Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference[1]

Jay Magidson

Statistical Innovations Inc.

Jeroen K. Vermunt

Tilburg University

Introduction and Goal

Multinomial logit choice models based on latent class (LC) or HB methods are utilized in marketing research today along with simulators which predict choices and market shares. However, a weakness in these models creates a potential interpretability and validity problem. The problem is that the part-worth preference (utility) parameters that are used to make such predictions are generally confounded with a *scale parameter* which reflects the amount of uncertainty by different respondents.

This scale parameter confound is assumed not to exist in current HB models under the common restrictive assumption that the level of uncertainty, inferred by the values of the scale parameter, is identical for each respondent. Similarly, no confound occurs in LC choice models under the assumption that the values of the scale factor are the same for all respondents regardless of the latent class to which they belong (Louviere et. al., 2000, page 206). If the equality assumption is violated, the predictions contain additional amounts of error as well as potential bias.

We begin this paper by introducing the scale factor issue in conjunction with some simple hypothetical LC choice models. We then propose some extended LC choice models that can be used to test the equal scale factor assumption, and when violated, to estimate and thus remove the effects of the different scale parameters that exist for different respondent subgroups. By not allowing differences in the scale factor to be a determinant in the formation of different latent classes, LC segments resulting from this new approach more clearly differentiate respondents based on their true preference differences by revealing their scale free part-worth preference utilities. Thus, isolation of the scale factors should result in more *pure* measures of part-worth preference utilities.

[1] The authors wish to thank John Wurst for his helpful comments.

Applications of these new extended models[2] in 2 real world choice experiments find that in both cases the commonly made assumption of equal scale factors is unwarranted. Comparing results from the standard models to the extended models suggests that the former leads to some faulty conclusions regarding the actual part-worths as well as misclassification of a substantial number of respondents into the *wrong* LC segment. The results also suggest that discrete factor (DFactor) choice models, with and without a scale-factor adjustment, may well provide additional improvements in both data fit as well as interpretation of results over the more traditional LC choice models.

What is the Scale Parameter?

The scale parameter λ_i relates to the amount of *certainty* in respondent i's *expected choices*. Swait and Louviere (1993) pointed out for a given respondent i, standard choice modeling estimates the product $\lambda_i \beta$ rather than β. Thus, differences in the scale parameters should be isolated in order to obtain meaningful comparisons of preference part-worth β-parameters across individuals or groups.

The scale parameter can also be explained in terms of the variance in observed responses. Response variance is inversely related to λ (variance = $\pi^2/6\lambda_i^2$). Thus response variance may be viewed as a measure of *un*certainty (lack of certainty). For persons with λ approaching 0, the response variance approaches infinity, which reflects *complete* uncertainty. In this case, choice probabilities derived from the multinomial logit model are equal for all alternatives (Ben-Akiva and Lerman,1985).

The magnitude and importance of such confounds are not well understood. Clearly, Louviere and Eagle (2006) believe the scale factor issue is very important. In their 2006 Sawtooth Conference presentation they state:

> "All choice models confound scale and [preference part-worth] parameter estimates. The confound is particularly problematic in complex models like random coefficients [HB-like] models and latent class models if one cannot separate scale and [preference] parameters."
>
> "Thus, it is likely that random coefficient models are biased and misleading."

---

[2] These models, which include extensions of standard LC choice models as well as discrete factor choice models, have been implemented in the syntax version of the Latent GOLD Choice program (Vermunt and Magidson, 2008).

> "So, the bottom line is that one cannot estimate individual-level parameters from choice models unless one can separate scale and model parameter estimates." …

and

> "The field needs research that leads to new models that can capture both scale and systematic component (mean) effects."

As a simple illustration of the scale parameter, consider first the situation where there is complete homogeneity with regards to brand preference. Specifically, assume that all respondents prefer brand A over B, and B over C and have the following identical BRAND preference part-worth utilities: $\beta_A = 0.5$, $\beta_B = 0.1$, and $\beta_C = -0.6$.

Further suppose that 2 respondent subgroups exist, the first expressing less certainty than the other in several pair-wise brand comparisons. The scale factors for these two groups are denoted by λ[1] and λ[2] respectively. Without loss of generality, for purposes of parameter identification, we fix λ[1]= 1 for the first subgroup, and for concreteness assume that λ[2] = 2 for the second subgroup. This situation is reflected in Table 1.

