And if $\alpha<0$, he behaves as if he would consider the shorter travel times as "more likely" than what their probability is.

Continuous Distribution of Travel Times

The form I have discussed up till now applies to the case of discrete outcomes. It applies to a situation when there is a finite number of observations of the travel time and the travel time distribution is not known. If the distribution is known or assumed known, through a large

[4] Fishburn (1988) p. 62

amount of measurements or on theoretical grounds, the discrete form can be replaced with a continuous form: an integral over the domain of the density function f(t) of travel times.

$$V[f(t)] = \frac{\int f(t)\,w(t)\,U(t)\,d(t)}{\int f(t)\,w(t)\,d(t)}$$

If one assumes that the travel time has a normal distribution and that the weight function is exponential, the utility functional reduces to a linear function of mean and variance of the travel time[5]. This is remarkable because it justifies the inclusion of travel time variance as a fully separate explanatory variable. It also shows the special assumptions one has to make in order to justify the previously ad hoc variables.

Assumptions:

$$t \sim N(\mu, \sigma^2)$$

$$U(t) = -b_1 * t$$

$$w(t) = \exp(\alpha * t)$$

The utility functional in the continuous case can be written

[5] I am grateful to Scott Richman for helping me with this derivation.

$$V(\cdot) = b_0 - b_1 \frac{\frac{1}{\sqrt{2\pi\sigma}} \int t * \exp(\alpha t) * \exp(\frac{-(t - \mu)^2}{2\sigma^2}) dt}{\frac{1}{\sqrt{2\pi\sigma}} \int \exp(\alpha t) * \exp(\frac{-(t - \mu)^2}{2\sigma^2}) dt} - b_2 * c$$

The common term of the two integrals can be manipulated as follows

$$\exp(\alpha t) * \exp(-\frac{(t - \mu)^2}{2\sigma^2})$$

$$= \exp(\frac{-t^2 + 2\mu t + 2\alpha\sigma^2 t - \mu^2}{2\sigma^2})$$

$$= \exp(-\frac{(t - (\mu + \alpha\sigma^2))^2}{2\sigma^2}) * \exp(\frac{(\mu + \alpha\sigma^2)^2 - \mu^2}{2\sigma^2})$$

The second term does not depend on t and can be placed in front of the integral. Because the second term appears in both the numerator and denominator, it cancels out and the utility functional becomes

$$V(\cdot) = b_0 - b_1 \frac{\int t * \exp(-\frac{(t - (\mu + \alpha\sigma^2))^2}{2\sigma^2}) dt}{\int \exp(-\frac{(t - (\mu + \alpha\sigma^2))^2}{2\sigma^2}) dt} - b_2 * c$$

which further reduces to

$$V(\cdot) = b_0 - b_1 (\mu + \alpha\sigma^2) - b_2 * c$$

Note that the only case when the person maximizes regular expected utility in this formulation is with risk neutrality when α=0, and the expression reduces to the traditional additive utility function without the variance term. This applies only when one uses the additive utility function, the exponential weight function, and only one variable has a non-degenerate normal distribution. When several variables have probability distributions the joint density function will have cross terms of different standard deviations which may not vanish. For practical applications the assumption of normally distributed travel times is rather restrictive: usually the travel times follow a skewed distribution much like a log-normal distribution.

The indifference curves of expected utility and weighted utility look very different in mean - variance coordinates. Consider the earlier discussed **expected utility** model with risk aversion from a polynomial approximation:

$$EU[t] = b_0 - b_1\mu_t - b_2(\sigma_t^2 - \mu_t^2).$$

Ignore the constant term (and the monetary cost). The expression can be made comparable to weighted utility by rewriting it:

$$K = b_0 - b_1\mu - \alpha b_1(\sigma^2 - \mu^2)$$

where K represents an arbitrary level of utility. To derive the slope of an indifference curve, solve first for variance:

$$\sigma^2 = -(\frac{K + b_1\mu}{\alpha b_1}) + \mu^2$$

and for the inverse of the slope:

$$\frac{\partial\sigma^2}{\partial\mu} = -\frac{1}{\alpha} + 2\mu$$

This sets the "minimum" of the expected utility indifference curve at $\mu=1/2\alpha$. Note that it is not dependent on b_1 or the level of utility K.

Now consider the **weighted utility**.

$$K = -b_1\mu - b_1\alpha\sigma^2$$

The variance equals

$$\sigma^2 = -\frac{K + b_1\mu}{b_1\alpha}$$

And the inverse of the slope of the indifference curve is

$$\frac{\partial\sigma^2}{\partial\mu} = -\frac{1}{\alpha}$$

Weighted utility has straight indifference curves in mean - variance space. When $\mu=0$, the indifference curves have the same slope, but the expected utility indifference curves start to curve to a parabola when μ differs from zero. The indifference curves of both expected and weighted utility are depicted in Figures 2-4 for α-values 0.03, 0.06, and 0.1 respectively. The pictures are drawn to compare utility levels which coincide at $\sigma^2=0$, and cover the space practical applications are concerned with. A curve for $\mu=3\sigma^2$ limits the area to distributions that are not likely to have negative travel times.

Comparing the three pictures one can see differences between expected and weighted utility: when α is close to zero, the expected utility indifference curves are straighter and both expected utility and weighted utility curves become more vertical and the angle between the curves is small. As the α gets larger, the discrepancy between the indifference curves increases. Also, when the mean of travel time grows, the slopes of expected utility indifference curves change whereas the straight slope of weighted utility indifference curves remains constant.

When the models are estimated, the parameters α, b_0 and b_1 will not have the same value for expected and weighted utility. The μ^2 term has a direct effect on the parameter for variance and consequently shifts also the other parameters. Graphically this means that the tangency point between the indifference curves shifts from $\mu=0$. This has to be kept in mind when the parameters of empirical models are interpreted.

Figure 2.

