Research note:

A logistic regression based method to uncover the Value-of-Travel-Time distribution

Sander van Cranenburgh[a], Marco Kouwenhoven[a]
[a]Transport and Logistics Group, Delft University of Technology

1 Motivation

Van Cranenburgh and Kouwenhoven (under review) propose a new nonparametric method to uncover the Value-of-Travel-Time (VTT) distribution using data from a binary within-mode experiment (in which decision makers have to choose between a slow and cheap alternative and a fast and expensive alternative). In the core of this method – which we henceforth refer to as CK method – is an Artificial Neural Network (ANN). The ANN aims to approximate individual level choice models. It does so by capturing associations between the tuple of explanatory variables (the series of T choices, the received Boundary Value-of-Travel-Times[1] (BVTTs), and a set of socio-demographic variables), and the dependent variable: the probability of choosing the fast and expensive alternative in the 'next' T+1[th] choice task. The CK method is appealing as it uncovers the VTT distribution (and its moments) without making strong assumptions on the underlying behaviour. Moreover, the method incorporates covariates, accounts for panel effects and yields a distribution right of the maximum received BVTT.

But, while the empirical performance of ANNs is often found to be superior over those of theory-driven statistical techniques (Paliwal and Kumar 2009), a severe limitation of ANNs –and by extension of the CK method– is their intractability. ANNs are widely considered black boxes (Castelvecchi 2016). Amongst other things, this is because it is impossible to interpret or diagnose ANNs by looking at the weights obtained after training the network. In fact, the weights will tell the analyst nothing about whether the ANN has learned intuitively correct relationships, as opposed to spurious ones, or about the importance of attributes. Not even the signs of the weights can meaningfully be interpreted. This relates to the indeterminacy of ANNs. Even in single-layer ANNs there are many symmetric solutions (Sussmann 1992). This limitation hampers (1) learning from the ANN, (2) improving the ANN, and (3) diagnosing the ANN (e.g. can we trust the model's predictions?).

In the way the ANN is used in the CK method, it can be perceived as a juiced-up logistic regression model. The juicing-up comes from the hidden layers, which allow the ANN to capture nonlinearities and interactions between the explanatory variables. In case we would remove the hidden layers, what is left is a 'standard' logistic regression model (Bishop 1995). After all, the explanatory variables in the input layer then directly enter the output layer, where a softmax

[1] In binary two-attribute VTT choice data, the boundary value of time is the implicit price of the time difference between the two alternatives.

function is applied. The softmax function is a logit transformation. While a logistic regression model is a data-driven model, a clear advantage of a logistic regression model over an ANN is that it is not a black box; the model and its parameters can readily be interpreted. But, doing so comes at a cost. It is not capable to capture nonlinearities and complex interactions, i.e. unless they are explicitly programmed by the analyst.

In this research note we explore using a standard logistic regression model in CK's method, instead of the ANN. The logistic regression model can be seen as a simplified, but tractable proxy of the ANN. Thereby, we aim to shed light on *how* the ANN recovers the VTT distribution in CK's method. Although the regression model is a proxy for the ANN, it may be able to provide a deeper theoretical understanding on how the ANN works. Furthermore, it may reveal empirical insights. For instance, the sign and strength of the main effects could be made transparent.

To conduct this exploration, we take the following steps. We start with the ANN as depicted in Van Cranenburgh and Kouwenhoven (under review) Figure 2, and remove the hidden layers. What is left is a logistic regression model. However, this model does not immediately work. Therefore, we make a number of modifications to this regression model. Next, we estimate the parameters of the regression model using the same data as in Van Cranenburgh and Kouwenhoven (under review) and recover the individual level VTTs. We interpret the regression parameters and compare the recovered VTT distribution to that recovered using the ANN and to those of other parametric and nonparametric methods. Finally, we show that the logistic regression model can be casted in the form of a Random Valuation (RV) model.

The remainder of this research note is organised as follows. Section 2 builds the logistic regression model. Section 3 presents the empirical results and compares the recovered VTT distribution by the logistic regression model to those of other methods. Section 4 examines how the logistic regression model relates to RV models. Section 5 provides conclusions.