Table 1: Assumed values for the preference parameter β, and the product λ[s]*β for each of the 2 subgroups.

	Preference Parameter (β)		λ[s]*β	
BRAND	Subgroup s=1	Subgroup s=2	λ[1] = 1 Subgroup s=1	λ[2] = 2 Subgroup s=2
A	+ 0.5	+ 0.5	+ 0.5	+ 1.0
B	+ 0.1	+ 0.1	+ 0.1	+ 0.2
C	- 0.6	- 0.6	- 0.6	- 1.2

These parameters yield choice probabilities (shown in Table 2) which are obtained from the multinomial (conditional) logit model equations:

$$\text{Prob(Brand = A| Subgroup s)} = \exp(\lambda[s]*\beta_A)/\text{SUM}(s),$$

$$\text{Prob(Brand = B| Subgroup s)} = \exp(\lambda[s]*\beta_B)/\text{SUM}(s),$$

$$\text{Prob(Brand = C| Subgroup s)} = \exp(\lambda[s]*\beta_C)/\text{SUM}(s),$$

where

$$\text{SUM}(s) = \exp(\lambda[s]*\beta_A) + \exp(\lambda[s]*\beta_B) + \exp(\lambda[s]*\beta_C).$$

From this example, it should be clear that a higher value for $\lambda[s]$ (as occurs in subgroup 2) translates into more extreme choice probabilities.

Table 2: Choice probabilities for each subgroup obtained from the multinomial logit model equations under the parameter values given in Table 1.

	Choice Probabilities	
BRAND	Subgroup s=1	Subgroup s=2
A	.50	.64
B	.33	.29
C	.17	.07

Scale Parameter as a Confound

LC choice analysis identifies $K \geq 1$ different latent classes, each of which is associated with a unique set of part-worth parameters. Since a *standard* LC choice analysis assumes $\lambda = 1$ for all respondents, if data were generated according to the probabilities in Table 2, this standard analysis would tend to mistakenly identify each of the 2 subgroups as different LC segments with *different* preference part-worths as shown in Table 3 below:

Table 3: Expected results under a standard LC choice analysis which mistakenly assumes λ[1] = λ[2] = 1 (True values shown in parenthesis).

	Estimated (True) part-worth preference parameters (β) under assumption λ[s] = 1	
BRAND	Subgroup 1	Subgroup 2
A	+ 0.5 (+ 0.5)	+ 1.0 (+ 0.5)
B	+ 0.1 (+ 0.1)	+ 0.2 (+ 0.1)
C	- 0.6 (- 0.6)	- 1.2 (- 0.6)

Table 3 shows the incorrect results where the part-worth parameter is confounded by the scale parameter. It is incorrect to infer that subgroup 2 (labeled 'Segment 2' in Table 3) prefers Brand A *more* than subgroup 1, or that subgroup 2 prefers brand C *less* than subgroup 1 because the true brand preference part-worths are in fact identical for both subgroups 1 and 2. The only true difference between respondents in Segments 1 and 2 is that the responses obtained from the former group reflect greater amounts of uncertainty.

Scale-Extended LC Choice Models

As mentioned above, *standard* LC choice models contain a K-category latent variable X representing K homogeneous LC segments. Each segment k has its own unique preference part-worths which are estimated under the assumption λ = 1 for all respondents.

LC Choice Model extension #1 contains 2 categorical latent variables, X* and S: X* denotes LC segments differing in their preference part-worth utilities. S denotes latent subgroups that differ in their scale parameter, with λ[1] fixed to the value 1 for purposes of identification. That is, the λ corresponding to the first subgroup, λ[1], is set to 1. The values for the other λ[s] parameters are assumed to be nonnegative.

In the example described above, the standard LC choice model results in 2 LC segments that appear to differ in their part-worth utilities for each of the attributes, but upon further inspection it can be seen that these part-worth utilities are proportional to each other. This is an indication that the λ equality assumption does not hold true.