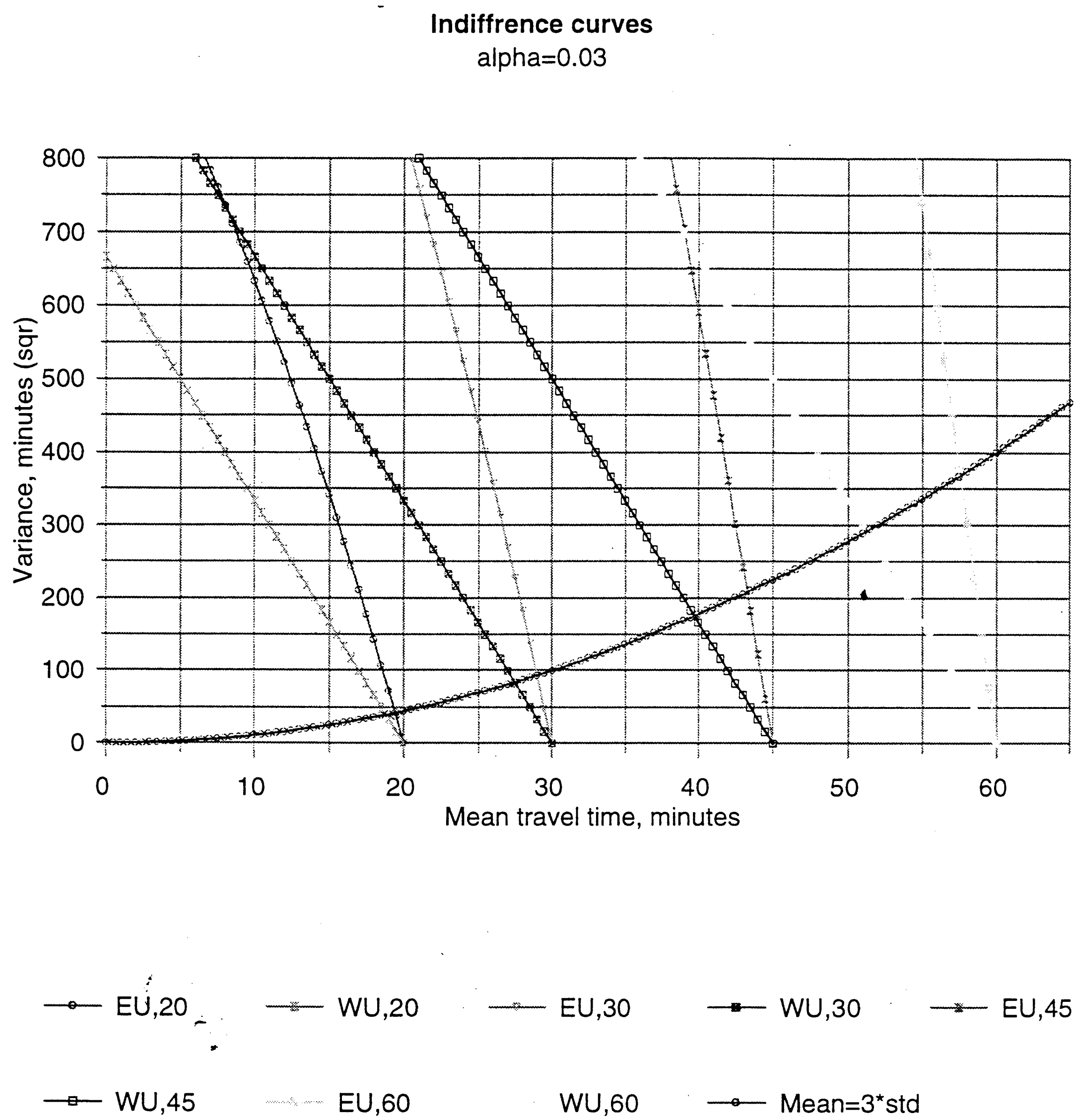

Indiffrence curves
alpha=0.03
Variance, minutes (sqr)
800
700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
Mean travel time, minutes
EU,20
WU,20
EU,30
WU,30
EU,45
WU,45
EU,60
WU,60
Mean=3*std

Figure 3.

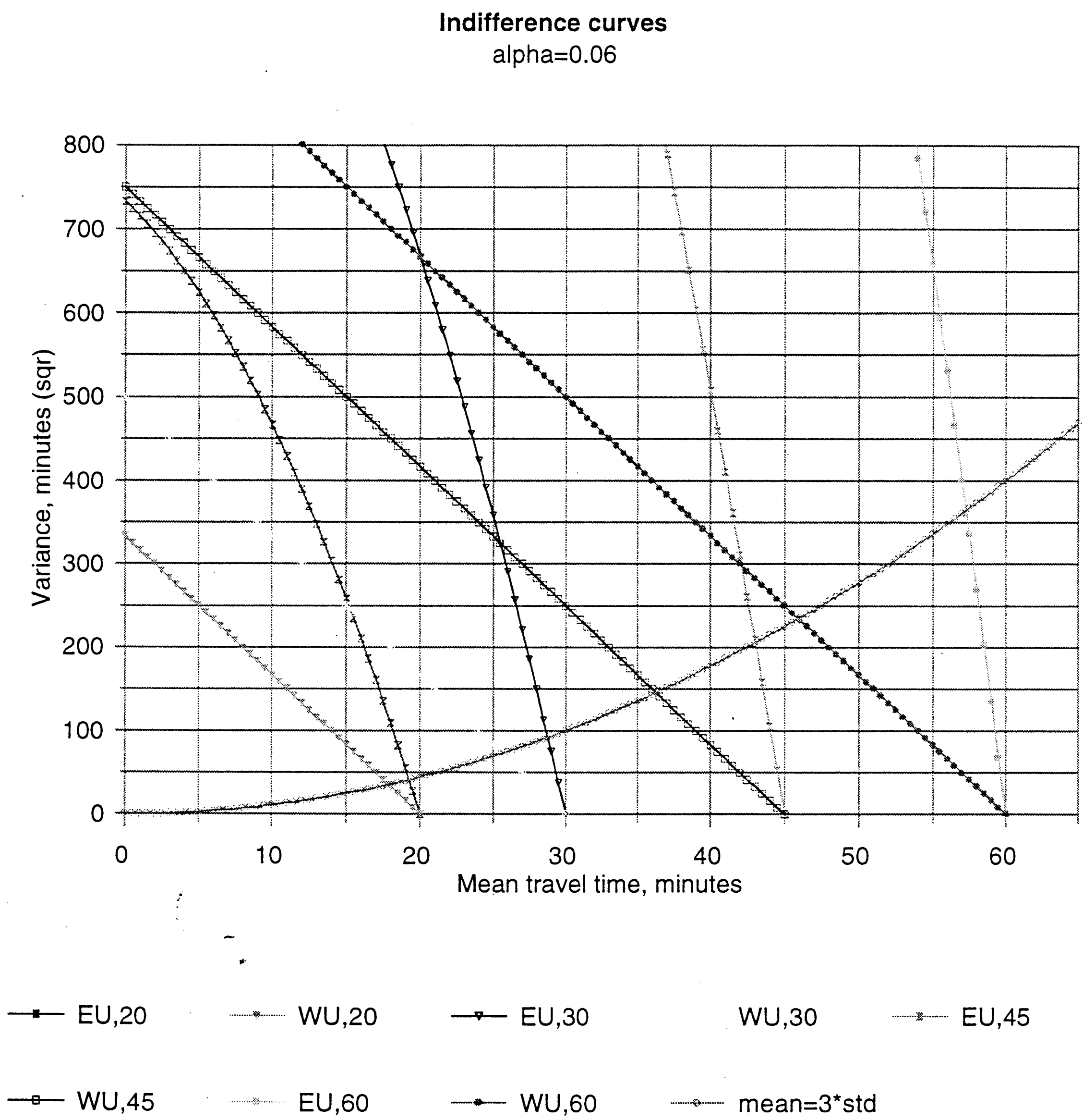
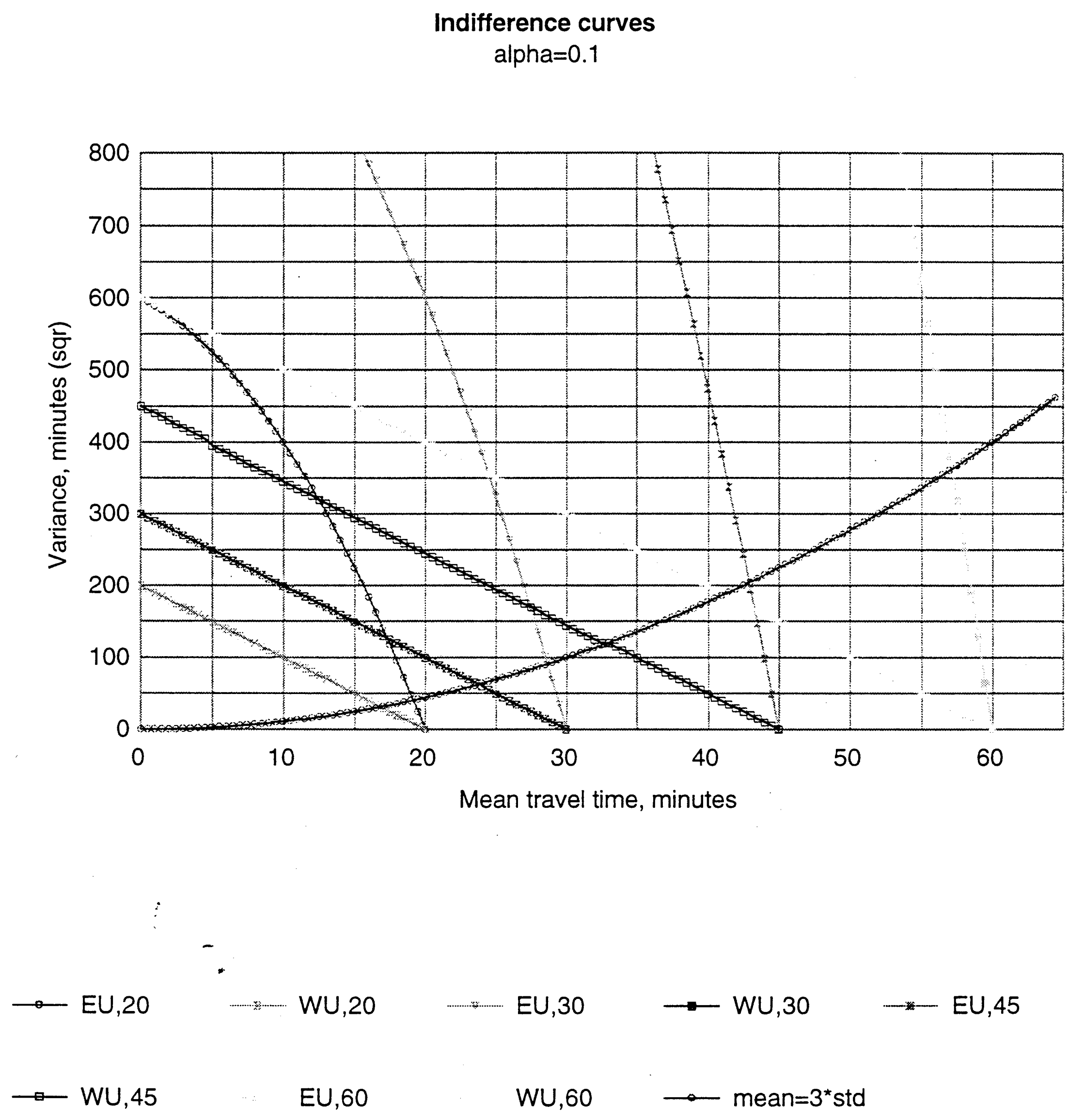

