The Journal of Open Source Software

validateHOT - Validate your Holdout Task

Joshua Schramm[1] **and Marcel Lichters**[1, 2]

1 Chemnitz University of Technology, Germany **2** Otto von Guericke University of Magdeburg, Germany

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC-BY).

Summary

validateHOT is a package that provides functions to both validate a validation/holdout task and run market simulations for results obtained in a (adaptive) choice-based conjoint analysis (hereafter ACBC and CBC, respectively) and maximum difference scaling (hereafter MaxDiff) using Sawtooth Software.[1]

Preference measurement techniques', such as (A)CBC or MaxDiff, ultimate goal is to predict future behavior (Green and Srinivasan 1990). Hence, it is essential for both academics and practitioners to ensure that the collected data is valid and can also predict outside tasks (i.e., the model has external validity). In terms of external validity, we refer to the generalizations to different settings (see, Calder, Phillips, and Tybout 1982, 240). The easiest way to test this is to include so-called validation or holdout task (Orme 2015), which are tasks that are fixed (i.e., same across participant) and are usually not used for estimating the part-worth utilities. Despite the important role of validation tasks, practitioners often do not include them (Yang, Toubia, and Jong 2018), which is unsatisfactory, given the fact that the model is used to estimate market shares which poses the basis for relevant marketing decisions.

validateHOT combines both validation and market simulation in one package. validateHOT has three key advantages: a) it helps you to decide which is the best model to proceed by validating it, b) it runs relevant market simulations that help to find the right product combinations, and finally, c) it is an open source tool including functions that are usually implemented in paid software, and therefore, remain a black-box for researchers and practitioners.

Statement of need

validateHOT is a practical tool for Sawtooth Software users in industry as well as academia. It provides an open source solution for a) validating a validation/holdout task and therefore ensuring that the model has predictive validity; b) running market simulations (e.g., **T**otal **U**nduplicated **R**each and **F**requency, hereafter TURF). Other packages, for example, Metrics (Hamner and Frasco 2018) provide functions to run validation metrics such as *mean absolute error*, *root mean squared error*, or the five metrics of the confusion matrix (see Table 1). However, to put the Sawtooth export into the right format, the user needs some data wrangling which could pose a barrier. Moreover, packages mainly focus on the analysis of conjoint analysis (e.g., ChoiceModelR

[1] validateHOT also works with MaxDiff raw logit utilities, see, for example, Chapman and Rodden (2023). For CBC, it should also work with other sofware, as long as all attributes are part-worth coded. Linear and piecewise coding is implemented to work with Sawtooth Software.

(Sermas 2022), choicetools (Chapman et al. 2023), logitR (Helveston 2023), bayesm (Rossi 2023) etc.). To the best of our knowledge, a package that converts raw utility scores into validation metrics or running a variety of marketing simulations (especially TURF) is still missing. validateHOT is especially helpful for researchers who plan to report results for research articles according to open science standards. In addition, since the former package turfR is no longer available on the Comprehensive R Archive Network (see CRAN), academics and practitioners do not have an alternative function in *R* to run this method. Therefore, currently practitioners and academics mainly have to stick to paid solutions.

Key functions

validateHOT's functions can be categorized into four main components, see Table 1. To bring the data into the right format, users can run the `createHOT` function, which creates the total utility of each alternative by applying the additive utility model (Rao 2014, 82). `turf` as well as the 2 rescaling functions, however, are not dependent on `createHOT`, and can be run using the raw logit scores.

Table 1: Overview of main four components of validateHOT and their corresponding functions

Validation metrics	Confusion matrix	Market simulations	Rescaling scores
hitrate()	accuracy()	freqassort()	att_imp()
kl()	f1()	marksim()	prob_scores()
mae()	precision()	reach()	
medae()	recall()	turf()	
mhp()	specificity()		
rmse()			

Typical workflow

In the following, we provide the workflow for a MaxDiff study (the vignette also provides detailed examples for a CBC as well as an ACBC).

After running the Hierarchical Bayes estimation in, for example, Sawtooth Software, the **raw** utility scores have to be exported and afterwards read into *R* as a data frame. This data frame also has to include the actual choice in the validation/holdout task.