2 The logistic regression model

2.1 Building the logistic regression model

We start by replacing Equation 3 of Van Cranenburgh and Kouwenhoven (under review) with a logistic regression model, see Equation 1. That is, we remove the hidden layers. As such, this equation is a 'one-to-one' conversion of the ANN into a logistic regression model, where P_2^n denotes the probability that decision maker *n* chooses the fast and expensive alternative. The first term of v_2^n, δ, is the regression intercept. The second term of v_2^n captures the effect of presented BVTTs in choice tasks $t = 1...T$; the third term captures the effect of the experimental covariates in choice tasks $t = 1...T$; the fourth term captures the effect of the choices in choice tasks $t = 1...T$; the fifth term captures the effect of the generic covariates d_r, and the sixth and seventh terms capture respectively the effects of the BVTT and experimental covariates in the choice task T+1. Note that in our data we have just one experimental covariate, the quadrant; hence no summation is used there. Furthermore, choice task R, see Figure 2 in Van Cranenburgh and Kouwenhoven (under review) is not included. We will discuss this in section 2.3.

$$
\begin{aligned}
&P_2^n = \frac{1}{1+e^{v^n}}, \quad \textit{where} \\
&v_1^n = 0, \\
&v_2^n = \delta + \sum_{t=1}^{T} \beta_{bvtt}^t bvtt_t^n + \sum_{t=1}^{T} \beta_s^t s_t^n + \sum_{t=1}^{T} \beta_y^t y_t^n + \sum_{r=1}^{R} \beta_r d_r^n + \beta_{bvtt}^{T+1} bvtt_{T+1}^n + \beta_s^{T+1} s_{T+1}^n
\end{aligned}
$$

Equation 1

However, the regression model in Equation 1 will not immediately work since in this form it does not ‘see’ which BVTTs are accepted, and which are not. To let the regression model see this, rather than using the BVTTs (second term) and the choices (fourth term) separately, we use the interaction between the two, see Equation 2. In this way the interaction parameter, denoted $\beta_{y \cdot bvtt}$, captures the effect of accepted propositions conditional on the received BVTTs. We expect a positive sign for $\beta_{y \cdot bvtt}$ as higher accepted BVTTs is expected to correlate positively with the probability of choosing the fast and expensive alternative in choice task T+1. Likewise, the third term in Equation 1 will not work in this form as the regression model does not see the effect of the experimental covariates (the quadrants) on the choice made in particular choice tasks. While it is also possible to incorporate these experimental covariates by means of interactions (like is done for the choices), for clarity of exposition we drop this term here. However, note that the effect of the quadrants on the T+1 choice task is directly accounted for (last term).

$$
v_2^n = \delta + \sum_{t=1}^{T} \beta_{y \cdot bvtt}^t y_t^n bvtt_t^n + \sum_{r=1}^{R} \beta_r d_r^n + \beta_{bvtt}^{T+1} bvtt_{T+1}^n + \beta_s^{T+1} s_{T+1}^n
$$

Equation 2

We can simplify the regression model, by reducing the number of estimable weights. In particular, we expect the weights associated with the interaction between y and $bvtt$ to be equal across tasks t = 1...T, i.e. $\beta_{y \cdot bvtt}^t = \beta_{y \cdot bvtt} \; \forall t$. After all, order effects are taken out since the data are shuffled in random order (see Section 2.2). In Van Cranenburgh and Kouwenhoven (under review), section 2.3, it is also noted that the ANN consumes more weights than strictly needed. However, in the context of the ANN equal weights are not imposed as this would make the CK method less accessible: it would then require specialised software. But, in the context of the logistic regression model imposing equal parameters can readily be done. This simplifies the second term of Equation 2 to Equation 3.

$$
v_2^n = \delta + \beta_{y \cdot bvtt} \sum_{t=1}^{T} y_t^n bvtt_t^n + \sum_{r=1}^{R} \beta_r d_r^n + \beta_{bvtt}^{T+1} bvtt_{T+1}^n + \beta_s^{T+1} s_{T+1}^n
$$

Equation 3

2.2 Model estimation

To estimate the regression model, we take the following steps. First, for each respondent we randomly select one choice task, which serves as the dependent choice task (T+1). After that, we estimate the logistic regression model using a standard Maximum likelihood based approach. Note that as a result of our modifications there is no need to permute the data numerous times, as is the case for the CK method. After all, for the logistic regression model we are able to directly estimate the generic $\beta_{y \cdot bvtt}$ parameter.

