Using Artificial Neural Networks for Recovering the Value-of-Travel-Time Distribution

van Cranenburgh, Sander; Kouwenhoven, Marco

DOI
10.1007/978-3-030-20521-8_8

Publication date
2019

Document Version
Final published version

Published in
Advances in Computational Intelligence - 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Proceedings

Using Artificial Neural Networks for Recovering the Value-of-Travel-Time Distribution

Sander van Cranenburgh[1(✉)] and Marco Kouwenhoven[1,2]

[1] Delft University of Technology, Jaffalaan 5, 2286 BX Delft, The Netherlands
{s.vancranenburgh,m.l.a.kouwenhoven}@tudelft.nl
[2] Significance, Grote Marktstraat 47, 2511 BH Den Haag, The Netherlands

Abstract. The Value-of-Travel-Time (VTT) expresses travel time gains into monetary benefits. In the field of transport, this measure plays a decisive role in the Cost-Benefit Analyses of transport policies and infrastructure projects as well as in travel demand modelling. Traditionally, theory-driven discrete choice models are used to infer the VTT distribution from choice data. This study proposes an alternative data–driven method to infer the VTT distribution based on Artificial Neural Networks (ANNs). The strength of the proposed method is that it is possible to uncover the VTT distribution (and its moments) without making strong assumptions about the shape of the distribution or the error terms, while being able to incorporate covariates and account for panel effects. We apply our method to data from the 2009 Norwegian VTT study. Finally, we cross-validate our method by comparing it with a series of state-of-the-art discrete choice models and other nonparametric methods used in the VTT literature. Based on the very encouraging results we have obtained, we believe that there is a place for ANN-based methods in future VTT studies.

Keywords: Artificial Neural Network · Value of Travel Time · Random Valuation · Nonparametric methods · Discrete choice modelling

1 Introduction

The Value-of-Travel Time (VTT) expresses travel time gains into monetary benefits [1] and plays a decisive role in the Cost-Benefit Analyses (CBA) of transport policies and infrastructure projects as well as in travel demand modelling. Not surprisingly in this regard, the VTT is one of the most researched notions in transport economics [2]. Most Western societies conduct studies to determine VTTs on a regular basis. But, despite decades of experience with data collection and VTT inference, the best way to obtain the VTT is still under debate. Early studies predominantly used Revealed Preference (RP) data in combination with Multinomial Logit (MNL) models [3]. However, despite the well-known advantages of RP data over data collected via Stated Choice (SC) experiments, nowadays RP data are seldom used for VTT studies. The main reason is that while the travellers' choices are observable (in a real-life setting), their actual trade-offs across alternatives are not – which hampers estimation of the VTT using RP data. More recent VTT studies therefore favour using SC data in combination with sophisticated

I. Rojas et al. (Eds.): IWANN 2019, LNCS 11506, pp. 88–102, 2019.
https://doi.org/10.1007/978-3-030-20521-8_8

discrete choice models that account for (some of the) potential artefacts of SC experiments (notably so-called size and sign effects) [4–8].

Besides discrete choice models, nowadays nonparametric methods are increasingly used in VTT studies [9, 10]. These methods are methodologically appealing as they do not make assumptions regarding the shape of the VTT distribution and the structure of the error terms. However, despite their methodological elegance they are typically not used to derive VTTs for appraisal. Rather, they are used as a first, complementary, step to learn about the shape of the distribution of the VTT, after which parametric discrete choice models are estimated to derive VTTs for appraisal. Börjesson and Eliasson [4] argue that nonparametric methods are not suitable to compute VTTs for appraisal for three reasons. First, they (often) cannot incorporate covariates. Second, they (often) cannot account for panel effects. Third, they do not recover the VTT distribution over its entire domain. That is, the distribution right of the highest VTT bid is not recovered, which hinders computation of the mean VTT.

Very recently, Artificial Neural Networks (ANNs) are gaining ground in the travel behaviour research arena [e.g. 11, 12–20]. A fundamental difference between discrete choice models and ANNs is the modelling paradigm to which they belong. Discrete choice models are theory-driven, while ANNs are data-driven. Theory-driven models work from the principle that the true Data Generating Process (DGP) is a (stochastic) function, which can be uncovered. To do so, the analyst imposes structure on the model. In the context of discrete choice models this is done by prescribing the utility function, the decision rule, the error term structure, etc. Then, the analyst estimates the model's parameters, usually compares competing models, and interprets the results. A drawback of this approach is that it heavily relies on potentially erroneous assumptions regarding choice behaviour, i.e. the assumptions may not accurately describe the true underlying DGP – potentially leading to erroneous inferences. Data-driven methods work from the principle that the true underlying process is complex and inherently unknown. In a data-driven modelling paradigm the aim is not to uncover the DGP, but rather to learn a function that accurately approximates the underlying DGP. The typical outcome in a data-driven modelling paradigm is a network which has very good prediction performance [18]. A major drawback of data-driven methods is that – without further intervention – they provide very limited (behavioural) insights on the underlying DGP, such as the relative importance of attributes, Willingness-to-Pay, or VTT. Yet, these behavioural insights are typically most valuable to travel behaviour researchers and for transport policy-making.