Indifference curves
alpha=0.06
800
700
600
500
400
300
200
100
0
Variance, minutes (sqr)
0
10
20
30
40
50
60
Mean travel time, minutes
EU,20
WU,20
EU,30
WU,30
EU,45
WU,45
EU,60
WU,60
mean=3*std

Figure 4.


Indifference curves
alpha=0.1
800
700
600
500
400
300
200
100
0
Variance, minutes (sqr)
0
10
20
30
40
50
60
Mean travel time, minutes
EU,20
WU,20
EU,30
WU,30
EU,45
WU,45
EU,60
WU,60
mean=3*std

Parameter restrictions implied by risk aversion and monotonicity

It is customary to require that a utility function exhibits risk aversion and monotonicity. I will discuss the implications of these requirements on the parameter values for the discontinuous model.

In the model the utility of time outcomes is $U(t) = - b_1 t$ and the weight function is $w(t) = \exp(\alpha t)$. Risk aversion is defined to mean that the utility of the expected outcome is preferred to the utility of the gamble[6]:

$$V\left(\frac{t_1 + t_2}{2}\right) > \frac{w(t_1)p_1 t_1 + w(t_2)(1-p_1)t_2}{w(t_1)p_1 + w(t_2)(1-p_1)}.$$

The left side simplifies to:

$$V\left(\frac{t_1 + t_2}{2}\right) = \frac{w\left(\frac{t_1 + t_2)}{2}\right) * (-b_1) * \left(\frac{t_1 + t_2}{2}\right)}{\frac{w(t_1 + t_2)}{2}} = - b_1 * \left(\frac{t_1 + t_2}{2}\right).$$

And the inequality can be written out:

$$- b_1 * \left(\frac{t_1 + t_2}{2}\right) [w(t_1)p_1 + w(t_2)(1 - p_1)]$$

$$> [w(t_1)p_1 t_1 + w(t_2)(1 - p_1)t_2] * (- b_1).$$

[6]This is a simplification of the definition that a risk averse person prefers less variation to more, other things held constant.

If $t_1 < t_2$, this leads to a risk aversion requirement $w(t_1) < w(t_2)$, which sets the parameter requirement $\alpha>0$.

Monotonicity of utility function in outcomes[7] generalizes into requirement that the utility functional exhibits first order stochastic dominance (FSD). FSD states that if an alternative A gives a better outcome in each state of the world than alternative B, A is preferred to B. An other way to state FSD is to say that if outcomes are ordered from worst to best, p FSD q if the cumulative distribution function of p is below the cdf of q throughout the outcome space. This means that p has higher probabilities for better outcomes than q. FSD is analogous to positive marginal utility or negative marginal cost. In this model the first order dominance is written

$$\frac{\partial \sum_i p_i w(t_i)\, U(t_i) / \sum_k w(t_k) p_k}{\partial t_i} < 0,$$

$$\forall \sum_i p_i U(t_i) \in rng\ V[\{p(t_i),\quad i = 1,\ldots I\}]$$

Using the quotient rule of derivatives this condition can be written out

$$\Rightarrow \frac{1}{\sum_k w(t_k) p_k} * [\frac{\partial}{\partial t_i}(p_i w(t_i)\, U(t_i) - V[\{p(t_i)\}]\frac{\partial}{\partial t_i}(\sum_k p_k w(t_k))] < 0$$

The summation term is always positive. Substituting $U(t_i) = - b_1 t_i$ and $w(t_i) = \exp(\alpha t_i)$, the expression reduces to

[7]Monotonicity in regular utility function requires that more of a good thing is preferred to less of a good thing. Thus the utility function has to be either constant or increasing in all parts of its domain. It cannot have downward 'bumps'.

$$\Rightarrow [p_i(\exp(\alpha t_i) * \alpha * (-b_1 t_i) + (-b_1) * \exp(\alpha t_i)) - V[\{p(t_i)\}] * p_i * \exp(\alpha t_i) * \alpha] < 0$$

$$\Rightarrow p_i * \exp(\alpha t_i)[-\alpha b_1 t_i - b_1 - V[\{p(t_i)\}]\alpha] < 0$$

Which further reduces to

$$\alpha(-b_1 t_i - V[\{p(t_i)\}]) < b_1$$

The range for outcomes t_i can be set to arbitrary [l,L] and thus the range for $V[\{p(t_i)\}]$ is $[-b_1L, -b_1l]$. This leads to four corner conditions:

1) If $V[\{p(t_i)\}] = -b_1L$ and $t_i = l$, $\alpha(-b_1l + b_1L) < b_1$ or

$$\alpha < \frac{1}{L-l}$$

2) If $V[\{p(t_i)\}] = -b_1L$ and $t_i = L$, $\alpha(-L + L) < 1$, which is always true.

