Value of travel time changes

Theory and simulation to understand the connection between Random Valuation and Random Utility methods

Ojeda-Cabral, Manuel; Chorus, Caspar G.

DOI
10.1016/j.tranpol.2016.03.006

Publication date
2016

Document Version
Accepted author manuscript

Published in
Transport Policy

Value of travel time changes: theory and simulation to understand the connection between random valuation and random utility methods

Manuel Ojeda-Cabral, *Corresponding Author*
Institute for Transport Studies, University of Leeds
30-40 University Rd, Leeds LS2 9JT,
UK
Tel: +44 (0)7834 738 505; Email: M.A.OjedaCabral@leeds.ac.uk

Caspar G. Chorus
Transport and Logistics Group, Delft University of Technology
Faculteit Techniek Bestuur en Management
Kamer B3.120, Jaffalaan 5
2628 BX Delft
Netherlands
Tel. +31 (0) 15 27 88546; Email: c.g.chorus@tudelft.nl

Abstract

This paper identifies and illustrates the theoretical connection between the Random Valuation (RV) and Random Utility (RU) methods for Value of Travel Time Changes (VTTC) analysis. The RV method has become more and more popular recently, and has been found to lead to very different estimation results than conventional RU models. Previous studies have reported these differences but did not explain them, which limited the confidence in the RV model as a useful foundation for transport policy analysis. In this paper, we first analytically show in what way exactly the two models are different and why they may generate different estimation results. Based on this deeper understanding of the connection and difference between the two models, we formulate hypotheses regarding the conditions under which differences in estimation results are expected to be smaller or larger. Using synthetic data, we empirically test these expectations. Results provide strong support for our hypotheses, allowing us to derive a number of practical recommendations for analysts interested in using the RV and RU models in their VTTC-analysis.

Keywords: random utility, random valuation, value of time, value of travel time changes

1. Introduction

The value of travel time changes (VTTC), which measures how people trade off travel time changes against changes in travel costs[1], is a crucial component of cost-benefit analyses and plays an important role in transport policy design and evaluation studies (Small, 2012; Börjesson and Eliasson, 2014). The large majority of VTTC-studies infer this trade off by means of estimating discrete choice models on data obtained from Stated Preference (SP) experiments, where participants to the experiment are asked to choose between a slower but cheaper, and a faster but more expensive route or travel mode (e.g. Mackie et al., 2003; Fosgerau et al., 2007; Börjesson and Eliasson, 2014). Traditionally, the adopted discrete choice model is of the Random Utility (RU) type (McFadden, 1974).

However, quite recently an interesting alternative to RU has emerged: this so-called Random Valuation (RV) model has been gaining attention lately, after several empirical studies have found it to be superior to RU in terms of explaining respondents' preferences (as measured in model fit). The RV model differs from the RU model in terms of how it conceptualizes behavior. The RV approach, in a context where a person can choose between a cheap but slow and a fast but expensive travel option, postulates that people decide as if they were in a "time market": they choose the fast option when their valuation of the presented travel gain is larger than the implicit price of the travel gain which is embedded in the choice situation. The RV-method[2] was suggested by Cameron and James (1987) in an environmental economics context, although the use of the term "RV" can be attributed to Hultkranz et al. (1996). Fosgerau et al. (2007b) were the first to formally introduce the method in a VTTC-context. Since then, a number of studies have shown that there may be large differences in the VTTCs estimated by RU and RV respectively, on a given dataset; model fit differences have been found to be substantial as well (e.g., Ojeda-Cabral et al., 2016, Daly and Tsang, 2009)). These studies reported VTTCs that, in comparison with a VTTC from a RV model, were often around 1.5 or 2 times greater when a RU model was estimated. Ojeda-Cabral et al. (2016) reported an extreme case where the RU estimate tripled the RV estimate. It goes without saying, that such differences have potentially very large implications for the evaluation of transport policies and infrastructure investments.

Although the theoretical relationship between the RU and RV models has been discussed in previous papers (Fosgerau et al., 2007b; Börjesson and Eliasson, 2014; Hultkranz et al., 1996, Ojeda-Cabral et al., 2016), this discussion is not complete, as we will argue below. As a consequence, the observed non-trivial empirical differences in model fit and estimated VTTC have so far come as a surprise, for which no full explanation is yet provided. Given that the RV approach is growing in popularity in the field of transport economics, we believe that a rigorous assessment of the connection and differences between the RU and RV approaches is needed. This paper provides such an in-depth exploration and interpretation of the connection between RU and RV through the use of analytical derivations and analyses on simulated data. Note that although at first sight, exploration of the differences between the two models might come across as a methodological exercise, it has clear and substantial policy relevance. More specifically, given that the differences and similarities between the two approaches have so far been ill understood at a conceptual level, there has been a hesitation to use the VTTC estimates produced by the relatively new and

[1] Most of the literature uses the term *travel time savings*. However, since many transport projects lead to travel time losses and, in fact, most studies do consider savings as well as losses, we use the more generic term *travel time changes*; see Ojeda-Cabral et al. (2016) for a more detailed overview of terminology.
[2] In this paper, we will use the terms 'model', 'method' and 'approach' when referring to RU or RV.

unknown RV model in cases where its empirical performance (e.g. model fit) turned out to be superior to that of the well-known RU model. As a consequence, the RV's penetration in the transport policy discourse has been severely limited by the absence of a clear and unambiguous understanding of how and when the model and its VTTC output differ from RU and its VTTC. This goal of this paper is to lift the confusion which so far has surrounded the RV model, and as such provide a more solid foundation based on which researchers and analysts can make safe and well informed decisions regarding which model and VTTC estimates to use for transport policy analyses, based on the model's empirical performance.

