EXTRACTING MEANINGFUL SEGMENTS FROM HB UTILITIES

JAY MAGIDSON
STATISTICAL INNOVATIONS INC.
JEROEN K. VERMUNT
TILBURG UNIVERSITY

OVERVIEW

The goal of latent class (LC) modeling is to separate respondents into homogeneous groups (latent classes) that differ in meaningful ways. When data consists of Best-Worst (MaxDiff) choice responses, or HB utilities derived from these responses, *meaningful* means that classes differ with respect to respondent *preferences*.

The existence of not only *preference* heterogeneity but also *scale* heterogeneity in choice data presents a methodological challenge to avoid scale confounds which make the resulting segments difficult to interpret (Swait and Louviere, 1993; Louviere and Eagle, 2006).

This paper investigates under what circumstances it is possible to extract meaningful segments from MaxDiff choices or the related HB utilities. It also addresses concerns raised by Lyon (2019) and others who pointed out that LC clustering of HB utilities can result in very different segments depending on how the utilities are coded. Specifically, we use both a real-world MaxDiff dataset as well as simulated data to illustrate and shed light on:

1. the relative performance of various segmentation models that work directly with MaxDiff responses versus working with related HB utilities,
2. the effect of failing to account for the random utility theory scale factor in the segmentation, and
3. the effect of various coding of HB utilities on segmentation results and a clear recommendation for which coding to use.

After presenting the results, we discuss various implications for obtaining meaningful segments from MaxDiff data and propose some further research.

BACKGROUND

Magidson and Vermunt (2007) proposed a variant of the LC Choice model called Scale Adjusted Latent Class (SALC) modeling to deal with potential confounds caused by the presence of scale heterogeneity (see also Magidson, 2018). While this SALC model can be applied directly to MaxDiff choice responses (the "1-Step approach"), Eagle and Magidson (2019) showed how a SALC variant of the LC Cluster model can be used to cluster HB utilities derived from the MaxDiff choice responses (the "2-Step approach").

Regardless whether the 1-Step or 2-Step approach were used, results from these papers showed that segments obtained from *standard LC* models confounded Preference and Scale, while Preference segments obtained from *SALC* models were free from such confounds.[1] Moreover, Eagle and Magidson (2019) obtained the surprising result that 88% of respondents were classified into the same preference segment regardless of whether 1-Step or 2-Step SALC modeling was used.

In light of equally promising results obtained from SALC segmentation modeling based on either MaxDiff choice responses or HB utilities derived from these utilities, as well as new results presented in this paper regarding the proper coding of HB utilities, we reconsider whether the 1-Step approach to segmentation should be considered the gold standard.

AUSTRALIAN HEALTH CARE REFORM STUDY

For concreteness, MaxDiff data from the Australian Health Care Reform Study (Louviere and Flynn, 2010) will be used to compare results from various LC segmentations. These data consist of 15 Principles (MaxDiff items) considered more or less important for use in reforming health care. A sample MaxDiff Scenario from that study is shown in Figure 1.

Figure 1: MaxDiff Scenario

Principles	Most Important	Least Important
Promoting wellness & strengthening prevention	○	○
Providing for future generations	○	○
Value for money	○	○
Recognize social & environmental influences shape our health	○	○
Comprehensiveness	○	○
Shared responsibility	○	○
Equity	○	○
Taking the long term view	○	○

Based on his analysis of these data, Flynn stated:

> "In health economics you usually find people separate out into 3 classes—
> those who prefer **Equity,**
> those who prefer investment in future health/**People and family centered**, and
> those who prefer **Efficiency** /**Value for money.**"

Throughout this paper we will refer back to these particular principles in our own descriptions of meaningful vs confounded segments.

---

[1] Adjustment for scale confounds via SALC is similar conceptually to adjustment for response level in ratings data using a random intercept regression model (see e.g., Magidson and Vermunt, 2006; Popper et al., 2004).

Tables 1 and 2 below compare results from standard 3 class LC models vs. SALC variants based on the 1-Step (Table 1) and 2-Step approaches (Table 2) respectively. In both tables, the LC model contains evidence of a preference scale confound in the form of a *low scale* class, the associated parameter estimates from the low scale class tending to be much closer to zero than the other classes. In contrast, the SALC variants in both tables lack such confounds, and more clearly correspond to Flynn's 3 preferences.

Table 1: Comparison of Results of LC and SALC Choice Models (1-Step Approach)

Principles	***1-Step Approach*** 3-Class **LC Choice**			***1-Step Approach*** 3-Class **SALC Choice**		
	Value for $	People & Family	Low scale class	Value for $	People & Family	Equity
1: A culture of reflective improvement & innovation	-1.6	-0.5	-0.5	-2.2	-0.6	-3.1
2: A respectful, ethical system	-0.2	0.5	0.3	-0.6	1.4	1.2
3: Comprehensiveness	-0.2	-1.1	0.3	-0.2	-1.9	0.5
4: Equity	-0.1	-1.6	0.5	-1.0	-2.9	2.6
5: People & family centered	0.5	1.6	-0.2	-0.3	2.9	1.8
6: Promoting wellness & strengthening prevention	0.3	1.3	0.2	0.9	2.6	-0.1
7: Providing for future generations	0.0	1.0	0.1	0.1	2.0	-0.2
8: Public voice & community engagement	-1.7	-0.5	-0.3	-2.8	-1.3	-1.7
9: Quality & safety	2.1	0.9	0.4	2.9	1.6	3.8
10: Recognize social & environ influences shape health	-1.1	0.6	-0.1	-1.3	1.3	-2.1
11: Responsible spending	0.9	-0.3	0.1	2.1	-1.2	-0.2
12: Shared responsibility	-0.5	-0.8	-0.5	-1.0	-1.3	-1.5
13: Taking the long term view	-0.1	0.2	-0.2	0.3	0.0	-1.2
14: Transparency & accountability	0.0	-0.2	0.0	-0.3	-0.9	1.0
15: Value for money	1.8	-1.1	0.0	3.4	-2.0	-0.8
Class Size	0.35	0.31	0.34	0.39	0.35	0.26

Table 2: Comparison of Results of LC and SALC Cluster Models (2-Step Approach)

Principles	*2-Step Approach* 3-class **LC Cluster**			*2-Step Approach* 3-class **SALC Cluster**		
	Equity/ Value for $	People & Family	Low scale	Value for $	People & Family	Equity
1: A culture of reflective improvement & innovation	-3.1	-0.4	-1.0	-2.9	-0.7	-3.7
2: A respectful, ethical system	0.4	1.1	-0.1	-0.5	1.3	1.2
3: Comprehensiveness	0.2	-1.4	-0.4	-0.2	-2.2	0.4
4: Equity	0.7	-1.6	-0.6	-1.0	-2.8	1.8
5: People & family centered	0.5	2.2	0.6	0.1	3.3	1.7
6: Promoting wellness & strengthening prevention	-0.2	2.1	0.9	0.9	2.8	0.4
7: Providing for future generations	-0.3	1.2	0.6	0.2	2.2	-0.2
8: Public voice & community engagement	-2.1	-0.8	-1.3	-3.5	-1.2	-2.2
9: Quality & safety	3.8	1.5	0.9	3.4	1.8	4.9
10: Recognize social & environ influences shape health	-2.3	1.5	-0.4	-2.1	1.8	-2.6
11: Responsible spending	1.1	-1.5	0.8	2.7	-1.7	-0.1
12: Shared responsibility	-1.0	-0.9	-0.9	-1.3	-1.3	-1.4
13: Taking the long term view	-0.5	-0.2	0.3	0.3	0.2	-1.0
14: Transparency & accountability	1.1	-0.7	-0.4	0.1	-1.0	1.2
15: Value for money	1.7	-2.1	0.9	4.0	-2.6	-0.3
Class Size	0.27	0.23	0.49	0.42	0.32	0.25

Moreover, the SALC model segments for both the 1-Step and 2-Step approaches are highly consistent with each other, in that 88% of the 204 respondents are classified into the same SALC preference class: (75 + 64 + 40)/ 204 = 88% agreement (see Table 3).

