Data-driven assisted model specification for complex choice experiments data
Association rules learning and random forests for Participatory Value Evaluation experiments

Hernandez, Jose Ignacio; van Cranenburgh, Sander; Chorus, Caspar; Mouter, Niek

DOI
10.1016/j.jocm.2022.100397

Publication date
2023

Document Version
Final published version

Published in
Journal of Choice Modelling

Journal of Choice Modelling 46 (2023) 100397

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Data-driven assisted model specification for complex choice experiments data: Association rules learning and random forests for Participatory Value Evaluation experiments

Jose Ignacio Hernandez *, Sander van Cranenburgh, Caspar Chorus, Niek Mouter

Department of Engineering Systems and Services, Delft University of Technology. Jaffalaan 5, 2628BX Delft, The Netherlands

A R T I C L E I N F O

Keywords:
Machine learning
Choice experiments
Participatory value evaluation
Association rules
Random forests

A B S T R A C T

We propose three procedures based on association rules (AR) learning and random forests (RF) to support the specification of a portfolio choice model applied in data from complex choice experiment data, specifically a Participatory Value Evaluation (PVE) choice experiment. In a PVE choice experiment, respondents choose a combination of alternatives, subject to a resource constraint. We combine a methodological-iterative (MI) procedure with AR learning and RF models to support the specification of parameters of a portfolio choice model. Additionally, we use RF model predictions to contrast the validity of the behavioural assumptions of different specifications of the portfolio choice model. We use data of a PVE choice experiment conducted to elicit the preferences of Dutch citizens for lifting COVID-19 measures. Our results show model fit and interpretation improvements in the portfolio choice model, compared with conventional model specifications. Additionally, we provide guidelines on the use of outcomes from AR learning and RF models from a choice modelling perspective.

1. Introduction

In the last years, Participatory Value Evaluation (PVE) choice experiments have become an alternative to capture more complex and realistic forms of human decision making in diverse fields (Mouter et al., 2021b; Rotteveel et al., 2022; Mulderij et al., 2021). PVE is a preference elicitation framework based in a portfolio choice experiment (Wiley and Timmermans, 2009), in which respondents select their preferred set of alternatives, subject to one or more resource constraints (Mouter et al., 2021a). In the PVE choice experiment, respondents face a set of available alternatives, the attributes and costs of each alternative, and the available resources. Then, respondents must choose a combination of alternatives (if any), without violating the constraints. As in recently developed experiments (Caputo and Lusk, 2022; Carson et al., 2022; Neill and Lahne, 2022), a PVE choice experiment is an extension of the discrete choice experiment (DCE) approach that provides a more realistic experimental setting for choice situations where a multiple choice subject to constraints is required (e.g., policy makers deciding to fund certain policies with a scarce budget).

While PVE choice experiments offer a more realistic experimental setting than a conventional DCE, specifying choice models to analyse data from such experiments is challenging. Hitherto, choice models developed to analyse PVE choice experiments data (Dekker et al., 2019; Bahamonde-Birke and Mouter, 2019) have been built to address multiple-discrete (portfolio) choices, the presence of resource constraints and interaction effects when two or more alternatives are chosen together (Bahamonde-Birke and Mouter, 2019). However, the specification process of these models usually relies on prior knowledge from the analyst concerning, for example, how respondents derive utility, how attributes interact (e.g., linear-in-parameters specification), the respondents' decision

* Corresponding author.
E-mail address: j.i.hernandez@tudelft.nl (J.I. Hernandez).

https://doi.org/10.1016/j.jocm.2022.100397
Received 10 March 2022; Received in revised form 21 September 2022; Accepted 15 November 2022
Available online 18 November 2022

rule, what interactions between alternatives are relevant to include, etc. Furthermore, finding a proper model specification usually involves a trial-and-error procedure, in which several candidate specifications are tested and the most parsimonious or plausible model is chosen. This process is already cumbersome for discrete choice models (Ortelli et al., 2021), but for more complex choice models, and models for PVE choice experiments data in particular, even more so. The presence of considerably more variables, possible combinations of chosen alternatives, and potential interactions effects between alternatives impose more complexity in the specification of a choice model for PVE choice experiments data, with the consequently longer estimation times than for a discrete choice model, namely from the range of minutes for a simple specification, to an hour in more complex cases.

In the last years, there has been an increasing interest on assisting the specification of choice models with data-driven methods. Data-driven methods (e.g., machine learning, data mining) are methodological approaches that aim to identify relevant patterns and/or learn the underlying data-generating process (DGP) directly from the data. Recent studies have shown that data-driven methods can complement the toolbox of choice modellers (see, for example van Cranenburgh et al., 2022; Sifringer et al., 2020; Wang et al., 2020), or provide further insights for researchers, without explicitly using choice models (e.g., Keuleers et al., 2001; van Cranenburgh and Kouwenhoven, 2020). Furthermore, specific approaches based in data-driven methods to assist the specification of discrete choice models have been recently proposed in literature (Ortelli et al., 2021; Hillel et al., 2019; Shiftan and Bekhor, 2020). However, to the authors' knowledge, no studies have explored methods to assist the specification of choice models to analyse data from more complex type of choice experiments than a DCE, and particularly from PVE choice experiments, or they explored potential insights obtained from using this methods with PVE choice experiments data.

In this paper, we propose three procedures to assist the specification of choice models for PVE choice experiments based in two data-driven methods, and we provide insights on the interpretation of the outcomes of such methods a choice modelling perspective. The first method is Association Rules (AR) learning (Agrawal et al., 1993); a data mining approach used to identify frequent interactions between the variables of a dataset in terms of a set of empirical relational statistics. Applications of AR learning in areas where choice models are standard methods can be seen in the works of Keuleers et al. (2001), Geurts et al. (2003) and Kaur and Kang (2016), but solely focused on gathering association rules between explanatory variables of choice data. We use AR learning to gather association rules between chosen alternatives of the PVE choice experiment that can be interpreted as relevant interactions made by respondents. The second method is a Random Forest (RF) model (Breiman, 2001); a predictive machine learning model built from an ensemble of decision tree models. RF models can model complex relationships from the data, while yet providing a degree of interpretability through the computation of variable importances. We build upon the works of Hillel et al. (2019), Yao and Bekhor (2020) and Shiftan and Bekhor (2020), and we propose two methodological-iterative (MI) procedures to specify the parameters of the utility functions of a portfolio choice model applied in PVE experiments data (Bahamonde-Birke and Mouter, 2019) in a structured way, based on the outcomes of AR learning and RF models, respectively. Finally, we propose a procedure to test the validity of the behavioural assumptions of different specifications of the portfolio choice model, based on comparing their ranking of combinations of alternatives with highest choice probability with the ranking obtained from a RF model.

For our analyses we use data from a PVE choice experiment to elicit the preferences of Dutch citizens for relaxing COVID-19 restrictions after the first wave of the Coronavirus pandemic (Mouter et al., 2021a). In this experiment, respondents were asked to choose their preferred package of COVID-19 restrictions to be relaxed from eight options, such that a constraint of pressure to the healthcare system is not violated. On the one hand, relaxing COVID-19 restrictions may lead to increasing deaths due to COVID-19; on the other hand, it can provide psychological relief and reduce economic losses. Interactions between individual relaxations are reasonably expectable in this PVE choice experiment, as well as differences in terms of the preferences for an impact among different options. Furthermore, it is reasonable to expect the existence of complex interactions that are difficult to uncover from a choice model. In fact, analyses of written arguments to make a choice in this PVE choice experiment suggest the existence of semi-compensatory and lexicographic choice behaviour in a significant amount (Mouter et al., 2021a). In that sense, a more agnostic approach (in terms of behavioural assumptions), such as a RF model can be more appropriate for prediction purposes than a choice model.

This paper is organised as follows. Section 2 details the PVE choice experiment data preparation and data description. Section 3 formalises AR learning, RF models, the portfolio choice model and the procedure to assist the specification of choice models. Section 4 presents the results. Section 5 concludes and provides a discussion of our findings and further research directions.

2. Data

2.1. The COVID-19 PVE choice experiment data

We use data from a PVE choice experiment conducted to elicit the preferences of Dutch citizens to relax COVID-19 measures in the Netherlands (Hernandez et al., 2021),[1] henceforth the COVID-19 PVE choice experiment. In this experiment, respondents were asked to choose which COVID-19 restrictions should be relaxed, without surpassing a maximum level of pressure increase to the healthcare system. Respondents faced eight relaxation options (alternatives):

1. Nursing and care homes allow visitors (NH),
2. Re-open businesses, other than contact professions and hospitality industry (RB),

[1] The dataset is available from https://doi.org/10.4121/14413958.v1

Table 1
Example of a choice matrix of a PVE choice experiment.

Participant ID	NH	RB	RC	YP	...	RH
1	1	1	0	0	...	1
2	1	1	1	0	...	1
3	0	0	0	1	...	1
...						
N	0	0	0	0	...	0

Table 2
Variables of the COVID-19 PVE choice experiment dataset.

Variable	Description
`Choice_` 1 to 8	Binary choice indicator (= 1 if alternative is chosen)
`Y_index`	Unique index of chosen combination of alternatives (from 0 to 181)
`Pressure_` 1 to 8	Additional pressure to the healthcare system
`Deaths_70plus_` 1 to 8	Additional deaths of people of 70+ years old
`Deaths_less70_` 1 to 8	Additional deaths of people of less than 70 years old
`Plus_physical_injury_` 1 to 8	Additional people with (permanent) physical injury
`Minus_mental_injury_` 1 to 8	Decrease of people with (permanent) mental injury
`Minus_HH_incloss_` 1 to 8	Decrease of households with severe income loss

3. Re-open contact professions (RC),
4. Young people may come together in small groups (YP),
5. All restrictions lifted for people with immunity (LI),
6. All restrictions lifted in Northern provinces (LN),
7. Direct family members from other households can have social contact (DF),
8. Re-open hospitality and entertainment industry (RH).

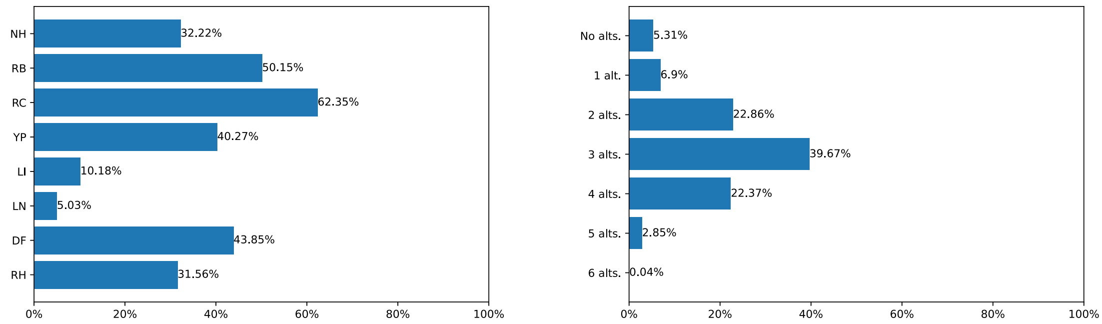
Choosing an alternative generated an additional (percentage) pressure to the healthcare system. Respondents cannot surpass an increase of 50% of pressure to the healthcare system. In addition, each alternative was characterised by five attributes: a) additional deaths of people with 70 or more years old, b) additional deaths of people with less than 70 years old, c) additional cases of (permanent) physical injury, d) reduction of cases of (permanent) mental injury, and e) reduction of households with severe income loss. A more detailed description of the design of this experiment is presented in the work of Mouter et al. (2021a). The choices of a PVE choice experiment can be represented in a matrix of size $N \times J$, as illustrated in Table 1. Each row is a choice situation (respondent) from 1 to N, while each column is an alternative from 1 to J. A choice in a PVE choice experiment is a combination of choices among the J alternatives, represented by ones (if chosen) and zeros (if not chosen).