In Section 2, we highlight the importance of an element which has been missing in previous studies: whereas those studies have argued that the two methods are equivalent in the deterministic domain (i.e., when error terms are excluded), we show that this equivalence only applies in an ordinal sense (i.e., preference orderings between two alternatives are the same in both models), but not in a cardinal sense (i.e., the extent to which an alternative is preferred over another one may vary substantially across the two model types). Since, in a discrete choice context, cardinal differences determine choice probabilities (after error terms have been included), this cardinal inequivalence between RU and RV causes differences in terms of model fit and VTTC estimates. Based on this insight, we are able to formulate hypotheses about the size of the difference between the RU and RV models that one would expect for various types of data, i.e., various types of SP designs and different levels of randomness in choice behavior. These hypotheses are subsequently tested based on empirical analyses on synthetic data.

In section 3, we formulate hypotheses concerning their differences in terms of model fit and obtained VTTCs, for different types of data. We also present the construction of the simulated data sets, estimation of the RU and RV models, and the interpretation of estimation results. In section 4 we present overall conclusions, and we provide recommendations for future research; in addition, we discuss practical implications of the obtained insights.

2. Random utility and random valuation: the theoretical connection

The RU model assumes that a person faced with a choice between multiple options, chooses the option that offers the greatest total utility. This total utility is usually conceived in term of a summation of a deterministic (or: 'systematic', 'observed') utility and a random error. For sake of exposition, we initially focus only on this deterministic part of utility. Deterministic utility V_i of each option i is a usually linear-additive function of its observable characteristics (in our case, travel time and cost) and associated parameters: $V_i = \beta_c c_i + \beta_t t_i$; here, β_t and β_c are the estimable marginal utilities of travel time (t) and cost (c), respectively. The value of travel time changes (VTTC) is equal to the marginal rate of substitution between time and cost, which is of a convenient form when systematic utility is specified linearly, as above: $VTTC = \frac{\partial V}{\partial t} \Big/ \frac{\partial V}{\partial c} = \beta_t \big/ \beta_c$.

The Random Valuation (RV) model (Cameron and James, 1987; Hultkranz et al,, 1996, Fosgerau et al., 2007b) is applicable when, in the choice context, there is an implicit 'price' for the good we want to value such as in our case a change in travel time. This is the case in a binary choice context where alternatives are described in terms of a price attribute and a quality attribute (in our case travel time); note that many recent SP-experiments have adopted such a binary, two attribute choice context, including several European national VTTC studies, including those in the UK, Denmark, Sweden and Norway (Mackie et al., 2003; Fosgerau et al., 2007; Ramjerdi et al, 2010; Börjesson and Eliasson, 2014). The implicit price (denoted Boundary VTTC or BVTTC) can then be defined as follows. Throughout the paper, we will assume a choice context in which option 1 is slower but cheaper than option 2 (i.e. faster and more expensive): i.e. $t_1>t_2$ and $c_1<c_2$. Then, the price threshold or BVTTC, is equal to: $\text{BVTTC} = \frac{-(c1-c2)}{(t1-t2)} = -\frac{\Delta c}{\Delta t}$, where Δt and Δc are the differences in travel time and cost, respectively, between options 1 and 2. The RV model assumes that people choose whether they accept the price of time (BVTTC) which is implicitly embedded in the choice situation, or not. If the individual's VTTC is larger than the BVVTC, the faster but more expensive option is chosen. As in the RU model, additive errors are introduced in the RV model to accommodate randomness; hence the individual's choice probabilities will be driven by the difference between the VTTC and the BVTTC, such that $y = 1\{VTTC < \text{BVTTC} + \varepsilon\}$ (see further below for details).

The RV model has been said to be equivalent to the RU model in the deterministic domain, i.e. before randomness in the form of errors is introduced (Fosgerau, 2007; Ojeda-Cabral et al., 2016). However, these studies implicitly referred to *ordinal* equivalence. Indeed, in the deterministic domain, the two models can easily be shown to be equivalent in an ordinal sense. To see this, consider an individual whose VTTC equals $\frac{\beta_t}{\beta_c}$. Take the above described binary choice situation involving a cheap and slow alternative (1) and a fast but expensive alternative (2), with an implicit price that equals $\frac{-(c_1-c_2)}{(t_1-t_2)}$. Now it can be easily seen that $\frac{-(c_1-c_2)}{(t_1-t_2)} > \frac{\beta_t}{\beta_c}$ if and only if $\beta_t t_1 + \beta_c c_1 > \beta_t t_2 + \beta_c c_2$. In other words, if BVTTC > VTTC in the RV model this necessarily implies that $V_1 > V_2$ in the RU model; both inequalities imply that the cheaper but slower option is chosen. This makes the two models equivalent in an ordinal sense.