Table 1 shows the regression results.[2] Based on Table 1 a number of observations can be made. First, the ρ^2 of 0.40 indicates the model fits the data well, although not as good as the ANN, which achieved a ρ^2 of 0.49 (based on out-of-sample data). The model fit is roughly the same as the RV models which include random parameters; see Van Cranenburgh and Kouwenhoven (under review). But, it should be noted that the rho squares cannot one-to-one be compared, as the fit of the logistic regression model is conditional on the previous choices made while the RV models are unconditional. Second, the explanatory variables have the expected and intuitive correct signs (apart from a few generic covariates that are found to be insignificant). In particular, we see that $\beta_{y \cdot bvtt}$ is positive and highly significant. This means that accepting high BVTTs increases the probability of choosing the fast and expensive alternative in task T+1.

Looking at the explanatory variables associated with choice task T+1, we see that β_{bvtt}^{T+1} is negative and highly significantly different from one. This is also expected since a higher BVTT in choice task T+1 is expected to reduce the probability of choosing the fast and expensive alternative. With regard to the effect of the quadrants in choice task T+1, in line with expectations, we see the estimates are of EL, EG and WTP are all negative. This means the probability of choosing the fast and expensive alternative decreases when the T+1*th* choice task is presented in the EL, EG, or WTP domain instead of in the WTA domain. Finally, most of the covariates are insignificantly different from zero, except for a few covariates for which we expect strong effects on the VTT, such as the highest income levels. Apparently, the choices made in choice tasks 1 to T explain a large portion of the heterogeneity in the VTT, leaving less to be explained by the covariates.

Table 1: Logistic regression results

[2] Note that we used dummy coding for socio demographic variables and the quadrants

	Est.	SE	t-stat	p-val
No. observations		5832		
No. parameters		24		
Null LogLikelihood		-4042.4		
Final LogLikelihood		-2435.4		
ρ^2		0.40		
Intercept				
δ	-0.572	0.288	-1.99	0.05
t = 1 ... T				
$\beta_{y \cdot bvtt}$	0.038	0.002	24.14	0.00
t = T+1				
β^{T+1}_{bvtt}	-0.142	0.005	-30.02	0.00
β_{WTA}	0.000	fixed		
β_{EL}	-0.218	0.096	-2.27	0.02
β_{EG}	-0.356	0.099	-3.60	0.00
β_{WTP}	-0.704	0.109	-6.46	0.00
Mode				
Car	0.000			
Public transport	-0.284	0.124	-2.29	0.02
Bus	-0.239	0.171	-1.40	0.16
Train	-0.002	0.168	-0.01	0.99
Gender				
Male	0.000			
Female	0.002	0.077	0.02	0.98
Age				
18-20	0.000			
21-35	0.277	0.276	1.00	0.32
36-50	0.100	0.279	0.36	0.72
51-64	-0.147	0.280	-0.53	0.60
65+	-0.116	0.292	-0.40	0.69
Purpose (short distance)				
Return home	0.000			
Travel to the workplace	0.058	0.110	0.52	0.60
Leisure/excerise activities	-0.238	0.143	-1.66	0.10
Commute	-0.003	0.114	-0.03	0.98
Long/short distance				
Short	0.000			
Long	0.409	0.147	2.78	0.01
Personal gross income				
Under 300 000 NOK/year	0.000			
300 001 - 400 000 NOK/yr.	0.234	0.107	2.20	0.03
400 001 - 500 000 NOK/yr.	0.453	0.113	4.01	0.00
+500 001 NOK/yr	0.621	0.114	5.44	0.00
Do not know / No answer	0.216	0.174	1.24	0.22
Current trip characteristics				
Travel time	-0.001	0.000	-1.89	0.06
Travel cost	0.006	0.001	4.33	0.00

2.3 Recovery of individual level VTTs

Recovery of individual level VTTs from the estimated logistic regression model is fairly straightforward. Unlike in the case of the ANN, we can derive the VTTs analytically. In other words, there is no need to simulate choice probabilities. Specifically, from Equation 1 it can be seen that an individual is indifferent between the fast and expensive alternative and the slow and cheap alternative in case v_2^n equals zero. Therefore, setting v_2^n to zero in Equation 3 and solving it for $bvtt_n^{T+1}$ gives us the individual level VTT (Equation 4).

$$VTT^n = -\frac{1}{\beta_{bvtt}^{T+1}}\left(\delta + \beta_{y\cdot bvtt}\sum_{t=1}^{T} y_t^n bvtt_t^n + \sum_{r=1}^{R}\beta_r d_r^n + \sum_{q=1}^{Q}\beta_{sq}^{T+1} s_{qT+1}^n\right) \quad \text{Equation 4}$$