3) If $V[\{p(t_i)\}] = -b_1l$ and $t_i = l$, $\alpha(-l + l) < 1$, which is always true.

4) If $V[\{p(t_i)\}] = -b_1l$ and $t_i = L$, $\alpha(-L + l) < 1$, which is true if $\alpha > 0$, i.e. if the functional exhibits risk aversion.

The first condition is the most likely to be violated. If the travel time has a normal distribution with an infinite range of outcomes, the decision maker violates monotonicity if she is risk averse (α is a positive number). Of course the assumption of normal distribution is restricted by the fact that travel times are always positive. Furthermore, with actual data one can always find the boundary observations. On the other hand, it is also plausible that people are not monotonic in the whole range of outcomes.

Simulations

I studied three model specifications. All three models were binomial logit mode choice models, where the attributes affecting the choice were travel time, monetary cost, and a mode specific constant for one of the modes. The monetary cost was assumed known in advance for the decision maker, but the travel time was uncertain. The three specifications differed by their treatment of travel time.

In the benchmark model the choice was modeled to depend on the mean value of travel times. This reflects the common practice of first estimating the travel time and then using the estimated travel time in an additive utility function as a sure and known characteristic of the alternative. The second specification assumed that the decision maker knows that the travel times are distributed normally and that he also knows the mean and variance of that distribution. The third specification used ten discrete observations from that normal distribution, selected so that there was 10% of the probability mass between each observation. The decision maker has knowledge or experience of these 10 observations, and he treats the distribution as if these observations would form the whole discrete travel time distribution. Note that as the number of discrete observations increases, the assumption of decision maker's travel time information approaches the information the decision maker has in the second model with continuously distributed time.

Each model used the same travel time and cost data in five estimations, where the sample sizes were 100, 200, 500, 1000, and 4000. The monetary cost was drawn from a uniform distribution from the interval [4,5]. The travel time means were drawn from a uniform distribution from the interval [18, 20], and the standard deviations of travel times were drawn from a uniform distribution from the interval [0,2]. The choices were created according to the postulated decision rules and parameter values $\alpha=0.15$, $b_0=0.5$, $b_1=1.5$, and $b_2=2$. The data creation procedure and parameter values and are arbitrary. They were chosen to create a situation where the predicted probabilities would not be 0, 1 or half. Only the value of α was set intentionally low. This was chosen to reflect small deviations from expected utility, which is the most plausible case for weighted utility. It is also the hardest to identify in the case where the true data is produced by weighted utility behavior. The critical value for FSD violation in the discrete time observations weighted utility model is $\alpha<=0.1$ (using the fact that the observations are within two standard deviations from the mean). Thus the true behavior is assumed to violate FSD.

Given the choices, the models were estimated to retrieve the parameters. The estimation results are in tables 1-3. I have indicated with two stars the estimates within one standard deviation from the true value, with one star the estimates within two standard deviations, and with a minus sign those that cannot be distinguished from zero on statistical grounds. These results are only indicative, because the simulations have to be repeated several times in order to get meaningful statistics for the parameters.

I experimented with different initial values and discovered that the likelihood surface is not smoothly curved. With a starting point far from the true values the procedure does not converge, as one or more partial derivatives gets a value of infinity. However, in the cases I studied further, the value of likelihood-function was worse in this region than at the true parameter values.

A few observations can be made. The continuous time weighted utility (CWU) and discrete time observations weighted utility (DWU) are asymptotically same. Therefore, when the true model is either one of the weighted utility models, one would assume that the estimated parameters should be close to each other. This is true for the parameters b_0, b_1, and b_2, in the two upper sections of tables 1 and 2. But the weight parameter seems to be sensitive to misspecification: misspecifying the model to be CWU when it is DWU produces downward biased estimates of α, whereas misspecifying the model to be DWU when it is CWU biases the weight parameter upwards. Randomness of the data can still be a larger factor in determining the bias than the misspecification, offsetting the result back to the true value. In the four cases studied this seems particularly true in the case of misspecified CWU estimation.

The lowest sections of tables 1 and 2 display the estimation results when the model is estimated as mean value utility, the common practice in planning models. The bias does not have any particular pattern in smaller sample sizes, but with sample sizes 1000 and 4000 all the estimated parameters are downward biased when compared to the correct model specification. In these large samples the law of large numbers overcomes the effect of random data and the small systematic effect of misspecification starts to show. When the weight parameter is in effect restricted to zero, this restriction forces the other parameters to adjust downwards[8].

[8]See Appendix B for fuller discussion of bias mechanism.

In the top and middle parts of Table 3 the choice is based on mean value, but the estimated models assume weighted utility. Both WU models consistently estimate the risk attitude parameter to be statistically zero and replicate the other parameters from the correct model specification.

In order to get statistics about the estimated parameters I did a Monte Carlo simulation. I created the random data set and ran the models 50 times, starting at point [1,1,1,1]. The size of each data set was 1000, since that seemed to be the smallest data set to give plausible estimates. The data creating parameters and parameters of the choice model were unchanged, except for α. The true value of α was changed from 0.15 to 0.5. This higher value of α creates clear differences between the models, but it is less plausible as it assumes very risk averse behavior to the extent of grossly violating first order dominance.

The usual number of iterations was around 20. If the model did not converge soon after that, it was highly likely that one of the derivatives had reached a value of infinity. If that happens, the process goes to an infinite loop and terminates only when the maximum number of iterations is reached. I would interrupt the procedure at 50 iterations. Tables 4-6 contain data from the 50 estimations of each model.

The simulation shows that the discontinuous time weighted utility is hardest to estimate reliably when α deviates significantly from zero and the behavior doesn't exhibit first order dominance.

In table 4 the parameters are nicely retrieved in the case of continuous time weighted utility, and the rather small standard deviations show that the results are statistically significant. However, when the model is misspecified as discontinuous time weighted utility, the risk parameter α is estimated to be totally out of range. The other parameters don't seem to suffer from the misspecification very much. The mean value utility estimates show a clear downward bias. This results from the fact that the risk parameter α was in effect restricted to equal zero and the other parameters scale down to maintain their relative magnitudes.