Table 3: Comparison of Respondent Classifications Under 1-Step and 2-Step Approaches

SALC (1-Step) \ SALC (2-Step)	Segment 1	Segment 2	Segment 3
Segment 1	75	10	5
Segment 2	1	64	5
Segment 3	2	2	40

Thus, regardless of whether the segmentation is performed directly on the MaxDiff choices or on the HB utilities derived from such choices, the resulting 3 SALC segments turn out to be quite similar to each other and are consistent with Flynn's conjecture as to the preferences each of the 3 segments share.

Before using simulation to explore these results further, in the next section we will explore how the clusters described in Table 2 would change if different codings were used for the HB utilities.

Coding of HB Utilities

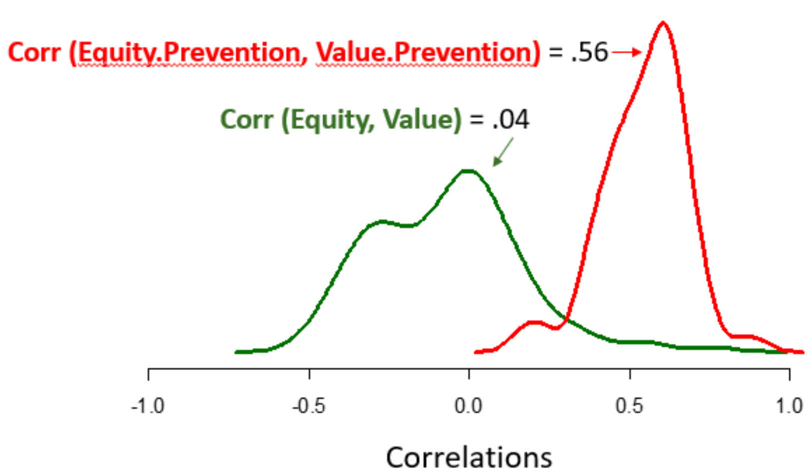
As discussant for Eagle and Magidson (2019), Lyon (2019) pointed out that the results of 2-Step LC and SALC models (reproduced in Table 2 above) were obtained using zero-centered (ZC) HB utilities. Lyon demonstrated that very *different* results would be obtained if zero-referenced (ZR) utilities were used for such analyses. He also alluded to strange results that have been reported occasionally in previous analyses where ZR was used to identify HB utilities (see e.g., Lee and Brazell, 2019).

We confirmed Lyon's results and determined that the reason for such differences is that ZR induces positive (spurious) correlations in the data which distort the results obtained from latent class. To see this, we note that ZR utilities can be obtained from ZC utilities by subtracting the reference utility. For example, taking utility 6 (Prevention) as the reference, we have:

$$
\begin{aligned}
&R1.6 = Z1 - Z6\\
&R2.6 = Z2 - Z6\\
&\ldots\\
&R6.6 = Z6 - Z6 = 0\\
&\ldots\\
&R15.6 = Z15 - Z6
\end{aligned}
$$

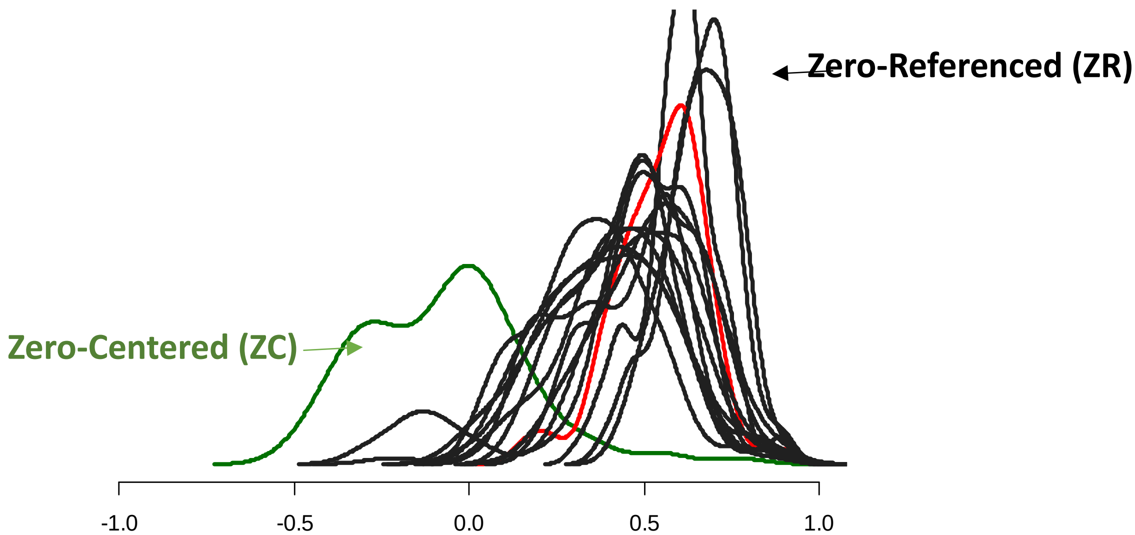
Comparing the distribution of the 14*13/2 = 91 correlations computed for all pairs of zero-referenced utilities R1.6–R15.6 (ignoring the reference R6.6 = 0), to correlations computed for all pairs of the associated *zero-centered* utilities Z1–Z15 (ignoring the reference Z6), we see a clear shift to the right (higher correlations) for the *zero-referenced* correlations (plotted in red). For example, while the correlation between items Z4 (Equity) and Z15 (Value) = .04, the corresponding value when zero-referencing is used is Corr(Z4.6, Z15.6) = .56.

Figure 2: Comparison of Correlation Distributions Based on ZC (green) and ZR.6 (red) Utilities



As shown in Figure 3, zero-referencing induces spurious positive correlations regardless which utility is used as reference!

Figure 3: Comparison of Correlation Distributions Based on ZC (green) and All ZR Utilities



Since standard LC Cluster models utilize correlations along with means and variances as part of their estimation criteria, different correlations associated with different HB coding yield different segments. To explore the quality of the segments obtained from the different coding, in the next section we simulate data from known "true" populations and assess the accuracy of the segments obtained using ZC vs. ZR coding.

