The Journal of Open Source Software

validateHOT - an R package for holdout task validation and market simulations

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

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

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 23 January 2024
Published: unpublished

Summary

validateHOT is an R 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, for example, ChoiceModelR (Sermas, 2022) or Sawtooth's Lighthouse Studio.

Statement of need

Preference measurement techniques' (e.g., (A)CBC or MaxDiff) aim is to predict behavior (Green & Srinivasan, 1990). Hence, it is essential to ensure that the collected data is valid and predicts outside tasks (i.e., the model has external validity) well. The easiest way for testing validity is by including validation tasks (Orme, 2015; e.g., Rao, 2014), which are fixed tasks (i.e., same across participants) and not used for estimating the part-worth utilities (raw logit utilities) in hierarchical Bayes (HB) estimation. Despite their importance, practitioners don't always include them (Yang et al., 2018). This 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 and has three key advantages, it (1) helps opting for the best model and (2) runs relevant market simulations that help finding the right product combinations or assortments, and (3) is an open source tool which helps especially researchers reporting accompanied scripts for their research papers.

State of the field in R

Other packages provide functions to calculate validation metrics, however, these are not always specified for individual part-worth utilities. Metrics (Hamner & Frasco, 2018), for example, provide functions to run validation metrics such as *mean absolute error*, *root mean squared error*, or the five metrics of the confusion matrix. However, to get the output of, for example, Sawtooth Software or ChoiceModelR (Sermas, 2022) into the right format, the user needs some data wrangling. The package conjoint (Bak & Bartlomowicz, 2012) provides functions that are most similar to validateHOT's ones, but no validation functions are included and the package focuses on classical conjoint analysis, thus it is limited when applying more common conjoint methods. support.BWS (Aizaki & Fogarty, 2023) only covers best-worst scaling case 1 (i.e., MaxDiff). logitr (Helveston, 2023) provides market simulations tools, however, no validation metrics such as mean hit probability (Voleti et al., 2017) or hit rate (Netzer & Srinivasan, 2011). Figure 1 shows a comparison of validateHOT's functions with current R packages. 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 missing.

Key functionalities	validateHOT (v 1.0.0)	Metrics (v 0.1.4)	caret (v 6.0-94)	conjoint (v 1.4)	philentropy (v 0.7.0)	logitr (v 1.1.1)	mlogit (v 1.1-1)	support.bws (v 0.4-6)
Confusion matrix	✓	✓	✓					
Creating design matrix				✓		✓	✓	✓
Creating holdout / market scenario	✓			✓		✓		
Estimate utilities				✓		✓	✓	✓
Estimate WTP						✓		
Hit rate	✓							
Kullblack-Leibler-Divergence	✓				✓			
MAE, MedAe, RMSE	✓	✓	✓					
Market Shares	✓		✓	✓		✓		✓
Mean hit probability	✓							
TURF	✓							

Figure 1: Comparison of validateHOT's function to existing R packages

validateHOT is introduced with data estimated with Lighthouse Studio using effects-coding for creating the design matrix. It, however, can easily be used with data estimated with ChoiceModelR (Sermas, 2022), bayesm (Rossi, 2023), or STAN (2023), if used with similar settings (ChoiceModelR, for example, automatically implements effects-coding).

Key functions

validateHOT's functions can be categorized into four main components, see Table 1. To bring the data into the right format for some functions, the `createHOT()` function can be applied, which creates each alternatives' total utility by applying the additive utility model (Rao, 2014, p. 82).

Table 1: Overview of validateHOT's main components 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()	zc_diffs()
medae()	recall()	turf()	zero_anchored()
mhp()	specificity()		
rmse()			

Typical workflow

We provide the workflow for a MaxDiff study and a CBC study with only part-worth coded attributes (the vignette provides detailed examples for other CBCs and an ACBC).

MaxDiff

Creating Holdout Task / Market Scenario

After running the HB estimation (Allenby & Ginter, 1995; Lenk et al., 1996), the **raw** utility scores have to be exported and read into an *R* data frame. Assuming you included a validation task with seven alternatives plus the no-buy alternative (`none`). To create this validation task in *R*, we use the `createHOT()` function.

```
HOT <- createHOT(
  data = MaxDiff,
  id = "ID",
  none = "none",
  prod.levels = list(3, 10, 11, 15, 16, 17, 18),
  method = "MaxDiff",
  choice = "HOT",
  varskeep = "Group"
)
```

Validating Holdout Task

To get the relevant validation metrics that are reported in conjoint studies, for example, hit rate (e.g., Ding et al., 2005) or mean hit probability (mhp, Voleti et al., 2017), we provide the data, the alternatives in the validation task (`opts`), and the actual choice (`choice`). The function can be implemented using the tidyverse (Wickham et al., 2019) logic.

```
hitrate(
  data = HOT,
  opts = c(Option_1:None),
  choice = choice
) %>%
  round(2)
```

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

Market Simulations

We also introduce two functions for market simulations, namely `marksim()` and `turf()`. In the following example, the market share is calculated according to the multinomial logit model (McFadden, 1974).

```
marksim(
  data = HOT,
  opts = c(Option_1:None),
  method = "sop"
) %>%
  mutate_if(is.numeric, round, 2)
```