These 2 segments can be explained more simply as a single homogeneous segment having the same preference part-worths, but with 2 different latent subgroups which have different scale parameter values λ[2] ≠ λ[1]. Each respondent in subgroup 1 shares the same scale parameter λ[1], and each respondent in subgroup 2 shares the same scale parameter λ[2]. That is, in this example S consists of 2 subgroups differing only in their level of uncertainty, not their preference part-worth utilities. Thus, S is dichotomous, and X* has only a single category (i.e., there is only 1 LC segment).

Next assume that X* has 2 underlying LC segments that truly differ in brand preferences, segment 1 most preferring brand A, while segment 2 most prefers brand C. Further suppose that each of these segments consists of both more and less certain respondents. This situation can be represented in terms of the 4 cells in Table 4 below. LC segment 1 consists of the 2 cells (1,1) and (1,2), while LC segment 2 contains cells (2,1) and (2,2).

Table 4: Example with 2 LC segments and 2 scale parameter subgroups

	Less certain subgroup (s = 1): λ[s]=1	More certain subgroup (s = 2): λ[s]=2
Segment 1 (X*=1): β_1 = (0.5, 0.0, -0.5)	joint class (1,1)	joint class (1,2)
Segment 2 (X*=2): β_1 = (-0.2, -0.8, 1.0)	joint class (2,1)	joint class (2,2)

Our scale-extended choice model generalizes the latent class (LC) multinomial logit choice model from its standard log-*linear* form to a log-*bilinear* form. The standard LC choice model expresses the expected utility of the j-th alternative as a *linear* function of the part-worth parameters. For example, with attribute A1 at level l_1, and A2 at l_2, the expected utility of class X=k is $\beta_{l_1.k}^{A1.X*} + \beta_{l_2.k}^{A2.X*}$.

The scale-extended model incorporates the scale parameter λ[s], and thus is *bi*-linear in the parameters. This yields a logit model with a linear term of the form $\lambda_s \beta_{l_1.k}^{A1.X*} + \lambda_s \beta_{l_2.k}^{A2.X*}$, where the β and λ[s] parameters are estimated simultaneously. Note that each part-worth is multiplied by the same scale parameter λ[s].

Each member belonging to latent class segment k=1,2,…,K shares the same pure preference utilities, but some of these members differ in uncertainty (i.e., differ on their scale factor). For purposes of identification, we set $\lambda_1 = 1$ for the largest subscale group. For 2 or more latent subgroups, each of the K LC segments contains some respondents with scale parameter = 1, and some with a higher value (reflecting lower amounts of uncertainty) and/or a lower value (reflecting larger amounts of uncertainty. When there is only a single subgroup, the model reduces to a standard LC choice model, where all respondents are assumed to have a common scale factor (λ = 1.)

How do the results compare between a standard LC choice analysis designed to identify different latent segments, and a scale-extended LC model where within each class, subgroups of respondents are allowed to have different scale parameters? Theoretically, if the true underlying population conforms to Table 4, the standard LC choice modeling approach might identify 4 LC segments that might be more or less confounded, while the scale-extended approach would identify 2 segments, each consisting of more or less certain subgroups.

Below, two real-world choice data sets are used to illustrate the various models; the first is a 5-attribute choice experiment for coffee makers (Skrondal and Rabe-Hesketh, 2004), the second is a 50% random sample provided by Sawtooth software of TV choice data originally described by Huber et. al. (1998). The latter data has an additional variable reflecting the length of time to complete the choice task, information that might be expected to be related to one's scale factor.

We also introduce a new class of LC choice models, called discrete factor (DFactor) choice models. In clustering applications, DFactor models were shown to outperform traditional LC models (Magidson and Vermunt, 2001), so they might be expected to outperform traditional LC choice models as well. The DFactor models are illustrated in the first example.

Example 1: Coffee Makers – Standard vs. Scale-Extended Models

Consider data from a discrete choice study with 5 attributes – BRAND, CAPACITY, PRICE, FILTER and THERMOS (Skrondal and Rabe-Hesketh, 2004). For these data, models containing between K = 1 – 6 LC segments were estimated, each with between 1 and 3 scale subgroups. For the standard models (i.e., those containing only a single scale factor), the 5-class solution fit best (lowest BIC). For scale-extended models with S* dichotomous (i.e., 2 scale subgroups), again the 5-class model fit best. Table 5 provides the fit statistics for these models, which were estimated using the LG-Syntax module in Latent GOLD Choice. Appendix A1 provides the model specifications. For more technical details, see Vermunt and Magidson (2008).

