Statistical Innovations Online course:

Latent Class Discrete Choice Modeling with Scale Factors

Session 3: Advanced

SALC Topics

Outline:

A. Advanced Models for MaxDiff Data
B. Modeling the Dual Response None option
C. Fusion of Choice and Ratings Data
D. Problems with Hierarchical Bayes (HB) methods

A. Advanced Models for MaxDiff Data

In Session 2 we showed how to use the point-and-click interface in LG Choice 5.0 to estimate various Sequential BestWorst models. The Sydney Transport data were used to illustrate the analyses (recall pages 1-16 of Tutorial 8A). In this Session 3, we will continue with the analysis of these data using some advanced Best-Worst models.

On page 12 of the Session 2 Lecture Notes we also referred to the Joint Best-Worst model, which is also referred to by Marley and Louviere (2005) as the 'MaxDiff' model in the case of a single nominal attribute. We also refer to this model as the MaxDiff Quasi-Independence Model (MD-QI), to distinguish it from the MaxDiff Independence Model (MD-I) model that Sawtooth Software implements for the analysis of MaxDiff data. Estimation of Joint Best-Worst models requires the LG-Syntax add-on module for LG Choice 5.0.

In this session, we will show how to estimate an advanced Best-Worst model where a *continuous* latent variable is used to allow for *continuous* latent scale factors in a choice model, rather than the inclusion of latent scale classes to allow for separate scale factors for each latent scale class. We will also show how to estimate the Joint Choice (MD-QI) model.

Reading Material

Assigned Reading Material:

Pages 16-25 of LG Choice Tutorial 8A.

Exercise A1

Continue where you left off at the end of Exercise D1 in Session 2, as explained on pages 16-18 of the LG Choice Tutorial 8A.

Question A1: As shown in Fig. 22 of this reading material, allowing correlation between 8 classes and the 2 scale classes increases the number of model parameters from 74 to 81. Examine the estimated correlation parameters in the Parameters Output for Model 5. Why are 81-74 = 7 distinct parameters required to allow for correlation rather than just 1 correlation parameter?

Exercise A2

Continue reading this tutorial through the middle of page 21. Focus on the 2 models identified by boxes in Figure 25 on page 21.

Question A2.1: Are the 8 classes obtained in the first model similar to the 8 classes obtained in the second model? Hint: Compare the Parameters Output from both models and also compare the size of the classes as shown in the Profile Output.

Question A2.2: Why do you think that allowing separate scale factors for Best and Worst in the Best-Worst model improves the fit of this model?

B. Modeling the Dual Response None option

As an extension to the standard MaxDiff design where respondents select the Best and Worst alternative, the "Dual Response None" design asks the additional follow-up question which allows for the additional alternative 'None of the Above' to be selected instead of the option chosen as Best. The question might be phrased "Would you purchase the alternative you chose as Best?"

The main purpose of this extended design is to enable estimation of a "market share" rather than "preference share" model. Without the "no buy" (or "none") option we are only able to simulate preference share, where participants are forced to choose between a fixed number of profiles (newspaper offerings in the case study described below).

Dual Response None Case Study:

A Best-Worst experiment was conducted to select the most and least preferred newspaper offerings among 6 offerings displayed in sets of 4. After obtaining the responses from a given set, respondents were then exposed to a new choice set containing an additional option ('None of these'). That is, they were then asked whether or not they would actually *purchase* their most preferred option or not – that is, would they purchase their most preferred option or would they choose the 5th option 'None of these 4 papers'.

Attributes included:

1) Title: Newspaper name
2) price: price
3) promot: promotional offer
4) sunsup: Whether a Lifestyle Supplement was included
5) None

For more detail on the design, see 'Case_study1_Design_info_V2.xlsx' (included in the Dual Response None Example folder in Session 3 Materials)

It is better to use the best/worst (or most preferred / least preferred) choice experiment rather than the standard 'first-only' choice design, as this effectively "doubles up" on the number of observations relating to what drives choice between profiles (i.e. we get two observations per replication/question). The three part "dual response" question provides a further improvement as not only do we get this extra information on drivers of preference for the profiles but we also get to estimate a utility for a "no buy" option which enables us to simulate a market share rather than just preference share.