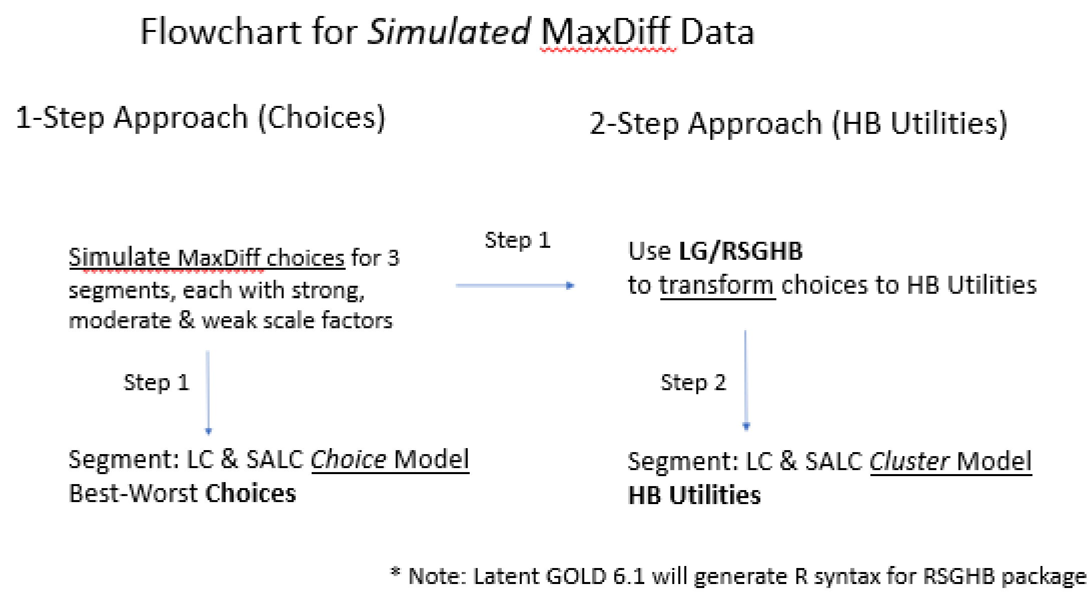
Simulating Data to Assess Accuracy

When segmentation is performed on real data, one is somewhat limited in drawing inferences regarding whether the resulting segments are "correct." A benefit of simulation is that one cannot only determine whether the *number of segments* is correct, but also the accuracy with which simulated respondents are assigned to their "true" segment. We can do this because we ***assume*** that there are, say, 3 meaningful segments that differ in their preferences and we generate respondents that belong to each of these segments according to the LC model parameters used to define that segment.

To be most useful, we will simulate 3 preference segments using parameters similar to those estimated using the Flynn dataset. For concreteness, we will refer to these 3 segments as those whose most important health care reform principle is Value for $ (Segment 1), People and Families (Segment 2), and Equity (Segment 3), respectively.

The following flowchart describes the process used to simulate 3 segments of respondents who differ in their *preferences* as described above. Since *real-world* respondents also differ in their *preference strength*, respondents within each segment are simulated such that they vary between Strong, Moderate and Weak preference strength.[2] (For further simulation details, see Appendix A.)

Figure 4: Flowchart of the Process Used to Simulate Respondents



Further Exploration of the HB Utility Coding Issue

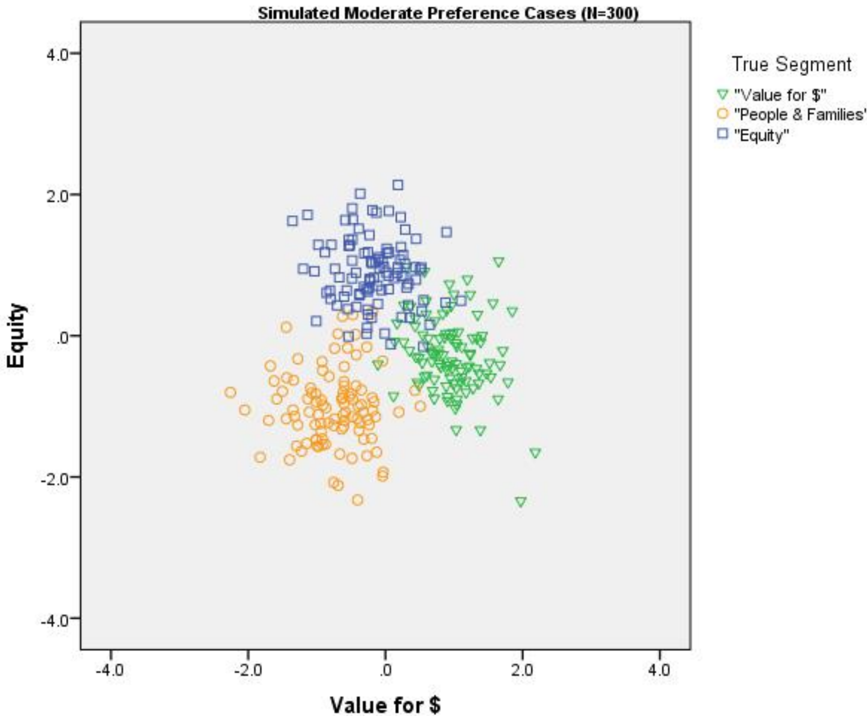
Lyon (2019) demonstrated that poor agreement exists between segmentations obtained using LC Clustering of zero-referenced HB utilities, with different reference points. Since this LC Clustering model assumes implicitly that all respondents have the same scale factor (i.e., the same preference strength), to further explore the effects of coding we use the subset of 300 simulated respondents that exhibited *Moderate* preference strength, consisting of 100 respondents from each of the 3 segments.

Fig. 5 is a plot of the respondents simulated to be in the Moderate preference group. These respondents are distinguished by color according to their true segment. As can be seen, there is moderately good separation between these 3 segments in the 2-dimensional space formed by the zero-centered utilities "Equity" and "Value for $."

---

[2] The Latent GOLD® syntax was used to assign each simulated respondent to one of 3 discrete scale factors (called "scale classes" or "sClasses"). Alternatively, the syntax could have generated *continuous* scale factors to be distributed among the respondents. If the latter approach had been used, the results would have been similar. For ease in explaining, we used the discrete approach.

Figure 5: Scatterplot of Simulated Respondents with Moderate Preference Strength According to ZC HB Utilities for Equity and Value for $.



ABOUT LC CLUSTERING OF HB UTILITIES

LC Clustering, also known as *LC Profile modeling* when the variables are continuous, utilizes information from means, variances and correlations of continuous variables to obtain latent class segments, correlation information taken into account implicitly through the use of the *local independence* criterion (see Vermunt and Magidson, 2004).

In this section we analyze simulated data to investigate the accuracy of assigning respondents to the correct preference class (i.e., to the true segment) when using

- the 1-Step vs. 2-Step approach to LC modeling

and, under the 2-Step approach:

- 1) How well the LC Clustering model reproduces the "true" segments
  - o when ZC coded HB utilities are used as input to LC, and
  - o when ZR coded HB utilities are used as input to LC.
- 2) We will also repeat 1) after relaxing the LC local independence assumption

We begin our analysis by using LC Clustering model with HB utilities (the 2-Step approach).

When utilities are zero-centered, LC Clustering achieves high accuracy, 94.3% of simulated respondents being classified into the correct segment. This is depicted in Fig. 5A. Alternatively, when z*ero-referencing* is used to identify the utilities with Utility 6 (Wellness/Prevention) as reference, a large positive spurious correlation is induced. This is illustrated in Fig. 5B where the correlation between the zero-referenced utilities Equity and Value for $ is .57 vs. .05 when zero-centering is used (as illustrated in Fig. 5A).