There is a general sense that ANNs (and other data-driven models), could complement existing (predominantly) theory-driven research efforts. In light of that spirit, this paper develops an ANN-based method to investigate the VTT distribution. Specifically, we develop a novel pattern recognition ANN which is able to estimate travellers' individual underlying VTTs. Our method capitalises on the strong prediction performance of ANNs (see [21] for a comprehensive review of articles that involve a comparative study of ANNs and statistical techniques). The strength of this method is that it is possible to uncover the VTT distribution (and its moments) without making strong assumptions on the underlying behaviour. For instance, it does not prescribe the

utility function, the shape of the VTT distribution, or the structure of the error terms. Moreover, the method can incorporate covariates, account for panel effects and does yield a distribution right of the maximum VTT bid. Thereby, it overcomes important limitations associated with other nonparametric methods. As such, this method can be used to derive VTTs for appraisal. Finally, the method does not require extensive software coding on the side of the analyst as the method is built on a standard MultiLayer Perceptron (MLP) architecture. Hence, the method can be applied using off-the-shelf (open-source) software.

The remainder of this paper is organised as follows. Section 2 develops the ANN-based method for uncovering the VTT distribution. Section 3 applies the method to an empirical VTT data set from a recent VTT study. Section 4 cross-validates the method by comparing its results with those obtained using a series of state-of-the-art discrete choice models and other nonparametric methods. Finally, Sect. 5 draws conclusions and provides a discussion.

2 Methodology

2.1 Preliminary

Panel Data Format (time series)
Throughout this paper we suppose that we deal with data from a classic binary SC experiment, consisting of $T + 1$ choice tasks per individual, in which within-mode trade-offs between travel cost TC and travel time TT are embedded. This data format is in line with standard VTT practice in many Western European countries, including the UK [22], The Netherlands [7, 23], Denmark [8], Norway [5] and Sweden [4]. Figure 1 shows a choice task from such a SC experiment. Choice tasks are pivoted around the respondents current travel time and travel cost, which are typically elicited prior to the SC experiment. In the SC experiment respondents are confronted with $T + 1$ choice tasks consisting of two alternatives, in each choice task one trip being their current one and the other one being either a faster and more expensive, or a slower and cheaper trip. In each choice task there is an implicit price of time which is commonly referred to as the Boundary VTT (BVTT). The BVTT is defined as:

$$BVTT = -\frac{\Delta TC}{\Delta TT} = \frac{-(TC_2 - TC_1)}{(TT_2 - TT_1)} \tag{1}$$

where alternative 1 denotes the fast and expensive alternative and the alternative 2 denotes the slow and cheap alternative. The BVTT can be perceived as a valuation threshold as a respondent choosing the fast and expensive alternative signals a VTT which is (most likely) above the BVTT, while a respondent choosing the slow and cheap alternative signals a VTT which is (most likely) below the BVTT.

Trip A	Trip B
Travel time: TT	Travel time: TT - ΔTT
Travel cost: TC	Travel cost: TC + ΔTC
Which trip do you prefer?	
□ Trip A	□ Trip B

Fig. 1. Example choice task

Covariates in VTT Studies

It is important to incorporate covariates in models that aim to infer the VTT. Börjesson and Eliasson [4] provide four reasons for this. Firstly, accounting for covariates in VTT models allows better extrapolating the VTT to new situations. Secondly, accounting for covariates in VTT models allows better understanding what trip characteristics influence the VTT. Thirdly, accounting for covariates in VTT models allows the analyst to remove the influence of undesirable factors, such as income or urbanisation level from the VTT used for appraisal. Fourthly, accounting for covariates in VTT models allows accounting for so-called size and sign effect stemming from the experimental design [24]. Size effects are due to the behavioural notion that the VTT is dependent on the size of the difference in travel time and travel cost across alternatives in the choice task [25]. Sign effects are due to the behavioural notion that losses (e.g. higher travel cost and longer travel time) loom larger than equivalently sized gains (e.g. lower travel cost and shorter travel time) [24, 26].

2.2 Uncovering Individual VTTs Using ANNs

The ANN-based method is based on three observations. The first observation is that ANNs are very good at making predictions [21]. Their good prediction performance stems from the versatile structure of ANNs, which allow them to capture non-linearity, interactions between variables, and other peculiarities in the DGP, for instance in this context relating to the set-up of the experimental design. The second observation is that we can use the ANN to find the BVTT where it is maximally uncertain on the choice of the decision maker. The third observation is that we can give a behavioural interpretation to this BVTT and recover the individual's VTT from it. That is, we can interpret the BVTT where the ANN is maximally uncertain as the point where the individual is indifferent between choosing the fast and expensive alternative and the slow and cheap alternative. From this behavioural perspective, this BVTT reflects the VTT of the individual. Taking these three observations together, we can develop an ANN-based method that recovers individual level VTTs and can be used to derive VTTs for appraisal.