Table 5: Fit measures for the LC choice models estimated with Coffee Makers data

	Standard LC Choice Models			LC Choice Models + 2 Scale Factors		
	LL	BIC(LL)	Npar	LL	BIC(LL)	Npar
1-class	-1298.7	2639.2	8	-1180.1	2412.4	10
2-class	-1110.6	2310.0	17	-1089.8	2278.7	19
3-class	-1073.6	2283.0	26	-1051.2	2248.6	28
4-class	-1043.2	2269.1	35	-1021.9	2236.9	37
5-class	-1013.6	2257.0	44	-996.9	2233.9	46
6-class	-992.9	2262.5	53	-974.3	2235.7	55

Compared to the standard models with K segments, the corresponding K-class scale-extended model contains 2 additional distinct parameters to be estimated –- λ[2], the scale factor for the 2nd subgroup and P[2], the proportion of the population falling into this 2nd subgroup.[3] Note that based on the BIC criteria, the scale-extended models is preferred over the standard models regardless of the number of segments.

For the scale-extended models, the 2 latent variables were assumed to be independent. This hypothesis was tested for the 5-class model by a likelihood ratio comparison with a model assuming dependence between the two latent variables and found to be supported. Also, models with different numbers of scale subgroups were estimated for the 5-class model and those containing 2 subgroups were found to fit best.

For the 5-class scale-extended model, each segment k consists of both lower (s=1) and higher (s=2) certainty subgroups. The larger S-class consisted of 52% of the respondents who expressed higher amounts of certainty (λ[2] = 11.2) than the other subgroup (λ[1] = 1). The scale-extended model segments differed considerably from those from the standard model. As shown in Table 6, only 62+16+20+9+9 = 118 of the 185 respondents remain grouped in the same class. For example, only 62 of the 83 respondents grouped into class X*[1] were grouped into the corresponding X[1]-class. Overall, 37% of the respondents were classified into an X-class that differed from the corresponding X* classification.

[3] Given P[2], P[1], the proportion in the first subgroup can be computed as 1 – P[2], and thus is not counted as a distinct parameter to be estimated.

Table 6: Cross-tabulation of classification with and without scale factor

	X*[1]	X*[2]	X*[3]	X*[4]	X*[5]	Total
X[1]	62	26	0	7	0	95
X[3]	4	16	0	5	0	25
X[4]	0	1	20	0	0	21
X[2]	17	2	6	9	1	35
X[5]	0	0	0	0	9	9
Total=	83	45	26	21	10	185

DFactor Models as an Alternative to LC Choice Models

Magidson and Vermunt (2001) suggested the use of Discrete Factor (DFactor) models as an alternative to traditional LC Cluster models and introduced a basic model containing dichotomous uncorrelated DFactors that generally provided a more parsimonious explanation of data. Thus, it might be expected that basic DFactor choice models might also outperform traditional LC choice models.

Discrete Factor (DFactor) Choice Models posit *M latent DFactors* denoted as $X_1, X_2, \ldots, X_M$ with $K_1, K_2, \ldots, K_M$ categories which yields a total of $K_1 \times K_2 \times \ldots \times K_M$ segments. A *basic* DFactor model contains *dichotomous* DFactors which are mutually *independent* of each other. These models impose a factor structure on the parameters of a LC choice model. For example, the part-worth utility parameters corresponding to segment $(X_1, X_2) = (k_1, k_2)$ of a *basic* 2-DFactor model[4] are expressed as:

$$\beta_{l_1.k_1,k_2}^{A_1.X_1,X_2} + \beta_{l_2.k_1,k_2}^{A_2.X_1,X_2} + \ldots$$

where:

$$\beta_{l_1.k1,k2}^{A_1.X_1,X_2} = a_{l_1} + b_{l_1 1}^{A_1} X_1 + b_{l_1 2}^{A_1} X_2$$

and

$$\beta_{l_2.k_1,k_2}^{A_2.X_1,X_2} = a_{l_2} + b_{l_2 1}^{A_2} X_1 + b_{l_2 2}^{A_2} X_2$$

[4] For dichotomous DFactors, dummy coding is used (X_1 = 0 or 1 and X_2 = 0 or 1) so that the constant a_{l_2} reflects the part-worth associated with attribute A_1 for the first LC segment (0, 0).