To define the validation/holdout task, which has a total of 7 items (`prod`) plus the no-buy alternative (`none`), we use the `createHOT` function. Here, the user can define the attributes as well as the method (here `MaxDiff`).

```r
data("MaxDiff") # read in the data
HOT <- createHOT(
  data = MaxDiff, # data frame
  id = 1, # index unique identifier
  none = 19, # index of none alternative
  prod = 7, # no of alternatives in HOT excluding none
  prod.levels = list(3, 10, 11, 15, 16, 17, 18), # index of alternatives
  method = "MaxDiff", # method applied
  choice = 20, # column index of choice alternative
  varskeep = 21
)
```

Next, to get the relevant validation metrics that are often reported in conjoint studies, for example, hit rate (e.g., Ding, Grewal, and Liechty 2005), mean hit probability (mhp, Voleti, Srinivasan, and Ghosh 2017), or mean absolute error (mae, Wlömert and Eggers 2014), we only need to provide the data, the alternatives in the validation/holdout task (`opts`), and the actual choice (`choice`). Everything can be implemented in the tidyverse (Wickham et al. 2019) logic.

```
hitrate(
  data = HOT, # data frame
  opts = c(Option_1:None), # column names of alternatives
  choice = choice # column name of choice
) %>%
  round(3)
```

```
## # A tibble: 1 x 5
##      HR    se chance   cor     n
##   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1  55.7  5.98   12.5    39    70
```

validateHOT also provides the five metrics for the confusion matrix. The underlying logic hereby is that the user has to provide a no-buy alternative (`none`). validateHOT calculates, for example, how often a buy or no-buy was correctly predicted, therefore, it is testing whether the model correctly predicts general demand (exemplary showed by applying the `accuracy` function and results split by `Group`).

```
accuracy(
  data = HOT, # data frame
  group = Group, # optional grouping variable
  opts = c(Option_1:None), # column names of alternatives
  choice = choice, # column name of choice
  none = None # column name
)
```

```
## # A tibble: 3 x 2
##   Group accuracy
##   <int>    <dbl>
## 1     1     73.9
## 2     2     72
## 3     3     63.6
```

Finally, we show two functions for market simulations, namely `marksim` as well as `turf`. First, we calculate the market shares based on the multinomial logit model (McFadden 1974). Besides the aggregated shares, `marksim` also provides standard errors and the 95% confidence interval.

```
marksim(
  data = HOT,
  opts = c(Option_1:None),
  method = "sop"
)
```

```
## # A tibble: 8 x 5
##   Option      mw    se  lo.ci up.ci
```

```
##   <chr>    <dbl> <dbl>  <dbl> <dbl>
## 1 Option_1 18.3  4.12  10.2   26.3
## 2 Option_2 11.3  2.69   6.05  16.6
## 3 Option_3  4.08 1.49   1.16   6.99
## 4 Option_4 32.5  4.45  23.8   41.2
## 5 Option_5  1.93 0.916  0.131  3.72
## 6 Option_6 10.4  2.68   5.12  15.6
## 7 Option_7  5.58 1.75   2.15   9.01
## 8 None     16.0  3.29   9.53  22.4
```

`turf`, a "product line extension model" (Miaoulis, Parsons, and Free 1990, 29) is a tool to find the perfect assortment that creates the highest reach and a powerful tool especially for MaxDiff studies (Chrzan and Orme 2019, 108). To optimize the search for the optimal bundle, we also include the arguments `fixed`, to define alternatives that have to be part of the assortment, as well as `prohib`, to define prohibitions of combinations of items that should not be part of the assortment (please see the vignette for more details as well as how to apply `turf` with data obtained using a likert scale).