Figure 5A: LC Clustering with ZC Utilities

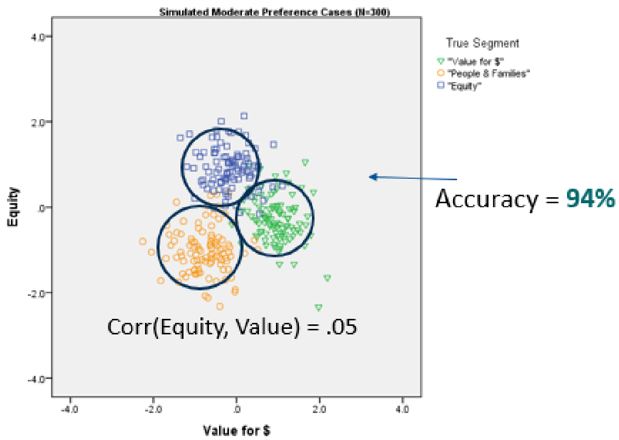
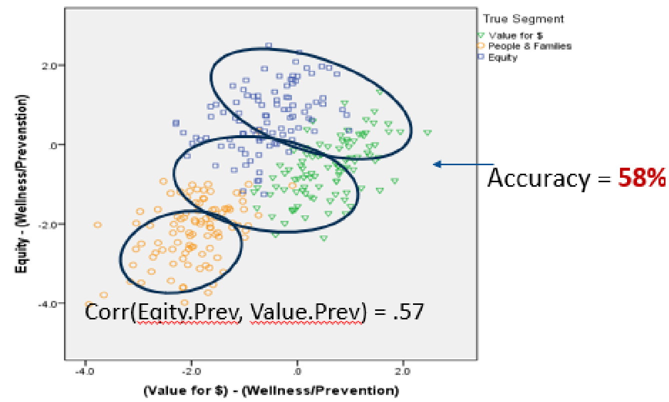


Figure 5B: LC Clustering with ZR Utilities



Column 2 of Table 4 shows that the 3-cluster solution obtained by clustering on ZC utilities achieves high accuracy (94.3%) compared to ZC utilities, which range from 57.8% to 94.3% depending on the reference. The lowest accuracy (57.8%) is obtained with Utility 6 as reference, as illustrated earlier (recall Fig. 2).

Table 4: Summary of Accuracy Obtained under LC Clustering of HB Utilities for 3-Class LC Models Estimated under Local Independence (Column 2) vs. Local Dependence (Column 3)

	% Classified Correctly	
Reference	+ local independence	local dependence
Centered	94.3%	89.9%
12	94.3%	89.9%
13	93.2%	89.9%
14	93.1%	89.9%
2	92.2%	89.9%
7	91.9%	89.9%
9	89.9%	89.9%
1	87.9%	89.9%
10	87.5%	89.9%
4	80.6%	89.9%
11	79.6%	89.9%
5	72.0%	89.9%
15	70.0%	89.9%
8	68.6%	89.9%
3	62.2%	89.9%
6	57.8%	89.9%

As mentioned above, the effect of the local independence assumption is to allow utility correlations to affect the cluster solution. When ZR coding is used, these correlations contain a *spurious* component, which results in clusters that are less accurate and thus less meaningful. For example, when utility 6 is used as reference, the correlation between the corresponding ZR utilities **Equity** and **Value for $** is .57, which contains a large spurious component.

The local independence assumption causes clusters to be chosen that "explain" this large .57 correlation, resulting in the clusters being stacked from the lower left to the upper right, tracing out a moderate positive correlation (see Fig. 5B). Since most of this correlation is spurious, the result is a low accuracy (.58). When the local independence assumption is relaxed[3], removing the correlations from being part of the cluster solution criteria, the result is shown in Fig. 5C, with the higher accuracy of .90.

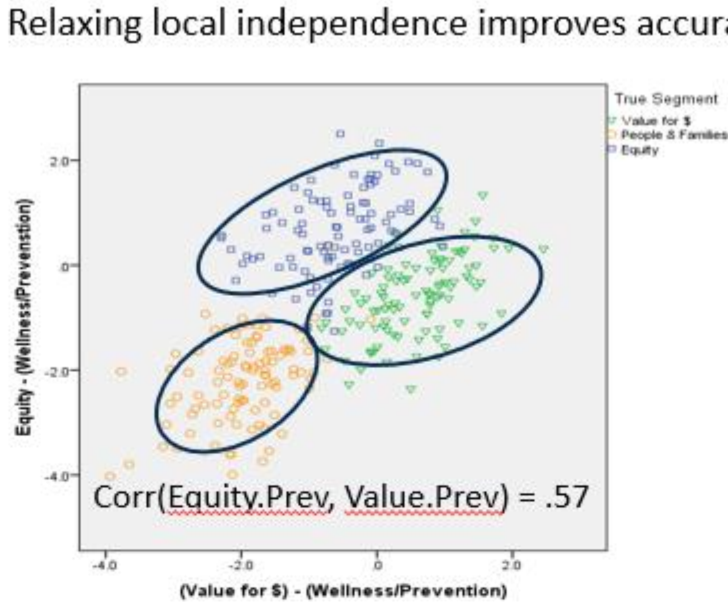
The right-most column in Table 4 shows that regardless of whether ZC coding is used, or whether ZR coding with any utility is used as reference, the resulting clusters will be identical (with 89.9% accuracy) when the local independence assumption is relaxed. Why does this occur? Relaxing local independence removes the correlations from being used as a criterion for determining the clusters. Removing the correlations means removing both the *spurious* as well as the *non-spurious* portion of the correlation, leaving the means and variances as the only two remaining criteria.

Fig. 5C illustrates the improved accuracy resulting from relaxing the local independence assumption. The resulting clusters no longer stack up from the lower left to the upper right (recall Fig. 5B) in order to explain the large observed correlation of .57. As a result, the accuracy improves as each ellipse is now able to capture more respondents belonging to the same true segment.

---

[3] Relaxing local independence is performed in the Latent GOLD® syntax by including *direct effects* between each utility pair which capture and utilize the *observed* utility correlations as explicit *external* model parameters instead of requiring *internal* model parameters to be estimated. (See Appendix B.) Since a portion of the correlation is spurious, the resulting distortion exhibited under the local independence restriction no longer occurs, resulting in higher accuracy. However, because not all of the correlation is spurious, the accuracy is not as high as it would have been had the non-spurious portion of correlation been allowed to be explained by the clusters. Thus, the accuracy increases to .90, which falls somewhat short of the .94 which occurs through the use of ZC, ZC containing no spurious correlation.
Regardless of what coding is used for the HB utilities, the fifteen *mean values* change only by a constant (the mean of the reference attribute in the class concerned)—so the means retain their effect in determining the clusters. While the *variances* of the utilities, as can be seen in Fig. 5C, *do* change, this is not a problem since variances are included explicitly in the Latent GOLD syntax as part of the model parameters. Thus, relaxing local independence removes only the correlations from the criteria. The fact that the accuracy is identical when the correlations are removed from the criteria shows that the low accuracy is caused by *spurious* correlations.

Figure 5C: LC Clustering with ZR Utilities with Direct Effects Included to Relax the Local Independence Assumption



Conclusion: In practice, one should *not* relax local independence, since the net effect is a reduction in accuracy from 94.3% when ZC coding is used, to 89.9%, a reduction of 4.5%. Instead, ***one should always use ZC coding when segmenting on HB utilities.***

Using Model Fit in Determining the Number of Classes

In this section we continue with the analysis of N=300 simulated respondents with Moderate preference strength (scale factor = .37) and examine how well various model fit criteria work in discovering that the true number of Preference Classes is 3.

For the *1-Step* approach, Table 5A shows that BIC, AIC3 and CHull all correctly select the 3-class model as the best fit (see Table 5A).

Table 5A: Model Fit Statistics for LC Choice Models Fit to 300 Simulated Respondents with Moderate Preference Strength

Moderate Scale (N = 300)				
LC Choice	LL	BIC	AIC3	CHull
1-Class	-16744	33568	33530	
2-Class	-16459	33083	33005	1.8
3-Class	-16298	32848	32729	10.2
4-Class	-16283	32902	32742	1.2
5-Class	-16270	32961	32761	1.0
6-Class	-16256	33020	32780	

Table 5B provides the corresponding model fit statistics for the *2-Step* approach, where HB utilities are analyzed using the LC Cluster model. In contrast to models based on the 1-Step approach, the BIC and AIC3 do not work correctly with the 2-Step approach, always selecting too many clusters (i.e., BIC continues to decrease as the number of classes increase). This suggests that the complexity associated with the additional step of estimating HB utilities cannot be properly accounted for by the Information Criteria (BIC, AIC3) assumptions.