Since this is a restricted (structured) version of a LC choice model with $K = K_1 \times K_2 \times \ldots \times K_M$ classes, it is a parsimonious model, containing relatively few parameters. In fact, a *basic* DFactor choice model with M DFactors has the same number of parameters as a LC choice model with only K=M+1 latent class segments. Thus, for example, a *basic* 2 DFactor model is a restricted 4-class model that contains the same number of parameters as a 3-class model.

LC Choice Model Extension #2 utilizes a Scale-Extended (DFactor) Choice Model which posits M DFactors $X_1^*, X_2^*, \ldots, X_M^*$ and S. By default, each X_m^* is dichotomous. As before, the categories of S represent subgroups with different scale factors, and for identification purposes, the restriction $\lambda[1] = 1$ is made for the 1st category of S, which is the category having the *lower* scale factor.

Example 1: Model Results for DFactor Models

For the coffee-maker data, models containing between 1 and 4 DFactors were estimated, with and without the scale parameter. As in the standard choice models, the scale-extended form of the model contains 2 additional parameters. (See Appendices A1 and A2 for the syntax equations corresponding to the standard LC Choice and DFactor models.) The fit statistics for these models are given in Table 7.

Table 7: Fit measures for the DFactor choice models estimated with Coffee Makers data

	D-Factor Choice Models				Scale Adjusted D-Factor Choice Models			
Model	LL	BIC(LL)	Npar			LL	BIC(LL)	Npar
1-dfac	-1110.6	2310.0	17		1-dfac w scale	-1089.8	2278.7	19
2-dfac	-1056.6	2249.0	26		2-dfac w scale	-1039.4	2224.9	28
3-dfac	-1021.7	2226.1	35		3-dfac w scale	-1001.6	2196.4	37
4-dfac	-988.1	2205.9	44		4-dfac w scale	-971.3	2182.7	46

First, comparing Table 7 with Table 5 we see that each *standard* DFactor model fits these data better (lower BIC) than the corresponding *standard* LC choice model containing the same number of parameters (Npar). Also, the *scale-extended* DFactor models outperform (lower BIC) the *standard* DFactor models. Among the models estimated, the scale-adjusted 4 DFactor model fit best.

As in the Extension #1, 2 scale subgroups were found to be best for 4-DFactor models. The larger S-class consisted of 61% of the respondents who expressed higher amounts of certainty

($\lambda[2] = 9.0$) than the other subgroup ($\lambda[1] = 1$). Again, S was found to be independent of X*. That is, about 61% of the members of each X* segment were classified into this S-subgroup.

Compared to the 5-Class Cluster solution, the 4-DFactor Scale-Extended Model classifies more respondents into the higher certainty group, 77% of whom are classified into the corresponding group by the Cluster approach. Similarly, 91% of those classified into the less certain group by the DFactor approach are classified into the corresponding group by the Cluster approach.

Overall, the S-subgroups identified by the different types of scale-extended choice models show strong agreement -- both models classify the same 99 respondents into the *more* certain group and the same 52 respondents into the *less* certain group. Overall, a somewhat higher number of respondents were classified into the more certain subgroup, consistent with the fact the scale factor for this at this subgroup was somewhat lower than obtained under the traditional approach (9.0 vs. 11.2).

Table 8: Fit measures for the DFactor choice models estimated with Coffee Makers data

Comparison of S-subgroups obtained from Cluster and Dfactor Models				
		4-DFactor Model Classifications		
5-Class Cluster Model		λ = 1	λ = 9.0	Total
1	λ = 1	52	29	81
2	λ = 11.2	5	99	104
Total		57	128	185

Example 2: Sawtooth Software TV Choice Data

It is possible to include covariates in a LC model to better predict/explain/describe the latent variable. In example #2 we include the time to complete the choice tasks as a numeric covariate, specify linear and quadratic time effects, and examine whether completion time is related to latent variables S and/or X*. Overall, the time to complete the 18 choice tasks ranged from 1 minute to 22 minutes with a mean time of 6.4 minutes.

This choice experiment consists of N = 176 respondents, who selected from various choice sets TVs that differed on their levels across 6 attributes – Brand, Screen Size, Sound, Channel Blocking, Picture-in-Picture availability and Price. The data were obtained as a random 50% sample of all respondents analyzed originally by Huber et. al. (1998).