To do so, we take the following 5 steps:

(1) **Data preparation and training**

The aim of this step is to train an ANN to (probabilistically) predict, for decision maker n the choice in the hold-out choice task $T + 1$, based on the BVTTs ($BVTT^n$) and

the choices made (Y^n) in choice tasks 1 to T, the probed BVTT in choice task $T + 1$ ($bvtt^n_{T+1}$), experimental covariates in choice task $T + 1$ (s^n_{T+1}), and a set of generic and experimental covariates, denoted D^n and S^n, respectively. In other words, we train the ANN to learn the relationships f, see (2, where P^n_{T+1} denotes the probability of observing a choice for the fast and expensive alternative in choice task $T + 1$ for decision maker n.

$$P^n_{T+1} = f\left(BVTT^n, Y^n, bvtt^n_{T+1}, s^n_{T+1}, D^n, S^n\right) \tag{2}$$

$$\begin{aligned} where\, BVTT^n &= \left\{bvtt^n_1, bvtt^n_2, \ldots, bvtt^n_T\right\} \\ Y^n &= \left\{y^n_1, y^n_2, \ldots, y^n_T\right\} \\ S^n &= \left\{s^n_1, s^n_2, \ldots, s^n_T\right\} \end{aligned}$$

Figure 2 shows the proposed architecture of the ANN. At the input layer, the independent variables enter the network. At the top, there are the generic covariates (green). Typical generic covariates encountered in VTT studies are mode, purpose, age, income, distance, etc. Below the generic covariates are the variables associated with choice tasks 1 to T (red). These include the BVTTs, the choices y and experimental covariates s (e.g. sizes and signs). Below the variables for choice tasks 1 to T is an extra set of input nodes for choice task R (blue). Choice task R is a replication of one choice task, randomly picked from the set choice tasks 1 to T. These input nodes come in handy later when the ANN is used for simulation (they make it possible to use all $T + 1$ observations instead of only T observations in the simulation). Finally, at the bottom are the variables associated with hold-out choice task $T + 1$ (yellow). These are essentially the 'knobs' of the model that can be used for simulation. The output layer consists of the dependent variable, which is the probability for choosing the fast and expensive alternative in choice task $T + 1$. One or multiple hidden layers can be used. In our analyses we find two layers to work well. However, the optimal number of hidden layer and the number of nodes depends on the complexity of the DGP that needs to be learned from the data, and hence may vary across applications.

To train the network in Fig. 2, we need to prepare the data. To do so, for each decision maker in the data we randomly draw T explanatory choice tasks from the $T + 1$ choice tasks that are available in the data for each decision maker. These T choice tasks are used as independent variables to predict the remaining choice. To avoid that the network undesirably learns a particular structure in the data, rather than the explanatory power of the variables it is crucial that the order in the set of T explanatory choice tasks is randomised.[1] We randomise the order in the set of explanatory choice tasks K times, each time creating a 'new' observation. The idea behind this is that the weights associated with the choice tasks attain (roughly) similar sizes. By doing so, we create a network that produces stable predictions, which is insensitive to the order of the explanatory choice tasks. In each manifestation of the

[1] Unless the order of the choice tasks is randomised during the data collection. Note that by doing so the network becomes blind to potential learning effects on the side of the respondent when conducting the survey. We come back to this point in the discussion.