In the CK method an additional set of nodes is added to the network architecture for choice task R. Choice task R is a randomly selected replication of one of the *T* explanatory choice tasks. This is done to ensure that all information on choices and BVTTs for a respondent can be used when simulating the choice probabilities, which is the intermediate step towards recovering the individual level VTTs. From Equation 4 it can readily be seen why it is necessary to add these extra nodes. In the term $\sum_{t=1}^{T} y_t^n bvtt_t^n$ the sum across choice tasks t = 1...*T* is taken. Since there are *T*+1 choice tasks in total and one is randomly taken out to serve as the dependent choice task, these sums depend on the particular choice task that is left out. As a result, the recovered VTT depends on the random manifestation of the selection of the *T*+1[th] choice task. Clearly, this is undesirable. For the logistic regression model, there is no need to add additional input variables (resembling the set of nodes associated with the choice task R). This issue can easily be resolved by using the average of the received bids interacted with the choices, see Equation 5 where $\widehat{ybvtt}^n$ respectively denotes the average of the accepted BVTTs for decision maker *n* across all *T*+1 choice tasks.

$$bvtt_{T+1}^n = -\frac{1}{\beta_{bvtt}^{T+1}}\left(\delta + \beta_{y\cdot bvtt} T \cdot \widehat{ybvtt}^n + \sum_{r=1}^{R}\beta_r d_r^n + \sum_{q=1}^{Q}\beta_{sq}^{T+1} s_{T+1}^n\right)$$
$$where,\ \widehat{ybvtt}^n = \frac{1}{T+1}\sum_{t=1}^{T+1} y_t^n bvtt_t^n \quad \text{Equation 5}$$

3 Empirical results

3.1 VTT distribution

To recover the individual level VTTs we apply Equation 5 for each respondent. Figure 1 shows histograms of the recovered VTT distributions. First, we look at the left-hand side plot. In this plot we see the VTT distributions for the WTP and WTA domains. It shows that a substantial number of respondents have a negative VTT, especially in the WTP domain. Furthermore, as expected (see section 2.3) the WTP and WTA distributions are similar: the WTA distribution is the same as the WTP distribution only shifted to the right by a little less than 5 euro per hour. This was expected given Equation 4, where it can be seen that the WTP-WTA gap, and given our choice of normalisation, equals $\beta_{WTP} / \beta_{bvtt}^{T+1}$. Using the estimates of Table 1 this boils down to a constant 4.94 euro/h.

To obtain the reference free VTT we compute the arithmetic mean across the VTTs of the WTP and WTA domains. Here, we deviate from the approach taken in Van Cranenburgh and Kouwenhoven (under review) who use the geometric mean. We use the arithmetic mean because of the substantial share of respondents for whom a negative VTT is recovered, especially in the WTP domain. Specifically, in case we would use the geometric mean, we would not be able to derive a reference free VTT for 23% of the respondents. By using the arithmetic mean, we partly resolve this problem, although still for 9.6% of the respondents also the arithmetic mean is negative. For the remaining analyses, we set their VTTs to zero (hence the spike at $x = 0$ in the right-hand side plot in Figure 1).

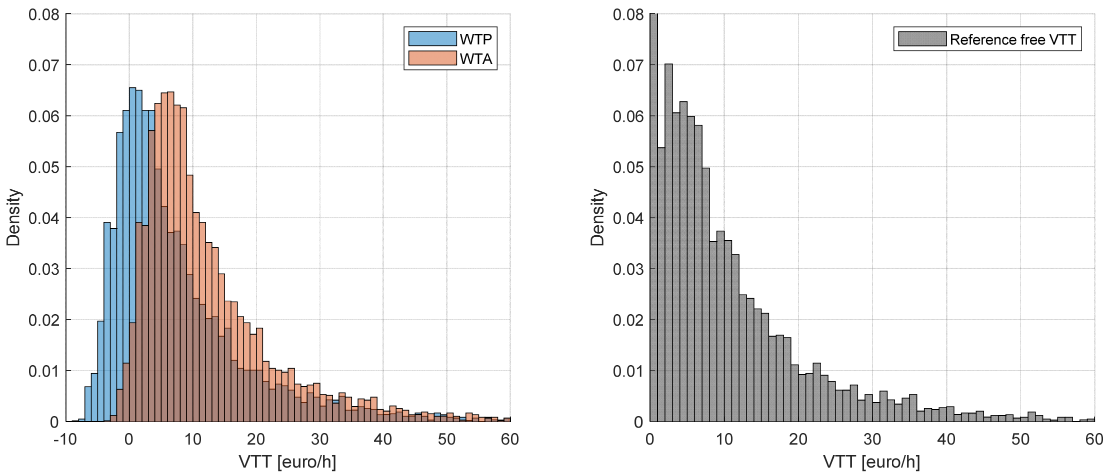

Figure 1: WTP and WTA VTT distribution (left), reference free VTT distribution (right)

