# Web Appendix

# Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model

Joost van Rosmalen, Hester van Herk, and Patrick J.F. Groenen

## *PARAMETER RESTRICTIONS*

Several parameter restrictions are necessary for identification of the model parameters. Here, we describe the restrictions that we impose. First, we consider location restrictions, that is, restrictions that are necessary because adding a constant to a set of parameters would not change the estimated probabilities in the model. Location restrictions are needed for $a_{j|r}$, **F**, $\mathbf{G}_k$, and $\mathbf{H}_s$. We impose sum-to-zero constraints for all these parameters per set, that is, we require that $\sum_{j=1}^{J} a_{j|r} = 0$, $\sum_{j=1}^{J} f_{jp} = 0$, $\sum_{t=1}^{T} h_{tp|s} = 0$, and that the elements of $\mathbf{g}_{kp}$ sum to zero. As a result, the centroid of the points referring to a single variable (Rating, Item, or a background variable) is the origin in the plot.

Scale and rotation constraints are required for **F**, $\mathbf{G}_k$, and $\mathbf{H}_s$. For notational convenience, let $\mathbf{B} = [\mathbf{B}_1 \,|\, ... \,|\, \mathbf{B}_K]$, $\mathbf{C} = [\mathbf{C}_1 \,|\, ... \,|\, \mathbf{C}_S]$, $\mathbf{G} = [\mathbf{G}'_1 \,|\, ... \,|\, \mathbf{G}'_K]'$, and $\mathbf{H} = [\mathbf{H}'_1 \,|\, ... \,|\, \mathbf{H}'_S]'$, where $\mathbf{C}_s$ contains the effects of the ratings on the items in item segment s. Then, the rank restrictions in Equation 2 can be written as

(W1) $$[\mathbf{B}|\mathbf{C}] = \mathbf{F}[\mathbf{G}'|\mathbf{H}'].$$

Without loss of generality, we can transform the parameter matrices as **F** = **FT**, $\mathbf{G} = (\mathbf{T}^{-1}\mathbf{G}')'$, and $\mathbf{H}=(\mathbf{T}^{-1}\mathbf{H}')'$ for any nonsingular $P \times P$ matrix **T**, because

$$\mathbf{F}[\mathbf{G}'|\,\mathbf{H}'] = \mathbf{F}\mathbf{T}\mathbf{T}^{-1}[\mathbf{G}'|\,\mathbf{H}'].$$

This freedom of scaling also occurs in principal components analysis and correspondence analysis (Gifi 1990). We compute a constrained solution as follows. Let [**B**|**C**] be obtained by some unconstrained **F**, **G**, and **H**. In addition, let $\mathbf{P\Phi Q}'$ = [**B**|**C**] be a compact singular value decomposition with **P** and **Q** orthogonal rotation matrices with P columns (with $\mathbf{P}'\mathbf{P} = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$) and $\mathbf{\Phi}$ a $P \times P$ diagonal matrix with positive monotonically decreasing values. As the rank of [**B**|**C**] is not greater than P, matrices **P**, $\mathbf{\Phi}$, and **Q** that meet the requirements above must exist and typically are unique up to a reflection per dimension. We set $\mathbf{F} = \omega\mathbf{P\Phi}^{1/2}$ and $[\mathbf{G}'|\mathbf{H}']' = 1/\omega\mathbf{Q\Phi}^{1/2}$, where ω is a constant that determines the relative scaling of the points of the ratings compared to the points of the items and the background characteristics. Choosing **F** and [**G**'|**H**']' in this way ensures that the equality in Equation W1 is preserved, as

$$\mathbf{F}[\mathbf{G}'|\,\mathbf{H}'] = \omega\mathbf{P\Phi}^{1/2}\left(\frac{1}{\omega}\mathbf{Q\Phi}^{1/2}\right)' = \mathbf{P\Phi Q}' = [\mathbf{B}\,|\,\mathbf{C}].$$

The value of ω can be adapted without altering the general results. We set

$$\omega = \left( \frac{J}{ST + \sum_{k=1}^{K} m_k} \right)^{1/4},$$

so that the average squared Euclidean distance of the points to the origin is the same for both **F** and [**G**'|**H**']'. For the LOV data set, this choice amounts to setting ω≈.65. We also simultaneously reflect the columns of **F**, **G**, and **H** in such a way that the first row of **F** only has positive values.