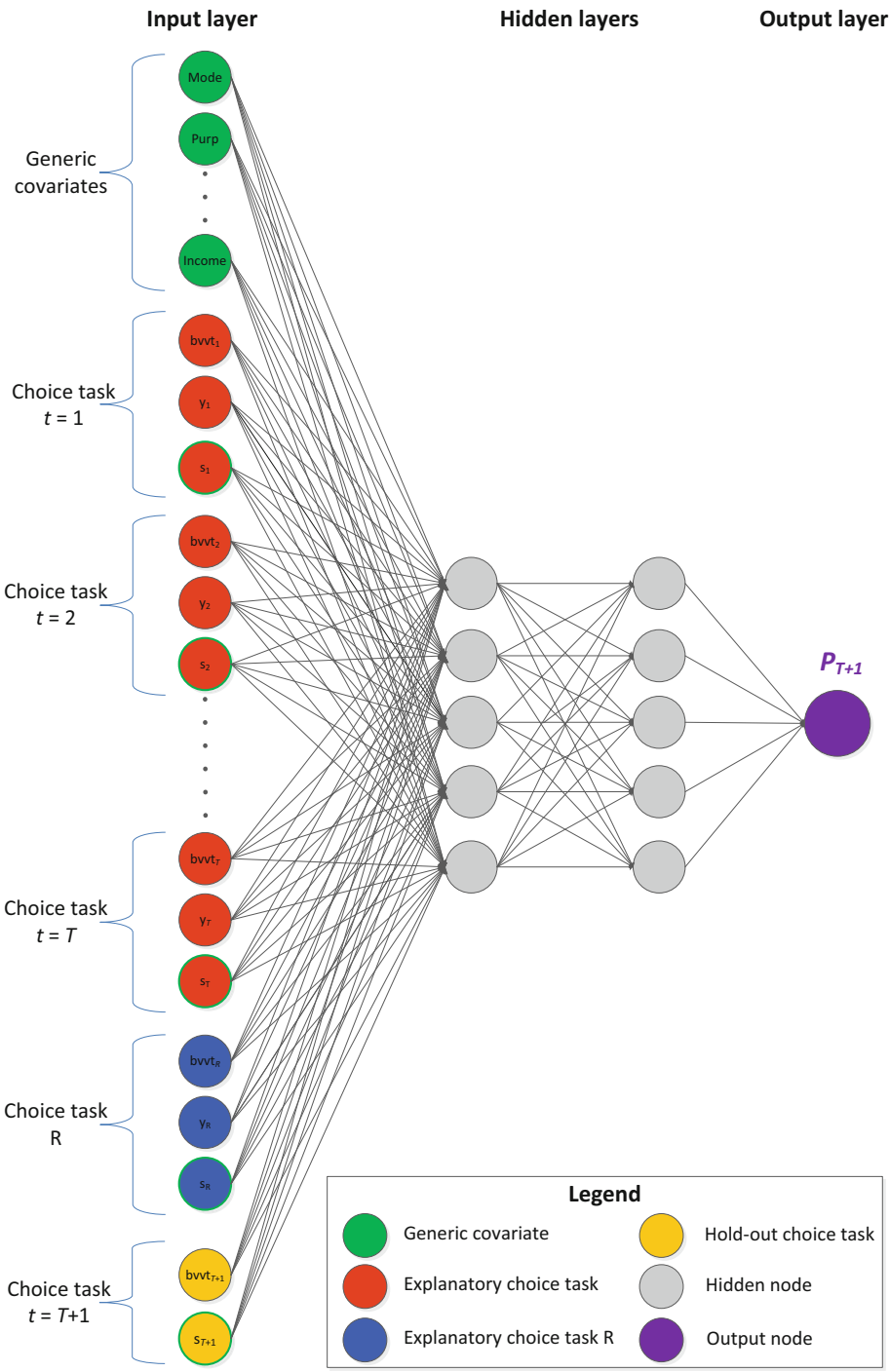

Fig. 2. ANN architecture (Color figure online)

K randomisations, choice task R is a randomly selected replication of one of the T explanatory choice tasks. By selecting a random choice task, we make sure that no single choice task weights more heavily in the training and ensure that the weights of the network are generic across all choice tasks.

(2) **Simulate**

After having trained the ANN, we can use the network to simulate choice probabilities in order to find the point where the ANN is maximally uncertain, and hence the decision maker is indifferent between choosing the fast alternative and the cheap alternative. Specifically, we simulate P^n_{T+1} while letting $bvvt^n_{T+1}$ run from 0 to a maximum BVTT value set by the analyst using a finite step size.[2] For simulation, we can use all $T + 1$ choice observations of a decision maker as explanatory choice tasks. This is possible because we created the extra choice task R in the network (see step 1). Thus, this 'trick' allows using all available information on a decision maker's preference for predicting his or her response to a given probed BVTT in the simulation in an elegant way. Moreover, it circumvents the need to randomly draw T explanatory choice tasks from the $T + 1$ available choice tasks – which would lead to increased variance in the predictions for P^n_{T+1}.

(3) **Recovery of individual VTTs**

Figure 3 illustrates how the simulated choice probabilities (y-axis) for an individual decision maker could look like as a function of $bvvt^n_{T+1}$ (x-axis).[3] The next step is to infer from these simulated probabilities for each decision maker his or her VTT. To do so, we need to find the BVTT which makes the ANN maximally uncertain, which from a behavioural perspective we interpret as the point where the decision maker is indifferent between the fast and expensive and the slow and cheap alternative. In our binary choice context, technically this is where the choice probabilities are equal to 0.5. Since the learned function f cannot easily be solved analytically, we have to determine this point numerically. Several options are available to do so. A simple and effective approach is to first determine the last $bvvt^n_{T+1}$ above $P = 0.5$ and the first $bvvt^n_{T+1}$ below $P = 0.5$, and then make a linear interpolation between those two points and to solve for the BVTT which makes the individual indifferent.

(4) **Repeat steps 2 and 3**

We repeat steps 2 and 3 numerous times. In each repetition we shuffle the order of the $T + 1$ explanatory choice tasks. This step is not strictly obligatory, but it helps to improve the stability of the outcomes. In particular, it is helpful to take out the effect of the order in which the explanatory choice tasks are presented to the network. Hence, for each decision maker his/her VTT is computed numerous times. After that, we compute each decision maker's VTT by taking the mean across all repetitions.

[2] Note that technically it is not necessary to simulate any further than the point where $P < 0.5$.

[3] Note that the plot is deliberately made a bit quivering to highlight the notion that very little structure is imposed by the ANN on the functional form.

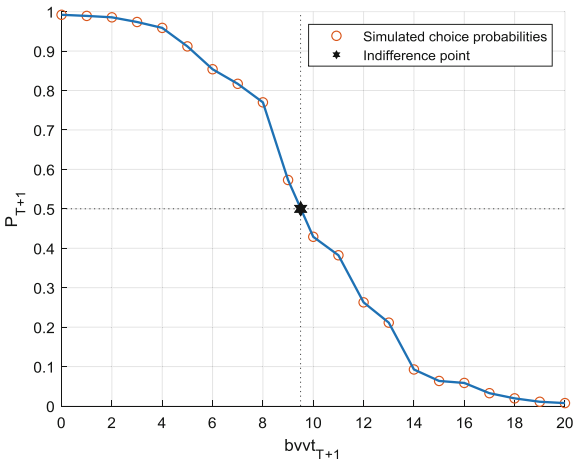

Fig. 3. Simulated choice probabilities

(5) **Construct the VTT distribution**

Having an estimate of the VTT for each decision maker, we can construct an empirical distribution of the VTT. Also, from the constructed empirical distribution we can readily compute the mean and standard deviation of the VTT.