Nevertheless, we note that the CHull scree statistic, proposed for use with complex models (Bulteel et al., 2013), correctly selects the 3-class model as best for both the 1-Step and 2-Step approaches. This suggests that a scree statistic, such as CHull should be used for the 2-Step approach, as well as other applications where clustering of random effects is performed (see e.g., Magidson and Vermunt, 2024).

Table 5B: Model Fit Statistics for LC Cluster Models Estimated on HB Utilities

Moderate Scale (N = 300)				
HB Cluster	LL	BIC	AIC3	CHull
1-Cluster	-2390	4951	4870	
2-Cluster	-1342	2947	2822	1.8
3-Cluster	-760	1874	1706	2.9
4-Cluster	-561	1567	1356	1.1
5-Cluster	-374	1284	1030	1.0
6-Cluster	-186	1000	702	1.0

In conclusion, using BIC to assess the number of classes for the 1-Step approach and the CHull for the 2-Step approach in both cases we get the correct number of classes 3, and in both cases the accuracy is over 90%. The accuracy for the 3-class Choice model is 92% and the corresponding accuracy for the 3-class Cluster model (the 2-Step approach), as mentioned earlier, is 94%.

While the 3-segments obtained from the 1-Step and 2-Step approaches are both meaningful in the sense that they extract the true segments fairly accurately, in the case that one clusters on HB utilities (2-Step approach), one should not rely on the BIC or other information statistic to assess the number of classes. Instead, one should consider using CHull[4], a scree-based heuristic (Bulteel et al., 2013).

We note that when attempting to estimate SALC models, the number of scale classes was correctly determined to be 1 (based on the BIC statistic for the SALC Choice model and based on the CHull statistic for the SALC Cluster model).

[4] The CHull statistic will be added to Latent GOLD® 6.1 along with an interface to Jeff Dumont's R package RSGHB (see Vermunt and Magidson, 2021a; and Vermunt and Magidson, 2005).

Including Scale Heterogeneity in Our Simulation

Since respondents in the *real world* differ not only in their preferences, but also in their *preference strength*, a more realistic simulation would need to include respondents that *do* differ in preference strength. In this section we expand our analysis sample to include the additional 300 respondents simulated to have *Weak* preference and the additional 300 respondents simulated to have *Strong* preference (recall Figure 4).

Simulated respondents were equally distributed among 3 Preference Segments and 3 Scale Class groups within each Preference Segment (N=100 per cell), for a total of 3x3=9 joint classes as shown below.

Table 6: The 9 Joint Segments Comprised by Preference x Strength of Preference Segments (N=100 in each)

Strength of Preference (Scale)	Pref. Segment 1	Pref. Segment 2	Pref. Segment 3
Strong	1	2	3
Moderate	4	5	6
Weak	7	8	9

Comparing Simulation Results for the 1-Step and 2-Step SALC Models

Below are the parameter estimates obtained by estimating SALC models on the 900 simulated respondents using the 1-Step and 2-Step approaches.

Table 7: SALC Models Estimated on the N=900 Simulated Respondents

Principles	*1-Step Approach* 3-Class **SALC Choice** Value for $	People & Family	Equity		*2-Step Approach* 3-Class **SALC Cluster** Value for $	People & Family	Equity
1: A culture of reflective improvement & innovation	-2.1	-0.9	-3.0		-1.8	-0.7	-2.5
2: A respectful, ethical system	-0.7	0.8	1.0		-0.5	0.5	0.7
3: Comprehensiveness	0.0	-1.8	0.0		-0.1	-1.4	0.0
4: Equity	-0.9	-2.8	2.9		-0.4	-2.3	2.2
5: People & family centered	-0.1	2.9	1.8		0.1	2.3	1.4
6: Promoting wellness & strengthening prevention	1.0	2.9	0.1		0.7	2.4	0.2
7: Providing for future generations	0.0	2.0	0.1		0.0	1.5	0.0
8: Public voice & community engagement	-3.2	-0.8	-2.0		-2.7	-0.7	-1.7
9: Quality & safety	3.0	1.9	2.9		2.6	1.4	2.5
10: Recognize social & environ influences shape health	-0.9	0.9	-2.0		-0.8	0.6	-1.6
11: Responsible spending	2.0	-1.1	0.1		1.5	-0.7	0.2
12: Shared responsibility	-1.1	-1.0	-0.9		-0.9	-0.8	-0.7
13: Taking the long term view	-0.1	0.0	-1.0		-0.2	-0.1	-0.7
14: Transparency & accountability	0.2	-0.9	0.9		0.2	-0.6	0.6
15: Value for money	3.0	-2.0	-0.9		2.3	-1.5	-0.4
Class Size	0.33	0.34	0.33		0.35	0.33	0.32

Comparing the highlighted segment-specific parameter estimates for the 2-Step approach with the corresponding values from the 1-Step approach, we find that on average, the magnitude of the 2-Step parameters is about .8 times that of the 1-Step parameters, corresponding to a 20% falloff. This regression to the mean was expected due to the additional heterogeneity introduced under the Bayesian estimation of the HB utilities. Nevertheless, despite this shrinkage, the segmentations are virtually identical—overall, 96% of simulated respondents being classified into the same segment by the different approaches.

Using all N=900 simulated respondents, and knowing to which true preference segments each respondent belongs (as well as knowing their preference strength), we can also address the question:

- How do the SALC models differ in accuracy?
  - 1-Step approach: How accurate is the SALC Choice model segmentation?
  - 2-Step approach: How accurate is the SALC Cluster model segmentation?

As we might expect given that the segmentations are so similar, accuracy is also quite similar—the 1-Step approach achieves 85% accuracy compared to 84% for the 2-Step approach.

EFFECT OF PREFERENCE*SCALE CONFOUNDS

In this section we consider the extent to which *preference*scale confounds* interfere with the ability to uncover meaningful segments when using unstructured LC as opposed to SALC models. In particular, can we rely on model fit criteria such as BIC to guide us in choosing the number of classes? We continue to use our simulated data in making this assessment so we know the true underlying structure as we assess the extent to which meaning is lost.

RESULTS FROM THE 1-STEP APPROACH

When analyzing the MaxDiff choice data using (unstructured) LC modeling, Table 6 reminds us that the number of true (joint) segments is 3x3=9. Beginning with the 1-Step approach to segmentation modeling, LC Choice models with 6–8 classes are highlighted in Table 8 below because they are selected by the BIC and AIC3 criteria as providing the best fit to the data (i.e., they have the lowest values for BIC and AIC3).

Table 8: Model Fit Statistics for 10 Estimated LC Choice Models

	Simulated Data (N=900)		
LC Choice	LL	BIC	AIC3
1-Class	-49702	99499	99432
2-Class	-48084	96365	96226
3-Class	-46975	94249	94038
4-Class	-45825	92051	91768
5-Class	-45670	91843	91487
6-Class	-45550	91705	91278
7-Class	-45501	91709	91209
8-Class	-45483	91775	91203
9-Class	-45468	91848	91205
10-Class	-45455	91924	91209

Based on this model fit summary presented in Table 8, we select 7 classes, which is midway among the 3 models most preferred according to the BIC and AIC3. Table 8A shows how these 7 classes are structured with respect to their true preferences and true preference strength (scale class).