2.2. Data preparation and description

The COVID-19 PVE choice experiment dataset contains 29,669 responses and 57 variables. Table 2 provides a definition of the variables of this dataset. First, we define the choice indicators in two forms: eight individual indicators per alternative (`Choice_` from 1 to 8) used by AR learning and the choice model, and a single variable that uniquely identifies a chosen combination of alternatives used by the RF model (`Y_index`) ranging from 0 to 181. Second, we define the attributes of each alternative as a set of numeric variables, per alternative and per attribute (`Pressure_` to `Minus_HH_incloss_` 1 to 8). Except for pressure to the healthcare system, all attributes are scaled by 10,000 to avoid numerical overflow issues in the estimation/training routines of the portfolio choice model.

Fig. 1 summarises the market shares (a) and the distribution of the number of chosen alternatives (b) the dataset. The most chosen alternatives are re-opening contact professions (RC) and other businesses (RB), with 62.4% and 50.1%, respectively; in contrast, lifting all restrictions for immune people (LI) and in the Northern provinces (LN) are the least chosen alternatives with 10.2% and 5% respectively. The vast majority of respondents choose between two and four alternatives (more than 80% of respondents). As expected, no respondents choose more than six alternatives due to the existence of a resource constraint. On the other hand, 5.3% of respondents choose no alternative at all (no choice). While this percentage is rather low if taken as a dropout measure (i.e., respondents who did not answer the choice experiment), it is considerably higher than the probability of randomly choosing any combination of alternatives of the dataset ($1/182$).

In addition to the empirical data, we create four datasets with pseudo-synthetic choices. Pseudo-synthetic datasets are generated to corroborate if our proposed methods are able to recover the true data-generating process and/or identify interactions included *a priori* in the data. For instance, we use pseudo-synthetic data to test whether the metrics of AR learning are aligned with the inclusion of explicit interactions. Pseudo-synthetic datasets are generated by using the experimental design of the COVID-19 PVE choice experiment data to generate synthetic choices, assuming a previously known DGP and "true" parameters. We provide a detail of the dataset generation process and parametrisation in Appendix A.



(a) Market shares (b) Distribution of number of chosen alternatives.

Fig. 1. Descriptive statistics of choice variables, COVID-19 PVE choice experiment.

3. Methods

3.1. Association rules learning

AR learning is a data mining method that aims to identify frequent relationships between variables of a transactions dataset (Agrawal et al., 1993). In an AR learning application, an algorithm scans combinations of variables in the dataset named as *itemsets*, keeping only the combinations that satisfy a minimum *support* (relative frequency) threshold defined by the analyst. Then, a set of association rules of the form $A \Rightarrow B$, with A and B itemsets, are constructed, considering only those rules with a *confidence* (conditional frequency) higher than a threshold defined by the analyst. By doing so, AR learning rules out combinations of alternatives that scarcely appear in the dataset. In this paper, we use the seminal approach of AR learning provided by Agrawal et al. (1993) to identify association rules between combinations of discrete alternatives using the "Apriori" algorithm.

We proceed to formalise AR learning. Consider a transactions dataset D with N rows and J columns. Each row $n \in \{1, \ldots, N\}$ of the dataset is a transaction over J items (columns). Each transaction is represented as a vector $y_n = \{y_{n1}, y_{n2}, \ldots, y_{nJ}\}$, in which each variable y_{nj} is a binary indicator that is equal to one if item $j \in \{1, \ldots, J\}$ is selected, and zero otherwise. Some examples of transaction datasets are supermarket purchase data, accesses to webpages, etc. In this paper, we treat the choice data of the PVE choice experiment as a transaction dataset, in which each transaction is a choice situation over J alternatives.

Define an itemset as a subset of items of the dataset. For example, $A = \{y_{n1}, y_{n2}\}$, $B = \{y_{n3}, y_{n4}, y_{n5}\}$ and $C = \{y_{n1}, y_{n2}, y_{n4}\}$ are itemsets of the dataset D. An association rule between itemsets A and B is a directional implication of the form $A \Rightarrow B$, with $A \cap B = \emptyset$, in which A is defined as the *antecedent* and B is the *consequent*. If itemsets A and B are two (disjoint) combinations of alternatives, the rule $A \Rightarrow B$ can be interpreted as "if combination of alternatives A is chosen, then combination of alternatives B is chosen".

The problem of AR learning is to find all the itemsets that satisfy a minimum *support* threshold, and then generate all the association rules that satisfy a minimum *confidence* threshold. The *support* $supp(A)$ of an itemset A as the relative frequency that A appears in dataset D. The support of an itemset A can be interpreted as the probability $P(A)$ on the domain of the dataset.

The *confidence* of an association rule $A \Rightarrow B$ is defined as:

$$conf(A \Rightarrow B) = \frac{supp(A \cup B)}{supp(A)} \tag{1}$$

Confidence is the percentage of transactions of itemset A that also contain itemset B. In the context of PVE choice experiments data, a support of A equal to s is interpreted as "$s * 100\%$ of the choices of the dataset involve the combinations described in A", whereas a confidence of the rule $A \Rightarrow B$ equal to c is interpreted as "$c * 100\%$ of the choices that involve A also involve B". The confidence of $A \Rightarrow B$ can be interpreted as the conditional probability of B given A.

While support and confidence measure how often an itemset or rule appear in the dataset, they do not provide information about the degree of dependence of the components of a rule. Thus, AR learning can generate trivial rules with high confidence and support for itemsets A And B, even if such itemsets have a small or no dependence (Keuleers et al., 2001). In light of this, computing the *lift* of the association rules is recommended. The lift of an association rule $A \Rightarrow B$ measures the degree of dependence between A and B as:

$$lift(A \Rightarrow B) = \frac{supp(A \cup B)}{supp(A) \cdot supp(B)}. \tag{2}$$

The lift is a ratio between the support of A and B together, divided by the independent supports of each itemset. If $lift(A \Rightarrow B) > 1$ then A and B are more often to be found in the dataset than if A and B were independent, and viceversa for $lift(A \Rightarrow B) < 1$.

The final outcome of an AR learning application is a list of association rules described by their support, confidence and lift. The analyst can refine this list according to their research needs. For instance, the analyst may be interested only in rules that contain a certain set of variables in the antecedent, and other sets of variables in the consequent; or finding the rules that are more frequent to appear than expected, by sorting them in terms of lift.

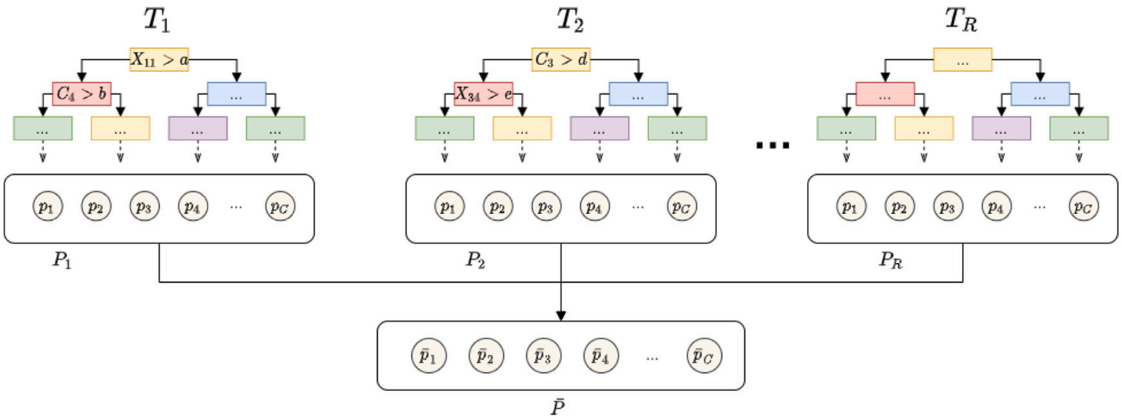


Fig. 2. Example of a RF model.

RF Model algorithm:

Training: Let $(\mathbf{X}, \mathbf{y})$ be the training sample.

1. For $r = \overline{1, R}$:

- **1a.** Draw a bootstrapped sample $(\mathbf{X^r}, \mathbf{y^r})$ from $(\mathbf{X}, \mathbf{y})$.
- **1b.** Train a DT model using $(\mathbf{X^r}, \mathbf{y^r})$, selecting random k variables from $\mathbf{X^r}$ to do each split of the tree.
- **1c.** Store the DT model as T_r.

Prediction: Let $(\mathbf{X}^*)$ be a subset of the original explanatory variables, different from the training data.

1. For $r = \overline{1, R}$, predict the choice probabilities on each decision tree.

2. Compute the final choice probabilities as the average among the R trees.

Fig. 3. RF model algorithm.

3.2. Random forests

A RF model (Breiman, 2001) is a supervised machine learning method that benefits of the strengths of three techniques from machine learning: an ensemble of decision tree (DT) models, bootstrap aggregation or bagging, and random feature selection. The process of constructing a RF model is illustrated in Fig. 2. Firstly, a set of R DT models is estimated. A DT is a supervised machine learning method based on partitioning the space of the explanatory variables into finite regions to construct a tree structure that best describes the response variable (Friedman et al., 2001). Secondly, to allow variability among trees, each DT model is trained (estimated) using a bootstrapped sample of the data. Thirdly, the partitions of each individual DT model are done using a random subset of explanatory variables, in order to reduce the correlation between trees. Finally, each tree generates a set of predictions that are averaged to provide the final prediction of the RF model.

We proceed to formalise the RF model used in this paper. Let y_n be a response variable that uniquely identifies a choice over J alternatives for respondent n. X_n is a set of explanatory variables, such as attributes and costs. Define $(\mathbf{y}, \mathbf{X})$ as a dataset of size N where $\mathbf{y} = \{y_1, \ldots, y_N\}$ and $\mathbf{X} = \{X_1, \ldots, X_N\}$. The goal of the RF model is to construct an ensemble of DT models that predicts $\mathbf{y}$ as a function of $\mathbf{X}$.