For these data we estimated a 4-class LC choice model where the 4th class has zero effects. That is, for segment #4, the choices are assumed to *not* be affected by the attributes (i.e., segment #4 is a random response segment). We compare models with and without a scale parameter. For the former, 2 scale subgroups were assumed. Both models included time as an active covariate. The model specifications are given in Appendix B.

For the standard LC choice model (no scale parameter), time was found to be a significant predictor of the classes (X), the 3rd and 4th classes having significantly lower mean completion time than classes 1 and 2. The scale-extended LC Choice model was preferred based on the BIC.

Results from the scale-extended model were as follows:

- The second subgroup of S consists of 69% of the cases. This subgroup is more certain (λ[2] = 3.7) than subgroup #1 (λ[1] = 1).
- Time was found to be a significant predictor of S, but not X*.

Table 9: Cross-tabulation of classification based on original and scale adjusted 4-class models

	Scale parameter = 1.0					
original classes	scale-adjusted classes				Total	Mean time to complete
	1*	2*	3*	4*		
1	7		1		8	4.6
2	0	1	1		2	8.0
3		2	21		23	4.4
4	13	9		0	22	4.2

	Scale parameter = 3.7					
original classes	scale-adjusted classes				Total	Mean time to complete
	1*	2*	3*	4*		
1	61		0		62	7.5
2	2	48	0		49	7.5
3		1	5		6	4.2
4	0	0		4	4	8.0

Table 9 shows the correspondence between respondent classifications based on the two models. For clarity, the original classes are denoted as 1, 2, 3 and 4 while the classes from the scale-extended model are denoted using an * (1*, 2*, 3* and 4*). The comparison can be summarized as follows:

- S and X* were correlated – Classes 1 and 2 tend to be in the more certain subgroup while class 3 tends to be in the less certain subgroup.
- The random responder class (class 4) is much smaller when the scale factor is estimated, as many of those classified as random responders by the standard model become reclassified into the uncertain subgroup of classes 1 or 2.
- Relationship between individual classifications:
    - Class 1 mostly corresponds to class 1*, mean time = 7.5, more certain (λ = 3.7)
    - Class 2 mostly corresponds to class 2*, mean time = 7.5, more certain (λ = 3.7)
    - Class 3 mostly corresponds to class 3*, mean time = 4.4, less certain (λ = 1)
    - Class 4 mostly redistributed to 1* and 2*, mean time = 4.2, less certain (λ = 1)

Note that since classes 1 and 2 tend to have higher λ than class 3, its part-worths would be expected to be *over*stated relative to class 3. Table 10 below shows that all of the part-worths are in fact higher than the corresponding scale-adjusted values.

Especially noteworthy is the comparison of the part-worths that address price sensitivity. Under the standard LC choice model, class 3 appears to be *less* price sensitive: 0.38 vs. -0.35 differs by *less* than either 0.87 vs. -0.85 (class 1) or 0.61 vs. -0.93 (class 2). However, given the same scale value (λ=1 shown) the scale-adjusted differences suggest *greater* price sensitivity for class 3.

Table 10. Comparison of Original vs. Scale-adjusted part-worths

	Traditional 4-class model					4-class model with scale adjustment			
	0.39	0.30	0.16	0.15		0.45	0.35	0.16	0.03
	Class1	Class2	Class3	Class4		Class1*	Class2*	Class3*	Class4*
brand									
JVC	-0.17	-0.39	-1.18	0		-0.05	-0.10	-1.16	0
RCA	-0.04	0.18	-0.38	0		-0.02	0.05	-0.24	0
Sony	0.21	0.21	1.56	0		0.06	0.05	1.40	0
size									
25" screen	-0.36	-0.28	-0.47	0		-0.10	-0.09	-0.29	0
26" screen	0.02	-0.09	0.10	0		0.00	-0.03	0.08	0
27" screen	0.34	0.37	0.37	0		0.10	0.11	0.21	0
sound									
Mono sound	-1.66	-0.83	-0.40	0		-0.50	-0.21	-0.43	0
Stereo sound	0.40	0.44	0.24	0		0.13	0.11	0.19	0
Surround sound	1.26	0.39	0.16	0		0.37	0.10	0.23	0
block									
No blockout	-0.15	-0.83	-0.25	0		-0.05	-0.24	-0.15	0
Channel blockout	0.15	0.83	0.25	0		0.05	0.24	0.15	0
pip									
No pip	-0.33	-1.05	-0.27	0		-0.11	-0.30	-0.18	0
Picture in picture	0.33	1.05	0.27	0		0.11	0.30	0.18	0
price									
$300	0.87	0.61	0.38	0		0.26	0.17	0.39	0
$350	0.40	0.50	0.13	0		0.12	0.14	0.13	0
$400	-0.42	-0.17	-0.17	0		-0.11	-0.05	-0.21	0
$450	-0.85	-0.93	-0.35	0		-0.27	-0.26	-0.32	0
mean time=	7.1	7.5	4.4	4.8		6.7	6.7	4.7	8
N =	69	52	29	26		83	61	28	4