Table 5 repeats the unreliable results for the risk parameter α. The other variables maintain their accuracy, and the mean value model exhibits the downward bias again. In table 6, where the true model is mean value, and α therefore truly equals zero, all models give accurate parameter values. This indicates that the use of weighted utility models will not distort the mean value utility model estimations when the true model is mean value utility.

Next the models were tested in the original, more likely situation, where α=0.15 and only slightly violates FSD. Tables 7-9 show the Monte Carlo statistics. Discrete time weighted utility model has still two convergence failures. All estimates are within one standard deviation of the true parameter values, except when a continuous weighted utility model is misspecified as discontinuous weighted utility. A smaller true α is accurately estimated when the model is correctly specified as continuous or discrete weighted utility model. Also the mean value utility model continues to exhibit the true value of α.

I wanted to see whether the model could distinguish the weighted utility behavior that complies with FSD and risk aversion. For the continuous time weighted utility model, where the support of the probability measure is the real line, any α>0 always violates FSD. But it is possible to find the non-violating range of α for discontinuous weighted utility. In the discontinuous time model considered here the border observations represent points of a normal distribution where 5% of probability mass belongs to the tails, so a range of two standard deviations on both sides of the mean ensures FSD. This sets the parameter requirement at $0 < \alpha < 0.1$.

First I estimated the models with a true value of α=0.06. The results are in tables 10-12. The parameters except α in all three models are so close to their counterparts that it is impossible to statistically distinguish between them, and α is estimated correctly. The misspecified mean value utility parameter estimates still show a slight downward bias. The important result is: when true α=0.06 and in the non-violating range of FSD, it can be estimated accurately to be significantly different from zero.

Tables 13-15 show estimation results when α=0.03. Even though the standard deviations for α are so big that α cannot be distinguished from zero on statistical grounds, the correct mean for α is estimated quite accurately.

Conclusions

This paper is a first attempt to apply weighted utility theory to a transportation choice model. It explicitly models in the choice between uncertain alternatives allowing the decision maker weight the probabilities of possible outcomes. The greater flexibility of the model also allows for some types of preference reversals that violate expected utility assumptions.
The paper also shows one set of assumptions one has to make to theoretically justify the ad hoc practice of adding variance of the uncertain characteristic as an additional variable to the mean value utility function.

The Monte Carlo simulations show that the weighted utility model is possible to estimate from a set of discrete data points or assuming a continuous, normally distributed uncertain attribute. The estimated model retrieves the true parameter values of a correctly specified utility function, whether it is a mean value or a weighted utility. It also shows that if the utility is misspecified as mean value utility, but in fact is weighted utility, the estimated parameters are downward biased, but their ratios stay true. This is something the practitioners should make a note of, since there is a tendency to assume that the value of time derived from transportation demand models includes a 'premium' for the uncertainty of travel time. At least when the true utility is weighted utility, this is not the case.

In general, these results have to be qualified to the extent that we do not know whether the true behavior can be characterized by weighted utility any better than expected utility. In this application the travel time is a monotonic 'bad', less of it is always better. It is outside the scope of this paper to discuss when one should use the competing theories using aspiration points and the assumption that time could be modeled as any other storable resource.

Table 1. Estimation results of three model specifications when the **choice is based on continuous time weighted utility**, and the true parameter values are α=0.15, b_0=0.5, b_1=1.5, and b_2=2.

Model	Sample size	$\bar{\alpha}$	b_0	b_1	b_2
Continuous time weighted utility **Correct Model Specification**	100	0.0896** (0.0615)	0.3716** (0.2789)	2.2231* (0.4702)	2.7624** (0.8162)
	200	0.1436** (0.0598)	0.3843** (0.1811)	1.6337** (0.2735)	2.8171* (0.5874)
	500	0.1085* (0.0386)	0.5662** (0.1111)	1.5264** (0.1674)	1.8847** (0.2992)
	1000	0.1481** (0.0273)	0.4509** (0.0776)	1.5236** (0.1161)	1.6332* (0.2039)
	4000	0.1589** (0.0139)	0.5100** (0.0395)	1.4580** (0.0575)	2.0993** (0.1058)
Discrete time weighted utility	100	0.1023** (0.0731)	0.3716** (0.2790)	2.2232* (0.4703)	2.7623** (0.8162)
	200	0.1687** (0.0730)	0.3842** (0.1811)	1.6334** (0.2723)	2.8170* (0.5873)
	500	0.1253** (0.0457)	0.5662** (0.1111)	1.5263** (0.1672)	1.8847** (0.2992)
	1000	0.1740** (0.0337)	0.4509** (0.0776)	1.5239** (0.1157)	1.6330* (0.2039)
	4000	0.1882 (0.0181)	0.5101** (0.0395)	1.4581** (0.0575)	2.0991** (0.1058)
Mean value utility	100		0.4205** (0.2736)	2.1886* (0.4629)	2.7646** (0.8078)
	200		0.3709** (0.1771)	1.6503** (0.2710)	2.6023* (0.5515)
	500		0.5803** (0.1101)	1.5320** (0.1657)	1.8624** (0.2945)
	1000		0.4399** (0.0760)	1.5069** (0.1140)	1.5927 (0.1994)
	4000		0.4915** (0.0385)	1.4157* (0.0560)	2.0211** (0.1031)
True parameter values		0.15	0.5	1.5	2

Table 2. Estimation results of three model specifications when the **choice is based on discrete time weighted utility**, and the true parameter values are α=0.15, b_0=0.5, b_1=1.5, and b_2=2.

Model	Sample size	α	b_0	b_1	b_2
Continuous time weighted utility	100	0.0896^{**-} (0.0615)	0.3716^{**-} (0.2789)	2.2231* (0.4702)	2.7624** (0.8162)
	200	0.1417** (0.0598)	0.3540^{**-} (0.1808)	1.6315** (0.2735)	2.8896* (0.5919)
	500	0.1015* (0.0385)	0.5682** (0.1113)	1.5286** (0.1678)	1.9527** (0.3017)
	1000	0.1271** (0.0265)	0.4273** (0.0774)	1.5397** (0.1165)	1.6856* (0.2047)
	4000	0.1334* (0.0134)	0.5296** (0.0395)	1.4878** (0.0579)	2.0770** (0.1057)
Discrete time weighted utility **Correct Model Specification**	100	0.1023^{**-} (0.0731)	0.3716^{**-} (0.2790)	2.2232* (0.4703)	2.7623** (0.8162)
	200	0.1662** (0.0726)	0.3539^{**-} (0.1808)	1.6313** (0.2722)	2.8895* (0.5918)
	500	0.1169** (0.0453)	0.5681** (0.1113)	1.5285** (0.1676)	1.9527** (0.3017)
	1000	0.1476** (0.0317)	0.4273** (0.0774)	1.5399** (0.1162)	1.6854* (0.2047)
	4000	0.1556** (0.0165)	0.5297** (0.0395)	1.4879** (0.0579)	2.0770** (0.1057)
Mean value utility	100		0.4205^{**-} (0.2736)	2.1886* (0.4629)	2.7646** (0.8078)
	200		0.3414^{**-} (0.1770)	1.6479** (0.2710)	2.6754* (0.5559)
	500		0.5819** (0.1104)	1.5357** (0.1663)	1.9325** (0.2973)
	1000		0.4203* (0.0762)	1.5310** (0.1150)	1.6565* (0.2012)
	4000		0.5153** (0.0388)	1.4577** (0.0568)	2.0209** (0.0474)
True parameter values		0.15	0.5	1.5	2

Table 3. Estimation results of three model specifications when the **choice is based on mean value utility**, and the true parameter values are α=0, b_0=0.5, b_1=1.5, and b_2=2.

