RSGHB R Package

Using R to Expand the Capabilities for Hierarchical Bayesian Model Estimation

Jeff Keller
Jeff Dumont
Nelson Whipple

the science of insight

Managerial Implications

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.

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[1] and Peter Rossi's bayesM R package[2]. In our experience, however, current packages lack the flexibility we need to achieve the rigor we envision, the transparency to be confident of exactly how they are working, and access to a broader set of related functions. Thus, the RSGHB package was born.

2 THE *R* LANGUAGE

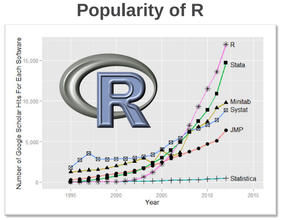
R was chosen as the venue for implementing RSGHB because it is free and open source, has a dynamic development community, a plethora of statistical routines, and continues to rise in popularity[3]. Its active developer and user base produces a robust library of graphical and statistical routines to support 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[4] to develop and distribute the RSGHB package. The package can be downloaded by visiting the RSGHB CRAN website or by installation via a remote repository connection from the R command line:

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

Popularity of R



3 THE RSGHB PACKAGE

Kenneth Train's MATLAB and Gauss code[5] 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 `doHB()` function, which takes three inputs:

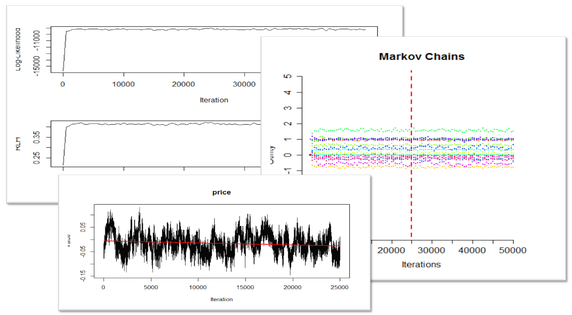
1. A likelihood function that returns the probability of choice at the observation-level given a set of utilities (part-worths)
2. A choice dataset
3. A list of control parameters to customize the estimation

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, MDCEV, ICLV, etc. Thus, RSGHB is not limited to linear-in-parameter models.

```
> doHB(likelihood_user = myFunction,
       choicedata = myData, control = list())
```

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

RSGHB Diagnostic Outputs



Relevant Software Features

	bayesM	CBC HB	RSGHB
Covariate Modeling	✓	✓	✓
Non-Normal Distributions			✓
Custom Prior Covariance Matrix	✓	✓	✓
Constraints		✓	✓
Independent Covariance Matrix			✓
Fixed Parameters			✓
Batch Model Runs	✓	✓	✓
Non-linear Utility Functions			✓
Graphical User Interface		✓	
Interactions	None	First Order	Unlimited

4 SYNTHETIC CASE STUDY

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 nested 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:

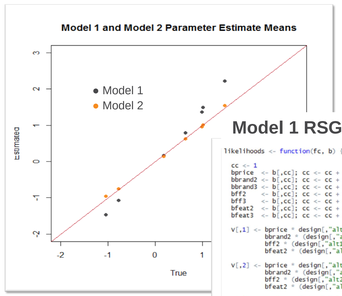
$$U_{price} = -e^{U_{price.norm} + \tau \cdot LV}$$

Where $LV = \gamma_1 \cdot demographic_1 + \gamma_2 \cdot demographic_2 + \varepsilon$

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 log-normal)

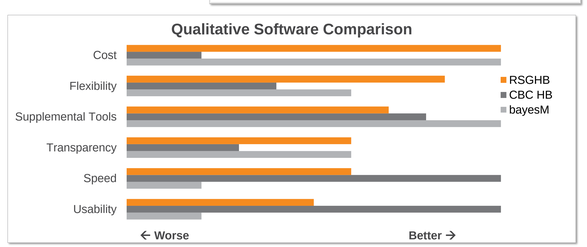
Model 1 was estimated using bayesM, CBC HB, and RSGHB with the same settings and number of iterations. It is only possible to estimate Model 2 using RSGHB. The three packages produced very similar results for Model 1, but RSGHB's ability to specify Model 2 allows for more accurate recovery of the true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