Summary and Conclusions

Louviere and Eagle (2006) argued that in choice modeling it is not valid to compare part-worths across individuals, groups or latent classes, without removing the potential confound with the scale parameters. In this presentation we introduced new LC models that can estimate and thus remove the scale parameters, and applied these models in 2 CBC examples with real data.

In both examples the equal scale parameter assumption used to justify standard LC choice modeling was found to be violated, different scale parameters reflecting different amounts of uncertainty being present in different subgroups. Thus, in both cases the confound existed.

In example 1, 2 subgroups that differed in their scale parameters were found. While these subgroups were approximately equally distributed among the classes, ignoring these subgroups resulted in misclassifying 37% of the cases into a different class. Similar misclassification rates were found in DFactor Choice models, which provided improved fits over the more traditional models.

In example 2, the 2 scale subgroups found were correlated with the LC segments. While segments #1 and #2 appeared to be *more* price sensitive than segment #3, they tended to have higher scale factors than segment #3 which masked the actual relationship. Removal of this confound resulted in the conclusion that these segtments were in fact *less* (not *more*) price-sensitive than segment #3.

Extensions of standard LC and discrete factor (DFactor) choice models are now available that can capture both scale and preference parameters, thus removing the scale confound from preference part-worth parameter estimates. We believe that application of these and related HB-like models based on CFactors (see e.g., Rabe Hesketh and Skrondal, 2004; Vermunt and Magidson, 2008) will result in improved estimates of preference part-worths and a better understanding of the effects associated with different respondent levels of uncertainty. This, in turn, can lead to improved targeting to relevant segments based on an improved understanding of segment preferences and levels of uncertainty.

References

Ben-Akiva and Lerman (1985) . Discrete Choice Analysis: Theory and application to travel demand, Cambridge, MA. MIT Press.

Huber, J., Arora, N and R. Johnson (1998) "Capturing Heterogeneity in Consumer Choices," *ART Forum*, American Marketing Association.Louviere, Jordan J., Thomas C. Eagle, 2006. "Confound it! That Pesky Little Scale Constant Messes Up," 2006 Sawtooth Software Conference Proceedings.

Louviere, Hensher and Swait (2000) Stated Choice Models, Cambridge University Press.

Magidson and Vermunt (2001) "Latent class factor and cluster models, bi-plots and related graphical displays" *Sociological Methodology*, 31, 223-264

Skrondal and Rabe-Hesketh (2004) Generalized Latent Variable Modeling. London: Chapman and Hall.

Vermunt and Magidson (2008) User's manual and Technical Guide for LG-Syntax™ – the syntax module for Latent GOLD and LG Choice 4.5.

Appendices

Below we present the Equation section and the Latent Variables portion of the Variables section of the LG-syntax that define the various models described in this article. The remaining portions of the syntax contain technical and output options, as well as the definition of the variables which specify the attributes, the dependent variable, and various id variables.

Appendix A1: LG-Syntax specifications for LC Choice Models estimated on the Coffee Maker Choice Data with and without a Scale Adjustment

A standard LC Choice Model is defined as follows:

```
variables
 latent Class nominal 5;
equations
   Class <- 1;
   choice <- brand |Class + capacity | Class + price | Class
          + filter | Class + thermos | Class;
```

Here, 'Class' is the name assigned to the categorical latent variable representing the 5 segments. One logit equation is for the class sizes, the other for the observed choices. Note that "| Class" means that the parameter concerned depends on Class.