Table 8A: 7-Class 1-Step LC Solution in Terms of True Class (Preference Segment) and True sClass (Scale Segment)

	Class						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Class Size	0.18	0.17	0.17	0.14	0.12	0.12	0.11
True Class							
1	0.18	**0.41**	0.05	0.03	0.00	0.00	**0.33**
2	0.21	0.04	0.05	**0.35**	0.00	**0.35**	0.00
3	0.16	0.07	**0.40**	0.02	**0.35**	0.00	0.00
True Scale Class							
1	0.00	0.01	0.00	0.01	**0.33**	**0.33**	**0.33**
2	0.08	**0.32**	**0.28**	**0.28**	0.02	0.02	0.00
3	**0.46**	0.20	0.22	0.12	0.00	0.00	0.00

As shown in Table 8A, note first that class 1 is the *largest* class, consisting of about 18% of the cases. Since this class contains approximately equal numbers from each of the three True Preference Classes, respondents in this class do not stand out as having different preferences from the other classes. Specifically, 18% of True Class 1, 21% of True Class 2, and 16% of True Class 3 belong to Class 1. This class can properly be referred to as a “low-scale” or less-certain class, as 46% of respondents with Weak preference (True Scale Class = 3) comprise this class.

The remaining 6 classes consist of 2 classes from each *Preference Class*, one of which comes from the *Strong Preference* group (Scale Class = 1) and one from the *Moderate Preference* class group (Scale Class = 2). In practice, one would not, of course, be privy to the information regarding True Class and True Scale Class.

Without such true class information, but with only parameter estimates (Table 8B) at our disposal, one might decide to maintain class 1 as a meaningful “low scale class” segment, and one might take on the difficult task of deciding whether to reduce the number of classes by possibly combining classes 2 and 7 (True Preference Class 1), classes 3 and 5 (True Preference Class 3) and classes 4 and 6 (True Preference Class 2). The task of obtaining meaningful segments would even be more daunting if one did not utilize the model fit statistics and selected fewer than 7 total classes.

Table 8B: Parameter Estimates for the 7-Class Choice Model for 900 Simulated Respondents

Principles	Class1	Class2	Class3	Class4	Class5	Class6	Class7
1: A culture of reflective improvement & innovation	-0.1	-0.5	-0.9	-0.4	-3.0	-0.9	-2.1
2: A respectful, ethical system	0.1	-0.2	0.2	0.3	1.0	0.7	-0.7
3: Comprehensiveness	0.0	-0.1	0.1	-0.7	0.0	-1.8	0.0
4: Equity	0.0	-0.3	0.9	-0.9	2.8	-2.9	-0.9
5: People & family centered	0.2	-0.1	0.5	1.0	1.8	2.9	-0.1
6: Promoting wellness & strengthening prevention	0.3	0.2	0.0	1.1	0.1	2.9	1.0
7: Providing for future generations	0.1	0.0	0.0	0.5	0.0	2.0	-0.1
8: Public voice & community engagement	-0.2	-0.9	-0.6	-0.4	-2.0	-0.7	-3.2
9: Quality & safety	0.2	1.0	0.9	0.6	2.9	1.9	3.0
10: Recognize social & environ influences shape health	0.0	-0.2	-0.6	0.5	-2.0	0.8	-0.9
11: Responsible spending	-0.1	0.7	0.0	-0.3	0.1	-1.0	2.0
12: Shared responsibility	-0.2	-0.4	-0.3	-0.3	-0.8	-1.0	-1.0
13: Taking the long term view	0.0	-0.1	-0.1	-0.1	-1.1	0.0	0.0
14: Transparency & accountability	-0.2	0.2	0.3	-0.1	0.9	-1.0	0.1
15: Value for money	0.0	0.9	-0.5	-0.8	-0.8	-2.0	3.0

For comparison, the model fit statistics for SALC models (Table 9 below) show that the 3-sClass/3-Class SALC model provides the best fit, the parameter estimates being provided in the left-most portion of Table 7.

Table 9: BIC Statistics for the SALC Choice Models Estimated on Simulated Data

SALC Choice	Model Description	LL	BIC(LL)	Npar
1-class SALC	3-sclass/1-Class	-49067	98256	18
2-class SALC	3-sclass/2-Class	-46836	93896	33
3-class SALC	3-sclass/3-Class	-45535	91396	48
4-class SALC	3-sclass/4-Class	-45521	91471	63

Although not shown in Table 7, it is noteworthy that SALC Choice models with 2 and 4 sClasses did not fit as well as SALC models with the true number of 3 scale classes.

Results from the 2-Step Approach

Similar to results reported earlier (recall tables 7 and 9), when LC Cluster and SALC Cluster models based on the *2-Step* approach are utilized, Tables 10 and 11 show once again that information statistics fail to select the correct number of classes as both BIC and AIC continue to decrease when more clusters are added. However, the CHull heuristic suggests 7 clusters, in agreement with the corresponding 1-Step approach.

Table 10: Various Model Fit Statistics for LC Cluster Models Estimated on Simulated Data

	Simulated Data (N=900)			
HB Cluster	LL	BIC(LL)	AIC3	CHull
1-Cluster	-12352	24909	24765	
2-Cluster	-9507	19328	19107	1.1
3-Cluster	-6942	14305	14007	1.3
4-Cluster	-4950	10431	10057	1.6
5-Cluster	-3720	8080	7628	1.6
6-Cluster	-2965	6679	6151	1.4
7-Cluster	-2426	5708	5103	1.8
8-Cluster	-2128	5223	4541	1.0
9-Cluster	-1845	4764	4005	1.2
10-Cluster	-1603	4390	3554	

Table 11 shows that the CHull statistic correctly selects the 3-Cluster 3-sClass model as best. (Although not shown in Table 11, we note that SALC Cluster models with 2 and 4 sClasses did not fit as well as SALC models with the true number of 3 scale classes.)

Table 11: Various Model Fit Statistics for SALC Cluster Models Estimated on Simulated Data

HB SALC	LL	BIC(LL)	CHull
1-Cluster 3-sClass	-10637	21506	
2-Cluster 3-sClass	-5546	11432	1.5
3-Cluster 3-sClass	-2249	4946	6.0
4-Cluster 3-sClass	-1701	3961	1.2
5-Cluster 3-sClass	-1240	3146	1.1
6-Cluster 3-sClass	-832	2440	0.5

Table 12: 7-Class 2-Step SALC Solution in Terms of True Class (Preference Segment) and True sClass (Scale Segment)

HB Cluster	**Cluster**						
SALC	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Cluster Size	0.20	0.17	0.15	0.14	0.11	0.11	0.11
True Class							
1	0.20	0.05	**0.38**	0.03	0.00	0.00	**0.33**
2	0.23	0.03	0.02	**0.38**	0.00	**0.34**	0.00
3	0.17	**0.42**	0.05	0.02	**0.34**	0.00	0.00
True Scale Class							
1	0.00	0.00	0.00	0.00	**0.33**	**0.33**	**0.33**
2	0.05	**0.31**	**0.30**	**0.31**	0.01	0.01	0.00
3	**0.55**	0.19	0.15	0.12	0.00	0.00	0.00

In summary, regarding use of the standard (unstructured) LC models:

- Appropriate fit statistics for the LC Choice model (Table 7) and the LC Cluster model (Table 10) both suggest 7 classes, one of which is a “low scale” class.
- In both of these cases, the LC models are “messy” in the sense that there are many classes, some of which should be combined to provide more meaningful segments.
- Overall, there was 90% agreement between the 7-class models obtained by the 1-Step and 2-Step approaches. In addition, these models yield similar scale confounds:
  - Class 1 consists mostly of respondents with Weak preferences (True Scale Class 3).
  - Each of the three True Preference Classes splits into separate classes for *Strong* and *Moderate* preferences, resulting in the remaining 6 classes (clusters).
  - These confounds are somewhat similar to what we saw in the original data.