The RF algorithm is described in Fig. 3. On each DT model r, a bootstrapped sample $(\mathbf{X^r}, \mathbf{y^r})$ of the original data is drawn and used to train the individual tree. Each split of the DT model r is done using a random subset of the variables contained in $\mathbf{X^r}$. Finally, each trained DT model r is stored. To make predictions with a RF model, a sample $(\mathbf{X}^*, \mathbf{y}^*)$ – that is different from the sample used for training – is used to predict the choice probabilities of each combination of alternatives among the R trees. The final choice probabilities of the RF model are computed by averaging the predictions of all trees.

In addition, RF models can be used to determine the importance of the explanatory variables used on the training process. This is done by computing the mean decrease of impurity of each explanatory variable among the splits (Friedman et al., 2001). The decrease of impurity is the contribution of an explanatory variable on reducing misclassifications in terms of the Gini index (Cheng et al., 2019) on a split of a DT:

$$G(X_i) = \sum_{j=1}^{J} P(X_i = L_j)(1 - P(X_i = L_j)), \tag{3}$$

where X_i is the candidate variable for making a split in the RF model, with a possible number of categories $L_1, \dots, L_J$, and $P(X_i = L_j)$ is the predicted probability of $X_i = L_j$. The mean decrease of impurity of a RF model is the average contribution of each explanatory variable on reducing misclassifications on each tree, and among trees of the RF model. Therefore, higher values of the mean decrease of impurity for a variable X_i imply a major importance of such variable in the RF model, and viceversa.

We identify two considerations on the training and use of results of RF models. The first consideration is the proper selection of so-called hyperparameters. The hyperparameters of the RF model (i.e., the number of DT models, the maximum depth of each tree and the number of variables used per split) can have a significant impact on the final predictions. In light of this, we followed a grid search process to determine the best hyperparameters of the RF model applied to our data. We describe in detail such procedure in Appendix B. The second consideration is that variable importances are computed from the training data, which can lead to difficulties to generalise their interpretations. In light of this, the importance measures employed in this paper are computed by using a cross-validation process based on training 100 RF models using random samples (with replacement) of the original data, and averaging the obtained importances of each explanatory variable among all repetitions.

3.3. The portfolio choice model for PVE choice experiments data

The model we aim to assist its specification is a portfolio choice model for PVE choice experiments data proposed by Bahamonde-Birke and Mouter (2019). This model is an extension to the joint choice model (Lerman, 1976), modified to only consider the choice probabilities of combinations that do not violate the resource constraint. In addition, this model can incorporate interaction parameters that address increases (decreases) of utility when two or more alternatives are chosen at once, interpreted as positive/negative synergies.

We proceed to formalise the portfolio choice model. Let be N respondents of a PVE choice experiment with J alternatives and an available amount of resources of B. Each alternative $j \in \{1, \dots, J\}$ that respondent $n \in \{1, \dots, N\}$ is characterised by the unitary cost of resources c_{nj} and the vector of K attributes X_{nj}. Each respondent perceives utility from their choice of a combination of alternatives p, where p is a number from one to $2^J - U_n$, i.e., the number of possible combinations between alternative choices, minus the number of unfeasible combinations. Additionally, each respondent perceives utility from the amount of non-spent resources.

Following (Bahamonde-Birke and Mouter, 2019) and assuming only interactions between two alternatives, the utility of choosing a combination p for respondent n is defined by Eq. (4):

$$U_{np} = \begin{cases} \sum_{j=1}^{J} y_{nj} \cdot U_{nj} + \delta_0 \cdot \left(B - \sum_{j=1}^{J} y_{nj} \cdot c_{nj} \right) + \sum_i \sum_j \theta_{ij} y_i y_j + \varepsilon_{np} & \text{, if } \left(B - \sum_{j=1}^{J} y_{nj} \cdot c_{nj} \right) \geq 0 \\ -\infty & \text{, if } \left(B - \sum_{j=1}^{J} y_{nj} \cdot c_{nj} \right) < 0 \end{cases} \tag{4}$$

where y_{nj} are binary indicators that are equal to one if respondent n selected alternative j and zero otherwise, U_{nj} is the utility of each individual alternative, δ_0 is a parameter that captures the preference for not spending resources, θ_{ij} is a parameter that captures the increase (or reduction) of utility when alternatives i and j are chosen together with $i \neq j$, and ε_{np} is a stochastic error term with an Extreme Value distribution. U_{nj} is defined by Eq. (5):

$$U_{nj} = \delta_j + \beta' X_{njk}, \tag{5}$$

where δ_j are alternative-specific constants and β is a vector of parameters associated with the attributes of each alternative.

Assuming that each individual choose the combination of alternatives that maximise his/her utility, the probability of choosing alternative i by respondent n is defined by Eq. (6):

$$P_i = P(U_{ni} \geq U_{np}, \forall p \neq i) = \frac{exp(V_{ni})}{\sum_p exp(V_{np})}, \tag{6}$$

where V_{ni} is the observed (non-stochastic) part of the utility function U_{ni}. Notice that the choice probabilities take the form of the MNL function, since the utility of a combination of alternatives incorporates an additive Extreme Value stochastic term. Furthermore, the choice probability of an unfeasible combination of alternatives collapse to zero, since $V_{ni} = -\infty$.

3.4. Assisted specification of the portfolio choice model: methodological-iterative approaches

Shiftan and Bekhor (2020) and Yao and Bekhor (2020)[2] propose a methodological-iterative (MI) approach to assist the specification of a discrete choice model using the variable importances of a RF model. We build upon these works, and we propose two separate variations of their MI approach, in which we assist the specification of the parameters of a portfolio choice model using the outcomes of AR learning and RF models, respectively.

[2] We appreciate the suggestion of one anonymous reviewer on considering this work.

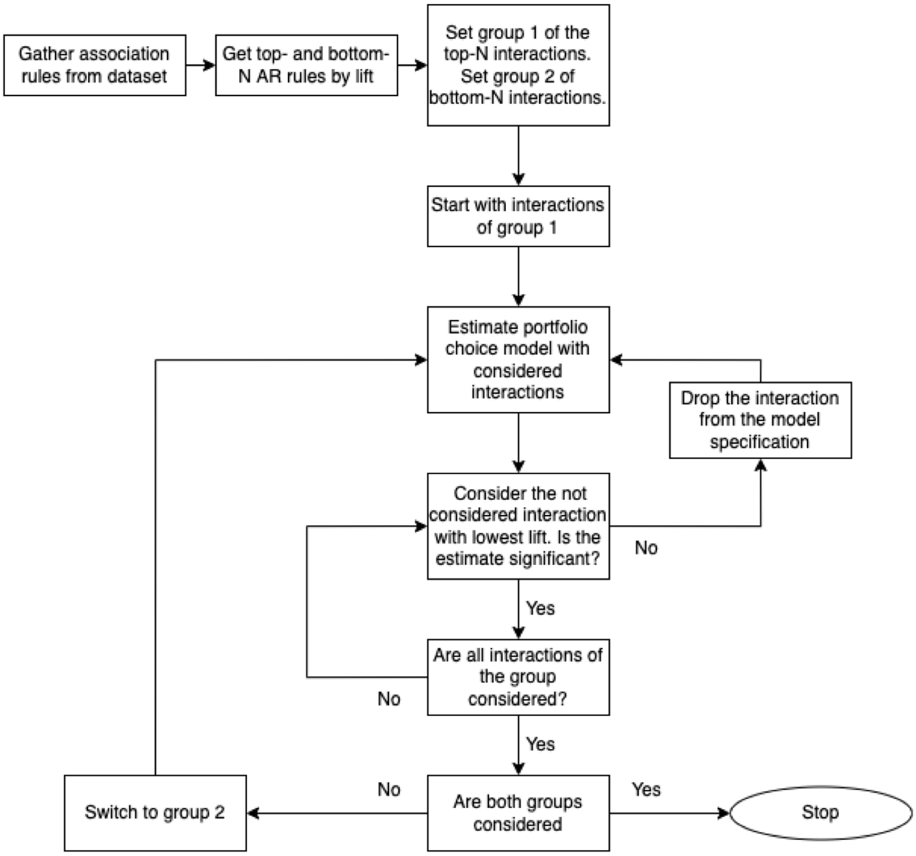


Fig. 4. Methodological-Iterative algorithm for AR learning, based on Shiftan and Bekhor (2020) and Yao and Bekhor (2020).

The first approach, named as MI/AR and detailed in Fig. 4, aims to use the set of association rules with highest and lowest lift values to specify alternative interaction parameters in the portfolio choice model. In the first step, we apply AR learning in the PVE choice experiment data, and we select the N rules with highest and lowest lift, with N chosen by the analyst. We name the set of rules with highest lift as "group 1", and the set of rules with lowest lift is named as "group 2". The algorithm starts by estimating a portfolio choice model with all the interactions of group 1 specified in the utility functions. Then, the algorithm selects the interaction with the lowest lift value of group 1 and evaluates whether the estimated parameter associated to such interaction is statistically significant. If the parameter is non-significant, the interaction is discarded from the model specification and a new portfolio choice model without the interaction is estimated, otherwise the interaction is kept and the process is repeated until all the interactions of group 1 are considered, in an increasing order in terms of lift. After all interactions of group 1 are considered, the process is repeated for group 2. The algorithm stops when all interactions of both groups are considered.

The second approach, named as MI/RF and illustrated in Fig. 5, aims to use the variable importances of a trained RF model to evaluate the inclusion/exclusion of attribute-specific parameters in the portfolio choice model. In the first step, we train the RF model with the PVE choice experiment data, we calculate the variable importances and we sort them in descending order. Then, we separate the variable importances in two groups: "group 1" contains the attributes with highest variable importances, and "group 2" contains the attributes with lowest variable importances. The algorithm starts by estimating a portfolio choice model with all the attribute-specific parameters of group 1 specified in the utility functions. Then, the exclusion of attribute-specific parameters is determined by their statistical significance, starting by the attributes with lowest importance from group 1. After all the attributes of group 1 are evaluated, the process is repeated for group 2. The algorithm stops when all the attributes of both groups are considered.

