RSGHB R Package
Using R to Expand the Capabilities for Hierarchical Bayesian Model Estimation

Jeff Keller, Jeff Dumont, Nelson Whipple

1 MOTIVATION

As choice-based conjoint analysis has continued to increase in use over the years, dozens of variants of the technique have been developed. In recent times, HB techniques have emerged that estimate individualized preferences by capturing heterogeneity from one consumer to the next.

A number of current software packages facilitate the estimation of HB models, including Sawtooth Software’s CBC HB and the bayesM R Package for HB model estimation.

2 THE R LANGUAGE

R was chosen as the venue for implementing RSGHB because it is free and open source, has a dynamic development community, a plethora of statistical routines, and continues to rise in popularity. Its active developer and user base produces a robust library of graphical and statistical routines to support the estimation and evaluation of models.

Furthermore, R allows transparency which eliminates the mystery of the “black box” approach, enabling greater flexibility, and customization of the internals of model estimation as well as pedagogical benefits.

RSG uses the R package system operated by CRAN to develop and distribute the RSGHB package. The package can be downloaded by visiting the CRAN R website or by installation via a remote repository connection from the R command line:

```R
> install.packages("RSGHB")
```

3 THE RSGHB PACKAGE

RSGHB increases the rigor behind Hierarchical Bayesian (HB) model estimation by enabling the modeler to specify parameters as either fixed or varying across individuals, non-linear-in-parameters utility functions, and non-normal distribution assumptions. As an open-source R package available on the Comprehensive R Archive Network (CRAN), it also provides transparency and access to supplemental analysis tools.

This poster 1 describes how RSGHB is used and 2) compares RSGHB to Sawtooth Software’s CBC HB and the bayesM R Package for HB model estimation.

4 SYNTHETIC CASE STUDY

Kenneth Train’s MATLAB and Gauss code for estimating HB models served as the basis for the RSGHB package. The code was ported to R, optimized, and extended for use with additional functions.

The estimation process is called from the doR () function, which consists of three inputs:

1. A likelihood function that returns the probability of choice at the observation-level given a set of utilities (part-worths)
2. A choice dataset
3. A list of control parameters to customize the estimation

Because the user supplies the likelihood function, RSGHB is capable of estimating any model type the user can code, including: nested logit, ordered logit, ordered probit, MDCVE, ICVL, etc. Thus, RSGHB is not limited to linear-in-parameter models.

A reference manual, How-To vignette, and function help pages are available in the package documentation.

To illustrate the features and flexibility of the RSGHB package, a synthetic dataset comprised of 1,000 individuals making 10 choices between 4 alternatives was generated. Each alternative is defined by price, brand, form factor, and a fourth “feature” attribute.

The error terms of the alternative utilities were correlated within form factors to create a nest structure that simulates increased substitutability between products of the same form factor.

Sensitivity to price is distributed negative log-normally and was interacted with a latent variable via the Integrated Choice and Latent Variable (ICLV) structural equation framework as a function of two demographic variables:

\[ \mu_{prices} = -\mu Price_{xovars} + \psi LN \]

Where \( \mu_{Price} \) - demographic1 + demographic2 + \( \epsilon \)

These preferences were applied to a completely random experimental design to generate synthetic choices. Two models were estimated from the synthetic choices:

1. Linear-in-parameters (normal distribution assumptions)
2. Nested ICLV (price assumed negative lognormal)

Model 1 was estimated using bayesM. CBC HB, and RSGHB with the same settings and number of iterations. It is only possible to estimate Model 2 using RSGHB. The three packages produced very similar results for Model 1, but RSGHB’s ability to specify Model 2 allows for more accurate recovery of the true parameters.

5 SUMMARY

We have presented an introduction and motivation for the RSGHB R package. The package benefits from the R venue by simultaneously eliminating use-costs and providing access to a vast library of supplemental features, functions, and tools.

The flexibility of the package was demonstrated with a synthetic case study using a Nested Integrated Choice and Latent Variable model structure. The RSGHB package is being continually updated with features and functionality to include more post-estimation analysis tools, improved performance, and better access to less technically savvy users.

REFERENCES