CONCLUSIONS AND DISCUSSION

The simulation confirmed several results from our original analyses regarding the usefulness of the SALC model to a) segment MaxDiff choice responses (the 1-Step approach) and b) segment HB utilities (the 2-Step approach), and the similarity of both approaches. Namely,

- The segments produced by 1-Step and 2-Step SALC models are similar so long as HB Utilities in the 2-Step approach are zero-centered. Overall, 88% of the 204 actual respondents were assigned to the same segment by the different approaches. With the simulated data, the agreement rate increased to 96%, and segments from both approaches proved meaningful—84% to 85% accuracy in uncovering the meaningful segments.
- While the 2-Step approach led to smaller class-specific preferences due to additional complexity in the second step, such expected shrinkage did not affect the clustering. Moreover, the fact that the HB utilities are contained in a simple rectangular data file makes it much easier to include additional variables in the segmentation, such as ratings, than the more complex data fusion that would be required to analyze choice responses and ratings (see e.g., Magidson et al., 2009).

From these results, it is clear that the parsimonious 2-dimensional SALC structure meets the theoretical challenge of handling the 2 different kinds of parameters—Preference and Scale—in a way that avoids confounds. Returning to the question of whether the 1-Step or 2-Step approaches should be called the “gold standard,” the answer is not so clear.

The current study was limited to MaxDiff data, and to some extent application to the Australian Health Care Reform Study. It is always useful to replicate the findings on additional data, and to explore the usefulness of the SALC model for choice applications beyond MaxDiff. In addition, a more extended simulation study might be undertaken to investigate the performance of the CHull model fit statistic.

In practice, one might not utilize model fit criteria at all, and simply examine LC solutions with say 2–5 classes to get a seemingly “interpretable” solution. But because we are uncovering *joint classes* based on both Preference and Strength of Preference, we might miss out in uncovering important segments by stopping with 5 classes. Alternatively, if we estimate say 6 to

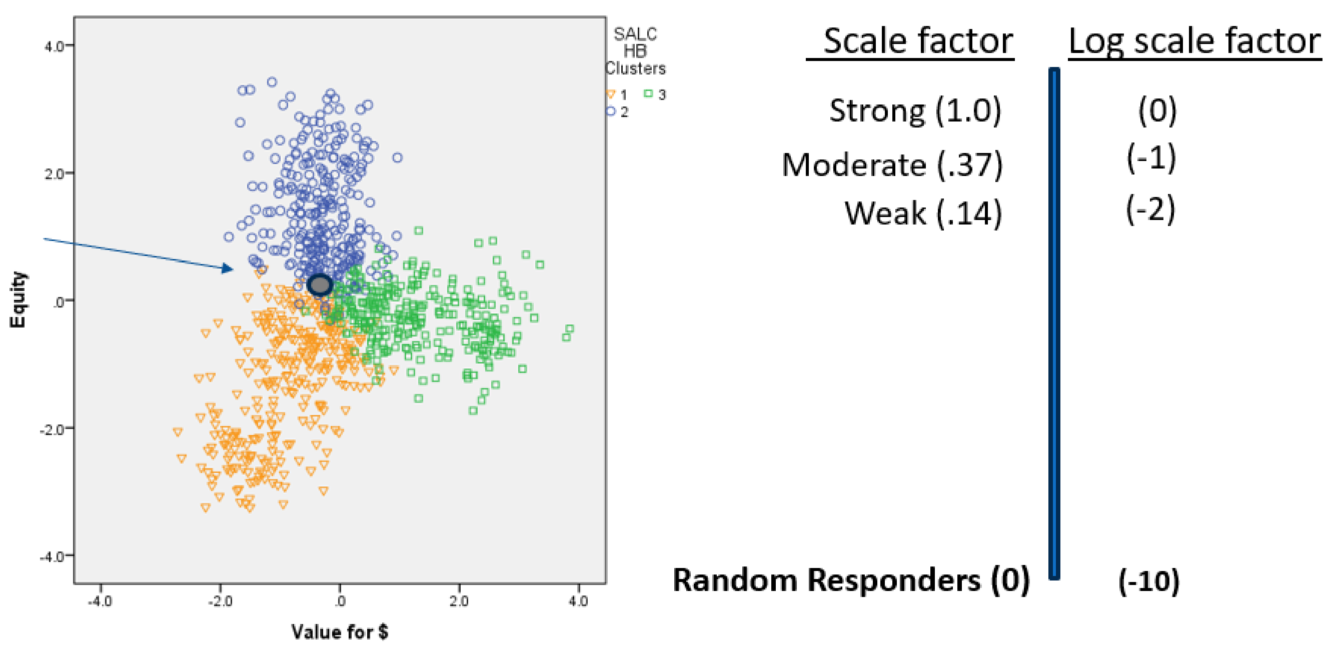
10 or more joint classes, we may well run into a messy situation where we would need to combine these in the "appropriate" way to get the most meaningful segments, and we may not know that we need to combine them and may not know how best to combine them.

Ongoing Research—Use SALC as Filter to Identify and Exclude Random Responders?

It should be noted that Scale can be modeled in the Latent GOLD® syntax (Vermunt and Magidson, 2021b) using a *discrete* (nominal or ordinal) latent variable (nominal or ordinal scale classes) or a *continuous* latent variable (sCFactor). While *nominal* scale classes were used here to simulate data, similar results would have been obtained had the continuous approach been used.

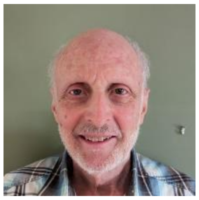
Another alternative is to extend the SALC model to include an additional scale class fixed at a very low value (say at a log-scale factor value of -10) to capture random responders. For example, a 4-category sClass variable can be fixed using log-scale factor values of 0, -1, -2 and -10, or the first 3 categories can be treated as nominal (freely estimated with the first category being the dummy coded 0-reference) and the last category fixed at -10. This model can then be estimated with the Latent GOLD 6.0 syntax (which would also require the Latent GOLD 6.0 choice module), and those respondents classified as random could then be eliminated from the analysis sample prior to re-estimating the model.

Figure 6: Scatterplot with Predicted Random Responders Identified



Acknowledgements

The authors wish to thank Tom Eagle, Dave Lyon and Tony Babinec for their helpful comments.



Jay Magidson

Jeroen K. Vermunt

APPENDIX A: PARAMETERS USED FOR SIMULATION

The assumed preference parameters were taken to be similar to those parameters estimated by the SALC model using the 1-Step approach (recall Table 1). See Table A1 for the parameters used to simulate Strong preference strength group (scale factor = 1). These were multiplied by .37 for the Moderate preference strength group (scale factor = .37), and multiplied by .14 to simulate the Weak preference group (Table A2). For each of these groups, N = 300 respondents were simulated, 100.