For the following example, let us assume that the user conducted an anchored MaxDiff analysis with 10 items (`opts`) and now wants to find the best assortment with a size of 3 (`size = 3`). As a threshold (`none`) the user uses the anchor (no-buy alternative).

```r
turf(
  data = MaxDiff, # define data
  opts = c(Option_01:Option_10), # define items
  none = none, # define threshold variable
  size = 3, # define size of assortment
  approach = "thres" # define approach
) %>%
  head(., n = 5) %>%
  mutate_if(is.numeric, round, 2) %>%
  t() %>%
  as.data.frame() %>%
  slice(-1) %>%
  rename_all(., ~ paste0("Combo ", c(1:5)))
```

```
##           Combo 1 Combo 2 Combo 3 Combo 4 Combo 5
## reach       82.86   81.43   81.43   81.43   80.00
## freq         1.46    1.57    1.43    1.41    1.44
## Option_01       1       1       1       1       1
## Option_02       0       0       1       0       0
## Option_03       0       1       0       0       0
## Option_04       1       0       1       1       0
## Option_05       0       0       0       0       0
## Option_06       1       1       0       0       1
## Option_07       0       0       0       0       0
## Option_08       0       0       0       0       1
## Option_09       0       0       0       0       0
## Option_10       0       0       0       1       0
```

Availability

validateHOT is available on Github.

References

Calder, Bobby J., Lynn W. Phillips, and Alice M. Tybout. 1982. "The Concept of External Validity." *Journal of Consumer Research* 9 (3): 240–44. https://doi.org/10.1086/208920.

Chapman, Chris, Eric Bahna, James Alford, and Steven Ellis. 2023. "Choicetools: Tools for Choice Modeling, Conjoint Analysis, and MaxDiff Analysis of Best-Worst Surveys."

Chapman, Chris, and Kerry Rodden. 2023. *Quantitative User Experience Research: Informing Product Decisions by Understanding Users at Scale*. Berkeley, CA: Apress; Springer Science+Business Media LLC. https://doi.org/10.1007/978-1-4842-9268-6.

Chrzan, Keith, and Bryan K. Orme. 2019. *Applied MaxDiff: A Practitioner's Guide to Best-Worst Scaling*. Provo, UT: Sawtooth Software.

Ding, Min, Rajdeep Grewal, and John Liechty. 2005. "Incentive-Aligned Conjoint Analysis." *Journal of Marketing Research* 42 (1): 67–82. https://doi.org/10.1509/jmkr.42.1.67.56890.

Green, Paul E., and V. Srinivasan. 1990. "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice." *Journal of Marketing* 54 (4): 3–19. https://doi.org/10.1177/002224299005400402.

Hamner, Ben, and Michael Frasco. 2018. "Metrics: Evaluation Metrics for Machine Learning." https://CRAN.R-project.org/package=Metrics.

Helveston, John Paul. 2023. "{Logitr}: Fast Estimation of Multinomial and Mixed Logit Models with Preference Space and Willingness-to-Pay Space Utility Parameterizations" 105. https://doi.org/10.18637/jss.v105.i10.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka, 105–42. Economic Theory and Mathematical Economics. New York: Academic Press.

Miaoulis, George, Henry Parsons, and Valerie Free. 1990. "Turf: A New Planning Approach for Product Line Extensions." *Marketing Research* 2 (1): 28–40.

Orme, Bryan K. 2015. "Including Holdout Choice Tasks in Conjoint Studies."

Rao, Vithala R. 2014. *Applied Conjoint Analysis*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87753-0.

Rossi, Peter. 2023. "Bayesm: Bayesian Inference for Marketing/Micro-Econometrics." https://CRAN.R-project.org/package=bayesm.

Sermas, Ryan. 2022. "ChoiceModelR: Choice Modeling in r." https://CRAN.R-project.org/package=ChoiceModelR.

Voleti, Sudhir, V. Srinivasan, and Pulak Ghosh. 2017. "An Approach to Improve the Predictive Power of Choice-Based Conjoint Analysis." *International Journal of Research in Marketing* 34 (2): 325–35. https://doi.org/10.1016/j.ijresmar.2016.08.007.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the {Tidyverse}" 4: 1686. https://doi.org/10.21105/joss.01686.

Wlömert, Nils, and Felix Eggers. 2014. "Predicting New Service Adoption with Conjoint Analysis: External Validity of BDM-Based Incentive-Aligned and Dual-Response Choice Designs." *Marketing Letters* 27 (1): 195–210. https://doi.org/10.1007/s11002-014-9326-x.

Yang, Liu (Cathy), Olivier Toubia, and Martijn G. de Jong. 2018. "Attention, Information Processing, and Choice in Incentive-Aligned Choice Experiments." *Journal of Marketing Research* 55 (6): 783–800. https://doi.org/10.1177/0022243718817004.