Due to these parameter restrictions, the spread of the points in the graphical representations decreases with the dimension, and the dimensions are orthogonal with respect to **F** and [**G**'|**H**']'. The point of Rating 1 must have positive values on all dimensions. A final parameter restriction is that the segment sizes must be nonnegative and should sum to one, so that $u_{rs} \geq 0$ and $\sum_{r=1}^{R} \sum_{s=1}^{S} u_{rs} = 1$.

*OPTIMIZATION ALGORITHM*

An Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) is used to estimate the model parameters by maximizing the likelihood function. The EM algorithm starts with initial parameter estimates and then iteratively performs an E-step and a M-step, until convergence has been achieved. In the E-step, the posterior segment membership probabilities $\Pr(i \in \Xi_{r,s})$ are computed using Equation 6, given the current parameter estimates. In the M-step, the expected complete log-likelihood is maximized with respect to the parameter estimates, given the segment membership probabilities computed in the E-step. In our implementation, every M-step consists of 10 iterations of the BFGS quasi-Newton optimization routine in the MATLAB Optimization Toolbox (version 3.0.4), with analytically computed gradients. Convergence is considered to have been achieved if the change in log-likelihood between two consecutive EM iterations is smaller than $2 \times 10^{-5}$.

It is possible that the EM algorithm converges to parameter estimates that are only locally optimal. To solve this problem, the EM algorithm was run 10 times for every value of R, S, and P with randomly chosen starting values, and the solution with the best likelihood value was retained.

*MODEL IDENTIFIABILITY*

An important issue is whether the model parameters are identifiable (see, for example, Teicher 1963). If there are few items in the study and many segments, the model may not be identifiable, so that unique parameter estimates that maximize the likelihood function may not exist. For example, with only one item in the study, it is not possible to distinguish more than one response style or item segment.

To ensure model identifiability, the parameters must be locally identifiable. In addition, we must account for trivial nonidentifiability, which occurs because the order of the segments can be interchanged without changing the likelihood value; this problem can be solved by imposing appropriate restrictions of the segment sizes. Local identifiability is guaranteed if the Hessian (that is, the matrix of second derivatives of the log-likelihood function at the final parameter estimates) has full rank, which can easily be checked after the estimating the model parameters. In principle, it may be possible to derive conditions in which the model is always identifiable, which can be checked before estimating the parameters. However, because we have two types of segments, deriving necessary and sufficient conditions for identifiability is very hard and beyond the scope of this paper.

Instead, we solve this problem empirically. After estimating the parameters of a model, we check whether the Hessian is of full rank (using the alternative parametrization described in the section ‘Confidence Ellipses’), and we impose a specific ordering of the segments. Based on numerical experimentation, we find that the model is almost always identifiable if there are more than three items, and the LC-BML model is also identified for the solution presented in this paper for the LOV data set. Most marketing segmentation studies use at least ten items and distinguish at most five segments; such studies will thus almost always result in identifiable models. Therefore, we do not believe that the identifiability of the LC-BML model poses a problem for marketing practitioners.

## *SIMULATION STUDY*

To apply the LC-BML model, the numbers of segments *R* and *S* and the dimensionality *P* need to be chosen. To do so, we use an information criterion. However, many information criteria exist, and their performances tend to depend on the characteristics of the model and the data set (Yang and Yang 2007). To determine how various information criteria perform for our model and to evaluate the performance of the optimization algorithm, we conduct a simulation study. In this simulation study, we generate data according to simulated model parameters and then apply the optimization algorithm.

To generate simulated data sets, we simulate model parameters as follows. The elements of $a_{.|1},\ldots,a_{.|R}$, $\mathbf{F}$, $\mathbf{G}_1,\ldots,\mathbf{G}_k$, and $\mathbf{H}_1,\ldots,\mathbf{H}_S$ are chosen as independently normally distributed variables, in such a way that the parameter restrictions that are described in the section ‘Parameter Restrictions’ are met (so that the columns of these parameter matrices must sum to zero). The variance of these parameters is one of the factors in the simulation study and equals either .6 or 1.2. The elements of U are chosen to be $1/(R \times S)$, so that all segments have the same size.