3.5. Using the RF model predictions to test the behavioural assumptions of the portfolio choice model

The final method proposed in this paper is a procedure to test the behavioural assumptions and specification of a portfolio choice model based on the predictions of a RF model. Specifically, we train the RF model with the PVE choice experiment data, and we compute the ranking of combinations of alternatives with the highest choice probability. This ranking consists of predicting the choice probabilities of all possible combinations of alternatives and sort them by their choice probability in decreasing order. The

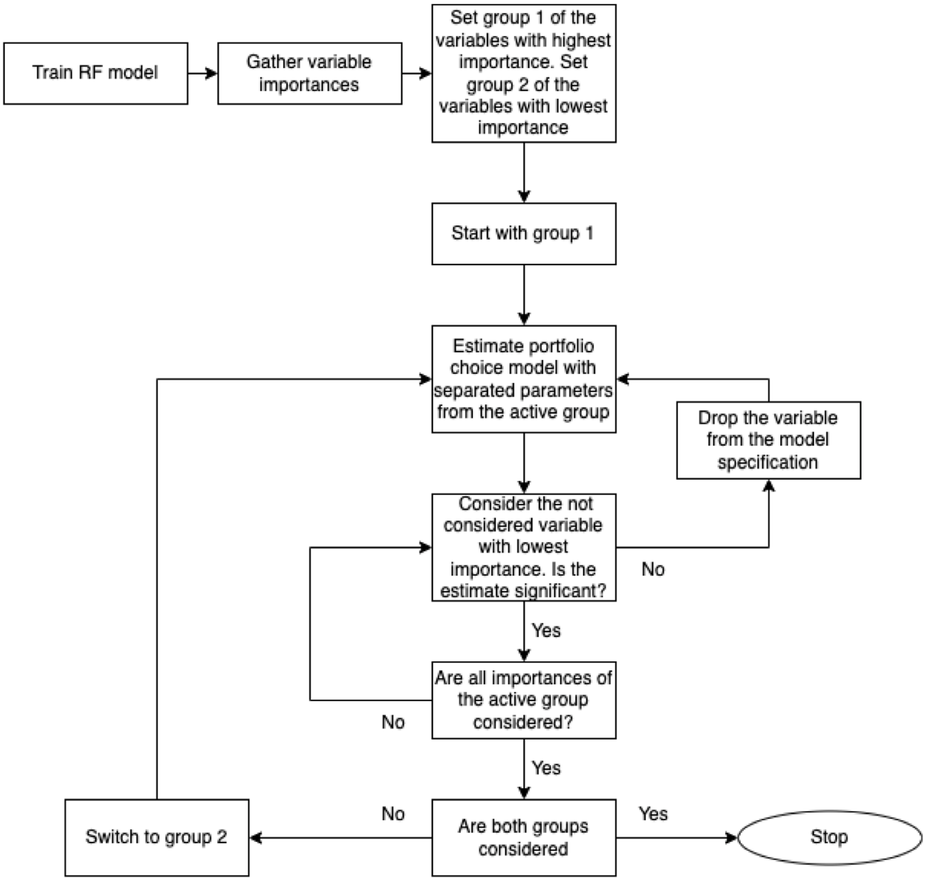


Fig. 5. Methodological-Iterative algorithm for RF models, based on Shiftan and Bekhor (2020) and Yao and Bekhor (2020).

same ranking is computed from the predictions of a portfolio choice model under different utility specifications, and we evaluate the (dis)similarity between these rankings and the ranking obtained from the RF model.

Fig. 6 details the procedure to compute the ranking of combinations of alternatives with highest choice probability using a RF model. The procedure is equivalent to compute the ranking using a portfolio choice model, but replacing the RF model by the portfolio choice model. We consider a RF model trained with a sample of the original data, and a prediction (test) sample that differs from the training data. First, we predict the choice probabilities for each combination of alternatives using the prediction sample using the trained RF model. Then, we average the choice probabilities among each choice situation to obtain a single vector of choice probabilities per combination of alternatives. This vector is sorted in descending order and the ranking is constructed by matching each combination of alternatives with their corresponding choice probability. Additionally, the cost of each combination is reported, and in case the portfolio choice experiment considers resource constraints, then combinations that violate such resource constraint are discarded.

To evaluate the (dis)similarity between predicted rankings, we use the Kendall's Tau correlation coefficient (Kendall, 1945). This statistic measures the correlation between two pairs of ranked lists. The statistic is defined as:

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T)(P + Q + U)}} \tag{7}$$

where P is the number of concordant pairs between lists, Q is the number of discordant pairs, T the number of ties of the first list and U the number of ties in the second list.

4. Results

4.1. Association rules learning

4.1.1. Gathering and interpreting association rules from the PVE choice experiment data

Table 3 summarises the found association rules from the COVID-19 PVE choice experiment data. We set low threshold values for support and confidence in order to avoid discarding rules that can be a sign of negative interactions between alternatives.

Ranking of combinations of alternatives with highest choice probability using RF models

Start: Consider a trained RF model. Let $(\mathbf{X}^*, \mathbf{y}^*)$ be a prediction sample.

1. Predict the choice probabilities $\hat{\mathbf{y}}$ with the prediction sample using the trained RF model.

2. Average the choice probabilities among choice situations (rows).

3. Sort the resulting average choice probabilities in descending order.

4. Construct the portfolio ranking by matching each combination with their respective choice probability.

5. Output the ranked combinations, their choice probabilities, and their total cost of resources.

Fig. 6. Algorithm for computing the ranking of combinations of alternatives with highest choice probability.

Table 3
Summary of found association rules.

	Value
Number of rules	2,100
Mean confidence	0.09
Confidence range	[0.0, 1.0]
Mean lift	0.458
Lift range	[0.013, 1.994]

Specifically, we set $minsupport \approx 0$ and $minconfidence = 0$ as support and confidence thresholds, respectively.[3] In total, we find 2100 association rules, with an average confidence of 9%, and average lift of 0.46, with a range between 0.01 and 2 approximately. We expected a low average confidence and large range of lift due to the low thresholds we specify for support an confidence. Additionally, we confirmed that confidence and lift values of association rules are aligned with the inclusion of interactions and unobserved correlation between chosen alternatives in the pseudo-synthetic data. We present the results of these analyses in Appendix C

For the purposes of an easier interpretation, we focus on binary (one antecedent and one consequent) association rules, and we discard rules with swapped antecedent and consequent, since their lift is the same than the kept rules. Table 4 summarises the support, confidence and lift of the top- and bottom-10 binary association rules sorted by lift. The interpretation of confidence for the found rules is the extent that the consequent is found in the antecedent. For instance, the confidence value of 0.74 of the second association rule means that 74% of respondents that choose to re-open the hospitality sector (RH) also choose to re-open contact professions (RC). In contrast, only 3% of respondents that choose re-opening the hospitality sector (RH) also choose to lift restrictions in Northern provinces (LN). The interpretation of lift is in terms of the extent that two alternatives are more (less) prone to be chosen together than each alternative separately, compared with the other rules after applying filter criteria. For instance, we find that choosing together to lift restrictions in the Northern provinces (LN) and for immune people (LI) is more prone to be chosen than independently, compared with the rest of binary association rules. Conversely, choosing to re-open the hospitality sector (RH) and lift restrictions in the Northern provinces (LN) together is found to be lower than choosing them independently, compared with the other binary rules.

4.1.2. Using association rules to assist the specification of the portfolio choice model

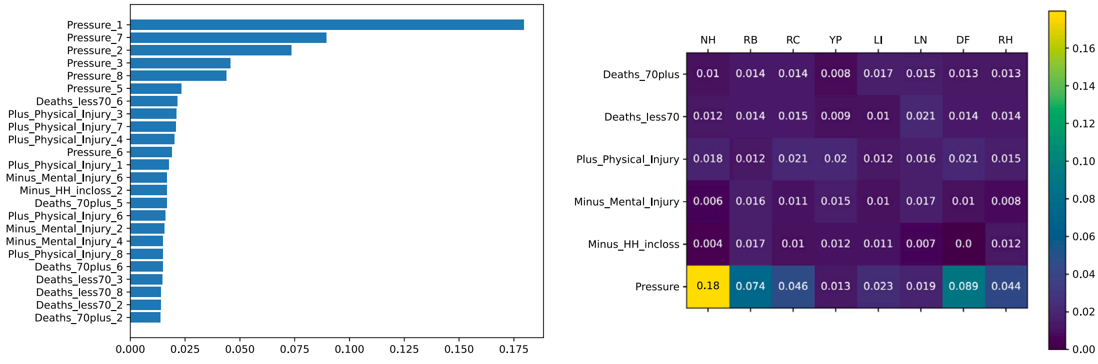
We use three different model specifications of the portfolio choice model. The first model is a baseline specification with alternative-specific constants, attribute parameters that do not vary across alternatives, and no alternative interaction parameters. The second model considers all interactions described in Table 4. The third model is specified with the MI/AR approach to discard non-significant alternative interaction parameters.

Table 5 details the estimation results of the portfolio choice model under the three different specifications. We find that the models that include interaction terms (last two columns) outperform the baseline specification in terms of log-likelihood and Akaike/Bayesian information criteria (AIC and BIC). Furthermore, the model specified using the MI/AR approach outperforms the other two specifications in terms of information criteria, which means that this model is more parsimonious. All interactions associated with high lift values are statistically significant and have a positive sign, in line with out expectations. On the other hand, we find that two of the interactions associated with lower lift (Interaction: ['YP', 'RH'] and Interaction: ['RH', 'LI']) have a positive sign and are statistically significant, against our expectations. However, we also observe that the inclusion of alternative interaction parameters induce a change of sign of some of the alternative-specific constants of the models. Furthermore, it is easy to

[3] We used a value of $minsupport = 1 * 10^{-16}$ since the Apriori algorithm only accepts support thresholds above zero.

Table 4
Top- and bottom-10 binary association rules ordered by lift.

Antecedents	Consequents	Support	Confidence	Lift
LN	LI	0.0067	0.1342	1.3192
RH	RC	0.2343	0.7423	1.1906
YP	DF	0.2082	0.5171	1.1792
RB	YP	0.2184	0.4354	1.0814
LI	YP	0.0419	0.4113	1.0216
YP	RC	0.2548	0.6328	1.0149
RC	RB	0.3130	0.5021	1.0011
DF	RC	0.2720	0.6202	0.9947
LI	DF	0.0443	0.4352	0.9924
RB	DF	0.2121	0.4230	0.9645
RH	YP	0.1110	0.3518	0.8738
LN	RB	0.0206	0.4107	0.8190
DF	LN	0.0170	0.0388	0.7729
RH	DF	0.1058	0.3353	0.7645
RC	LN	0.0226	0.0362	0.7201
LI	RH	0.0228	0.2237	0.7090
LN	NH	0.0097	0.1933	0.5998
RH	NH	0.0602	0.1907	0.5917
LI	NH	0.0190	0.1863	0.5781
RH	LN	0.0064	0.0202	0.4020



(a) Top-ranked variable importances (b) Heatmap view

Fig. 7. RF variable importances, empirical data.

see that the utility of choosing 'LI' and 'RH' together is lower in the model specified with the MI/AR approach than in the baseline model (−1.87 against −1.16, respectively). Thus, the interpretation of positive or negative interactions does not rely solely on the alternative interaction parameters, but also on the combination of such parameters with the alternative-specific constants.

4.2. Random forests

4.2.1. Obtaining and interpreting variable importances

Fig. 7 presents the top-half of variables ordered by their importance (a), as well as a heatmap view of the importance of all variables (b). We observe that the constrained attributes of the PVE choice experiment (i.e., pressure to the healthcare system) are among the most important variables. Visual inspection of the heatmap view allows to confirm that attributes other than pressure to the healthcare system play a rather minor role in terms of importance. In light of these results, we may expect that parameters associated with pressure to the healthcare system will predominate in a portfolio choice model specified with the MI/RF approach, and that most of the discards are focused on the remaining attribute-specific parameters.