Given the equivalence (in an ordinal sense) between RU and RV in the deterministic domain, previous research has related the observed differences between the two models in model fit and obtained VTTC-estimates, to the way in which randomness is introduced in the two models. However, here we show that the difference and connection between the two models in the deterministic domain is more subtle than the ordinal analysis directly above may suggest at first sight. Specifically, it has so far been overlooked that a difference between the two models arises when we consider a *cardinal* as opposed to ordinal perspective. To see this, consider again an individual whose VTTC equals $\frac{\beta_t}{\beta_c}$. Take again the above described binary choice situation involving a cheap and slow alternative (1) and a fast but expensive alternative (2), with an implicit price for the travel time difference that equals $\frac{-(c_1-c_2)}{(t_1-t_2)}$. Now, it can be seen that the cardinal difference between systematic utilities V_1 and V_2 in the RU model is *not* equal to the cardinal difference between price (BVTTC) and value (VTTC) in the RV model: $\beta_t t_1 + \beta_c c_1 - (\beta_t t_2 + \beta_c c_2) \neq \frac{-(c_1-c_2)}{(t_1-t_2)} - \frac{\beta_t}{\beta_c}$; or in other words: $V_1 - V_2 \neq BVVTC - VTTC$. Rather, one obtains $\frac{\beta_t t_1 + \beta_c c_1 - (\beta_t t_2 + \beta_c c_2)}{\beta_c (t_1 - t_2)} = \frac{\beta_t}{\beta_c} - \frac{-(c_1-c_2)}{(t_1-t_2)}$; or, equivalently, $V_1 - V_2 = \beta_c (t_1 - t_2) \cdot [BVVTC - VTTC]$. The factor $\beta_c (t_1 - t_2)$ is the product of the marginal utility of cost and the travel time difference between the two options.

If the utilities in the RU model are divided by this factor, it becomes a RV model[3]. Note that Börjesson and Eliasson (2014) and Ojeda-Cabral et al. (2016), in their comparisons of the RU and RV model, have also identified this factor as having role in scaling parameters and error terms. However, the factor's crucial property (i.e., that it determines the connection between the two models in the deterministic domain, from a cardinal perspective) has been overlooked until now.

In sum: both models, given a particular underlying value of travel time changes for an individual, always agree on *which* of the two alternatives (i.e., the cheap & slow or the expensive & fast alternative) is preferred by the individual. However, with the exception of some very specific conditions (see further below) the two models disagree on the *extent to which* one alternative is preferred over the other. To give one example for illustrative purposes: the RV model states that the extent to which one alternative is preferred over the other one, by an individual with a particular VTTC, remains constant as long as the implicit price (BVVTC) which is embedded in the choice situation remains the same. For example, for the RV model it does not matter if the fast alternative is 10 minutes faster and 2 pound more expensive than the slow one, or 5 minutes faster and 1 pound more expensive. In both cases, the BVVTC equals 0.2 pounds per minute, and the difference between this value and the individual's VTTC determines the extent to which the fast alternative is (not) preferred over the slow one. In contrast, the RU model postulates that when attribute differences between the alternatives become smaller, the extent to which one of the alternatives is preferred over the other one decreases as well, up to a point where the individual is assumed to become almost indifferent between the two alternatives when attribute differences become very small. So, in the above example the RU model predicts that – given a particular underlying VTTC – the extent to which the fast alternative is preferred by the individual over the slow one (or vice versa) is larger in the 10 minutes / 2 pound case than in the 5 minutes / 1 pound case. So, even though both models (RU and RV) would always agree on whether or not the fast alternative is to be preferred over the slow one, they may generate markedly different predictions in terms of the extent to which the most attractive alternative is preferred over the other one. It is this cardinal difference in preferences which gives rise to differences in choice probabilities in the stochastic domain. Although analysts may of course have theoretical preferences with respect to the different implicit behavioral premises underlying the two models (such as the ones discussed above), in the end it is of course an empirical question which of the two fits best with the collected choice data.