Model	Sample size	α	b_0	b_1	b_2
Continuous time weighted utility	100	0.0557**- (0.0653)	0.3834**- (0.2702)	1.9992* (0.4340)	2.7191** (0.7955)
	200	-0.0088**- (0.0524)	0.4521** (0.1834)	1.7073** (0.2838)	3.1564* (0.6031)
	500	-0.0372*- (0.0371)	0.5615** (0.1097)	1.4865** (0.1640)	1.9292** (0.2942)
	1000	0.0010**- (0.0263)	0.4436** (0.0757)	1.4359** (0.1120)	1.6943* (0.2002)
	4000	-0.0061**- (0.0123)	0.5011** (0.0392)	1.5314** (0.0584)	2.1527** (0.1062)
Discrete time weighted utility	100	0.0630**- (0.0748)	0.3835**- (0.2702)	1.9993* (0.4340)	2.7191** (0.7955)
	200	-0.0099**- (0.0580)	0.4521** (0.1834)	1.7073** (0.2838)	3.1564* (0.6030)
	500	-0.0420*- (0.0415)	0.5615** (0.1097)	1.4865** (0.1640)	1.9292** (0.2942)
	1000	0.0012**- (0.0288)	0.4436** (0.0757)	1.4359** (0.1119)	1.6943* (0.2002)
	4000	-0.0069**- (0.0138)	0.5011** (0.0392)	1.5314** (0.0584)	2.1527* (0.1062)
Mean value utility	100		0.4102**- (0.4331)	1.9963* (0.4331)	2.7337** (0.7933)
	200		0.4521** (0.1834)	1.7017** (0.2813)	3.1641* (0.6023)
Correct Model Specification	500		0.5520** (0.1091)	1.4707** (0.1627)	1.9182** (0.2940)
	1000		0.4437** (0.0756)	1.4362** (0.1118)	1.6945* (0.2002)
	4000		0.5010** (0.0392)	1.5305** (0.0583)	2.1521* (0.1061)
True parameter values		0	0.5	1.5	2

Table 4. Estimation results from Monte Carlo simulations when choice is based on **continuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.5058** (0.0447)	0.4983** (0.0566)	1.5006** (0.1300)	1.9193** (0.2336)
Discontinuous time weighted utility, N=50	0.9309 (0.0009)	0.4895** (0.0574)	1.7021* (0.1265)	1.8823** (0.2283)
Mean value utility, N=50		0.3535 (0.0566)	1.0577 (0.1030)	1.3495 (0.1679)
True parameter values	0.5	0.5	1.5	2

Table 5. Estimation results from Monte Carlo simulations when choice is based on **discontinuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.3315 (0.0361)	0.4912** (0.0640)	1.4962** (0.1158)	1.8914** (0.2383)
Discontinuous time weighted utility, N=31	1.8644 (1.6219)	0.4824** (0.0632)	1.4890* (0.1122)	1.8596** (0.2451)
Mean value utility, N=50		0.4139* (0.0656)	1.2528 (0.1030)	1.5811 (0.1977)
True parameter values	0.5	0.5	1.5	2

Table 6. Estimation results from Monte Carlo simulations when choice is based on **mean value utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0060** (0.0245)	0.4888** (0.0616)	1.5015** (0.1077)	1.9262** (0.1905)
Discontinuous time weighted utility, N=50	0.0068** (0.0265)	0.4888** (0.0616)	1.5015** (0.1077)	1.9262** (0.1905)
Mean value utility, N=50		0.4882** (0.0616)	1.4999** (0.1082)	1.9235** (0.1895)
True parameter values	0	0.5	1.5	2

Table 7. Estimation results from Monte Carlo simulations when choice is based on **continuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.1573** (0.0265)	0.4939** (0.0663)	1.5099** (0.1030)	1.9618** (0.1929)
Discontinuous time weighted utility, N=50	0.1864* (0.0346)	0.4879** (0.0608)	1.5131* (0.0108)	1.9609** (0.1934)
Mean value utility, N=50		0.4718 (0.0656)	1.4397** (0.0995)	1.8699** (0.1934)
True parameter values	0.15	0.5	1.5	2

Table 8. Estimation results from Monte Carlo simulations when choice is based on **discontinuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.1372** (0.0245)	0.4924** (0.0648)	1.5119** (0.1058)	1.9615** (0.1952)
Discontinuous time weighted utility, N=50	0.1604** (0.0316)	0.4861** (0.0583)	1.5143** (0.1072)	1.9596** (0.1987)
Mean value utility, N=50		0.4755** (0.0640)	1.4570 (0.1039)	1.8892** (0.1924)
True parameter values	0.15	0.5	1.5	2

Table 9. Estimation results from Monte Carlo simulations when choice is based on **mean value utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0072** (0.0245)	0.4963** (0.0640)	1.5037** (0.1100)	1.9472** (0.1889)
Discontinuous time weighted utility, N=50	0.0081** (0.0283)	0.4963** (0.0640)	1.5037** (0.1100)	1.9472** (0.1889)
Mean value utility, N=50		0.4954** (0.0632)	1.5016** (0.1100)	1.9448** (0.1873)
True parameter values	0	0.5	1.5	2

Table 10. Estimation results from Monte Carlo simulations when choice is based on **continuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0693** (0.0253)	0.4986** (0.0656)	1.5142** (0.1034)	1.9625** (0.1994)
Discontinuous time weighted utility, N=50	0.0791** (0.0293)	0.4986** (0.0656)	1.5142** (0.1034)	1.9625** (0.1994)
Mean value utility, N=50		0.4934** (0.0649)	1.4980** (0.1029)	1.9401** (0.1969)
True parameter values	0.06	0.5	1.5	2

Table 11. Estimation results from Monte Carlo simulations when choice is based on **discontinuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0625** (0.0244)	0.4985** (0.0667)	1.5116** (0.1028)	1.9574** (0.1983)
Discontinuous time weighted utility, N=31	0.0712** (0.0282)	0.4985** (0.0667)	1.5116** (0.1028)	1.9574** (0.1983)
Mean value utility, N=50		0.4942** (0.0656)	1.4981** (0.1030)	1.9392** (0.1958)
True parameter values	0.06	0.5	1.5	2

Table 12. Estimation results from Monte Carlo simulations when choice is based on **mean value utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0072** (0.0251)	0.4963** (0.0637)	1.5037** (0.1100)	1.9471** (0.1889)
Discontinuous time weighted utility, N=50	0.0082** (0.0285)	0.4963** (0.0637)	1.5037** (0.1100)	1.9472** (0.1889)
Mean value utility, N=50		0.4954** (0.0632)	1.5016** (0.1102)	1.9448** (0.1872)
True parameter values	0	0.5	1.5	2