4.2.2. Using variable importances to assist the specification of the portfolio choice model

Table 6 summarises the log-likelihood, rho-squared and information criteria of the baseline portfolio choice model detailed in Table 5 (first column), a model with all attribute-specific parameters separated per alternative (second column), and a model with attribute-specific parameters specified with the MI/RF approach (third column). We observe that the model fit measures of the three models are similar, with slightly better performance in the MI/RF specification. This result can be explained by the low contribution of attributes to explain the portfolio choice model, other than pressure to the healthcare system.

Table 5
Estimation results of portfolio choice models.

	Baseline model	All interactions	MI/AR
Remaining pressure	0.0442 (0.0014)***	0.0422 (0.0014)***	0.0421 (0.0013)***
Constant of NH	0.5185 (0.0415)***	0.5959 (0.0477)***	0.5932 (0.0471)***
Constant of RB	0.6956 (0.0311)***	0.3510 (0.0382)***	0.3604 (0.0340)***
Constant of RC	1.2643 (0.0351)***	0.5193 (0.0431)***	0.5197 (0.0397)***
Constant of YP	0.1102 (0.0172)***	−0.6477 (0.0339)***	−0.6497 (0.0319)***
Constant of LI	−0.9674 (0.0354)***	−1.2422 (0.0505)***	−1.2427 (0.0479)***
Constant of LN	−1.1536 (0.0547)***	−1.2205 (0.0807)***	−1.2166 (0.0563)***
Constant of DF	0.5025 (0.0389)***	0.0779 (0.0472)	0.0903 (0.0433)*
Constant of RH	0.4653 (0.0528)***	−0.2946 (0.0616)***	−0.2998 (0.0577)***
Additional deaths 70 y.o. or more	−0.0707 (0.0965)	−0.0890 (0.1245)	−0.0873 (0.1204)
Additional deaths less than 70 y.o.	−0.7498 (0.2242)***	−0.7653 (0.1884)***	−0.7686 (0.2174)***
Additional physical injury	−0.1109 (0.0220)***	−0.1049 (0.0213)***	−0.1049 (0.0208)***
Reduction of psychological injury	0.0204 (0.0046)***	0.0179 (0.0046)***	0.0180 (0.0045)***
Reduction of income losses	0.0211 (0.0032)***	0.0229 (0.0030)***	0.0229 (0.0029)***
Interaction: ['LI', 'LN']		0.8907 (0.0853)***	0.8942 (0.0852)***
Interaction: ['RH', 'RC']		1.2155 (0.0280)***	1.2171 (0.0288)***
Interaction: ['YP', 'DF']		0.6693 (0.0253)***	0.6701 (0.0248)***
Interaction: ['YP', 'RB']		0.4431 (0.0251)***	0.4454 (0.0246)***
Interaction: ['LI', 'YP']		0.3420 (0.0401)***	0.3419 (0.0390)***
Interaction: ['YP', 'RC']		0.2710 (0.0259)***	0.2708 (0.0258)***
Interaction: ['RC', 'RB']		0.2515 (0.0248)***	0.2508 (0.0244)***
Interaction: ['RC', 'DF']		0.2836 (0.0252)***	0.2824 (0.0250)***
Interaction: ['LI', 'DF']		0.2841 (0.0404)***	0.2822 (0.0412)***
Interaction: ['DF', 'RB']		0.0221 (0.0243)	
Interaction: ['YP', 'RH']		0.1255 (0.0270)***	0.1269 (0.0272)***
Interaction: ['LN', 'RB']		0.0089 (0.0582)	
Interaction: ['LN', 'DF']		0.0230 (0.0569)	
Interaction: ['RH', 'DF']		−0.1504 (0.0268)***	−0.1523 (0.0273)***
Interaction: ['LN', 'RC']		0.0066 (0.0540)	
Interaction: ['LI', 'RH']		0.2838 (0.0493)***	0.2861 (0.0494)***
Interaction: ['LN', 'NH']		−0.2198 (0.0672)**	−0.2147 (0.0633)***
Interaction: ['RH', 'NH']		−0.1082 (0.0317)***	−0.1043 (0.0316)***
Interaction: ['LI', 'NH']		−0.3192 (0.0550)***	−0.3186 (0.0495)***
Interaction: ['LN', 'RH']		−0.0611 (0.0854)	
Log-likelihood	−124,119.02	−122,336.75	−122,337.59
AIC	248,266.03	244,741.50	244,733.18
BIC	248,382.19	245,023.61	244,973.80
Rho-squared	0.0981	0.1110	0.1110

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Table 6
Model fit metrics of portfolio choice models.

	Baseline model	Separated parameters	MI/RF
Log-likelihood	−124,119.02	−124,003.24	−124,014.63
AIC	248,266.03	248,116.48	248,093.27
BIC	248,382.19	248,572.83	248,358.78
Rho-squared	0.0981	0.0989	0.0988

Table 7 summarises the estimation results of the portfolio choice model specified with the MI/RF approach. We observe that all alternative-specific constants are positive and statistically significant. The parameters associated with pressure to the healthcare system suggest mixed effects depending of each individual alternative. Notice that for this model specification (and the model with all attribute-specific parameters separated per alternative), pressure to the healthcare system is included as an additional attribute, hence it should be interpreted in terms of pressure increases, instead of remaining pressure such as in the results of Table 5. With respect to the remaining attributes, all but one of the attribute-specific parameters have the expected sign. Furthermore, the estimated parameters of additional deaths of people of 70 years old or older become statistically significant and have a negative sign. Finally, we find that the estimate of additional physical injury by choosing to allow visitors in nursing homes has a positive sign, against our expectations.

4.2.3. Testing the behavioural assumptions of the portfolio choice model

Table 8 summarises the model fit measures of the trained RF model compared with three specifications of the portfolio choice model: the baseline model, the specification based on the MI/AR approach used in Table 5 and the specification based on the MI/RF approach detailed in Table 7. We find that the RF model outperforms all the other choice modelling approaches, at least in terms of log-likelihood and rho-squared. We previously verified that the RF model is able to recover the true DGP of different specifications

Table 7
Estimation results, portfolio choice model specified with MI/RF approach.

	NH	RB	RC	YP	LI	LN	DF	RH
Constant	0.0383***	0.0476***	0.0418***	0.0850***	0.0511***	0.0427***	0.0677***	0.0190***
	(0.0023)	(0.0036)	(0.0043)	(0.0078)	(0.0054)	(0.0052)	(0.0038)	(0.0033)
Additional pressure	0.3043***	0.6364***	1.2694***	0.4563***	−0.9228***	−0.9068***	0.7593***	0.1239
	(0.0666)	(0.0427)	(0.0885)	(0.0542)	(0.0774)	(0.1541)	(0.0987)	(0.1061)
Additional deaths 70 y.o. or more				−2.1694***			−0.6655*	1.0220*
				(0.6454)			(0.2797)	(0.4918)
Additional deaths less than 70 y.o.			−1.1370**	−3.1763***				−1.7083**
			(0.4034)	(0.8106)				(0.5999)
Additional physical injury	0.8199*		−0.1210*			−0.3540**	−0.1324**	−0.1014*
	(0.3810)		(0.0524)			(0.1363)	(0.0463)	(0.0437)
Reduction of psychological injury	0.0284**						0.0240*	
	(0.0091)						(0.0096)	
Reduction of income losses		0.0261***	0.0235***					0.0143*
		(0.0049)	(0.0063)					(0.0064)
Log-likelihood	−124,014.63							
AIC	248,093.27							
BIC	248,358.78							
Rho-squared	0.0988							

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 8
Model fit measures, RF model compared with portfolio choice models.

	Baseline	MI/AR	RF/AR	RF
Log-likelihood	−124,119.02	−122,337.59	−124,014.63	−120,197.17
Rho-squared	0.0981	0.1110	0.0988	0.1266

Table 9
Comparison of Kendall's Tau of rankings of combinations of alternatives with highest choice probability.

	RF vs. Baseline	RF vs. MI/AR	RF vs. MI/RF
Top-5	−0.4000	−0.2000	−0.4000
Top-10	0.3778	0.4667+	0.3778

P-values of Kendalls Tau: + : $p < 0.1$.

of the portfolio choice model, and that the RF model can approximate the ranking of combinations of chosen alternatives with highest choice probability. A more detailed descriptions of such tests can be found in Appendix D.

Table 9 details the Kendall's Tau value obtained from contrasting the top-5 and top-10 rankings of combinations of alternatives with highest choice probability of the RF model with their respective top-5 and top-10 rankings obtained from the baseline portfolio choice model and the specifications specified with the MI/AR and MI/RF approaches. Among all the contrasts, the only case in which the hypothesis of no-correlation is rejected (at 90%) of confidence is between the top-10 rankings of the RF model and the portfolio choice model specified with the MI/AR approach.

Finally, Table 10 details the top-10 ranking of combinations of alternatives with highest choice probability of the RF model, the baseline portfolio choice model and the portfolio choice model specified with the MI/AR approach. We observe that not choosing any alternative is the combination with the highest probability from the predictions of a RF model. For the baseline portfolio choice model, the combination with highest choice probability is to re-open businesses, re-open contact professions and allow contact between direct family members of different households, which aligns with the results of Mouter et al. (2021a) despite a different choice model was used i.e., a Multiple Discrete-Continuous Extreme Value (MDCEV) model. The combination of alternatives with highest choice probability in the MI/AR portfolio choice model is to re-open businesses, re-open contact professions, allow young people to come together in groups and allow contact between direct family members of different households. None of the portfolio choice models include not choosing any alternative among the top-10 probability ranking, whereas the ranking of the RF model ranks as third-best the combination that ranks the first in the model specified with the MI/AR approach.

5. Conclusion and discussion

In this paper, we propose procedures based on AR learning and RF models to support the specification of a portfolio choice model applied in data from a PVE choice experiment, and we provide insights on the interpretation of the outcomes of the proposed models from a choice modelling perspective. We use data from a PVE choice experiment conducted to elicit the preferences of Dutch citizens to lift COVID-19 restrictions during the first wave of the Coronavirus pandemic in 2020. On the one hand, AR learning is used to identify relevant interactions between different combinations of alternatives chosen by respondents of the PVE choice experiment and support the specification of alternative interaction parameters in a portfolio choice model. On the other

Table 10
Top-10 ranking of combinations of alternatives with highest choice probability, RF and portfolio choice models.