3.2 Comparison to other VTT methods

In this section we compare the recovered VTT distribution using the logistic regression model to the VTT distribution recovered using the ANN based method (section 3.2.1) as well as to parametric and nonparametric methods presented in Van Cranenburgh and Kouwenhoven (under review) (section 3.2.2). Comparison of the results of the logistic regression model to those of the ANN will give insights on the extent to which the logistic regression model is a good proxy for the ANN. A large discrepancy between the two models indicates the logistic model is poorly able to capture the relations captured by the ANN, while a small discrepancy indicates the opposite. Hence, it can shed light on what is the cost of making the CK method more transparent by replacing the ANN with a logistic regression model. Comparisons of the logistic regression model to existing parametric and nonparametric methods can provide insight on how well the method performs relative to current VTT practice.

3.2.1 Logistic regression model vs. ANN

Figure 2 scatters the VTT of the logistic regression against the VTT of the ANN based method. The plot shows that the logistic regression model is a fair proxy for the ANN. The VTT predictions of the two methods are strongly correlated (the Pearson product-moment correlation coefficient is ρ = 0.88). To further investigate the correspondence between the two methods, a smoothing spline fit is added. This spline is close to the $y = x$ line, especially between VTT = 10 and 30 euro/h. This shows that in this domain the average of the predicted VTTs is fairly similar across models. Discrepancies are particularly noticeable at x <10 euro/h and at x > 30 euro/h. The discrepancy at $x = 0$ is due to the fact that the logistic regression model recovers a VTT of zero for 9.6% of the respondents, against 0.14% by the ANN-based CK method. The discrepancy at the x > 30 euro/h could be due to that the data becomes thinner populated, and hence the spline becomes more driven by outliers. Alternative, this is due to the fact that the ANN captures effects (e.g. nonlinearities) which the logistic regression model is blind for, resulting in systematic (downward) bias by the logistic regression model.

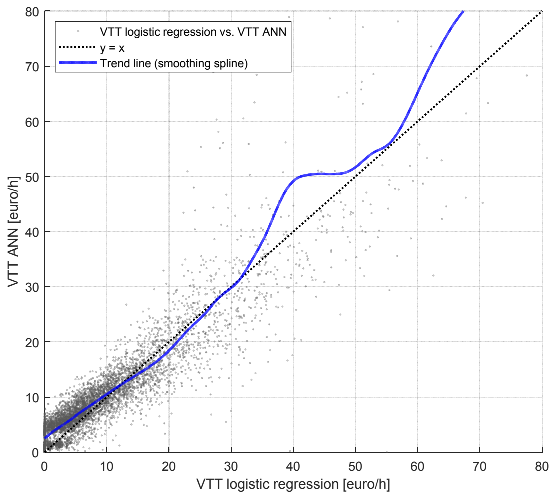

Figure 2: Scatter plot: VTT logistic regression model vs. VTT ANN

3.2.2 Logistic regression model vs. existing VTT methods

Figure 3 shows the Cumulative Density Functions (CDFs) of the VTT distribution, recovered using the Logistic regression model (purple, both plots), the ANN-based CK method (blue, both plots), an RV model with a lognormal distribution VTT distribution (orange, left plot), Rouwendal's method (Rouwendal et al. 2010) (green, right plot) and the semi nonparametric approach (cyan, right plot) developed by Fosgerau and Bierlaire (2007). The key result of Figure 3 is that the shape of the VTT distribution recovered by the logistic regression model is in line with that of the other methods. The most noticeable difference as compared to the other methods is that the logistic regression model predicts a relatively thin right-hand side tail.