In our simulations, we vary seven factors with two levels each, using a full-factorial design. These factors and the levels of these factors are shown in Table 1. The levels of these factors have been chosen in such a way, that the simulated data are roughly comparable to empirical data sets such as the LOV data set; to limit the computation time, the numbers of segments have been set lower than in the empirical study. We generate one data set for each of the $2^7 = 128$ combinations of the levels of these factors. For each data set, we apply the optimization method for $R = 1,\ldots,R_{sim}, R_{sim}+1$, $S = 1,\ldots,S_{sim}, S_{sim}+1$, and $P = 1,\ldots,P_{sim}, P_{sim}+1$, in which $R_{sim}$, $S_{sim}$, and $P_{sim}$ are the values of R, S, and P that have been used to simulate the data. The optimization method uses five random starts for each combination of R, S, and P in each data set, and the EM-algorithm is considered to have converged if the difference in log-likelihood between two iterations is smaller than $10^{-4}$. We then determine the preferred model in each data set for each of four information criteria. These criteria are: AIC, AIC-3, BIC, and CAIC (see Andrews and Currim 2003, for an overview).

Table 1
FACTORS IN SIMULATION STUDY

| Factor | Level 1 | Level 2 |
|---|---|---|
| Number of respondents (n) | n = 1000 | n = 4000 |
| Number of rating categories (J) | J = 5 | J = 9 |
| Background variables | no background variables | 2 variables with 2 categories each |
| Numbers of segments ($R_{sim}$, $S_{sim}$) | $R_{sim} = 2$, $S_{sim} = 2$ | $R_{sim} = 3$, $S_{sim} = 3$ |
| Dimensionality ($P_{sim}$) | $P_{sim} = 1$ | $P_{sim} = 2$ |
| Number of items (T) | T = 8 | T = 16 |
| Variance of parameters | 1.2 | .6 |

The first four rows of Table 2 show, for each information criterion, the proportion of the data sets in which the information criterion identified the correct model; these proportions are computed for each level of each factor, averaged over the six remaining factors. In each situation, either the AIC-3 or the BIC performs best. CAIC tends do slightly worse than BIC, and AIC seems to perform rather poorly (it often overestimates the numbers of segments). The differences in the performances between AIC-3, BIC, and CAIC appear to be rather small compared to the clearly inferior performance of AIC. For large data sets, BIC and CAIC are known to give good performances Andrews and Currim (2003), which is supported by the results in Table 2. As AIC-3 was only proposed as a criterion for determining numbers of segments Andrews and Currim (2003), and not for choosing a dimensionality, we decide to use BIC for the List of Values data set.

The results of the simulation study show that the optimization algorithm, in combination with a properly chosen information criterion, is capable of identifying the correct values for R, S, and P. We also wish to evaluate the parameter recovery of the optimization algorithm given that R, S, and P have been identified correctly. For each simulated data set, we compare the simulated parameters to the estimated parameters for the optimizations in which R, S, and P equal their correct values. The second part of Table 2 shows the mean absolute error for each parameter set, for each level of each factor. The mean absolute error is always smaller than .08 for each parameter set. As these estimated parameters (except for the elements of **P**) have a variance of .6 or 1.2, we believe that the parameter recovery is acceptable.

The last two rows of Table 2 show the average proportions with which the respondents are assigned to their simulated clusters, for both the response styles segments and the item segments. To assign each respondent to one segment, each respondent is placed in the segment in which his posterior membership probability is greatest. These proportions tend to be around 90% and are relatively mildly affected by the characteristics of the data set.

Finally, we have tested the effects of each factor in the simulation study on the recovery performance of the LC-BML model, and the results of these tests are also shown in Table 2. For the information criteria, these tests have been conducted using a binomial logit model, and for the other results in Table 2, an ANOVA model has been used; both types of models contained a constant term, main effects of all factors, and no interaction effects. The results show that the parameter recovery is better if the numbers of respondents and items are higher. Increasing the