We now proceed to the stochastic domain, by adding errors. We start with the RU model. To arrive at closed form Logit-type choice probabilities, the error term (ε_i) is assumed to follow a Gumbel distribution (type-I generalized extreme value distribution) with constant variance normalized at $\pi^2/6$, and is introduced additively (McFadden, 1974):

$$U_i = V_i + \varepsilon_i = \beta_c c_i + \beta_t t_i + \varepsilon_i \quad (1)$$

In the context of a binary choice set containing alternatives 1 and 2, (as noted earlier, the RV method only works in the context of binary choices), choice probabilities are then given by:

[3] Note that, while it is intuitive to think about a monetary price of time (i.e. RV model), there is no principled reason why one should not divide by the cost difference instead, giving an (inverse) RV model in e.g. minutes/pence terms. This alternative model would be worthy of investigation, but it is outside of the scope of this paper..

$$P(1) = \frac{\exp(V_1)}{\exp(V_1)+\exp(V_2)} = \frac{\exp(\beta_c c_1+\beta_t t_1)}{\exp(\beta_c c_1+\beta_t t_1)+\exp(\beta_c c_2+\beta_t t_2)} \text{ ; with } P(2) = 1 - P(1) \quad (2)$$

Note that the difference in systematic utilities (V_i) between travel alternatives determines the choice probabilities derived from the RU model.

In the RV model, like in the RU model, Gumbel errors with constant variance normalized at $\pi^2/6$ are added so as to allow for the derivation of closed form Logit type choice probabilities:

$$\begin{cases} U_1 = \mu \cdot \text{BVTTC} + \varepsilon_1 \\ U_2 = \mu \cdot \text{VTTC} + \varepsilon_2 \end{cases} \quad (3)^4,$$

$$P(1) = \frac{\exp(\mu\text{BVTTC})}{\exp(\mu\text{BVTTC})+\exp(\mu\text{VTTC})} \text{ ; with } P(2) = 1 - P(1) \quad (4),$$

Clearly, the difference between μ ·*BVTTC* and μ ·*VTTC* determines the choice probabilities derived from the RV model. Note that scale factor μ is estimated in the RV approach, together with VTTC. Importantly, the RU model can be rewritten in what has been called Willingness to Pay space, by dividing and multiplying the time-parameter by the cost-parameter. In notation, $V_i = \beta_c c_i + \beta_c(\beta_t/\beta_c)t_i$. In this case, cost-parameter β_c becomes a *de facto* scale parameter. This too would result in a model where scale of utility and VTTC are estimated. It is this variant of the RU model, which is fully equivalent to the formulation presented in (1) and (2), which we use in our empirical analysis, as it facilitates an easy comparison between RU and RV.

Having specified choice probabilities, we can now start exploring why the two models – which we have shown to be ordinally equivalent yet cardinally different in the deterministic domain – are expected to lead to different model estimation outcomes (i.e., model fit and estimated VTTC) in the stochastic domain. The key to understanding this lies in the obvious fact that choice probabilities are determined by the difference $V_1 - V_2$ in the RU model, and between μ ·*BVTTC* and μ ·*VTTC* in the RV model. Above, we have shown that $V_1 - V_2 = \beta_c(t_1 - t_2) \cdot [BVVTC - VTTC]$. Now, given that scale parameter μ is estimated in the RV model, the two models would become equivalent in the stochastic domain when $\mu = \beta_c(t_1 - t_2)$. However, when $t_1 - t_2$ differs between observations as is practically always the case in real life SP-experiments, it is impossible to find one estimate for μ which makes the choice probabilities derived from the two models equivalent for every single observation in the dataset. This argument lies at the core of the differences in estimation results reported in previous studies, and it allows us to formulate hypotheses as to when the difference between the RU and RV models should be expected to be substantial.

[4] In this equation, BVTTC is observed in the data, while VTTC is estimated.

3. Formulation of hypotheses and empirical analysis based on synthetic data

Previous work (Hultkranz et al., 1996; Daly and Tsang, 2009; Ojeda-Cabral et al., 2016) showed that there may be significant empirical differences between RU and RV model, both in the estimated VTTC as well as in model fit. In general, in these studies the RV model provided a much better model fit and a significantly lower valuation. However, as explained above, these sizeable differences remained not fully understood. It remained unclear if the RV model would often or always fit the data better or whether it would often or always provide lower VTTCs. Based on the derivations in the previous Section, explicit hypotheses can be formulated regarding what determines the differences in model estimation outcomes. More specifically, we identify two factors which determine the size of the difference between RU and RV estimation results (model fit and estimated VTTC):

1) The variation of Δt across cases, i.e., across choice tasks provided in the experiment: if only one level of Δt was used in the design (e.g. the fast route was always 10 minutes faster than the slow route), the RU and RV models will generate the same results. The reason for this lies in the fact that under this condition, there exists a single scale factor in the RV model which leads to identical behavior between RV and RU models: $\mu = \beta_c * \Delta t$. Under maximum likelihood estimation conditions, it is therefore impossible to obtain different model fits for the two models, or different VTTCs. To the extent that Δt differs across cases / choice tasks, the estimated value for μ will only be an imprecise proxy for $\beta_c * \Delta t$ for most cases. This implies that to the extent that Δt differs across cases / choice tasks, there may be a better or worse model fit for the RV model compared to RU (depending of course on which of them mimics best the underlying data generating process); and both models will lead to different VTTCs.
2) Level of randomness in choice behavior: when choices are such that in most cases there is always a very strong preference for one of the two options[5], then both the RU and RV model will generate very high choice probabilities for the most attractive alternative, and there will be only small differences in model fit and estimated VTTC between RU and RV. The reason behind this, is that in such a situation, the ordinal equivalence of the two models is what counts (i.e., both will always agree on which alternative in a choice task is the most attractive one). Even if for example the RU model predicts a substantially larger or smaller utility difference than the RV model, this will hardly impact choice probabilities as these are close to 0/1 anyway. A different situation occurs when, from the analyst's viewpoint, choices are more random in the sense that choices are more evenly distributed across the fast and slow routes. In that case, where choice probabilities generated by the two models are closer to 0.5, the fact that $V_1 - V_2 \neq [BVVTC - VTTC]$ does translate into relatively large choice probability differences between RU and RV, due to the steeper slope of the Logit-curve around choice probabilities of 0.5.