2.3 ANN Development

In Sect. 2.2 we presented the ANN without going into much detail on its architecture or on underlying design choices. In this subsection we discuss these in more detail. To develop an ANN capable of learning function (2), we have tested numerous different architectures, including fully and semi-connected networks, different numbers of hidden layers, the presence or absence of bias nodes, and we have tried several different activation functions. The two-hidden layer architecture presented in Fig. 2 with ten nodes at each hidden layer is found to work particularly well for our data.[4] The proposed architecture is a so-called Multilayer Perceptron (MLP). This is one of the most widely used ANNs architectures and is implemented in virtually all off-the-shelf machine learning software packages. For the activation functions in the network we find good results using a softmax function both at the nodes of the hidden layers as well as at the nodes of the output layer. Using a softmax function at the output layer ensures that the sum of the predicted choice probabilities across the two alternatives add up to 1. Finally, note that while no bias nodes are depicted in Fig. 2 bias nodes are present as they are found to improve the classification performance.

The fact that off-the-shelf software can be used is a desirable feature of this method, as it makes the method accessible for a wide research community. Admittedly, from a methodological perspective our network consumes more weights than is strictly

[4] The network consumes 491 weights in total.

needed, in the sense that in the input layer there are $T + 1$ weights for *bvtt*, *y* and *s*, while just one set of weights to be used across all the $T + 1$ choice tasks would suffice and hence would yield a more parsimonious network. However, while it is possible to create an architecture with shared weights across inputs variables, this would substantially hinder other researchers from using this method as most off-the-shelf software does not allow weight sharing, meaning that the analyst needs to write customised codes.

3 Application to Real VTT Data

3.1 Training and Simulation

In this study we use the Norwegian 2009 VTT data set, see [5] for details on the experimental design and the data collection. After cleaning, this data set consists of 5832 valid respondents. For each respondent, 9 binary choices are observed. While the currency in the SC experiment was Norwegian Kronor, for reasons of communication we converted all costs into euros. To train the network on these empirical data, 70% of the data were used for training, 15% for validation and 15% for testing. The observations were randomly allocated to these subsets. We use $K = 20$ randomisations (see Sect. 2.2). The trained ANN acquires a cross-entropy of 0.36. Table 1 shows the confusion plot. It shows that overall about 85% of the choices are correctly predicted (based on highest probability). To obtain the VTT distribution, we use the network to simulate choice probabilities and search for the BVTTs where the ANN is maximally uncertain. We do this 20 times[5] for each respondent (i.e., steps 2 to 4, see Sect. 2.2).

Table 1. Confusion plot (based on validation and test data)

	Target 1 (fast and expensive)	Target 2 (slow and cheap)	Σ
Output class 1 (fast and expensive)	**26.7%**	6.9%	79.4% *(Positive predictive value)*
Output class 2 (slow and cheap)	8.3%	**58.1%**	87.5% *(Negative predictive value)*
Σ	76.3% *(Sensitivity)*	89.4% *(Specificity)*	**84.8%** *(Overall accuracy)*

3.2 Results

Figure 4 shows the recovered distribution of the VTT (step 5). For eight respondents, it has not been possible to obtain a VTT estimate. For these respondents, the ANN predicts choice probabilities below 0.5, even for very small BVTTs, suggesting a zero or even a negative VTT. While this seems behaviourally unrealistic, from a data-driven

[5] We find that after 20 times the results are stable.

modelling perspective it can be well understood why it is not possible to obtain a VTT for each and every respondent, especially considering that over 13% of the respondents in the data always choose the slow and cheap alternative. The ANN may have learned that some respondents just never choose the fast and expensive alternative, even if it is just a fraction more expensive than the slow and cheap alternative. Close inspection of the eight respondents for which we have not been able to obtain a VTT estimate, shows that indeed these respondents never chose the expensive and fast alternative and that they all had low income levels. In the remainder of our analyses these eight respondents are given a VTT of zero. About 2% of the respondents always chose the fast and expensive alternative in each choice task. For all these respondents a VTT has been recovered, in between €20 and €123 per hour, with a median VTT of €85 per hour.

Figure 4 shows that the shape of the VTT distribution is positively skewed. The lognormal-like shape is behaviourally intuitive and has occasionally been found in previous VTT studies. However, when fitting the lognormal distribution onto these data (not shown), we find that it does not fit the data very well: in particular, it cannot accommodate for the spike at around VTT = €2/h and the drop at VTT = €16/h. Close inspection of the bins around VTT = €2/h reveal that they are predominantly populated with those respondents that always choose the slow and cheap alternative (for clarity, non-traders are depicted in red in the right-hand side plot). The bimodal shape of this distribution essentially emphasises the need for flexible methods to uncover the distribution of the VTT.