An alternative simpler method would be to always include the "no buy" option in a first choice model. The problem with this is that it is possible that an individual might always pick “no buy” in which case we learn nothing about what attributes drive their preferences between the profiles in which we are interested. Using dual response we cover both bases: we get maximum information on what attributes / attribute levels drive preference for the profiles AND estimate a part-worth for the “no buy” option. In that sense it is a best of all worlds and at minimal extra cost, given that all profiles have already been evaluated in the first part of the question (most preferred).

This three part approach is widely used in other choice packages such as Sawtooth’s program, but LG Choice provides much greater flexibility as you can now allow for scale differences between the Best choices (most preferred), the worst choices (least preferred) and the “no buy” option. As seen for this application, these provide better fitting models, allowing for the fact that level of certainty/consistency in choices can differ for each of these three types of choice. In this example we find that respondents are much less certain/consistent in their “worst” choices relative to their “best” choices, but slightly more certain in their choice when a “no buy” option is available.

Technical note –We are estimating an extended form of the Best-Worst model by adding the None component. Thus, we must use the syntax module with the general “replication scale” feature of LG choice to set up this model (and use the “ranking” rather than the new ‘BestWorst’ keyword to define the type of the dependent variable). The replication scale is set to +1 for the best (most preferred) and dual response (best with “no buy”) choice and -1 for the Worst (least preferred) choice. This ensures that we reverse the effects for estimating worst (least preferred) choices.

Exercise B1

Estimate the four 6-class models saved in ‘Dual_Resp_None.lgf’:

Model 1: 6-class
Model 2: 6-class,2-sclass – A 6-class model that accounts for scale heterogeneity by including 2 scale classes
Model 3: Same as Model 2 plus separate scale factors for best and worst
Model 4: Same as Model 3 but also separate scale factors for the Best-Worst responses and the None response

Which model fits best?

C. Fusion of Choice and Ratings Data

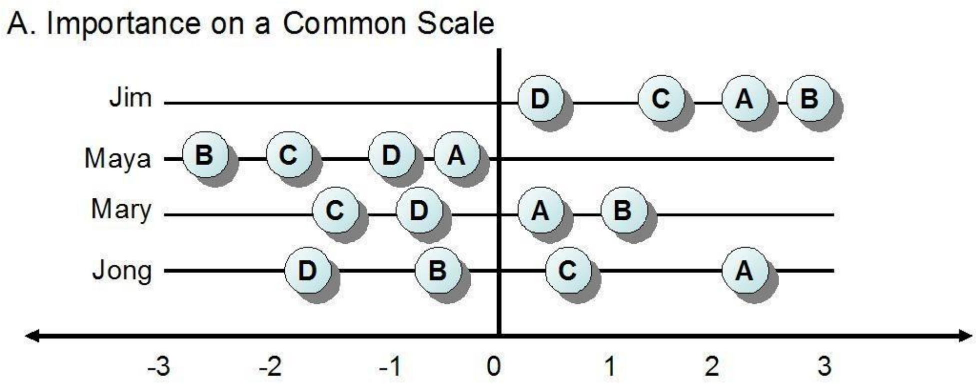
One limitation of MaxDiff scaling and other discrete-choice modeling approaches is that they estimate worth (part-worth) coefficients based on *relative* (ipsative scale) as opposed to absolute judgments. For a given subject or segment, this allows the options to be ranked on *relative* importance, but it does not allow segments to be compared on their inferred importance for any particular item. This issue is discussed further below.

One way to compensate for this limitation is to add one or more absolute ratings to the choice responses, and use the rating information as an absolute zero point to 'anchor' the class-specific ipsative scales to a common threshold, thus obtaining an absolute scale. For example, in the case of MaxDiff data, this is referred to as "Max-Diff with threshold".

The LG Choice program allows estimation of data fusion models where segments are obtained that differ not only with respect to their relative choices but also their ratings. The basic idea is that the standard Response file in the 3-file data format for a given respondent is supplemented with J additional records associated with the J ratings provided by that respondent. An example setup illustrating the response file structure for responses to 4 choice sets plus a single rating is shown in Figure 2-1 on page 7 of LG Choice Modeling Extensions via the LG-Syntax Module.