Model 1 RSGHB Likelihood Function

```
likelihoods <- function(fc, b) {

  cc <- 1
  bprice  <- b[,cc]; cc <- cc + 1
  bbrand2 <- b[,cc]; cc <- cc + 1
  bbrand3 <- b[,cc]; cc <- cc + 1
  bff2    <- b[,cc]; cc <- cc + 1
  bff3    <- b[,cc]; cc <- cc + 1
  bfeat2  <- b[,cc]; cc <- cc + 1
  bfeat3  <- b[,cc]; cc <- cc + 1

  v[,1] <- bprice * design[,"alt1.price"] +
           bbrand2 * (design[,"alt1.brand"] == 2) + bbrand3 * (design[,"alt1.brand"] == 3) +
           bff2 * (design[,"alt1.ff"] == 2) + bff3 * (design[,"alt1.ff"] == 3) +
           bfeat2 * (design[,"alt1.feat"] == 2) + bfeat3 * (design[,"alt1.feat"] == 3)

  v[,2] <- bprice * design[,"alt2.price"] +
           bbrand2 * (design[,"alt2.brand"] == 2) + bbrand3 * (design[,"alt2.brand"] == 3) +
           bff2 * (design[,"alt2.ff"] == 2) + bff3 * (design[,"alt2.ff"] == 3) +
           bfeat2 * (design[,"alt2.feat"] == 2) + bfeat3 * (design[,"alt2.feat"] == 3)

  v[,3] <- bprice * design[,"alt3.price"] +
           bbrand2 * (design[,"alt3.brand"] == 2) + bbrand3 * (design[,"alt3.brand"] == 3) +
           bff2 * (design[,"alt3.ff"] == 2) + bff3 * (design[,"alt3.ff"] == 3) +
           bfeat2 * (design[,"alt3.feat"] == 2) + bfeat3 * (design[,"alt3.feat"] == 3)

  v[,4] <- bprice * design[,"alt4.price"] +
           bbrand2 * (design[,"alt4.brand"] == 2) + bbrand3 * (design[,"alt4.brand"] == 3) +
           bff2 * (design[,"alt4.ff"] == 2) + bff3 * (design[,"alt4.ff"] == 3) +
           bfeat2 * (design[,"alt4.feat"] == 2) + bfeat3 * (design[,"alt4.feat"] == 3)

  p <- exp(v[cbind(1:nrow(design), design[,"choice"])])/rowSums(exp(v))
  return(p)
}
```

Qualitative Software Comparison



	TRUE VALUE	Model 1 (bayesM)	Model 1 (CBC HB)	Model 1 (RSGHB)	Model 2 (RSGHB)
RMSE	-	**0.403**	**0.467**	**0.414**	**0.05**
$price$	-1.04	-1.46	-1.52	-1.47	-1.03
$brand_1$	0 (fixed)	-	-	-	-
$brand_2$	1.00	1.36	1.42	1.36	**0.96**
$brand_3$	-0.80	-1.06	-1.10	-1.07	**-0.75**
ff_1	0 (fixed)	-	-	-	-
ff_2	0.25	0.15	0.16	0.16	**0.14**
ff_3	0.60	0.77	0.81	0.80	**0.62**
$feature_1$	0 (fixed)	-	-	-	-
$feature_2$	1.00	1.48	1.55	1.49	**1.01**
$feature_3$	1.50	2.20	2.31	2.22	**1.54**
ε	0.00	-	-	-	0.00
λ_{ff}	0.37	-	-	-	0.35
γ_1	0.20	-	-	-	0.23
γ_2	-0.45	-	-	-	-0.46
σ_1	1.00	-	-	-	1.01
σ_2	1.00	-	-	-	0.98
ζ_1	0.50	-	-	-	0.53
ζ_2	0.25	-	-	-	0.13
τ	0.25	-	-	-	0.25

Code to recreate the synthetic choice data and models is available.[7]

5 SUMMARY

We have presented an introduction and motivation for the RSGHB R package. The package benefits from the R venue – 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

1. Sawtooth Software, Inc. *CBC Hierarchical Bayes Module*. Version 5.5.2. http://www.sawtoothsoftware.com/products/conjoint-choice-analysis
2. Peter Rossi. (2012). *bayesM: Bayesian Inference for Marketing/Micro-econometrics.* R package version 2.2-5 http://CRAN.R-project.org/package=bayesm
3. Robert Muenchen. (2014). *The Popularity of Data Analysis Software.* http://R4stats.com/articles/popularity/
4. The R Project. *The Comprehensive R Archive Network*. http://cran.us.r-project.org
5. Kenneth Train. (2009). *Discrete Choice Methods with Simulation, 2nd edition.*
6. J. Dumont, J. Keller, and C. Carpenter. (2014). *RSGHB: Functions for Hierarchical Bayesian Estimation*. R package version 1.0.2. http://cran.r-project.org/web/packages/RSGHB/index.html
7. J. Keller, J. Dumont, N. Whipple. (2014). *RSGHB: Using R to Expand the Capabilities for Hierarchical Bayesian Model Estimation (Case Study)* https://github.com/jeffdumont/RSGHB/tree/master/2014_ART_Forum_Poster