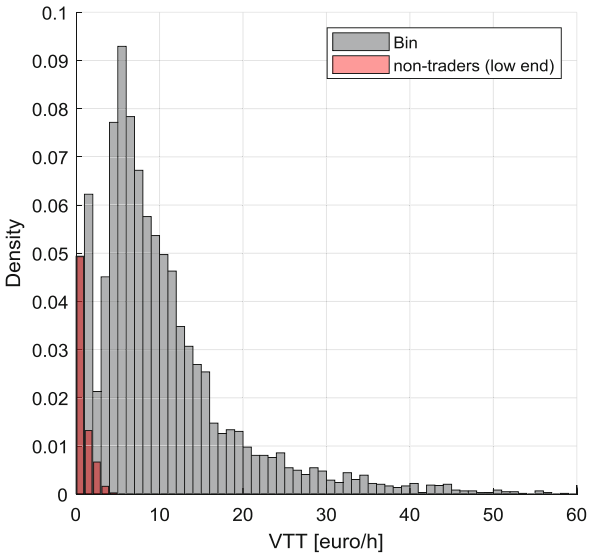

Fig. 4. VTT distribution

4 Cross-validation

To cross-validate the shape and mean of the recovered VTT distribution by the ANN-based method, we compare with state-of-the-art (parametric) choice models as well as with three (semi) nonparametric methods that have been used in recent VTT studies. The parametric models that we use in this cross-validation study are Random Valuation (RV) models [27, 28], with two types of distributions, namely the lognormal and the log-uniform distributions. The log-normal distribution has been used in the most recent Swedish VTT study; the log-uniform has been used in the most recent UK VTT study. Note that we also have estimated more conventional Random Utility Maximisation (RUM) models [29], but the RV models are found to outperform their random utility counterparts [30]. Therefore, we report only on the RV models. Regarding the non-parametric methods, the first nonparametric method that we consider is called local-logit. This method is developed by Fan, Heckman and Wand [31], pioneered in the VTT research literature by Fosgerau [10] and further extended by [32]. The local-logit method essentially involves estimation of logit models at 'each' value of the BVTT using a kernel with some shape and bandwidth. In our application we use a triangular shaped kernel with a bandwidth of 10 euro. The second nonparametric method is developed by Rouwendal, de Blaeij, Rietveld and Verhoef [33]. Henceforth, we refer this method as 'The Rouwendal method'. This method assumes that everybody has a unique VTT and makes consistent choices accordingly. But, at each choice there is a fixed probability that the decision maker makes a mistake and hence chooses the alternative that is inconsistent with his/her VTT. The third nonparametric method is put forward by Fosgerau and Bierlaire [34]. This is actually a semi-nonparametric method which approximates the VTT distribution using series approximations. We apply the method – which we henceforth refer to as 'SNP' – to the RV model that we also used in the parametric case.

The left-hand side plot in Fig. 5 shows the Cumulative Density Function (CDF) of the VTT recovered using the ANN-based method (blue) and the parametric RV models. The right-hand side plot in Fig. 5 shows, besides the CDF of the ANN VTT (blue), the CDFs created using the local-logit (orange), the Rouwendal method (green) and the SNP method (turquoise). A number of findings emerge from Fig. 5. A first general observation is that all methods roughly recover the same shape of the VTT distribution, except for the local-logit. But, there are non-trivial differences between the shapes too. Looking at the parametric methods, we see that between VTT = €3/h and VTT = €10/h, the VTT distribution recovered by the ANN is shifted by about 2 euros to the left. Furthermore, we see that in the tail the CDFs of the ANN and of the lognormal neatly coincide (but they do not before). The tail of the log-uniform seems to be substantially underestimated, at least as compared to the CDFs recovered using the other methods. Looking at the nonparametric methods, we see that the CDF of the Rouwendal method coincides with that of the ANN very well, especially up until VTT = €14/h and in the tail above €55/h. The CDF of the SNP method coincides well with that of the ANN for VTTs of €5/h and higher. The local-logit CDF deviates most from the other CDFs, in particular below VTT = €30/h. Possibly, this is caused by its inability to account for the panel nature of the data and its inability to disentangle

unobserved heterogeneity from irreducible noise in the data. After all, the local-logit method only considers choices from several respondents around the same BVTT, without considering the other choices made by these (or other) respondents.

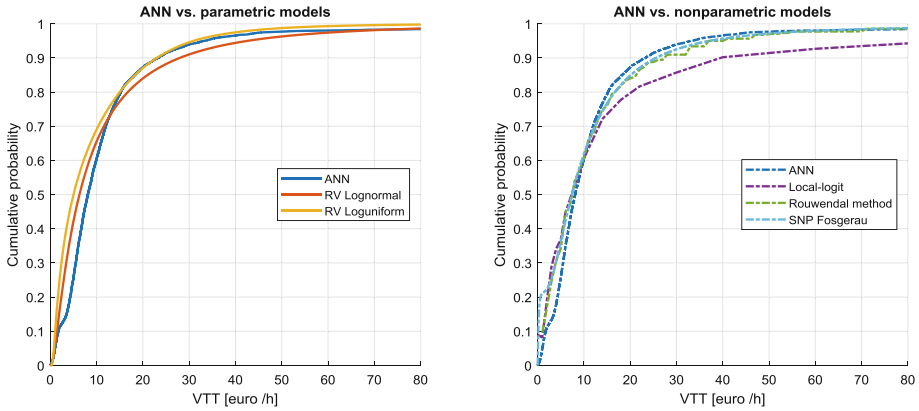

Fig. 5. Cross-validation of shape (Color figure online)