It goes without saying that most actual datasets will include substantial variation of Δt across cases, and will feature fairly dispersed choice behavior in the sense that observed choice frequencies close to 0/1 are rare in SP-data. As a consequence, the above discussion already indicates that one should expect relatively substantial differences between RU- and RV-based model estimation results in the context of real data. In the remainder of this section, we will put the above two hypotheses to the test empirically, making use of synthetic data, as such

[5] This can be due to either a particular combination of times and costs in the choice task, which makes one of the alternatives clearly superior to its competitor; or it can be due to a very strong dislike for times and costs in the population; or a combination of these two factors.

data allow us efficiently, effectively and independently to control the variation of Δt across cases and the level of randomness in choice behavior. Furthermore, in contrast to a real experiment, the synthetic set up allows us to control the true data generating process (DGP) in terms of decision rule (RU versus RV) and true underlying VTTC. That way, we can explain model fit differences in favor of one of the two models, and differences in VTTC, effectively.

The structure of this synthetic data experiment is in the matrix shown directly below:

Variation in Δt across cases	Much variation in Δt across cases	*C1*	*C2*	*C3*
	Some variation in Δt across cases	*B1*	*B2*	*B3*
	No variation in Δt across cases	*A1*	*A2*	*A3*
		Almost no randomness in choice behavior	Some randomness in choice behavior	Much randomness in choice behavior
		Degree of randomness in choice behavior		

Figure 1: Design of the synthetic data experiment

In line with the discussion above, we hypothesize to find larger differences between the RU and RV models, when moving away from the lower left hand area or 'origin' (the extreme case being A1) to the upper right hand area (C3 being the extreme case). The ordering of the table can be interpreted as a coordinate system where we have two axes *x* (randomness) and *y* (Δt), whose magnitudes increase from the origin (A1). For each cell of the matrix, we generate choices using RU and RV respectively as the true DGP; and then we estimate both models (i.e., RU model estimated on RU data, RU model estimated on RV data, RV model estimated on RU data, and RV model estimated on RV data). This implies that we generate 9*2=18 different datasets, and that we report a total of 9*4=36 model estimation results. Without loss of general applicability, each data set contains 10,004 choices made by as many individuals (i.e., each individual is assumed to make one choice). The reason for the rather odd number 10,004 is that, for the first simulated design, we removed all design rows where the BVTTC was greater than 100, retaining a total of 10,004 cases; we then adhered to that number for the other designs as well.

The SP-design we use to generate choice data builds on two major national VTTC studies: the UK VTTC study (Mackie et al., 2003) and the Danish VTTC study (Fosgerau et al., 2007). This facilitates drawing comparisons with these real datasets. Both studies used a simple design where only two options and two attributes (time and cost) were presented in each choice scenario, allowing for application of the RV method. The Danish study was a pioneer in implementing a form of the RV model to estimate official VTTC measures for national level transport policy evaluation. Each choice task is designed to make sure that there is always a faster but more expensive option and a cheaper but slower one. The following design rules were applied (note that letters A, B, and C refer to Figure 1):

i) Δt:

a. for design A, we used a travel time difference between the slow and fast option, of 10 minutes; and kept this constant for all cases.
b. For design B, travel time differences between the slow and fast option are randomly drawn, for each case, from a uniform distribution between 0 and 20 minutes[6].
c. For design C, travel time differences between the slow and fast option are randomly drawn, for each case, from a uniform distribution between 0 and 60 minutes[7].

ii) Δc: For all designs A, B and C, travel cost differences between the cheap and expensive option are randomly drawn, for each case, from a uniform distribution between 0 and 300 pence[8].

Note that in the context of designs B and C, the combination of random draws for Δt and Δc generated a wide variation in BVTTCs. To avoid numerical issues, we *ex post* restricted the range of BVTTC to an upper limit of 100 pence per minute. Also note that these random draws did not influence choice behavior: each design (A, B and C) is a fixed input prior to the simulation of more or less random choices, just as it is in a real life choice experiment.