Table 13. Estimation results from Monte-Carlo simulations when choice is based on **continuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0380**- (0.0240)	0.4951** (0.0662)	1.5053** (0.1091)	1.948** (0.2042)
Discontinuous time weighted utility, N=50	0.0430**- (0.0273)	0.4951** (0.0662)	1.5053** (0.1091)	1.9478** (0.2042)
Mean value utility, N=50		0.4932** (0.0655)	1.4991** (0.1093)	1.9395** (0.2015)
True parameter values	0.03	0.5	1.5	2

Table 14. Estimation results from Monte Carlo simulations when choice is based on **discontinuous time weighted utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0345**- (0.0241)	0.4952** (0.0655)	1.5014** (0.1108)	1.9484** (0.2016)
Discontinuous time weighted utility, N=31	0.0391**- (0.0274)	0.4952** (0.0655)	1.5014** (0.1108)	1.9484** (0.2016)
Mean value utility, N=50		0.4935** (0.0648)	1.4960** (0.1110)	1.9412** (0.1991)
True parameter values	0.03	0.5	1.5	2

Table 15. Estimation results from Monte Carlo simulations when choice is based on **mean value utility**. Sample size = 1000. Numbers in parenthesis are the standard deviations of estimates.

Model	α	b_0	b_1	b_2
Continuous time weighted utility, N=50	0.0072** (0.0251)	0.4963** (0.0637)	1.5037** (0.1100)	1.9471** (0.1889)
Discontinuous time weighted utility, N=50	0.0081** (0.0285)	0.4963** (0.0637)	1.5037** (0.1100)	1.9472** (0.1889)
Mean value utility, N=50		0.4954** (0.0632)	1.5016** (0.1102)	1.9448** (0.1872)
True parameter values	0	0.5	1.5	2

References

Camerer, C. F.(1989): An Experimental Test of Several Generalized Utility Theories. Journal of Risk and Uncertainty, V2, 61-104.

Chew, S. H. (1982): A Mixture set axiomatization of weighted utility theory. Discussion Paper 82-4, College of Business and Public Administration, University of Arizona, Tuscon.

Chew, S.H. (1983): A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. Econometrica, V51, 1065-92.

Chew, S. H. and MacCrimmon, K. R. (1979): Alpha-nu choice theory: a generalization of expected choice theory. Working paper 669, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, Canada.

Chew, S. H. and Waller, W. S.(1986): Empirical Tests of Weighted Utility Theory. Journal of Mathematical Psychology, V30, 55-72.

Conlisk, J. (1989): Three variants on the Allais Example. American Economic Review, V79, 392-407.

Fishburn (1981): An axiomatic characterization of skew-symmetric bilinear functionals, with applications to utility theory. Economic Letters, V8, 311-313.

Fishburn (1983): Transitive measurable utility. Journal of Economic Theory, V31, 293-317.

Fishburn (1988): Nonlinear Preference and Utility Theory. Baltimore: Johns Hopkins University Press.

Gaudry,Marc J. I. and Michael J. Wills (1978): Estimating the functional form of travel demand models, Transportation Research, V12, 257-289.

Lee, Lung-Fei (1982): Specification error in Multinomial Logit models - Analysis of the Omitted Variables Bias, Journal of Econometrics, V20, 197-209.

The MVA Consultancy, Institute of Transport Studies University of Leeds, and Transport Studies Unit University of Oxford (1987): The Value of Travel Time Savings, Berks, United Kingdom: Policy Journals.

Nakamura, Y. (1984): Nonlinear measurable utility analysis. Ph.D. dissertation, University of California, Davis.

Nakamura, Y. (1985): Weighted linear utility. Preprint, Department of Precision Engineering, Osaka University, Osaka, Japan.

Senna, L.A.D.S.(1991): Risk of delays, uncertainty, and travelers' valuation of travel time variability. In: Planning and Transport Research and Computation (International) Co. Meeting. Proceedings of Seminar G held at PTRC summer annual meeting. Vol. P347.

Senna, L.A.D.S. (1994): The influence of travel time variability on the value of time, Transportation, V 21, 203-228.

Winston, C. (1981): A Disaggregate Model of the Demand for Intercity Freight Transportation, Econometrica, V49, 981-1006.

Yatchew, Adonis and Zvi Griliches (1985): Specification error in Probit models, The Review of Economics and Statistics, V67, Feb:134-139.

APPENDIX A

Demonstration that indifference curves are parallel in probability space Expected Utility

To demonstrate that the behaviours characterized by weighted utility and expected utility are different one has to show that the indifference curves in probability space are different. If this is true, transforming the utility expressions of individual outcomes will not compensate for the difference.

The independence axiom of expected utility requires parallel straight indifference curves in probability space, where as the corresponding axioms in weighted utility require them to be straight lines originating from one common point outside the probability simplex.

Now, let the utilities of probability measures h, q, and r be U(h), U(q), and U(r). In expected utility framework the functional form of the utility function U has to be linear in probability i.e. $U(ph + (1-p)q) = pU(h) + (1-p)U(q)$.

Consider two arbitrary utility levels K and K'. If the indifference curves are not parallel, they will have to intersect at some point. To demonstrate that they don't, lets write the expected utility expressions for K, K' and the technical constraint that probabilities of all possible outcomes have to add up to one.

$$p_h U_h + p_q U_q + p_r U_r = K$$

$$p_h U_h + p_q U_q + p_r U_r = K'$$

$$p_h + p_q + p_r = 1$$

In matrix form

$$\underbrace{\begin{bmatrix} U_h & U_q & U_r \\ U_h & U_q & U_r \\ 1 & 1 & 1 \end{bmatrix}}_{A} \begin{bmatrix} p_h \\ p_q \\ p_r \end{bmatrix} = \begin{bmatrix} K \\ K' \\ 1 \end{bmatrix}$$

The first matrix A is clearly singular and therefore not invertible. This shows that there exists no solution to the problem and indifference curves are parallel.

The first matrix A is clearly singular and therefore not invertible. This shows that there exists no solution to the problem and indifference curves are parallel.

Now, let's consider similar situation with a weighted utility functions:

$$\frac{p_h w_h U_h + p_q w_q U_q + p_r w_r U_r}{p_h w_h + p_q w_q + p_r w_r} = K$$

$$\Leftrightarrow p_h(U_h w_h - K w_h) + p_q(U_q w_q - K w_q) + p_r(U_r w_r - K w_r) = 0$$

which is simplified with a change of terms into:

$$\Leftrightarrow p_h(s_h - K w_h) + p_q(s_q - K w_q) + p_r(s_r - K w_r) = 0$$

The matrix form is now

$$\underbrace{\begin{bmatrix} s_h - Kw_h & s_q - Kw_q & s_r - Kw_r \\ s_h - K'w_h & s_q - K'w_q & s_r - K'w_r \\ 1 & 1 & 1 \end{bmatrix}}_{A} \begin{bmatrix} p_h \\ p_q \\ p_r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

And the solution to p_h is

$$p_h = \frac{|A_1|}{|A|} = \frac{(K - K')(w_r s_q - w_q s_r)}{(K - K')(w_r s_q - w_q s_r + w_h s_r - w_r s_h + w_q s_h - w_h s_q)}$$

This shows that p_h does not depend on utility levels K and K', but only on utilities of individual outcomes U_h, U_q, and U_r, and their weights w_h, w_q, and w_r. Because the solution point p_h for two arbitrary utility levels does not depend on the utility levels, all indifference lines have to meet in the same point. This shows that one cannot transform the independent variables in the expected utility expression and get the same behavior as in weighted utility.