A Scale-Adjusted version of the LC Choice Model may be defined as follows:

```
latent
   Class nominal 5, sClass nominal 2, scale continuous;
equations
   Class <- 1 ; sClass <- 1;
   scale <- (s) 1 | sClass;
   (0) scale;
   choice <- brand scale |Class + capacity scale | Class
             + price scale | Class + filter scale | Class
             + thermos scale | Class ;
   s[1] = 1; s[2] = +;
```

The scale factor named 'scale' is defined as a continuous latent variable with a mean (intercept) depending on sClasses and a residual variance equal to zero. The restrictions set the scale factor for the first sClass to 1, and insure that the scale factor for the second sClass is nonnegative.

Appendix A2: LG-Syntax specifications for DFactor Choice Models estimated on the Coffee Maker Choice Data with and without a Scale Adjustment

A DFactor Choice Model is defined as follows:

```
latent
   dfac1 ordinal 2, dfac2 ordinal 2, dfac3 ordinal 2, dfac4 ordinal 2;
equations
   dfac1 <- 1; dfac2 <- 1; dfac3 <- 1; dfac4 <- 1; class <- 1;
   choice <- brand + capacity + price + filter + thermos
       + brand dfac1 + brand dfac2 + brand dfac3 + brand dfac4
       + capacity dfac1 + capacity dfac2 + capacity dfac3 + capacity dfac4
       + price dfac1 + price dfac2 + price dfac3 + price dfac4
       + filter dfac1 + filter dfac2 + filter dfac3 + filter dfac4
       + thermos dfac1 + thermos dfac2  + thermos dfac3 + thermos dfac4;
```

and a Scale-Adjusted DFactor Choice Model as

```
latent
   sClass nominal 2, dfac1 ordinal 2, dfac2 ordinal 2, dfac3 ordinal 2,
   dfac4 ordinal 2,  scale continuous;
equations
   scale <- (s) 1 | sClass;
   (0) scale;
   dfac1 <- 1; dfac2 <- 1; dfac3 <- 1; dfac4 <- 1; sClass <- 1;
   choice <- brand scale + capacity scale + price scale + filter scale
            + thermos scale + brand dfac1 scale + brand dfac2 scale
            + brand dfac3 scale + brand dfac4 scale + capacity dfac1 scale
            + capacity dfac2 scale + capacity dfac3 scale
            + capacity dfac4 scale + price dfac1 scale + price dfac2 scale
            + price dfac3 scale + price dfac4 scale + filter dfac1 scale
            + filter dfac2 scale + filter dfac3 scale + filter dfac4 scale
            + thermos dfac1 scale + thermos dfac2 scale
            + thermos dfac3 scale + thermos dfac4 scale;
   s[1] = 1; s[2] = +;
```

Appendix B: LG-Syntax specifications for 4-class models estimated on the TV Choice Data

A standard 4-class model with the 4th class being restricted to be a random class and the covariate 'TimeR' affecting the classes and monotonic non-increasing price effect is obtained by:

```
latent
   Class nominal 4;
equations
   Class <- 1 + (b) TimeR;
   Choice <- (a1) brand | Class + (a2) size | Class
            + (a3) sound | Class + (a4) block | Class
            + (a5) pip | Class + (b2) price | Class;
   a1[4]=0; a2[4]=0; a3[4]=0; a4[4]=0; a5[4]=0; b2[4]=0;
   b2 = -;
```

A 4-class Scale-Adjusted model with the 4th class being restricted to be a random class and the covariate 'TimeR' affecting the scale factor and monotonic non-increasing price effect:

```
latent
   scale  continuous, sClass nominal 2, Class nominal 4 ;
equations
   sClass <- 1 + (b) TimeR;
   Class <- 1;
   Class <-> sClass;
   scale <- (s) 1 | sClass;
   (0) scale;
   Choice <- (a1) brand scale | Class + (a2) size scale | Class
            + (a3) sound scale | Class + (a4) block scale | Class
            + (a5) pip scale | Class + (b2) price scale | Class;
   a1[4]=0; a2[4]=0; a3[4]=0; a4[4]=0; a5[4]=0; b2[4]=0;
   b2 = -; s[1] = 1; s[2] = +;
```