Table 2
RESULTS OF SIMULATION STUDY

| | Factor 1 respondents | | Factor 2 rating cats. | | Factor 3 backgr. vars. | | Factor 4 segments | | Factor 5 dimensions | | Factor 6 items | | Factor 7 param. var. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1000 | 4000 | 5 | 9 | 0 | 2 | 2 | 3 | 1 | 2 | 8 | 16 | 1.2 | 0.6 |
| Information criterion | Proportion of data sets in which are R, S, and P are chosen correctly | | | | | | | | | | | | | |
| AIC | .23[a] | .09 | .23[a] | .09 | .19 | .14 | .17 | .16 | .17 | .16 | .13 | .20 | .20 | .13 |
| AIC-3 | .94 | .91 | .92 | .92 | .92 | .92 | .94 | .91 | .94 | .91 | .88 | .97 | .94 | .91 |
| BIC | .89 | .92 | .88 | .94 | .92 | .89 | .98[b] | .83 | .94 | .88 | .86 | .95 | .95 | .86 |
| CAIC | .88 | .92 | .88 | .92 | .91 | .89 | .98[b] | .81 | .92 | .88 | .84[a] | .95 | .95[a] | .84 |
| Parameter set | Average absolute error | | | | | | | | | | | | | |
| $a_{.\|1},\ldots,a_{.\|R}$ | .078[b] | .038 | .056 | .060 | .057 | .059 | .044[b] | .073 | .048[b] | .068 | .078[b] | .039 | .065[a] | .051 |
| $\mathbf{F}$ | .054[b] | .022 | .033 | .043 | .046 | .031 | .030 | .046 | .025[b] | .052 | .049[a] | .028 | .034 | .043 |
| $\mathbf{G}_1,\ldots,\mathbf{G}_k$ | .033 | .013 | .014 | .032 | -[c] | .023 | .013 | .034 | .011 | .035 | .033 | .014 | .020 | .026 |
| $\mathbf{H}_1,\ldots,\mathbf{H}_S$ | .051[b] | .024 | .035 | .041 | .026[a] | .050 | .031 | .045 | .027[a] | .049 | .045 | .030 | .033 | .043 |
| $\mathbf{P}$ | .016[b] | .007 | .011 | .012 | .011 | .012 | .012 | .012 | .011 | .012 | .015[b] | .009 | .010[a] | .014 |
| Type of segment | Average proportion of respondents placed in correct segment | | | | | | | | | | | | | |
| Response style segment | .891 | .907 | .885 | .912 | .913 | .885 | .940[b] | .858 | .922[b] | .876 | .857[b] | .941 | .918[b] | .879 |
| Item segment | .913 | .912 | .912 | .913 | .946[b] | .879 | .944[b] | .881 | .874[b] | .951 | .873[b] | .952 | .947[b] | .878 |

a: The factor has a significant effect at the 5% level.

b: The factor has a significant effect at the 1% level.

c: The parameter set $\mathbf{G}_1,\ldots,\mathbf{G}_k$ is not used for data sets without background variables.

Notes: For the information criteria, the statistical significance of the effects of factors in the simulation study is determined using a binary logit model. For the measures of parameter recovery and segment membership recovery, the statistical significance of factors has been tested using analysis of variance with only main effects.

numbers of segments and dimensions may negatively affect the parameter recovery and the probability that the correct model is identified, though the effects of these factors are often not statistically significant. For the other factors in the simulation study, no consistent results are found.

### *BIC VALUES*

The BIC values for P = 1,…,3 dimensions, R = 1,…,13 response style segments, and S = 1,…,6 item segments are shown in Table 3. The lowest BIC value (130,685) is attained with P = 2 dimensions, R = 11 response styles segments, and S = 5 item segments.

### *CONFIDENCE ELLIPSES*

The confidence ellipses in the Results section are based on maximum likelihood theory. Under regularity conditions, a maximum likelihood parameter estimator is normally distributed with covariance matrix equal to the inverse of the information matrix. The information matrix can be estimated as the negative of the matrix of second-order partial derivatives of the likelihood function with respect to all model parameters, evaluated at the maximum likelihood parameter estimates. However, as the model parameters are not identified without the parameter constraints described in the section ‘Parameter Constraints’, this matrix of second-order partial derivatives is not invertible. To circumvent this problem, we created an alternative model parametrization that does not require any parameter constraints for parameter identification and is equivalent to our original model. Several estimated parameters in the original parametrization represent probabilities of almost exactly 0. These parameters were fixed at these values for the alternative parametrization, so that the uncertainty in the associated parameter estimates is not taken into account in the confidence ellipses.