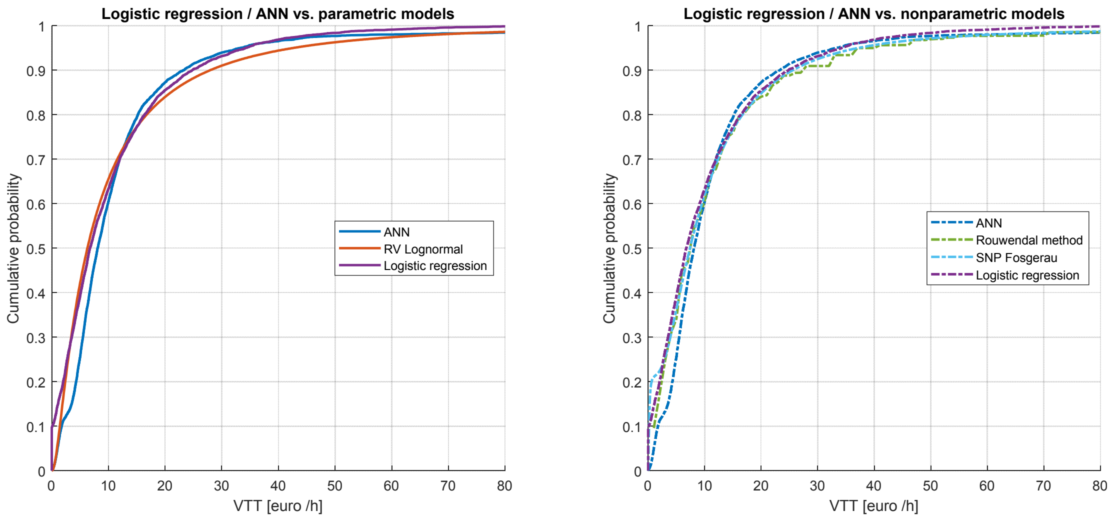

Figure 3: Cross-validation of shape

4 Relation to Random Valuation

The logistic regression model in this approach can be casted in an RV model form. This is possible since a logit transformation is in the core of both the logistic regression model and the RV model. However, it is crucial to notice that while it is mathematically possible to specify the logistic regression model in RV form, we do not conceive the logistic regression model, and by extension the ANN, to be a disaggregate choice model. Given the data-driven nature of CK's method, they try to *approximate* the individual specific choice models f_n.

To show how the logistic regression model can mathematically be specified as an RV model recall that in our notation subscript 1 and 2 respectively denote the slow and cheap alternative and the fast and expensive alternatives. This is similar to conventional notation for RV models. In RV models the utility of alternative 1 and 2 are respectively given by $V_1 = \mu \cdot \text{BVTT}$ and $V_2 = \mu \cdot \text{VTT}$, where μ is interpreted as a scale parameters. To specify the logistic regression model in this form, we subtract $-\beta_{bvtt}^{T+1} bvtt_{T+1}^{n}$ from v_1^n and v_2^n in Equation 3 (see Equation 6). This is inconsequential because only the difference between v_1^n and v_2^n matters for logit transformations.

$$
\begin{aligned}
v_1^n &= -\beta_{bvtt}^{T+1} bvtt_{T+1}^{n} \\
v_2^n &= \delta + \beta_{y \cdot bvtt} \sum_{t=1}^{T} y_t^n bvtt_t^n + \sum_{r=1}^{R} \beta_r d_r^n + \sum_{q=1}^{Q} \beta_{sq}^{T+1} s_{T+1}^{n}
\end{aligned}
\qquad \text{Equation 6}
$$

Next, we substitute $-\beta_{bvtt}^{T+1} = \mu$, and multiply v_2^n by μ / μ, which yields Equation 7.

$$v_1^n = \mu \cdot bvtt_{T+1}^n$$
$$v_2^n = \mu\left(\frac{\delta}{\mu} + \frac{\beta_{y \cdot bvtt}}{\mu}\sum_{t=1}^{T} y_t^n bvtt_t^n + \sum_{r=1}^{R}\frac{\beta_r}{\mu} d_r^n + \sum_{q=1}^{Q}\frac{\beta_{sq}^{T+1}}{\mu} s_{T+1}^n\right)$$

Equation 7

Finally, we obtain our logistic regression model in RV form, where ~ indicate that the estimates are rescaled (Equation 8).

$$v_1^n = \mu \cdot bvtt_{T+1}^n$$
$$v_2^n = \mu \cdot VTT^n$$
$$where\ VTT^n = \left(\tilde{\delta} + \tilde{\beta}_{y \cdot bvtt}\sum_{t=1}^{T} y_t^n bvtt_t^n + \sum_{r=1}^{R}\tilde{\beta}_r d_r^n + \sum_{q=1}^{Q}\tilde{\beta}_{sq}^{T+1} s_{T+1}^n\right)$$