APPENDIX B

Lee (1982) studied bias in multinomial logit models due to omitted dichotomous variables. He concludes that as long as the omitted explaining variables z are independent of included variables x, conditional on the response variable y, the coefficient b1 of x will not be biased. However, the constant term b0 will be biased.

When the response variable y, included explaining variable x and omitted explaining variable z are not independent, the coefficients will be biased.

Lee presents a multinomial logistic probability model

$$\ln\frac{P(y = i|x,z)}{P(y = 0|x,z)} = \alpha_{i_0} + x\alpha_{i_1} + z_1\beta_{i_1} + \ldots + z_M\beta_{i_M}, \qquad i = 1,\ldots,L$$

where $z_1,\ldots,z_M$ are dichotomous variables and

$$P(z = j|x) = F_j(x) \qquad j=1,..,M$$

is the (unspecified) probability function of z conditional on x.

It follows that

$$\ln\frac{P(z = j|x,y)}{P(z = 0|x,y)} = J_j(x) + y_1\beta_{i_j} + \ldots + y_L\beta_{L_j} \qquad j=1,\ldots,M$$

where

$$J_j(x) = \ln\frac{F_j(x)}{F_0(x)} + \ln\left(\frac{1 + \sum_{i=1}^{L}\exp(\alpha_{i_0} + x\alpha_{i_1})}{1 + \sum_{i=1}^{L}\exp(\alpha_{i_0} + x\alpha_{i_1} + \beta_{i_j})}\right)$$

Since x and z are not independent conditional on y, $J_j(x)$ are not constant functions for at least some j. The bias for omitted variables will be

$$\ln \frac{P(y = i|x)}{P(y = 0|x)} = \alpha_{i_0} + x\alpha_{i_1} - G_i(x)$$

where

$$G_i(x) = \ln\left(\frac{1+\sum_{j=1}^{M} \exp J_j(x)}{1+\sum_{j=1}^{M} \exp(J_j(x) + \beta_{i_j})}\right).$$

If the functions $G_i(x)$ are not constants, and the misspecified model omits the $G_i(x)$ functions, both the coefficient of x and the constant term will be affected.

Lee is able to derive more specific results by further restricting the included variable to be discrete. In that case

$$\ln \frac{P(z = j|x,y)}{P(z = 0|x,y)} = \delta_{j_0} + x_1\delta_{j_1} + \ldots + x_K\delta_{j_K}$$

$$+ y_1\beta_{1_j} + \ldots + y_L\beta_{l_j} \qquad j = 1,\ldots,M$$

Now the misspecified model is

$$\ln \frac{P(y^{*} = i|x)}{P(y = 0|x)} = \alpha_{i_0} + x_1\alpha_{i_1} + \ldots + x_K\alpha_{i_K}$$

$$- \ln\left(\frac{1+\sum_{j=1}^{M}\exp(\delta_{j_0} + x_1\delta_{j_1} + \ldots + x_K\delta_{j_K})}{1 + \sum_{j=1}^{M}\exp(\delta_{j_0} + x_1\delta_{j_1} + \ldots + x_K\delta_{j_K} + \beta_{i_j})}\right)$$

$$= \alpha_{i_0^*} + x_1\alpha_{i_1^*} + \ldots + x_K\alpha_{i_K^*}, \qquad i = 1,\ldots,L$$

where

$$\begin{aligned} \alpha_{i_0^*} &= \alpha_{i_0} - G_i(k) + G_i(0) \\ \alpha_{i_k^*} &= \alpha_{i_k} - G_i(k) + G_i(0) \end{aligned} \qquad k = 1,\ldots,K$$

with

$$G_i(0) = \ln \frac{1 + \sum_{j=1}^{M}\exp\delta_{j_0}}{1+\sum_{j=1}^{M}\exp(\delta_{j_0} + \beta_{i_j})}$$

$$k = 1,\ldots,K$$

$$G_i(k) = \ln \frac{1+\sum_{j=1}^{M}\exp(\delta_{j_0} + \delta_{j_k})}{1 + \sum_{j=1}^{M}\exp(\delta_{j_0} + \delta_{j_k} + \beta_{i_j})}$$

When the omitted variable z is dichotomous, i.e. M=1, and the included variable x is discrete, the coefficient α_{ik} of x_k in the misspecified model will be

i) biased upward if either $\beta_i > 0$ and $\delta_k > 0$, or $\beta_i < 0$ and $\delta_k < 0$,

ii) biased downward if either $\beta_i > 0$ and $\delta_k < 0$, or $\beta_i < 0$ and $\delta_k > 0$,

iii) unaffected if either $\beta_i = 0$ or $\delta_k = 0$.

For the models estimated in this paper the true model is one of the weighted utility models when α is not equal to zero. Both of the models represent same behavior, but it is easier to consider the continuous time model:

$$\begin{aligned} V &= b_0 - b_1(\mu + \alpha\sigma^2) - b_2 c \\ &= b_0 + b_1\mu - b_1\alpha\sigma^2 - b_2 c \end{aligned}$$

In the misspecification α is restricted to equal zero and the travel time variance term is omitted. However, the variance is continuous in [0,4], and therefore the results do not apply directly. But because of the signs of the variables in regression

$$\sigma^2 = \underset{(-)}{\delta_0} + \underset{(-)}{\delta_1\mu} + \underset{(-)}{\delta_2 c} + \underset{(+)}{\beta_1 y}$$

we can expect the estimated coefficients to be biased downwards.

On the same theme, Yatchew and Griliches (1985) derived a bias formula for omitted variables in a probit model:

$$\hat{b}_0 = \frac{b_0 + \beta\delta_0}{\sqrt{\beta^2\sigma_v^2 + \sigma_e^2}}$$

$$\hat{b}_1 = \frac{b_1 + \beta\delta_1}{\sqrt{\beta^2\sigma_v^2 + \sigma_e^2}}$$

where

$b_{0,1}$ = The true parameter of included variables in the correctly specified model

β = The true parameter of the omitted variable

$\delta_{0,1}$ = Parameters from a regression where the omitted variable is regressed on the included variables and the choice variable

σ_v^2 = Variance of the error term when the omitted variable is regressed on the included and choice variables

σ_e^2 = Variance of the error term of the true model

The authors elaborate: "The impact of omitting the variable z on the estimate of α and β is twofold. There is the usual effect, familiar from the linear case, where the bias in the coefficient of the omitted variable equals the coefficient of the omitted variable times the coefficient of the included variable from the regression of the omitted on the included variables.

In addition there is a rescaling effect determined by the denominators [.], so that even if the omitted variable is uncorrelated with included variables, there is bias in the coefficients."

The simulation models considered in this paper demonstrated some correlation between the variables. If the logit model can be considered an approximation of the probit model, both of the above sources of bias are present.