Table 2 summarises key statistics of the recovered VTT distributions for the methods that we have used. Since the nonparametric methods do not recover the VTT distribution beyond the maximum VTT bid, the presented statistics for these methods can be considered as lower bounds. However, in these data the maximum VTT is set very high (see [5]) and only 2% of the respondents in these data always choose for the fast and expensive alternative, many of whom did not receive a very high VTT bid. As such, the unrecovered tail for the Rouwendal method is very small, representing less than 0.05% of the density. But, the unrecovered tail for the local-logit still represents about 5% of the density (rendering computation of its moments unreliable). The statics for the SNP method are computed from the CDF. In line with previous practice using this method, we censored the right-hand side tail above VTT = €200/h. Not doing so, would substantially inflate the recovered standard deviation of this distribution. The overview shows that the mean recovered by the ANN-based method is close to those of all other methods, except the RV-log-uniform. The median VTT recovered by the ANN is higher than those of the parametric methods. This is presumably due to the limited flexibility of the latter methods to account for the substantial number of respondents having a very low VTT (13% of the respondents always choose the slow and cheap alternative), while still covering the VTT distribution over a large range. Altogether, it can be concluded that the shape, mean and median recovered by the ANN seem very plausible.

Table 2. Mean, median and standard deviations of recovered VTT distributions

	ANN	RV Lognormal	RV Log-uniform	Rouwendal method	Local-logit	SNP[a]
Mean	11.75	12.13	9.34	12.45	12.16	12.34
Median	8.09	6.30	5.01	7.74	7.33	7.40
Std deviation	13.68	17.57	11.41	15.54	15.24	15.64

[a]Censored at VTT = €200/h

5 Conclusions and Discussion

This study proposes a novel ANN-based method to study the VTT. The method is highly flexible, in the sense that it does not impose strong assumptions regarding the specification of the utility function, the VTT distribution, or the structure of the error terms. Moreover, the method can incorporate covariates, account for panel effects and does yield a distribution right of the maximum VTT. Thereby, it overcomes important limitations associated with nonparametric methods that are put forward in the VTT literature. In this study we have cross-validated the proposed method by comparing it with a series of state-of-the-art discrete choice models and nonparametric methods. Based on the encouraging results of this study, we believe that there is a place for ANN-based methods in future VTT studies.

The method proposed in this study provides ample scope for further research. A first direction for further research involves acquiring a good understanding regarding the data requirements for this method to work well. For instance, how many respondents are at least needed for this method? A commonly used rule-of-thumb in the Machine Learning field is that the number of observations needs to be at least ten times more than the number of estimable weights. However, a recent study on this topic in the context of choice data suggest a more conservative factor of 50 times more observations than weights [35]. Likewise, what is the 'minimum' number of choice tasks per respondents that is needed? In our study we found good results with nine choice tasks per respondents. But, will the method also work with just five choice tasks per respondent, or will it work even better with fifteen choice tasks? A second, related, direction for further research concerns the design of the SC experiment. Current SC experiments are optimised for estimation of discrete choice models. However, data from these experiments may actually be suboptimal for the ANN-based method. A question that remains to be answered therefore is how to design experiments optimised for this method? A third direction for further research concerns the generalisation of the method to work with choice tasks having three or more attributes. While it is clear that it becomes more difficult to recover a VTT from a choice task consisting of three or more alternatives using this method, there are – as far as we can see tell – no fundamental reasons why the method would be confined to data from two-attribute experiments only. A fourth interesting direction is investigating whether it is possible to also capture and incorporate learning and ordering effects. Some empirical studies suggest that respondents are subject to learning effects and ordering anomalies [36].

A fifth research direction for further research is application of this method to other VTT data sets, as well as applying the method to other areas of application, such as inference of the distribution of the value of reliability. Finally, a drawback of the ANN-based method relates to the opaque nature of ANNs: they cannot easily be diagnosed, e.g. by looking at its weights. Future research may be directed to illuminate the black boxes of ANNs, especially in contexts where they are used for behavioural analysis [37].