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

The Journal of Open Source Software

```
##   <chr>     <dbl> <dbl> <dbl> <dbl>
## 1 Option_1 18.3   4.12 10.2  26.4
## 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  0.92  0.13  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
```

Next, `turf()`, a "product line extension model" (Miaoulis et al., 1990, p. 29), is a tool to find the perfect assortment that creates the highest reach and is especially powerful for MaxDiff studies (Chrzan & Orme, 2019, p. 108). To optimize the search for the optimal assortment, we also include the arguments `fixed`, to define alternatives that have to be part of the assortment, and `prohib`, to prohibit certain item combinations in the assortment (see the vignette for more details and how to apply `turf()` with data obtained using a likert scale).

For the following example, let's 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 items. The user uses the anchor (no-buy alternative) as a threshold.

```
turf(
  data = MaxDiff,
  opts = c(Option_01:Option_10),
  none = none,
  size = 3,
  approach = "thres"
) %>%
  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
```

CBC

Creating Holdout Task / Market Scenario

The setup is almost the same, only the arguments `prod.levels`, `coding`, and `method` are different or new, respectively.

```
HOT_CBC <- createHOT(
  data = CBC,
```

```r
  id = "ID",
  none = "none",
  prod.levels = list(c(4, 9, 19), c(8, 12, 17), c(5, 10, 17)),
  coding = c(0, 0, 0),
  method = "CBC",
  choice = "HOT"
)
```

Rescaling Scores

We can also display the attributes importance scores. Therefore, we need to define the attribute levels as well as the coding of the attributes.

```r
att_imp(
  data = CBC,
  attrib = list(
    c(4:8),
    c(9:13),
    c(14:20)
  ),
  coding = c(rep(0, 3)),
  res = "agg"
) %>%
  mutate_if(is.numeric, round, 2)
```

```
## # A tibble: 3 x 3
##   Option        mw   std
##   <chr>      <dbl> <dbl>
## 1 att_imp_1   35.7 11.3
## 2 att_imp_2   27.7 10.0
## 3 att_imp_3   36.6  9.32
```

Availability

validateHOT is available on Github.

Acknowledgments

We would like to thank Sawtooth Software for their great transparent documentation.

References

Aizaki, H., & Fogarty, J. (2023). *R packages and tutorial for case 1 best-worst scaling*. *46*, 100394. https://doi.org/10.1016/j.jocm.2022.100394

Allenby, G. M., & Ginter, J. L. (1995). Using Extremes to Design Products and Segment Markets. *Journal of Marketing Research*, *32*(4), 392–403. https://doi.org/10.1177/002224379503200402

Bak, A., & Bartlomowicz, T. (2012). *Conjoint analysis method and its implementation in conjoint r package*. 239–248. http://keii.ue.wroc.pl/pracownicy/tb/Bak_A_and_Bartlomowicz_T_Conjoint_analysis_method_and_its_implementation_in_conjoint_R_package.pdf

Chrzan, K., & Orme, B. K. (2019). *Applied MaxDiff: A practitioner's guide to best-worst scaling*. Sawtooth Software. ISBN: 9780999367728

Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-Aligned Conjoint Analysis. *Journal of Marketing Research*, *42*(1), 67–82. https://doi.org/10.1509/jmkr.42.1.67.56890

Green, P. E., & Srinivasan, V. (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, B., & Frasco, M. (2018). *Metrics: Evaluation metrics for machine learning*. https://CRAN.R-project.org/package=Metrics

Helveston, J. P. (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

Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science*, *15*(2), 173–191. https://doi.org/10.1287/mksc.15.2.173

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press. ISBN: 0127761500

Miaoulis, G., Parsons, H., & Free, V. (1990). Turf: A new planning approach for product line extensions. *Marketing Research*, *2*(1), 28–40.

Netzer, O., & Srinivasan, V. (2011). Adaptive Self-Explication of Multiattribute Preferences. *Journal of Marketing Research*, *48*(1), 140–156. https://doi.org/10.1509/jmkr.48.1.140

Orme, B. K. (2015). *Including holdout choice tasks in conjoint studies*.

Rao, V. R. (2014). *Applied conjoint analysis*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87753-0

Rossi, P. (2023). *Bayesm: Bayesian inference for marketing/micro-econometrics*. https://CRAN.R-project.org/package=bayesm

Sermas, R. (2022). *ChoiceModelR: Choice modeling in r*. https://CRAN.R-project.org/package=ChoiceModelR

Stan Development Team. (2023). *{RStan}: The {r} interface to {stan}*. https://mc-stan.org/

Voleti, S., Srinivasan, V., & Ghosh, P. (2017). An approach to improve the predictive power of choice-based conjoint analysis. *International Journal of Research in Marketing*, *34*(2), 325–335. https://doi.org/10.1016/j.ijresmar.2016.08.007

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). *Welcome to the {tidyverse}*. *4*, 1686. https://doi.org/10.21105/joss.01686

Yang, L. (Cathy)., Toubia, O., & Jong, M. G. de. (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