Random forest

	Rk.1	Rk.2	Rk.3	Rk.4	Rk.5	Rk.6	Rk.7	Rk.8	Rk.9	Rk.10
Comb. ID	0	134	78	6	70	5	69	7	76	196
NH						X	X	X		
RB		X	X	X	X			X		
RC		X	X	X	X	X	X	X	X	X
YP			X						X	
LI										
LN										
DF			X		X		X		X	X
RH		X								X
Choice probability	5.34%	4.94%	3.71%	2.7%	2.6%	2.6%	2.53%	2.53%	2.46%	2.44%
Pressure	0.0%	40.41%	38.22%	21.5%	31.71%	29.54%	39.75%	39.49%	28.26%	40.66%

Baseline portfolio choice model

	Rk.1	Rk.2	Rk.3	Rk.4	Rk.5	Rk.6	Rk.7	Rk.8	Rk.9	Rk.10
Comb. ID	70	6	78	14	68	134	4	7	76	12
NH								X		
RB	X	X	X	X		X		X		
RC	X	X	X	X	X	X	X	X	X	X
YP			X	X					X	X
LI										
LN										
DF	X		X		X				X	
RH						X				
Choice probability	3.68%	3.56%	2.93%	2.9%	2.87%	2.82%	2.78%	2.67%	2.34%	2.27%
Pressure	31.71%	21.5%	38.22%	28.01%	21.75%	40.41%	11.54%	39.49%	28.26%	18.05%

MI/AR portfolio choice model

	Rk.1	Rk.2	Rk.3	Rk.4	Rk.5	Rk.6	Rk.7	Rk.8	Rk.9	Rk.10
Comb. ID	78	134	132	196	6	70	4	7	204	14
NH								X		
RB	X	X			X	X		X		X
RC	X	X	X	X	X	X	X	X	X	X
YP	X								X	X
LI										
LN										
DF	X			X		X			X	
RH		X	X	X					X	
Choice probability	3.98%	3.79%	3.2%	3.18%	2.92%	2.64%	2.44%	2.41%	2.37%	2.31%
Pressure	38.22%	40.41%	30.45%	40.66%	21.5%	31.71%	11.54%	39.49%	47.17%	28.01%

hand, RF models are used to identify the most (least) relevant attributes of the PVE choice experiment, and with this information assist the inclusion/exclusion of attribute-specific estimates of the portfolio choice model. Finally, RF models are used to predict the combinations of alternatives with the highest choice probability, and use that information to test the validity of the behavioural assumptions of several specifications of the PVE choice model.

5.1. Main findings

Firstly, we show that AR learning successfully identifies relevant interactions between chosen alternatives of a PVE choice experiment. For instance, we find that choosing to lift all restrictions for the immune people (LI) and in the Northern provinces (LN) together have the highest lift among binary association rules, despite both alternatives are the least chosen independently. Furthermore, we find that the MI/AR approach to specify interactions in the portfolio choice model leads to model fit improvements in terms of log-likelihood and information criteria, compared with a baseline portfolio choice model. Additionally, we find that directly interpreting the sign of the interaction parameters does not indicate whether an interaction is positive or negative. Instead, a comparison of the utilities of the baseline model and the model with specified interactions can shed light on the positive or negative effect of interactions.

Secondly, our data analyses with RF models show that respondents of the PVE choice experiment mostly care about the constrained attribute (additional pressure to the healthcare system) across almost all alternatives, whereas they put considerable lower relevance to the other attributes of the PVE choice experiment. We find that the MI/RF approach leads to modest improvements of model fit compared with estimating the baseline portfolio choice model. This can be a consequence of the small relevance of the attributes, other than pressure to the healthcare system, found from the variable importances of the RF model. Despite the latter,

Table 11
DGP specification of pseudo-synthetic datasets.

Dataset	Utility specification
Dataset 1 and Dataset 2	$U_{np} = \sum_{j=1}^{J} y_{nj} \cdot (\delta_j + \beta' X_{njk}) + \delta_0 \cdot \left(B - \sum_{j=1}^{J} y_{nj} \cdot c_{nj} \right) + \sum_i \sum_j \theta_{ij} y_i y_j + \varepsilon_{np}$ $\varepsilon_{np} \sim Gumbel(0, 1)$ $\theta_{ij} = 0, \forall i, j$ (dataset 1)
Dataset 3 and Dataset 4	$U_n = \left(B - \sum_{j=1}^{J} y_{nj} \cdot c_{nj} \right) \cdot \exp(\delta_0 + \varepsilon_{n0}) + \sum_{j=1}^{J} y_{nj} \cdot \exp(\delta_j + \beta' X_{nj} + \varepsilon_{nj})$ $\varepsilon_{nj} \sim Gumbel$ (for dataset 3) $\varepsilon_{nj} \sim GEV$ (for dataset 4)

we find additional insights from the MI/RF approach in terms of interpretation of parameters, such as preference differences for the same attribute across individual alternatives. For instance, additional deaths of people of less than 70 years old due to COVID-19 has significantly different effects (in terms of magnitude) between re-opening contact professions (RC) and allowing young people to come together in groups (YP), despite both estimates having a negative sign.

Finally, we find that RF models are able to recover the true DGP from PVE choice experiment data under different specifications of pseudo-synthetic data, and they outperform portfolio choice models in terms of model fit using empirical data, under different model specifications. We find that portfolio choice model specified with the support of AR learning leads to a predicted ranking that tends to get closer to the ranking obtained from a RF model (in terms of the Kendall's Tau), compared with the other model specifications (i.e., baseline model and the model assisted with the variable importances of a RF model). Nevertheless, all portfolio choice models underestimate the choice probability of not choosing any alternative, which ranks as the highest in the ranking obtained with a RF model. Our findings evidence that the portfolio choice models still have misalignments between their behavioural assumptions and the actual DGP embedded in the data, but the procedures we propose in this paper can help to mitigate such misalignments.

5.2. Additional uses of the outcomes of AR learning and RF models

Besides assisting the specification of portfolio choice models, the outcomes of AR learning and RF models applied in PVE choice experiments data can be directly used. With respect to AR learning, this method provides beforehand information about frequent interactions between chosen (combinations of) alternatives, without the need of specifying and estimating a choice model or compute welfare measures. Such interactions can be used in policy making to, for example, recommend in favour of conducting combinations of policies that rank high in terms of lift. In that regard, when the aim is merely identifying frequent combinations of chosen alternatives, using AR learning is advantageous because it does not rely on strict behavioural assumptions that can restrict (or privilege) certain interactions over others. Additionally, the computation runtime of AR learning is shorter than the regular estimation time of a portfolio choice model.

With respect to the variable importances of RF models, this approach can be used as an alternative to estimating the attribute-specific parameters or marginal utilities of a portfolio choice model, without the need of explicitly specifying the form of the utility function. This information can be used to prioritise (or avoid) policy options that perform high on desirable (undesirable) attributes such as, for instance, not lifting restrictions to visits in nursing homes since the relevance of the pressure to the healthcare system is high for this option. As an additional advantage, the time dedicated to train a RF model and obtain the variable importances is generally shorter than the estimation time of a portfolio choice model, and such differences are bigger as the number of individual alternatives of the PVE choice experiment increase. However, variable importances only provide information about the relevance of an alternative, and they do not inform whether the effect of an attribute is positive (negative) for choosing an alternative, unlike the attribute-specific estimates of a portfolio choice model.

Finally, when the aim is solely prediction (i.e., no focus on behavioural interpretation), the RF model is advantageous to determine the ranking of combinations of alternatives with the highest choice probability without relying on *a priori* behavioural assumptions. We find that RF models outperform several specifications of a portfolio choice model, in terms of predictive performance. We argue in favour that the probability ranking of an RF model trained with PVE choice experiment data should be a closer reflection of the true ranking that is embedded in the data.

5.3. Considerations and further research

As a consideration of this work, we advice that, while we provide potential uses and interpretations of the outcomes of AR learning and RF models, we recall that such outcomes should not be treated as equivalent to the outcomes of a choice model. For instance, finding that an association rule has a high (low) lift value does not necessarily mean that the corresponding interaction specified in a choice model will be statistically significant. As shown in this paper, we recall that interactions with the highest (lowest) lift does not necessarily lead to interaction parameters with positive (negative) sign in the choice model. We emphasise that AR learning and RF models are used as supportive tools in this paper, whereas choice models are used as confirmatory tools.

In addition, we provide suggested interpretations of the outcomes that can be obtained with currently developed data-driven methods, but a potential step beyond is to develop outcomes that are particularly tailored to the particularities of the data that is analysed, such as in the case of a PVE choice experiment. For instance, the formulas of support, confidence and lift used in AR learning are built for analysing transaction datasets, but they were not thought for the case of a PVE choice experiment, in

Table 12
Parametrisation of synthetic datasets.

Type of parameter	Description	Parameter	Value
Marginal utility of non-spent resources	Datasets 1 and 2	δ_0	0.01
	Datasets 3 and 4		−3
Alternative-specific constants	ASC for NH	δ_1	0
	RB	δ_2	0.2
	RC	δ_3	−0.3
	YP	δ_4	0.4
	LI	δ_5	0.5
	LN	δ_6	0.4
	DF	δ_7	0.3
	RH	δ_8	−0.9
Attribute-specific parameters	Additional 70+ deaths	β_1	−0.6
	Additional < 70 deaths	β_2	−0.8
	Additional people w. physical injuries	β_3	−0.1
	Reduction people w. psychological injuries	β_4	0.03
	Reduction households w. income losses	β_5	0.03
Interaction parameters (only dataset 2)	Interaction LI & LN	$\theta_{5,6}$	2.5
	Interaction RB, RC & RH	$\theta_{2,3,8}$	−4.8
	Interaction NH & DF	$\theta_{1,7}$	−0.3
Dissimilarity parameters (only dataset 4)	NH & RB	$\lambda_{1,2}$	0.03
	RC & YP	$\lambda_{3,4}$	0.05
	LI & LN	$\lambda_{5,6}$	0.1
	DF & RH	$\lambda_{7,8}$	0.2

which choices have a resource constraint, and hence the interpretation of measures can be affected. In this regard, developing expressions that consider constrained choices, such as in a PVE choice experiment, can provide more strength to the use of AR learning in these contexts. Finally, this paper opens the door to new research directions in the field of bringing data-driven methods to the choice modelling field. For instance, based in recent research (Alwosheel et al., 2021) and our experience with obtaining variable importances from RF models, we see opportunities to integrate explainable AI techniques to analyse data from PVE choice experiments.

CRediT authorship contribution statement

Jose Ignacio Hernandez: Writing – original draft, Writing – review & editing, Data curation, Conceptualisation, Methodology, Formal analysis, Project administration. **Sander van Cranenburgh:** Conceptualisation, Methodology, Writing – review & editing, Supervision. **Caspar Chorus:** Conceptualisation, Methodology, Writing – review & editing, Supervision. **Niek Mouter:** Conceptualisation, Methodology, Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the Netherlands Organisation for Scientific Research (NWO) for funding part of this research (NWO Responsible Innovation grant – 313-99-333), the support of TU Delft AI Labs programme and from the ERC-BEHAVE group (ERC Consolidator grant BEHAVE—724431). We also acknowledge the valuable comments done by the two anonymous reviewers of this paper. Jose Ignacio Hernandez would like to also acknowledge the valuable feedback provided by his fellow PhD researchers to shape this work.

Appendix A. Pseudo-synthetic data generation and parametrisation

A.1. Data-generating processes

We generate four pseudo-synthetic datasets using the experimental design of the COVID-19 PVE choice experiment. The first two datasets are generated using the behavioural assumptions of the portfolio choice model proposed by Bahamonde-Birke and Mouter (2019), whereas the last two datasets are based on the MDCEV-based model proposed by Dekker et al. (2019). Table 11 summarises the utility and stochastic specification of each of the pseudo-synthetic datasets. For datasets 1 and 2, we use the utility specification

Table 13
Hyperparameter values for RF model specification.