References

1. Small, K.A.: Valuation of travel time. Econ. Transp. **1**, 2–14 (2012)
2. Abrantes, P.A.L., Wardman, M.R.: Meta-analysis of UK values of travel time: an update. Transp. Res. Part A Policy Pract. **45**, 1–17 (2011)
3. Wardman, M., Chintakayala, V.P.K., de Jong, G.: Values of travel time in Europe: review and meta-analysis. Transp. Res. Part A Policy Pract. **94**, 93–111 (2016)
4. Börjesson, M., Eliasson, J.: Experiences from the Swedish value of time study. Transp. Res. Part A Policy Pract. **59**, 144–158 (2014)
5. Ramjerdi, F., Flügel, S., Samstad, H., Killi, M.: Value of time, safety and environment in passenger transport–Time. TØI report 1053-B/2010. Institute of Transport Economics (TØI) (2010)
6. Hess, S., Daly, A., Dekker, T., Cabral, M.O., Batley, R.: A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. Transp. Res. Part B Methodol. **96**, 126–149 (2017)
7. Kouwenhoven, M., et al.: New values of time and reliability in passenger transport in The Netherlands. Res. Transp. Econ. **47**, 37–49 (2014)
8. Fosgerau, M., Hjorth, K., Lyk-Jensen, S.V.: The Danish value of time study: Final Report (2007)
9. Fosgerau, M.: Investigating the distribution of the value of travel time savings. Transp. Res. Part B Methodol. **40**, 688–707 (2006)
10. Fosgerau, M.: Using nonparametrics to specify a model to measure the value of travel time. Transp. Res. Part A Policy Pract. **41**, 842–856 (2007)
11. Alwosheel, A., Van Cranenburgh, S., Chorus, C.G.: Artificial neural networks as a means to accommodate decision rules in choice models. ICMC2017, Cape Town (2017)
12. Mohammadian, A., Miller, E.: Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. Transp. Res. Rec.: J. Transp. Res. Board **1807**, 92–100 (2002)
13. Omrani, H., Charif, O., Gerber, P., Awasthi, A., Trigano, P.: Prediction of individual travel mode with evidential neural network model. Transp. Res. Rec.: J. Transp. Res. Board **2399**, 1–8 (2013)
14. Wong, M., Farooq, B., Bilodeau, G.-A.: Discriminative conditional restricted Boltzmann machine for discrete choice and latent variable modelling. J. Choice Model. **29**, 152–168 (2017)
15. Sifringer, B., Lurkin, V., Alahi, A.: Enhancing discrete choice models with neural networks. In: hEART 2018–7th Symposium of the European Association for Research in Transportation Conference (2018)
16. Cantarella, G.E., de Luca, S.: Multilayer feedforward networks for transportation mode choice analysis: an analysis and a comparison with random utility models. Transp. Res. Part C Emerg. Technol. **13**, 121–155 (2005)

17. Golshani, N., Shabanpour, R., Mahmoudifard, S.M., Derrible, S., Mohammadian, A.: Modeling travel mode and timing decisions: comparison of artificial neural networks and copula-based joint model. Travel. Behav. Soc. **10**, 21–32 (2018)
18. Karlaftis, M.G., Vlahogianni, E.I.: Statistical methods versus neural networks in transportation research: differences, similarities and some insights. Transp. Res. Part C Emerg. Technol. **19**, 387–399 (2011)
19. Lee, D., Derrible, S., Pereira, F.C.: Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling. Transp. Res. Rec. **2672**, 101–112 (2018). 0361198118796971
20. Van Cranenburgh, S., Alwosheel, A.: An artificial neural network based approach to investigate travellers' decision rules. Transp. Res. Part C Emerg. Technol. **98**, 152–166 (2019)
21. Paliwal, M., Kumar, U.A.: Neural networks and statistical techniques: a review of applications. Expert Syst. Appl. **36**, 2–17 (2009)
22. Batley, R., et al.: New appraisal values of travel time saving and reliability in Great Britain. Transportation, 1–39 (2017). https://link.springer.com/article/10.1007/s11116-017-9798-7
23. HCG: The second Netherlands' value of time study - final report (1998)
24. De Borger, B., Fosgerau, M.: The trade-off between money and travel time: a test of the theory of reference-dependent preferences. J. Urban Econ. **64**, 101–115 (2008)
25. Daly, A., Tsang, F., Rohr, C.: The value of small time savings for non-business travel. J. Transp. Econ. Policy (JTEP) **48**, 205–218 (2014)
26. Ramjerdi, F., Lindqvist Dillén, J.: Gap between willingness-to-pay (WTP) and willingness-to-accept (WTA) measures of value of travel time: evidence from Norway and Sweden. Transp. Rev. **27**, 637–651 (2007)
27. Cameron, T.A., James, M.D.: Efficient estimation methods for "closed-ended" contingent valuation surveys. Rev. Econ. Stat. **69**, 269–276 (1987)
28. Fosgerau, M., Bierlaire, M.: Discrete choice models with multiplicative error terms. Transp. Res. Part B Methodol. **43**, 494–505 (2009)
29. McFadden, D.L.: Conditional logic analysis of qualitative choice behavior. In: Zarembka, P. (ed.) Frontiers in Econometrics, pp. 105–142. Academic Press, New York (1974)
30. Ojeda-Cabral, M., Hess, S., Batley, R.: Understanding valuation of travel time changes: are preferences different under different stated choice design settings? Transportation **45**, 1–21 (2018)
31. Fan, J., Heckman, N.E., Wand, M.P.: Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. J. Am. Stat. Assoc. **90**, 141–150 (1995)
32. Koster, P.R., Koster, H.R.A.: Commuters' preferences for fast and reliable travel: a semi-parametric estimation approach. Transp. Res. Part B Methodol. **81**, 289–301 (2015)
33. Rouwendal, J., de Blaeij, A., Rietveld, P., Verhoef, E.: The information content of a stated choice experiment: a new method and its application to the value of a statistical life. Transp. Res. Part B Methodol. **44**, 136–151 (2010)
34. Fosgerau, M., Bierlaire, M.: A practical test for the choice of mixing distribution in discrete choice models. Transp. Res. Part B Methodol. **41**, 784–794 (2007)
35. Alwosheel, A., van Cranenburgh, S., Chorus, C.G.: Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. J. Choice Model. **28**, 167–182 (2018)
36. Day, B., Pinto Prades, J.-L.: Ordering anomalies in choice experiments. J. Environ. Econ. Manag. **59**, 271–285 (2010)
37. Castelvecchi, D.: Can we open the black box of AI? Nat. News **538**, 20 (2016)