Equation 8

Looking at Equation 8 we see that the logistic regression model derived from the CK method can be casted in an RV form that looks fairly conventional. Actually, the only unconventional term in the VTT equation is the second term. This term lays bare the fundamental difference between theory and data-driven methods. From a theory-driven perspective, this term is odd as it goes against choice modeller's orthodoxy to use previous choices in utility functions. In a disaggregate model of choice behaviour this term does not make sense, as the utility difference between two alternatives should not depend on the choices made in earlier choice tasks. However, as noted before, we do not conceive the logistic regression model as a disaggregate model of choice behaviour: there is no underlying concept as utility. It tries to *approximate* the underlying choice models. Therefore, from a data-driven perspective it makes good sense to use the previous choices as a cue for the next ones in case they help the analyst to make more accurate predictions (in our case for the individual level VTTs).

For the sake of completeness we estimated the logistic regression model casted in RV form. Table 2 reports the estimation results for this model (Model I), alongside the estimation results of standard RV model (Model II). The latter model is the same as Model I, but with $\beta_{y \cdot bvtt}$ fixed to zero. Note that for clarity of exposition we here ignore the socio-demographic variables. The Pythonbiogeme syntax for Model I is reported in Appendix A. In Table 2 we see that, as expected, the model fit and model estimates of Model I are very similar to those reported in Table 1. The difference in model fits is due to the fact that here we ignored the socio-demographic variables; the difference in estimates across the two tables is mainly due to differences in parametrisation related to the scale (which is inconsequential).

Table 2: RV estimation results

Model	(I) CK method with Logistic regression model				(II) Random Valuation MNL model			
No. observations	5832				5832			
No. parameters	6				5			
Null LogLikelihood	-4042.4				-4042.4			
Final LogLikelihood	-2506.8				-3162.8			
ρ^2	0.38				0.22			
	Est.	SE	t-stat	p-val	Est.	SE	t-stat	p-val
Intercept								
δ	3.510	0.588	5.97	0.00	11.30	0.726	15.58	0.00
$t = 1 \ldots T$								
$\beta_{y \cdot bvtt}$	0.319	0.011	30.16	0.00	0.000	fixed		
$t = T+1$								
mu	0.132	0.004	29.65	0.00	0.085	0.003	26.53	0.00
β_{WTA}	0.000	fixed			0.000	fixed		
β_{EL}	-1.610	0.716	-2.25	0.02	-3.270	0.983	-3.32	0.00
β_{EG}	-2.590	0.739	-3.50	0.00	-4.160	1.020	-4.10	0.00
β_{WTP}	-5.250	0.821	-6.39	0.00	-6.870	1.110	-6.16	0.00

5 Conclusions

In this research note we replaced the ANN in CK's method by a logistic regression model. Thereby, we shed light on how the ANN in this method recovers the VTT distribution. Our results show that the logistic regression model is a fair proxy for the ANN, in the sense that we find a strong correlation in terms of the predicted VTTs by both methods. Although the ANN clearly outperforms the logistic regression model in terms of prediction performance, there are two advantages for using a logistic regression model over an ANN. First, the logistic regression model is tractable, while the ANN is largely a black box. Second, it is relatively easy to apply: the logistic regression can be conducted in software packages choice modellers are familiar with, like Biogeme (Bierlaire 2016) and STATA. Nonetheless, given the decisive role of the VTT in Cost Benefit Analysis, it is worthwhile to invest considerable efforts to obtain an as accurate as possible VTT estimate. Therefore, the enhanced performance provided by the ANN over the logistic regression model can be worth the extra efforts. In light of the black box nature of ANNs, we believe it is good practice when applying CK's method to start with a logistic regression model, and built it into an ANN after having gained understanding on the data and confidence in the method.

Finally, this research note explored the fundamental relation between CK's method with a logistic regression model and an RV model. We find that the key difference between CK's method with a logistic regression model and an RV model can be traced back to one term. This term is the interaction between the previous choices made and the received BVTTs. This term lays bare the fundamental difference between theory and data-driven methods. In a theory driven perspective, the analyst aims to understand behaviour by creating models of disaggregate choice behaviour. In this perspective, this term does not make sense as the utility difference between two alternatives should not depend on the choices made in earlier choice tasks. In data-driven perspective, on the other hand, the analyst aims to best explain the observed choices in the data in order to make as accurate as possible future predictions. In this perspective, this terms makes good sense, i.e. as long as is helps to improve the predictions of the model. Therefore, in the context of VTT research, ultimately the choice between using a theory-driven or data-driven approach boils down to what is considered more important: having an elegant behavioural underpinned model or having more accurate VTT estimates.