Parameter	Values
Number of trees	From 10 to 1,000, in multiples of 10.
Depth	3, 5, 10, max (default).
Maximum variables per split	4, 8, 16, $\sqrt{J * (K+1)}$ (auto)

of Bahamonde-Birke and Mouter (2019) that relies in a linear-in-parameters utility of each possible combination of alternatives, plus the addition of combination-specific stochastic errors (notice that errors are specified as ε_{np}) with a Gumbel (Extreme-Value type 1) distribution. Dataset 1 and 2 differs in the specification of explicit interactions: in dataset 1, we assume that no interactions between chosen alternatives are present (i.e., $\theta_{ij} = 0, \forall i, j$), whereas in dataset 2 we let these parameters free to be estimated. Datasets 3 and 4 are generated using the utility specification of Dekker et al. (2019), hence relying in the assumptions of the MDCEV-type choice model. Apart from differences in the specification of the utility function, the MDCEV-type datasets differ from the former approach in the specification of stochastic terms, which in this case correspond to alternative-specific terms (notice that ε_{nj} are at alternative-level). For dataset 3, we assume i.i.d. Gumbel-distributed terms, whereas for dataset 4 we incorporate unobserved correlation between alternatives by using a Generalised Extreme Value (GEV) distribution.

Table 12 details the values used to parametrise each of the pseudo-synthetic datasets. We define eight alternative-specific constants ranging from −0.9 to 0.5. The attribute-specific parameters are assumed equal across different alternatives and range from −0.8 to 0.03. The parameters associated to the marginal utility of non-spent resources as 0.01 for datasets 1 and 2, and −3 for datasets 3 and 4. In addition, we define positive and negative interactions between chosen alternatives for dataset 2. Specifically, we define a positive interaction when lifting all restrictions to immune people and from Northern provinces (LI and LN) are chosen together, a negative interaction when re-opening all types of businesses (RB, RC and RH) are chosen together, and a negative interaction when allowing visits in nursing homes and allowing contact between direct family members (NH and DF) are chosen together. In the same way, we explicitly define unobserved correlation between alternatives through different so-called dissimilarity parameters of the GEV distribution on dataset 4, varying across consecutive pairs of alternatives.

Appendix B. Hyperparameter tuning of the RF model

We conduct a grid search process to find the best combination of hyperparameters, and we keep the combination that reports the highest test (out-of-sample) log-likelihood. Table 13 presents the values considered for the tuning process. We tested trees ranging from 10 to 1,000 individual decision trees, increasing this number in multiples of ten. In terms of maximum depth, we used three, five, ten and the default setting (max) of the RF model optimisation algorithm. Finally, we fixed the maximum number of variables per split in power values of four, from four to 16, plus the default setting of $\sqrt{J * (K+1)}$, named as "auto". We constructed RF models using all possible combinations of parameters of Table 13.

To identify the best combination of hyperparameters, we proceed in two stages. First, we train each possible RF model under different combinations of hyperparameters. Second, we fix either the tree depth or the maximum number of variables per split, and we plot the (out-of-sample) log-likelihood for different specifications of the other parameter as a function of the maximum number of trees. Finally, we choose the combination of hyperparameters that reports the maximum log-likelihood.

Fig. 8 details the log-likelihood values for different number of variables per split and different number of trees, for a tree depth fixed in five layers, trained with the empirical data. We conclude that a RF model with maximum depth of five layers, 16 variables per split and 200 decision trees lead to the best log-likelihood. We conducted the same process in the pseudo-synthetic datasets, leading to the same combination of hyperparameters.

Appendix C. AR learning outcomes of pseudo-synthetic data

We show that the confidence and lift values of AR learning align with the specification of interactions and unobserved heterogeneity in pseudo-synthetic data. Specifically, we gather association rules from the pseudo-synthetic datasets, and we compare the confidence and lift between datasets without and with such interactions.

Table 14 summarises the support, confidence and lift of a selection of rules in which we explicitly defined interactions or correlations between errors. As expected, incorporating interactions between alternatives in the portfolio choice dataset induce a change of magnitude and direction of the confidence and lift values for a same association rule. For instance, the association rule of lifting restrictions for immune people (LI) and from Northern provinces (LN) has a confidence of 24% and a lift equal to 0.74 in dataset 1 (without interactions), whereas in dataset 2 (with interactions) these values increase to 74% and 1.18, respectively, in line with the positive interaction defined for these two alternatives in dataset 2. We observe the same behaviour for association rules in which we defined negative interactions. Similar patterns are observed in the case of correlated errors in MDCEV-type datasets, in which the incorporation of these correlations induce an increase on confidence and lift values, contrasted with the same rule in the dataset with i.i.d. errors. Notice that incorporating interactions or correlated errors does not necessarily mean that the lift of a rule will be above (below) one. Instead, we observe changes with respect to the lift value computed in the datasets without explicit interactions.

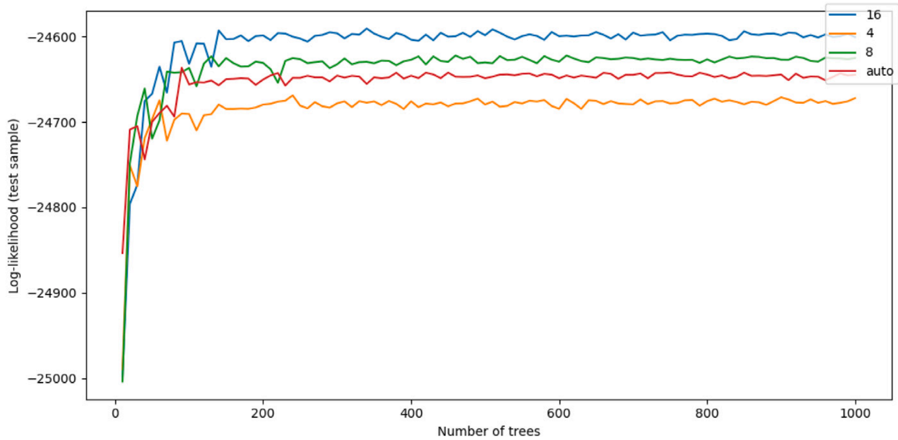


Fig. 8. Log-likelihood of RF models at different parameter specifications, empirical data.

Table 14
Effect of interactions in association rules.

Portfolio choice model							
Association rules		Dataset 1 Without interactions			Dataset 2 With interactions		
Antecedents	Consequents	Support	Confidence	Lift	Support	Confidence	Lift
LI	LN	0.0962	0.2415	0.7147	0.5018	0.7435	1.1881
RC	RB	0.1562	0.4451	0.94	0.0916	0.3519	0.937
RH	RB	0.0815	0.4097	0.8651	0.0354	0.3053	0.8128
RH	RC	0.0555	0.2789	0.7948	0.0219	0.1888	0.7254
RH, RC	RB	0.0199	0.3588	0.7578	–	–	–
DF	NH	0.0884	0.2124	0.8638	0.0383	0.1196	0.7656
NH	DF	0.0884	0.3597	0.8638	0.0383	0.2449	0.7656
MDCEV-type model							
Association rules		Dataset 3 i.i.d. errors			Dataset 4 GEV (correlated errors)		
Antecedents	Consequents	Support	Confidence	Lift	Support	Confidence	Lift
RB	NH	0.1479	0.248	0.9942	0.1857	0.2912	1.216
RC	YP	0.2822	0.6726	0.9867	0.3242	0.768	1.0677
LI	LN	0.1341	0.2897	0.7443	0.1542	0.3374	0.8611
DF	RH	0.0725	0.135	0.8504	0.0751	0.4992	0.927

Table 15
True and predicted log-likelihood values, pseudo-synthetic datasets.

DGP	Portfolio choice model		MDCEV-type model	
	Dataset 1 Without interactions	Dataset 2 With interactions	Dataset 3 i.i.d. errors	Dataset 4 GEV (correlated errors)
True log-likelihood	−134476.81	−110298.15	−113452.41	−97375.57
RF log-likelihood	−135041.43	−111779.66	−119345.83	−113775.94

Note: True log-likelihood of datasets 3 and 4 are computed using 10,000 simulations.

Appendix D. Probability rankings with pseudo-synthetic data

Table 15 shows the predicted log-likelihood of the RF models trained with each pseudo-synthetic datasets (RF log-likelihood), compared with their respective log-likelihood values obtained from the true DGP (True log-likelihood). The RF model is able to get close to the true DGP with a considerable precision in datasets 1 and 2, and with more distance in datasets 3 and 4. This distance between the true and predicted log-likelihood in the latter datasets can be attributed by slight differences in the choice probabilities across different combinations of alternatives, as well as the different error structures imposed in these datasets, compared with datasets 1 and 2.

Table 16 summarises the Kendall's Tau values resulting from comparing the top-5 and top-10 rankings of the trained RF models with their respective rankings obtained from the true DGP, for each pseudo-synthetic dataset. We observe that the Kendall's Tau of the top-5 portfolio is close to one in three out of four datasets (Datasets 2, 3 and 4), which means that the trained RF model

Table 16
Kendall's Tau correlation between most likely chosen portfolios and "true" rankings. Pseudo-synthetic datasets.

DGP	Portfolio choice model		MDCEV-type model	
	Dataset 1 Without interactions	Dataset 2 With interactions	Dataset 3 i.i.d. errors	Dataset 4 GEV (correlated errors)
Top-5	0.583**	$\sim$ 1.000**	$\sim$ 1.000**	$\sim$ 1.000**
Top-10	0.513**	$\sim$ 1.000**	0.800**	0.500**

P-values of Kendalls Tau: ** : $p < 0.001$, * : $p < 0.01$.

Table 17
Probability ranking, dataset 1.

		Rk. 1	Rk. 2	Rk. 3	Rk. 4	Rk. 5	Rk. 6	Rk. 7	Rk. 8	Rk. 9	Rk. 10
	NH										
	RB	X	X	X		X	X	X			
	RC										
	YP	X	X	X	X		X		X		X
True	LI	X		X		X			X	X	X
	LN		X		X			X	X	X	
	DF	X			X	X	X				X
	RH										
	Choice prob.	0.018	0.017	0.017	0.016	0.015	0.014	0.014	0.014	0.014	0.014
	Pressure	42.0%	38.49%	31.65%	38.99%	35.46%	26.74%	31.96%	43.91%	37.37%	32.15%
	NH										
	RB	X	X	X		X		X			X
	RC										
	YP	X	X	X	X	X	X		X		
RF	LI	X	X				X		X	X	X
	LN			X	X			X	X	X	
	DF		X		X	X	X				X
	RH										
	Choice prob.	0.018	0.018	0.017	0.016	0.015	0.015	0.015	0.014	0.014	0.013
	Pressure	31.65%	42.0%	38.49%	38.99%	26.74%	32.15%	31.96%	43.91%	37.37%	35.46%

Table 18
Probability ranking, dataset 2.