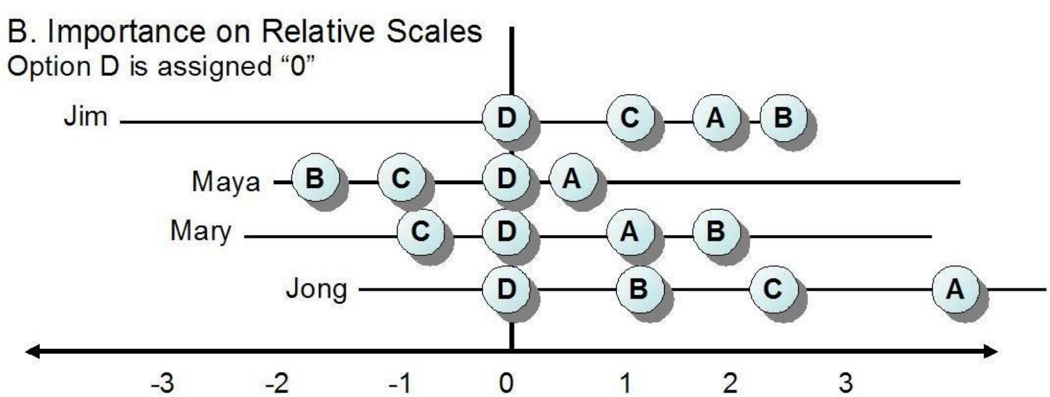
Ipsative vs. Absolute Scale

Suppose that we obtained 4 LC segments from an analysis of MaxDiff Data, and named these segments 'Jim', 'Maya', 'Mary' and 'Jong'. From the results we might see that Mary judges attribute D to be more important than C, but we would not know that she does not consider either to be very important. Similarly, our results might indicate that Jim considers C to be more important than D, but we would not know that Jim considers both C and D to be very important. Given only their relative judgments, it may be tempting but it is not valid to infer that Mary considers D to be more important than Jim does (see e.g., Bacon et al. 2007, 2008). For example, Plot 'A' describes the results of a "Max-Diff with threshold" analysis where '0' represents the absolute 'neutral' threshold measuring importance.



Typically, effect coding is employed separately for each LC segment in a MaxDiff analysis, in which case the mean worth for each segment would be ‘0’. Plot A above is obtained by plotting such effect-coded results after adding a class-specific constant that increases the worth estimates for Jim, decreases them for Maya, etc., to take into account how they differ in terms of a measure of absolute importance.

Alternatively, Plot B displays the same results from a MaxDiff analysis where dummy coding is used instead of effect coding to display the results. It would not be correct to state that all 4 segments agree on the worth of attribute D because the ‘0’ worth assigned to attribute D can only be interpreted relative to the worths of the other attributes. Neither dummy nor effect coding will capture the fact that Jim values all 4 attributes more highly than Maya.



This issue raises serious questions about the meaning of ‘individual coefficients’ obtained from Hierarchical Bayes (HB) or LC choice models, as it is a very tricky proposition to interpret the fact that a particular individual coefficient for one respondent is higher than that for another respondent.

In order to convert from an ipsative to an absolute scale, one must supplement the choice information provided by a choice experiment with absolute information that can be obtained for example from ratings. This additional information serves as an anchor for each segment so that Plot A is possible.

The reading material describes estimation of fusion models with response data from best-worst choices and ratings.

Reading Material

Assigned Reading Material:

Pages 83-89 in:

Magidson, J., D. Thomas, & J.K. Vermunt (2009). "ANewModelfortheFusionof MaxDiffScalingandRatingsData", *2009 Sawtooth Software Proceedings*, 83-103.

Sections 4 and 7.2 in:

Vermunt, J.K. (2013). Categorical response data. In: M.A. Scott, J.S. Simonoff, and B.D. Marx (eds.), The SAGE Handbook of Multilevel Modeling, 287-298. Thousand Oaks, CA: Sage

Optional Additional Reading

Pages 89-103 in:

Magidson, J., D. Thomas, & J.K. Vermunt (2009). "ANewModelfortheFusionof MaxDiffScalingandRatingsData", *2009 Sawtooth Software Proceedings*, 83-103.

For more detail on the design, see 'Case_study1_Design_info_V2.xlsx'

Exercise C1

Open 'MaxDiffFused.lgs' and estimate the 5 saved 'data fusion' models. Compare the LL or BIC values you obtain with those reported in Table 2 on Page 14 of Vermunt (2013) to make sure that they match.

