The Journal of Open Source Software

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

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

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

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

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, for example, `ChoiceModelR` (Sermas 2022) or Sawtooth's Lighthouse Studio.

Statement of need

Preference measurement techniques', such as (A)CBC or MaxDiff, aim is to predict behavior (Green and Srinivasan 1990). Hence, it is essential for both academics and practitioners to ensure that the collected data is valid and predicts outside tasks (i.e., the model has external validity) well.[1] The easiest way for testing validity is by including so-called validation or holdout tasks (Orme 2015), which are tasks that are fixed (i.e., same across participants) and are usually not used for estimating the part-worth utilities in hierarchical Bayes estimation. 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 and has three key advantages, it (1) helps to opt for the best model, (2) runs relevant market simulations that help to find the right product combinations or assortments, and finally, (3) is an open source tool including functions that are usually implemented in property commercial, and therefore, remain a black-box for researchers and practitioners.

State of the field in R

Other packages provide functions to calculate validation metrics, however, these are not specified for individual raw logit coefficients which are usually the output when running random parameter logit / hierarchical Bayes models. Metrics (Hamner and 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

[1] In terms of external validity, we refer to the generalizations to different settings (see, Calder, Phillips, and Tybout 1982, 240).

right format, the user needs some data wrangling. The package conjoint (Bak and Bartlomowicz 2012) provides functions that are most similar to validateHOT's ones. However, no functions for validation are included and moreover, conjoint (Bak and Bartlomowicz 2012) mainly focuses on classical conjoint analysis. support.BWS (Aizaki and Fogarty 2023) only covers best-worst scaling case 1 (also known as Maximum Difference Scaling) and only provides market simulations based on conditional logit rule. logitr (Helveston 2023) provides market simulations tools, however, no validation metrics such as mean hit probability (Voleti, Srinivasan, and Ghosh 2017) or hit rate (Netzer and Srinivasan 2011). Lastly, conjoint (Bak and Bartlomowicz 2012) also provides functions for both market simulations, A comparison of validateHOT's functions with current R packages is provided in Figure 1. 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.

validateHOT is introduced with data estimated with Lighthouse Studio using effects-coded data for the design matrix, 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 functionalities	validateHOT (v 0.0.0.9)	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

Key functions

validateHOT's functions can be categorized into four main areas, 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 four rescaling functions, however, are not dependent on `createHOT()`, and can be run using the raw logit scores.

Table 1: Overview of main four areas 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()	zc_diffs()
medae()	recall()	turf()	zero_anchored()
mhp()	specificity()		
rmse()			

Typical workflow

In the following, we provide the workflow for a MaxDiff study and a CBC study with only part-worth coded attributes (the vignette also provides detailed examples for a CBC including linear-coded attributes as well as an ACBC).

MaxDiff

After running the hierarchical Bayes estimation (Allenby and Ginter 1995; Lenk et al. 1996), the **raw** utility scores have to be exported and read into an *R* data frame. This data frame has to include the actual choice in the validation/holdout task.

Assuming you included a validation/holdout task with a total of 7 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"
)
```

To get the relevant validation metrics that are 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 provide the data, the alternatives in the validation/holdout task (`opts`), and the actual choice (`choice`), which 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
```

The underlying logic of the confusion matrix is that the user has to provide a no-buy alternative (`none`). validateHOT calculates how often a buy or no-buy was correctly predicted, therefore, it is testing whether the model correctly predicts general demand (here by applying `accuracy()`).

```
accuracy(
  data = HOT,
  group = Group,
  opts = c(Option_1:None),
  choice = choice,
  none = None
) %>%
  round(2)
```

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

Finally, we show 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
##   <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
```

Finally, `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 is especially powerful 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, and `prohib`, to prohibit certain item combinations of being part of the assortment (see the vignette for more details and how to apply `turf()` with data obtained using a likert scale).

For the following example, we 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. As a threshold (`none`), the user uses the anchor (no-buy alternative).

```
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

Again, we start by creating the validation scenario.

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

This time we calculate the mean hit probability.

```
HOT_CBC %>%
  mhp(
```

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

```
## # A tibble: 1 x 2
##     MHP    se
##   <dbl> <dbl>
## 1  40.6  3.53
```

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

```
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.

References

Aizaki, Hideo, and James Fogarty. 2023. "R Packages and Tutorial for Case 1 Best-Worst Scaling" 46: 100394. https://doi.org/10.1016/j.jocm.2022.100394.

Allenby, Greg M., and James L. Ginter. 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., and T. Bartlomowicz. 2012. "Conjoint Analysis Method and Its Implementation in Conjoint r Package," 239–48. http://keii.ue.wroc.pl/pracownicy/tb/Bak_A_and_Bartlomowicz_T_Conjoint_analysis_method_and_its_implementation_in_conjoint_R_package.pdf.

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.

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.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young. 1996. "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs." *Marketing Science* 15 (2): 173–91. https://doi.org/10.1287/mksc.15.2.173.

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.

Netzer, Oded, and V. Srinivasan. 2011. "Adaptive Self-Explication of Multiattribute Preferences." *Journal of Marketing Research* 48 (1): 140–56. https://doi.org/10.1509/jmkr.48.1.140.

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.

Stan Development Team. 2023. "{RStan}: The {r} Interface to {Stan}." https://mc-stan.org/.

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.