		Rk. 1	Rk. 2	Rk. 3	Rk. 4	Rk. 5	Rk. 6	Rk. 7	Rk. 8	Rk. 9	Rk. 10
	NH								X		
	RB			X			X			X	
	RC					X					
	YP		X				X	X			
True	LI	X	X	X	X	X	X	X	X	X	X
	LN	X	X	X	X	X	X	X	X	X	X
	DF				X			X		X	
	RH										X
	Choice prob.	0.098	0.081	0.059	0.05	0.042	0.034	0.025	0.02	0.015	0.012
	Pressure	37.37%	43.91%	47.22%	47.72%	48.86%	53.76%	54.26%	55.33%	57.57%	56.44%
	NH								X		
	RB			X			X			X	
	RC					X					
	YP		X				X	X			
RF	LI	X	X	X	X	X	X	X	X	X	X
	LN	X	X	X	X	X	X	X	X	X	X
	DF				X			X		X	
	RH										X
	Choice prob.	0.098	0.082	0.058	0.05	0.043	0.035	0.025	0.02	0.016	0.012
	Pressure	37.37%	43.91%	47.22%	47.72%	48.86%	53.76%	54.26%	55.33%	57.57%	56.44%

is able to retrieve the true ranking in this case, whereas in Dataset 1 the Kendall's Tau is of 58,3%, but still statistically different from zero, which suggests a correlation between the prediction of the RF model and the true DGP. The Kendall's Tau values of the top-10 portfolios show a decrease of predictive power on this ranking, which can be explained due to slight changes of position of some combinations. Despite the latter, all correlation values are statistically different from zero, and thus still suggesting the existence of correlation between the predicted and true probability rankings. We provide a detail of the rankings for each dataset in Tables 17 20.

Table 19
Probability ranking, dataset 3.

		Rk. 1	Rk. 2	Rk. 3	Rk. 4	Rk. 5	Rk. 6	Rk. 7	Rk. 8	Rk. 9	Rk. 10
	NH										
	RB	X	X	X	X	X		X	X		X
	RC			X	X	X	X				X
	YP	X	X	X	X	X	X	X	X	X	X
True	LI	X			X	X	X		X		
	LN		X					X	X	X	X
	DF	X	X	X	X		X			X	
	RH										
	Choice prob.	0.041	0.032	0.028	0.025	0.024	0.021	0.019	0.018	0.018	0.017
	Pressure	42.0%	48.84%	38.23%	53.5%	43.15%	43.65%	38.49%	53.76%	38.99%	49.99%
	NH										
	RB	X	X	X	X	X	X		X	X	
	RC			X	X	X		X		X	
	YP	X	X	X	X	X	X	X	X	X	X
RF	LI	X			X	X		X	X		X
	LN		X				X		X	X	X
	DF	X	X	X	X			X			
	RH										
	Choice prob.	0.042	0.032	0.029	0.025	0.025	0.02	0.02	0.018	0.018	0.017
	Pressure	42.0%	48.84%	38.23%	53.5%	43.15%	38.49%	43.65%	53.76%	49.99%	43.91%

Table 20
Probability ranking, dataset 4.

		Rk. 1	Rk. 2	Rk. 3	Rk. 4	Rk. 5	Rk. 6	Rk. 7	Rk. 8	Rk. 9	Rk. 10
	NH								X		
	RB	X	X	X	X	X		X	X	X	
	RC		X		X	X	X		X	X	
	YP	X	X	X	X	X	X	X	X	X	X
True	LI	X			X	X	X	X			X
	LN			X				X		X	X
	DF	X	X	X	X		X		X		
	RH										
	Choice prob.	0.04	0.035	0.031	0.029	0.028	0.024	0.022	0.021	0.02	0.019
	Pressure	42.0%	38.23%	48.84%	53.5%	43.15%	43.65%	53.76%	56.19%	49.99%	43.91%
	NH							X			
	RB	X	X	X	X	X		X	X	X	
	RC		X		X	X	X	X		X	
	YP	X	X	X	X	X	X	X	X	X	X
RF	LI	X			X	X	X		X		X
	LN			X					X	X	X
	DF	X	X	X	X		X	X			
	RH										
	Choice prob.	0.041	0.037	0.032	0.03	0.027	0.023	0.023	0.022	0.021	0.019
	Pressure	42.0%	38.23%	48.84%	53.5%	43.15%	43.65%	56.19%	53.76%	49.99%	43.91%

References

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. pp. 207–216. http://dx.doi.org/10.1145/170036.170072.

Alwosheel, A., van Cranenburgh, S., Chorus, C.G., 2021. Why did you predict that? Towards explainable artificial neural networks for travel demand analysis. Transp. Res. C 128, 103143. http://dx.doi.org/10.1016/j.trc.2021.103143.

Bahamonde-Birke, F.J., Mouter, N., 2019. About positive and negative synergies of social projects: treating correlation in participatory value evaluation. In: HEART 2019: 8th Symposium of the European Association for Research in Transportation. Budapest, Hungary. URL https://transp-or.epfl.ch/heart/2019/abstracts/hEART_2019_paper_166.pdf.

Breiman, L., 2001. Random forests. Mach. learn. 45 (1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Caputo, V., Lusk, J.L., 2022. The basket-based choice experiment: A method for food demand policy analysis. Food Policy 109, 102252. http://dx.doi.org/10.1016/j.foodpol.2022.102252.

Carson, R.T., Eagle, T.C., Islam, T., Louviere, J.J., 2022. Volumetric choice experiments (VCEs). J. Choice Model. 42, 100343. http://dx.doi.org/10.1016/j.jocm.2022.100343.

Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behav. Soc. 14, 1–10. http://dx.doi.org/10.1016/j.tbs.2018.09.002.

Dekker, T., Koster, P., Mouter, N., 2019. The economics of participatory value evaluation. Tinbergen Institute Discussion Paper 2019-008/VIII, http://dx.doi.org/10.2139/ssrn.3323645.

Friedman, J., Hastie, T., Tibshirani, R., et al., 2001. The elements of statistical learning, Vol. 1, no.10. Springer series in statistics New York, http://dx.doi.org/10.1007/978-0-387-21606-5.

Geurts, K., Wets, G., Brijs, T., Vanhoof, K., 2003. Profiling of high-frequency accident locations by use of association rules. Transp. Res. Rec. 1840 (1), 123–130. http://dx.doi.org/10.3141/1840-14.

Hernandez, J.I., Mouter, N., Itten, A., 2021. Participatory value evaluation for relaxation of COVID-19 measures. In: 4TU.ResearchData. http://dx.doi.org/10.4121/14413958.v1.

Hillel, T., Bierlaire, M., Elshafie, M., Jin, Y., 2019. Weak teachers: Assisted specification of discrete choice models using ensemble learning. In: HEART 2019: 8th Symposium of the European Association for Research in Transportation. Budapest, Hungary, URL https://transp-or.epfl.ch/documents/proceedings/HilBieElsJin_HEART2019.pdf.

Kaur, M., Kang, S., 2016. Market basket analysis: Identify the changing trends of market data using association rule mining. Procedia Comput. Sci. 85, 78–85. http://dx.doi.org/10.1016/j.procs.2016.05.180.

Kendall, M.G., 1945. The treatment of ties in ranking problems. Biometrika 33 (3), 239–251. http://dx.doi.org/10.1093/biomet/33.3.239.

Keuleers, B., Wets, G., Arentze, T., Timmermans, H., 2001. Association rules in identification of spatial-temporal patterns in multiday activity diary data. Transp. Res. Rec. 1752 (1), 32–37. http://dx.doi.org/10.3141/1752-05.

Lerman, S.R., 1976. Location, housing, automobile ownership, and mode to work: a joint choice model. Transp. Res. Rec. 610, 6–11.

Mouter, N., Hernandez, J.I., Itten, A.V., 2021a. Public participation in crisis policymaking. How 30,000 Dutch citizens advised their government on relaxing COVID-19 lockdown measures. PLoS One 16 (5), e0250614. http://dx.doi.org/10.1371/journal.pone.0250614.

Mouter, N., Koster, P., Dekker, T., 2021b. Contrasting the recommendations of participatory value evaluation and cost-benefit analysis in the context of urban mobility investments. Transp. Res. Part A: Policy Prac. 144, 54–73. http://dx.doi.org/10.1016/j.tra.2020.12.008.

Mulderij, L.S., Hernandez, J.I., Mouter, N., Verkooijen, K.T., Wagemakers, A., 2021. Citizen preferences regarding the public funding of projects promoting a healthy body weight among people with a low income. Soc. Sci. Med. 280, 114015. http://dx.doi.org/10.1016/j.socscimed.2021.114015.

Neill, C.L., Lahne, J., 2022. Matching reality: A basket and expenditure based choice experiment with sensory preferences. J. Choice Model. 44, 100369. http://dx.doi.org/10.1016/j.jocm.2022.100369.

Ortelli, N., Hillel, T., Pereira, F.C., de Lapparent, M., Bierlaire, M., 2021. Assisted specification of discrete choice models. J. Choice Model. 39, 100285. http://dx.doi.org/10.1016/j.jocm.2021.100285.

Rotteveel, A., Lambooij, M., Over, E., Hernandez, J., Suijkerbuijk, A., de Blaeij, A., de Wit, G., Mouter, N., 2022. If you were a policymaker, which treatment would you disinvest? A participatory value evaluation on public preferences for active disinvestment of health care interventions in the netherlands. Health Econ. Policy Law 1–16. http://dx.doi.org/10.1017/S174413312200010X.

Shiftan, Y., Bekhor, S., 2020. Utilizing a random forest classifier for a methodological-iterative discrete choice model specification and estimation. In: HEART 2020: 9th Symposium of the European Association for Research in Transportation. Lyon, France, URL https://transp-or.epfl.ch/heart/2020/abstracts/HEART_2020_paper_151.pdf.

Sifringer, B., Lurkin, V., Alahi, A., 2020. Enhancing discrete choice models with representation learning. Transp. Res. B 140, 236–261. http://dx.doi.org/10.1016/j.trb.2020.08.006.

van Cranenburgh, S., Kouwenhoven, M., 2020. An artificial neural network based method to uncover the value-of-travel-time distribution. Transportation 1–39. http://dx.doi.org/10.1007/s11116-020-10139-3.

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., Walker, J., 2022. Choice modelling in the age of machine learning-discussion paper. J. Choice Model. 42, 100340. http://dx.doi.org/10.1016/j.jocm.2021.100340.

Wang, S., Wang, Q., Zhao, J., 2020. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. Transp. Res. C 118, 102701. http://dx.doi.org/10.1016/j.trc.2020.102701.

Wiley, J.B., Timmermans, H.J., 2009. Modelling portfolio choice in transportation research. Transp. Rev. 29 (5), 569–586. http://dx.doi.org/10.1080/01441640902829462.

Yao, R., Bekhor, S., 2020. Data-driven choice set generation and estimation of route choice models. Transp. Res. C 121, 102832. http://dx.doi.org/10.1016/j.trc.2020.102832.