Appendix 1

```
# Logistic regression model to uncover the VTT distribution
# Sander van Cranenburgh - Delft University of Technology
from biogeme import *
from headers import *
from loglikelihood import *
from statistics import *

#Parameters to be estimated
mu              = Beta('mu',          0,    -10000, 10000,  0)
delta           = Beta('delta',       0,    -10000, 10000,  0)
Beta_ybvtt      = Beta('Beta_ybvtt',  0,    -10000, 10000,  0)
Beta_wtp        = Beta('Beta_wtp',    0,    -10000, 10000,  0)
Beta_el         = Beta('Beta_el',     0,    -10000, 10000,  0)
Beta_eg         = Beta('Beta_eg',     0,    -10000, 10000,  0)
Beta_wta        = Beta('Beta_wta',    0,    -10000, 10000,  1)

# Rescale parameters such that the estimated VTTs have units [euro per hour]
CostLx = DefineVariable('CostLx', CostL / 9)    # Convert from NOK to euro
CostRx = DefineVariable('CostRx', CostR / 9)    # Convert from NOK to euro
TimeLx = DefineVariable('TimeLx', TimeL / 60 )  # Convert from minutes to hours
TimeRx = DefineVariable('TimeRx', TimeR / 60 )  # Convert from minutes to hours

# Reorder alternatives such that Alt1 is the cheap and slow alternative and Alt2 is the fast and expensive alternative
Cost1 = min(CostLx,CostRx)
Time1 = max(TimeLx,TimeRx)
Cost2 = max(CostLx,CostRx)
Time2 = min(TimeLx,TimeRx)

# Compute boundary Value-of-Travel-Time
BVTT = -((Cost1 - Cost2)/(Time1 - Time2))

# Construct a new choice vector, given the re-ordering of the alternatives
YRV  = 1 + (CostL > CostR ) * (Chosen==1) + (CostR > CostL ) * (Chosen==2)

# Value-of-Travel-Time specification
VTT =   delta + Beta_ybvtt * sumYBVTT + Beta_wtp * (Quadrant == 1) + Beta_eg * (Quadrant == 2) + Beta_el * (Quadrant == 3) + Beta_wta * (Quadrant == 4)

# Random Valuation model
V1 = mu * BVTT
V2 = mu * VTT
```

```
# Associate utility functions with the numbering of alternatives
V = {1: V1,
     2: V2}

# Associate the availability conditions with the alternatives
av = {1: 1,
      2: 1}

# The choice model is a logit, with availability conditions
Prob = bioLogit(V,av,YRV)

# We only want to predict the (T+1)th choice task. The T+1)th choice task is indicated by the vector depTask in the data file
BIOGEME_OBJECT.EXCLUDE = (depTask == 0)

# For each item of personIter, iterates on the rows of the group.
rowIterator('obsIter')

# Likelihood function
BIOGEME_OBJECT.ESTIMATE = Sum(log(Prob),'obsIter')

# Statistics
BIOGEME_OBJECT.PARAMETERS['optimizationAlgorithm'] = "BIO"
BIOGEME_OBJECT.PARAMETERS['numberOfThreads'] = "8"
```

References

Bierlaire, M. (2016). PythonBiogeme: A Short Introduction, Technical Report TRANSP-OR 160706.

Bishop, C. M. (1995). *Neural networks for pattern recognition*: Oxford university press).

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, *538(7623)*, 20.

Fosgerau, M. & Bierlaire, M. (2007). A practical test for the choice of mixing distribution in discrete choice models. *Transportation Research Part B: Methodological*, *41(7)*, 784-794.

Paliwal, M. & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, *36(1)*, 2-17.

Rouwendal, J., de Blaeij, A., Rietveld, P. & Verhoef, E. (2010). The information content of a stated choice experiment: A new method and its application to the value of a statistical life. *Transportation Research Part B: Methodological*, *44(1)*, 136-151.

Sussmann, H. J. (1992). Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks*, *5(4)*, 589-593.

Van Cranenburgh, S. & Kouwenhoven, M. (under review). An Artificial Neural Network based method to uncover Value-of-Travel-Time.