For every design we simulated choices based on an RU- as well as based on an RV-based decision process. These decision processes assume values for β_t and β_c (RU model), as well as for for $\frac{\beta_t}{\beta_c}$ (i.e., VTTC) and μ (RV model). We made sure that both models were always based on the same underlying VTTC of 10 pence per minute, which holds for all simulation exercises (this homogeneity allows us to more easily interpret differences between the RU and RV model outcomes). By carefully selecting combinations of β_t, β_c and μ, while ensuring a constant ratio $\frac{\beta_t}{\beta_c}$ for both models, we were able to systematically vary the degree of randomness embedded in the simulated choices , while keeping constant the underlying VTTC (since the degree of randomness by definition decreases with the magnitude of the coefficients, *ceteris paribus*). In an iterative process, we obtained the following three levels of randomness (note that numbers 1, 2, and 3 refer to Figure 1):

1) Almost no randomness: for both models, more than 9,600 out of 10,004 cases come with a predicted choice probability for the most attractive alternative which is higher than 90%. In other words, in the vast majority of cases, both models assign a very high choice probability to the most attractive option, making the dataset almost deterministic from the analyst's viewpoint (and implying a very high rho-squared, i.e. implying a very good model fit, for both models).
2) Some randomness: for both models, between 800 and 900 (out of 10,004) cases come with a predicted choice probability for the most attractive alternative which is higher than 90%. In other words, in some cases, both models assign a very high choice probability to the most attractive option, while in many other cases, the difference in choice probabilities between the two options is less pronounced. Note that the associated rho-

[6] This in fact is based on the values used for the 2003-UK VTTC study, where 20 was the maximum level.
[7] This in fact is based on the values used for the 2007-Danish VTTC study, where 60 was the maximum level.
[8] This in fact is based on the values used for the 2003-UK VTTC study, where 300 was the maximum level.

squared of around 0.175 is about the same size of what one would expect in a real dataset in the context of VTTC-estimation.

3) Much randomness: for both models, less than 70 (out of 10,004) cases come with a predicted choice probability for the most attractive alternative which is higher than 90%. In other words, only in some rare cases, do both models assign a very high choice probability to the most attractive option, while in the vast majority of cases, the difference in choice probabilities between the two options is much less pronounced, leading to a highly random dataset and very low levels of model fit.

All models were estimated using Biogeme (Bierlaire, 2003). Table 1 shows estimation results for all 36 models, displaying parameter estimates and measures of model fit. Note that as discussed in the previous section, to estimate the RU model we have rearranged the parameters of the model to allow us to estimate VTTC directly instead of β_t (note that β_c becomes a scale parameter, consequently denoted by μ in the table). This does not affect model fit in the context of MNL and facilitates comparison between RU and RV estimates.

Table 1. Estimation results

		Almost no randomness in choice behavior				**Some randomness in choice behavior**				**Much randomness in choice behavior**			
Much variation in Δt across cases	**Preferences**	*True DGP: RU*		*True DGP: RV*		*True DGP: RU*		*True DGP: RV*		*True DGP: RU*		*True DGP: RV*	
	Model estimated	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*
	Null LL	-6934.24				-6934.24				-6934.24			
	LL	-113.9	-243.7	-200.1	-166.6	-5462.1	-5896.7	-5401.3	-4998.8	-6201.2	-6434.5	-6833.9	-6758.7
	Adj. ρ^2	0.98	0.97	0.97	0.98	0.21	0.15	0.22	0.28	0.11	0.07	0.01	0.03
	Parameters	**C1**				**C2**				**C3**			
	VTTC	9.99	10	9.93	9.96	10.1	21	7.59	9.97	9.75	22.4	6.03	10.3
	μ	0.48	3.48	0.29	5.24	0.01	0.06	0.01	0.20	0.00	0.04	0.00	0.03
	VTTC (s.e.)	0.02	0.04	0.03	0.03	0.26	0.89	0.12	0.16	0.36	1.16	0.33	0.63
	μ (s.e.)	0.04	0.91	0.02	0.39	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Some variation in Δt across cases	**Preferences**	*True DGP: RU*		*True DGP: RV*		*True DGP: RU*		*True DGP: RV*		*True DGP: RU*		*True DGP: RV*	
	Model estimated	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*
	Null LL	-6934.24				-6934.24				-6934.24			
	LL	-239.2	-391.4	-290.8	-252.3	-5418.4	-5907.4	-6233.9	-5976.1	-6771.7	-6828.6	-6742.2	-6666.6
	Adj. ρ^2	0.97	0.94	0.96	0.96	0.22	0.15	0.10	0.14	0.02	0.02	0.03	0.04
	Parameters	**B1**				**B2**				**B3**			
	VTTC	10	10	10	10	10.1	10.7	9.63	10.3	10.5	11.3	10.1	11
	μ	0.50	2.99	0.41	4.71	0.01	0.07	0.01	0.06	0.00	0.02	0.00	0.03
	VTTC (s.e.)	0.02	0.03	0.03	0.02	0.15	0.33	0.24	0.37	0.48	1.45	0.45	0.87
	μ (s.e.)	0.03	0.35	0.03	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No variation in Δt across cases	**Preferences**	*True DGP: RU*		*True DGP: RV*		*True DGP: RU*		*True DGP: RV*		*True DGP: RU*		*True DGP: RV*	
	Model estimated	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*	*RU*	*RV*
	Null LL	-6934.24				-6934.24				-6934.24			
	LL	-236.1	-236.1	-236.1	-236.1	-5473.9	-5473.9	-5473.9	-5473.9	-6787.6	-6787.6	-6787.6	-6787.6
	Adj. ρ^2	0.97	0.97	0.97	0.97	0.21	0.21	0.21	0.21	0.02	0.02	0.02	0.02
	Parameters	**A1**				**A2**				**A3**			
	VTTC	10	10	10	10	10.1	10.1	10.1	10.1	10.8	10.8	10.8	10.8
	μ	0.47	4.74	0.47	4.74	0.01	0.13	0.01	0.13	0.00	0.04	0.00	0.04
	VTTC (s.e.)	0.02	0.02	0.02	0.02	0.18	0.18	0.18	0.18	0.59	0.59	0.59	0.59
	μ (s.e.)	0.03	0.32	0.03	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The table shows the results of 36 models estimated, organized in 9 big cells (3x3); it thus corresponds exactly to the experimental scheme provided in Figure 1 presented earlier. Each row corresponds to one SP-design (A, B and C), while each column corresponds to a degree of randomness in choices (the adj. ρ^2 being an indicator of it). For each cell, we show 4 sets of results: two models (RU and RV) are estimated on a dataset where the DGP was RU, and on a dataset where the DGP was RV. If the estimated model matches the DGP, we will refer to this as the 'right' model; a 'wrong' model is an estimated model that does not match the DGP. The mean and robust standard error (s.e.) of the VTTC and scale parameters are displayed, together with model fit measures (final Log-Likelihood (LL) and adj. ρ^2). It is directly seen, that obtained results match our expectations:

Constant travel time differences ($\Delta t = 10$)

- In the simplest design (A), where we assume that in every case, the travel time difference between the fast and slow option equals 10 minutes, both models yield identical results irrespectively of the underlying DGP. In all these cases the estimation results show that $\mu = \beta_c * \Delta t$. The VTTC of 10 p/min. is recovered with great precision in A1 and A2. The great degree of randomness in A3 causes the VTTC estimation to deviate slightly (10.8 p/min.) from the underlying true value, as one may expect. However, also then both models result in the exact same estimate for VTTC (and exactly the same final-LL).

Hardly any randomness in choice behavior

- If Δt varies across cases, but choices are almost deterministic implying very high choice probabilities for the most attractive option, in almost every case – i.e., in cases B1 and C1 – the RU and RV models are almost equivalent, as hypothesized. They both identify the true underlying VTTC, although model fit differences are significant in designs B1 and C1, in favor of the model that corresponds to the DGP.

Entering the real world

Cells B2, B3, C2 and C3 represent what is typically observed in real life experiments: choices are relatively random (from the analyst's perspective) and experiments consider different levels of Δt for different cases.

- The right model is always able to recover the true underlying VTTC, although as expected the precision decreases (i.e., the Standard Error increases) as the level of randomness in the choices increases.
- The wrong model is now always much worse in terms of model fit compared to the right one, even when it does not perform too badly in terms of recovering the true VTTC (e.g. case B2, where the wrong models give VTTC of 10.7 and 9.63 p./min respectively).
- When the variation in Δt is larger (design C), the wrong models estimate VTTCs that are very far from the underlying 10p./min, even when choices are not very random (see the VTTCs of 21 and 7.59 p./min in C2).

4. Conclusions, discussion and directions for further research

This paper has identified the connection between the Random Valuation (RV) and Random Utility (RU) methods for Value of Travel Time Changes (VTTC) analysis. The RV method has become more and more popular recently, often leading to very different estimation results (i.e., model fit and estimated VTTC). Previous studies have reported these differences but did not explain their source; instead they pointed at the fact that the two models are equivalent in the deterministic domain, in the sense that they will always agree on which of the two options is the most attractive one in a given choice task. In this paper, we first analytically showed that the two models actually differ in the deterministic domain, from a cardinal perspective, in the sense that the extent to which one option is preferred over the other one may differ between RU and RV models. We then showed how this cardinal difference translates into differences in model estimation results. This deeper understanding of the connection and differences between the two models allowed us to formulate precise hypotheses regarding the conditions under which smaller or larger differences in estimation outcomes are to be found. We then employed a carefully constructed experiment based on synthetic data to test these hypotheses.

Taken together, results obtained from that synthetic data experiment provided strong support for our hypotheses, and were also found to be in line with – and help explain – findings obtained in previous studies based on real data. In sum: to the extent that the choice probabilities of the fast and slow options are somewhat similar (i.e., both are relatively close to 0.5), and to the extent that travel time differences between the two options vary across cases/choice tasks, the RU and RV model should generate different results in terms of model fit and estimated VTTC. Only under the fairly unrealistic assumption that choice probabilities of the fast and slow options are always very close to 0 or 1, and/or in a (yet unexplored) context where travel time differences between the two options are constant across cases/choice tasks, do the RU and RV model become equal.