Table A1: Assumed Population Parameters Used to Simulate Moderate and Strong Preference Groups

Principles	Scale factor = 1.0 (Strong)			Scale factor = 0.37 (*Moderate)*		
	Segment	Segment	Segment	Segment1	Segment2	Segment3
1: A culture of reflective improvement & innovation	-2	-1	-3	-0.7	-0.4	-1.1
2: A respectful, ethical system	-1	1	1	-0.4	0.4	0.4
3: Comprehensiveness	0	-2	0	0.0	-0.7	0.0
4: Equity	-1	-3	3	-0.4	-1.1	1.1
5: People & family centered	0	3	2	0.0	1.1	0.7
6: Promoting wellness & strengthening prevention	1	3	0	0.4	1.1	0.0
7: Providing for future generations	0	2	0	0.0	0.7	0.0
8: Public voice & community engagement	-3	-1	-2	-1.1	-0.4	-0.7
9: Quality & safety	3	2	3	1.1	0.7	1.1
10: Recognize social & environ influences shape health	-1	1	-2	-0.4	0.4	-0.7
11: Responsible spending	2	-1	0	0.7	-0.4	0.0
12: Shared responsibility	-1	-1	-1	-0.4	-0.4	-0.4
13: Taking the long term view	0	0	-1	0.0	0.0	-0.4
14: Transparency & accountability	0	-1	1	0.0	-0.4	0.4
15: Value for money	3	-2	-1	1.1	-0.7	-0.4

Table A2: Assumed Population Parameters Used to Simulate Weak Preference Group

Principles	Scale factor = 0.14 (Weak)		
	Segment2	Segment2	Segment3
1: A culture of reflective improvement & innovation	-0.3	-0.1	-0.4
2: A respectful, ethical system	-0.1	0.1	0.1
3: Comprehensiveness	0.0	-0.3	0.0
4: Equity	-0.1	-0.4	0.4
5: People & family centered	0.0	0.4	0.3
6: Promoting wellness & strengthening prevention	0.1	0.4	0.0
7: Providing for future generations	0.0	0.3	0.0
8: Public voice & community engagement	-0.4	-0.1	-0.3
9: Quality & safety	0.4	0.3	0.4
10: Recognize social & environ influences shape health	-0.1	0.1	-0.3
11: Responsible spending	0.3	-0.1	0.0
12: Shared responsibility	-0.1	-0.1	-0.1
13: Taking the long term view	0.0	0.0	-0.1
14: Transparency & accountability	0.0	-0.1	0.1
15: Value for money	0.4	-0.3	-0.1

Appendix B: Latent GOLD® Syntax to Relax Local Independence Assumption in LC Cluster Model Estimated on Simulated Respondents with Moderate Preference Strength

```
variables
   select sClassTrue =  2; // select only simulated respondents with Moderate preference strength
   dependent Jan8dum1.6 continuous, Jan8dum2.6 continuous, Jan8dum3.6 continuous,
      Jan8dum4.6 continuous, Jan8dum5.6 continuous, Jan8dum6.6 continuous,
      Jan8dum7.6 continuous, Jan8dum8.6 continuous, Jan8dum9.6 continuous,
      Jan8dum10.6 continuous, Jan8dum11.6 continuous, Jan8dum12.6 continuous,
      Jan8dum13.6 continuous, Jan8dum14.6 continuous, Jan8dum15.6 continuous;
   independent ClassTrue nominal;
   latent
      Cluster nominal 3;
equations
   Cluster <- 1;
   Jan8dum1.6 - Jan8dum15.6 <- 1 + Cluster;
  Jan8dum1.6 - Jan8dum15.6;  // variances
```

```
//Relax local ind. by adding direct effects
Jan8dum2.6 <-> Jan8dum1.6 ; // direct effects
Jan8dum3.6 <-> Jan8dum1.6 ;
Jan8dum3.6 <-> Jan8dum2.6 ;
Jan8dum4.6 <-> Jan8dum1.6 ;
Jan8dum4.6 <-> Jan8dum2.6 ;
Jan8dum4.6 <-> Jan8dum3.6 ;
Jan8dum5.6 <-> Jan8dum1.6 ;
Jan8dum5.6 <-> Jan8dum2.6 ;
Jan8dum5.6 <-> Jan8dum3.6 ;
Jan8dum5.6 <-> Jan8dum4.6 ;
Jan8dum7.6 <-> Jan8dum1.6 ;
Jan8dum7.6 <-> Jan8dum2.6 ;
Jan8dum7.6 <-> Jan8dum3.6 ;
Jan8dum7.6 <-> Jan8dum4.6 ;
Jan8dum7.6 <-> Jan8dum5.6 ;
Jan8dum8.6 <-> Jan8dum1.6 ;
...
Jan8dum15.6 <-> Jan8dum14.6 ;
```

References

Bulteel, K., T.F. Wilderjams, F. Tuerlinckx, W. Ceulemans (2013). "CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers," Behavioral Research 45:782–791.

Eagle, T. and Magidson, J. (2019). "Segmenting Choice and Non-choice Data Simultaneously: Part Deux" 2019 Sawtooth Software Conference Proceedings, pp. 231–250.

Lee, J. and J. D. Brazell (2019). "Multivariate analysis with MaxDiff: A curious look at the properties of MaxDiff utilities," AMA Advanced Research Techniques Forum, Brigham Young University, Provo, UT.

Louviere, Jordan J., Thomas C. Eagle, 2006. "Confound it! That Pesky Little Scale Constant Messes Up," 2006 Sawtooth Software Conference Proceedings.

Louviere, J. J., & Flynn, T. N. (2010). "Using best-worst choice experiments to measure public perceptions and preferences for health care reform in Australia." The Patient: Patient-Centered Outcomes Research, 3(4), 275–283.

Lyon, D. (2019) Comments on "Segmenting choice and non-choice data simultaneously: Part deux," 2019 Sawtooth Software Conference Proceedings, page 281.

Magidson, J. (2018) "An improved latent class (LC) paradigm to obtain meaningful segments in the presence of scale confounds: Scale Adjusted Latent Class (SALC) Tree modeling," 2018 Advanced Research Techniques Forum, American Marketing Association.

Magidson, J. Thomas, D., and Vermunt, J.K. (2009). A new model for the fusion of MaxDiff scaling and ratings data. 2009 Sawtooth Software Proceedings, 83–103.

Magidson, J. and Vermunt, J.K. (2006), “Use of latent class regression models with a random intercept to remove overall response level effects in ratings data.” In A.Rizzi & M.Vichi (Eds.), Proceedings in computational statistics (pp. 351–360). Heidelberg, Germany: Springer.

Magidson, J and Vermunt, J.K. (2007). “Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference,” 2007 Sawtooth Software Conference Proceedings, pp 139–154.

Magidson, J. and Vermunt, J.K. (2024) “Latent Class Clustering of Random Coefficient Estimates Obtained from a Multilevel Analysis,” presented at Modern Modeling Methods Conference, University of Connecticut.

Popper, R., Kroll, J., and Magidson, J. (2004). Applications of latent class models to food product development: a case study. *Sawtooth Software Proceedings*, 2004.

Swait, J. and Louviere, J.J. (1993) “The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models,” Journal of Marketing Research, 30, 305–314.

Vermunt, J.K. (2004). “Latent profile model.” In: M.S. Lewis-Beck, A. Bryman, and T.F. Liao (eds.), *The Sage Encyclopedia of Social Sciences Research Methods*, 554–555. Thousand Oaks, CA: Sage Publications.

Vermunt, J.K. and J. Magidson (2005). “Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced. Belmont Massachusetts: Statistical Innovations Inc.”

Vermunt, J.K., and Magidson, J. (2004). “Local independence.” In: M.S. Lewis-Beck, A. Bryman, and T.F. Liao (eds.), *The Sage Encyclopedia of Social Sciences Research Methods*, 580–581. Thousand Oaks, CA: Sage Publications.

Vermunt, J.K. and Magidson, J. (2021a). Upgrade Manual for Latent GOLD Basic, Advanced, Syntax, and Choice Version 6.0. Latent GOLD 6.0 upgrade manual. Arlington, MA: Statistical Innovations Inc.

Vermunt, J.K., and Magidson, J. (2021b). Latent GOLD Syntax User’s Guide: Manual for Latent GOLD Syntax Module Version 6.0. Arlington, MA: Statistical Innovations Inc.