The model named 'Full' contains a submodel for the choice responses and a submodel for the ratings responses. The equations for these submodels are:

```
CHOICE <- (b1) attFeature | Class;
RATING <- 1 + Class + u + (b2) Feature | Class;
b2=b1;
```

The restriction 'b2=b1' assures that the ranking of the features in importance and the ratings of the features are consistent with each other. Thus, the latent classes identified by this model differ with respect to the importance they place on these features (both in relative and absolute terms).

This model includes different scale factors for the 2 different scale classes, for both the choice responses and the ratings. You will find estimates for these parameters in the log-scale model at the bottom of the Parameters Output in a section labeled 'Scale Parameters'.

Answer these questions about Model 'Full':

Question C1.1: For the choice model, the scale factor is set to 1 (log-scale factor=0) for sClass(1), and the sClasses are ordered so that the higher scale factor is associated with sClass(1). What is the estimated scale factor for sClass(2)? Does it differ significantly from 1 (i.e., are the scale factors for the 2 scale classes significantly different from each other?)

Question C1.2: As mentioned above, respondents in sClass(2) are less consistent in their choice responses than those in sClass(1). For the Rating model, the log-scale effects are effect coded. Are these respondents less extreme in their ratings? That is, do they have a significantly lower ratings scale factor than those in sClass(2)?

Answer these questions about Model 'sClass=1':

Question C2.1: By eliminating sClass(2), Model 'sClass=1' has 3 fewer parameters than Model 'Full': these are associated with the 2 distinct scale factors in the Choice and Ratings submodel plus the relative size of the second scale class. Does this model fit better than Model 'Full' according to the BIC?

Question C2.2: Is the result from question C2.1 consistent with your answer from question C1.2?

D. Problems with Hierarchical Bayes (HB) methods

First, Magidson, Eagle and Vermunt (2005) make the case that HB overfits data. Secondly, while a common practice is to take individual coefficients from HB and subject them to clustering to obtain segments, Magidson (2003) makes the case that better, more meaningful segments can be obtained using the simultaneous approach offered by LC Choice modeling.

In this section we will focus primarily on how the issue of scale factors relates to HB modeling. Since individual coefficients are an integral part of HB choice modeling, the fact that these coefficients represent the preference (part) worth multiplied by the scale factor for a given individual is especially problematic for HB when attempting to obtain segments.

Hess and Rose investigated whether it is possible to disentangle the preference part-worth from the scale factor under the standard normal distributional assumptions common to random effects modeling (and HB). They make a strong case that under these assumptions, it is not possible to separate out these components. Their main point is that random effects models are always modeling heterogeneity in theta=alpha*beta, where alpha is the scale. Having a separate heterogeneity distribution for alpha and beta, only results in a more flexible distribution for theta. It is not at all clear or convincing that one is really separating alpha and beta, as is claimed by proponents of G-MLM. With certain distributional choices (all log normal), it not even possible to distinguish correlations between random effects from scale heterogeneity.

To the extent to which the scale factors differ across respondents, the individual HB coefficients are hopelessly confounded, and thus by subjecting these individual coefficients to a cluster analysis it is even more problematic to obtain homogeneous segments that differ in meaningful ways.

The arguments in Hess and Rose do not apply to SALC models since those arguments assume the normality (and log-normality) distributional assumptions which are not made in the SALC models discussed in this course. Our SALC models determine segments for which the parameters are NOT proportional. That is, the segments obtained differ in ways that cannot be interpreted as scale effects. It is for this reason that we claim that the segments differ in their preferences only.

We have seen in Sessions 1 and 2 that adjusting for scale heterogeneity makes a difference in the resulting segments. Determining the number of scale classes turns out not to be so important because increasing the number of scale classes does not make much difference in the definition of segments. Thus, we determine the number of scale classes to include in a model using the parsimony principle (lowest BIC).

Reading Material

Assigned Reading Material:

Magidson J., T. Eagle, & J. K. Vermunt (2005) Using Parsimonious Conjoint and Choice Models to Improve the Accuracy of Out-of-Sample Share Predictions (2005). Presented at the 2005 ART Forum.

Magidson (discussant of Steve Cohen paper). Sawtooth Software Proceedings, 2003.

References

Bacon, Lenk, Seryakova, and Veccia (2007). Making MaxDiff more informative: Statistical data fusion by way of latent variable modeling. *Sawtooth Software Conference Proceedings.*

Hess, S., and J.M. Rose (2012). Can scale and coefficient heterogeneity be separated in random coefficients models?. *Transportation*. DOI 10.1007/s11116-012-9394-9