**User’s guide for analyzing choice data using Markov chain consumer movement model**

**Adam Zeilinger, Dawn Olson, David Andow**

**May 2014**

This guide is designed to facilitate the use of the models and associated tests for the analysis of consumer choice data using the Markov chain consumer movement model described in Zeilinger et al. (2014). *A likelihood-based bio-statistical model for analyzing consumer movement in simultaneous choice experiments*. This guide is designed to be used in conjunction with R scripts available on the webpage <https://github.com/arzeilinger/Consumer-Choice-Model>, on which we will also provide future revisions and extension to the model and R code.

1. Assessing choice experiment data: Are the data appropriate for analysis?
   1. Were only two choices made available simultaneously to one individual consumer per trial?
   2. Were the choices and consumer replaced for each trial; i.e., were trials independent replicates?
   3. Was the location of each consumer recorded repeatedly over the course of the trials?
   4. Can consumer location data be binned into 3 discrete states: on choice 1, on choice 2, and in the neutral space within the experimental arena?
   5. Were observation times fixed, i.e., the same across all trials, and were the observation times, from the start of the experiment, recorded?
   6. If the answer is “yes” to questions 1a – 1e then the data are appropriate for analysis. If not, a different analytical approach should be considered.
   7. Example data sets, in two different formats titled “dat” and “dat.obs”, can be viewed in the R script *MLE Consumer Model* at the above website and in Supplementary Material, Appendix C.
2. Testing assumptions: Do the data meet the assumptions of the model?
   1. Are attraction and leaving rates constant for the duration of the experiment?
      1. *Test*: Survival analysis of settling and leaving events.
         1. First, conduct separate survival analyses for each parameter of the model, meaning the arrival times and leaving times for each choice. We recommend using the each consumer’s first choice as the “events” relating to arrival times and the first time that each consumer left a choice as the “events” relating to leaving times. If an arrival or leaving event occurred between observation times (i.e., the event was not directly observed), we recommend taking the median between the two times as the time the event occurred.
         2. Second, extract the survival estimates, *S(time)*, from the survival analysis and plot ln(*time*) against ln{-ln[S(*time*)]}. Assess whether the data points fall along the line-of-best-fit.
         3. See R script *Consumer Choice Survival Analysis* for R code for analyzing and extracting data from survival functions.
   2. Are sequential choices independent?
      1. *Test*: Contingency table analysis.
         1. Determine the series of movement events for each trial. Construct contingency tables of first choice vs. second choice, second choice vs. third choice, etc. until table cell values become too small for analysis. Contingency table cell values will be too small for analysis if a cell value = 0. Use Fisher’s Exact Test to test for independence between consecutive choices if any cell values are < 5. If the test is non-significant then the consecutive choices are independent and the assumption is met.
      2. *Alternative test:* MLE correlation matrix.
         1. If movement event frequencies are too small for contingency table analysis and MLE parameter estimates are not on a constraint boundary, correlations between parameter estimates can be calculated from the MLE model object.
         2. First, extract the Hessian matrix from the model object for the Free Model.
         3. Second, invert the Hessian matrix to calculate the variance-covariance matrix. Here, the diagonal entries indicate the variances for each parameter and the off-diagonal entries indicate the co-variances for each parameter pair. Scale the matrix according to the variances to calculate correlations between parameter pairs. If the correlations between *µ1* and *p2* and between *µ2* and *p­1* are small, then choices are independent.
         4. See the R script *MLE Consumer Model* for code on calculating the variance-covariance matrix and correlation matrix. See Bolker (2008) for more detail on the normal approximation method and the calculation of the correlation matrix.
   3. If either assumption is violated, this does not necessarily exclude model results. However, parameter estimates will be biased and care should be taken in interpretation of results. Tests of both assumptions should provide insight into the nature of parameter estimate bias.
3. Model selection: which model variant best fits the data?
   1. See R script *MLE Consumer Model* for code on running maximum likelihood estimation for each model variant and model selection using AIC or AICc.
4. Is model averaging warranted?
   1. Burnham et al. (2011) suggest that models with ΔAIC < 7 relative to the best model may still provide useful information. Note that other authors suggest a lower threshold of ΔAIC < 2 (Bolker 2008). Regardless, for all models within the chosen threshold, model averaging of parameter estimates and variances may be warranted. See R script *MLE Consumer Model* for model averaging code.
5. Estimating variance: Are any the parameter estimates on the boundary of an inequality constraint?
   1. If no parameter estimates are on an upper or lower constraint boundary, the profile or normal approximation methods can be used to estimate variances and confidence intervals—based on the normal distribution—for each parameter estimate (Bolker 2008, Millar 2011). See R script *MLE Consumer Model* for code on estimating variances using the normal approximation method.
   2. If one or more parameter estimates are on the upper constraint boundary, consider increasing the value of the upper constraint or removing the constraint, unless the value(s) set for the upper constraint have some biological meaning.
   3. If one or more parameter estimates are on a constraint boundary, estimate variances using the jackknife method. Because the distribution of jackknife estimates is not well-known (Efron and Tibshirani 1993), confidence intervals should not be calculated from jackknife variance estimates. See R script *MLE Consumer Model*.

**Literature Cited**

**Bolker, B. M.** **2008**. Ecological Models and Data in R. Princeton University Press, Princeton, NJ.

**Burnham, K. P., D. R. Anderson, and K. P. Huyvaert**. **2011**. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. 65: 23–35.

**Efron, B., and R. J. Tibshirani**. **1993**. An introduction to the bootstrap. Chapman & Hall/CRC.

**Millar, R. B.** **2011**. Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB, 1st ed. John Wiley & Sons, Ltd, West Sussex, England.