Of course, in real life experiments, we never know the true underlying choice processes of the individuals, making it impossible to *a priori* select one model's estimation results. Our results highlight the risk of getting completely wrong values if we fail to approximate the true underlying choice process by estimating a RU model when RV is much closer to the data generating process (DGP), or vice versa. The good news is that we can now safely argue in this RU-RV context that, if in real life a given model (RU or RV) gives better model fit, it is apparently a better explanation of the observed choices and we should prefer the VTTC estimate derived from it, even if it is very different from the other model's VTTC. This may to some extent appear to be obvious, but note that in previous studies, given the incomplete assumption that the two models were equivalent in the deterministic realm, large differences in model fit and valuation came as a surprise (Ojeda-Cabral et al., 2016), making it difficult to argue that the VTTC of the best fitting model should in fact be preferred for transport policy analysis. It is this observation that carries the policy relevance of our analyses: by lifting the confusion surrounding the RV model, we provide a more solid base for researchers and policy analysts to select and trust the RV model and its VTTC in case its empirical performance is better than that of RU.

Another source of policy relevance of this paper lies in the observation that evidence from previous studies on real data (Hultkranz et al., 1996; Daly and Tsang, 2009; Ojeda-Cabral et al., 2016) where RU and RV were compared empirically, suggested that RV consistently yielded lower VTTC-estimates. This turns out not to be the case in the context of our simulated datasets, where the RV often leads to higher VTTC estimates than those obtained by RU. Apparently, estimating the ‘wrong’ model can lead to failure in the recovery of the true underlying VTTC, but with our current knowledge it is not possible to state *a priori* the direction of the bias. Based on our analyses (including our analytical identification of the similarities and differences between the RU and RV models) we can safely advise analysts to select the model (RU or RV) with best empirical performance, and trust its VTTC-estimate for policy analysis.

In sum, this paper expands current knowledge concerning the RV model, being an alternative model to the classical RU model, which has been receiving increasing attention among scholars and practitioners during the last few years. Our work clarifies the relationship between these two models, thereby substantially increasing the scope for applying the RV model for transport policy analysis.

Obviously, our study leaves considerable opportunities for further research, of which we here identify two: firstly, our empirical exercises assumed a unique VTTC for the full (artificial) population of respondents. This is not a realistic representation of real life, where the VTTC varies across individuals and even for the same person, across choice tasks. The replication of this work introducing distributions for the underlying VTTC seems an important direction for future research. Secondly, whereas our study focused on linear specifications of the RU and RV models (which is in line with the fact that the large majority of VTTCs used for policy analysis are obtained from linear models), some previous studies have been experimenting with log-specifications. Extending our results to such non-linear models is also an interesting avenue for further study.

References

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models *Proceedings of the 3rd Swiss Transportation Research Conference,* Ascona, Switzerland.

Börjesson, M. and Eliasson, J. (2014). Experiences from the Swedish Value of Time study. *Transportation Research Part A, 59, pp.144-158.*

Cameron, T.A. and James, M.D. (1987). Efficient estimation methods for ‘closed-ended’ contingent valuation surveys. *The Review of Economic and Statistics, Vol.69, No.2, pp.269-276*

Daly, A. and Tsang, F. (2009). Improving understanding of choice experiments to estimate values of travel time. *Association for European Transport*.

Fosgerau, M. (2007). Using Non-parametrics to specify a model to measure the value of travel time. *Transportation Research A 41 (9), pp.842-856*

Fosgerau, M., Hjorth, K. and Lyk-Jensen, S.V. (2007). The Danish Value of Time Study: Results for Experiment 1. Report for the Ministry of Transport, *Danish Transport Research Institute*

Fosgerau, M., Hjorth, K. and Lyk-Jensen, S.V. (2007b). An approach to the estimation of the distribution of marginal valuations from discrete choice data. *Munich Personal RePEc Archive, Paper No.3907*

Hultkranz, L., Li, C., Lindberg, G. (1996). Some problems in the consumer preference approach to multimodal transport planning. *CTS Working Paper.*

Mackie, P., Wardman, M., Fowkes, A.S., Whelan, G., Nellthorp, J. & Bates J.J. (2003). Values of Travel Time Savings in the UK. Report to Department for Transport. *Leeds and Abingdon: Institute for Transport Studies, University of Leeds & John Bates Services.*

McFadden, D. (1974). Conditional logit analysis of quantitative choice behavior. *Frontiers in econometrics, ed by P.Zarembka, Academic Press, New York,* p.105-142.

Ojeda-Cabral, M. Hess, S. and Batley, R. (2016). The value of travel time: random utility versus random valuation. *Transportmetrica A: Transport Science*, Vol.12, pp.230-248

Ramjerdi, F., Flügel, S., Samstad, H., and Killi, M. (2010). Value of time, safety and environment in passenger transport – Time. *TØI report 1053B/2010, Institute of Transport Economics, Oslo.*

Small, K. (2012). Valuation of Travel Time. *Economics of Transportation 1, p.2-14*.