The confidence ellipses are constructed using the following procedure. First, the final parameter estimates of our original parametrization are transformed to the alternative one. Using these transformed parameter estimates, the matrix of second-order partial derivatives is calculated numerically. Then, 10,000 simulated parameter vectors are drawn from a multivariate normal distribution with covariance matrix equal to the negative inverse of the matrix of second-order derivatives. For each simulated parameter vector, the associated parameter vector in the original parametrization is computed. Analysis of the simulated parameters in the original parametrization shows that the locations of the points in the graphical representation have approximately a joint bivariate normal distribution. Finally, the simulated locations of the points are used to construct 95% normal theory confidence ellipses.

Table 3
BIC VALUES FOR R=1,…,13 RESPONSE STYLE SEGMENTS, S=1,…,6 ITEM SEGMENTS, AND DIMENSIONALITIES P=1,…,3 IN THE LOV DATA SET

| Dim. | Response style segments | Item segments 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| | 1 | 144,946 | 144,233 | 142,440 | 142,081 | 141,934 | 141,530 |
| | 2 | 139,964 | 138,791 | 138,379 | 137,668 | 137,705 | 137,547 |
| | 3 | 136,436 | 135,256 | 134,852 | 134,580 | 134,480 | 134,446 |
| | 4 | 135,428 | 134,158 | 133,666 | 133,397 | 133,175 | 133,154 |
| | 5 | 134,508 | 133,222 | 132,791 | 132,467 | 132,317 | 132,278 |
| | 6 | 134,063 | 132,694 | 132,156 | 131,858 | 131,622 | 131,595 |
| P = 1 | 7 | 133,727 | 132,321 | 131,822 | 131,498 | 131,318 | 131,277 |
| | 8 | 133,471 | 132,072 | 131,570 | 131,255 | 131,105 | 131,081 |
| | 9 | 133,340 | 131,918 | 131,406 | 131,094 | 130,932 | 130,918 |
| | 10 | 133,270 | 131,846 | 131,333 | 130,979 | 130,839 | 130,850 |
| | 11 | 133,218 | 131,789 | 131,263 | 130,944 | 130,792 | 130,819 |
| | 12 | 133,189 | 131,753 | 131,246 | 130,909 | 130,789 | 130,820 |
| | 13 | 133,185 | 131,738 | 131,229 | 130,916 | 130,809 | 130,842 |
| | 1 | 144,324 | 142,145 | 140,528 | 140,023 | 139,895 | 139,476 |
| | 2 | 139,435 | 137,861 | 137,448 | 135,611 | 135,158 | 134,987 |
| | 3 | 136,022 | 134,962 | 134,606 | 133,910 | 133,895 | 133,733 |
| | 4 | 134,937 | 133,842 | 133,405 | 133,058 | 132,800 | 132,828 |
| | 5 | 134,055 | 132,957 | 132,539 | 132,144 | 131,940 | 131,960 |
| | 6 | 133,645 | 132,462 | 131,972 | 131,710 | 131,610 | 131,579 |
| P = 2 | 7 | 133,329 | 132,149 | 131,631 | 131,361 | 131,186 | 131,313 |
| | 8 | 133,093 | 131,943 | 131,420 | 131,149 | 131,057 | 131,085 |
| | 9 | 132,980 | 131,786 | 131,247 | 130,963 | 130,870 | 130,967 |
| | 10 | 132,901 | 131,694 | 131,170 | 130,839 | 130,728 | 130,804 |
| | 11 | 132,825 | 131,613 | 131,098 | 130,791 | **130,685** | 130,793 |
| | 12 | 132,799 | 131,597 | 131,087 | 130,767 | 130,699 | 130,799 |
| | 13 | 132,808 | 131,590 | 131,061 | 130,765 | 130,705 | 130,813 |
| | 1 | 144,202 | 142,027 | 140,501 | 140,079 | 139,857 | 139,320 |
| | 2 | 139,309 | 137,761 | 137,429 | 135,661 | 135,200 | 135,149 |
| | 3 | 135,888 | 134,938 | 134,474 | 133,990 | 133,788 | 133,663 |
| | 4 | 134,820 | 133,787 | 133,361 | 132,954 | 132,855 | 132,374 |
| | 5 | 134,051 | 133,002 | 132,629 | 132,366 | 132,240 | 132,133 |
| | 6 | 133,642 | 132,544 | 132,121 | 131,914 | 131,855 | 131,898 |
| P = 3 | 7 | 133,329 | 132,226 | 131,775 | 131,599 | 131,571 | 131,601 |
| | 8 | 133,091 | 132,018 | 131,561 | 131,371 | 131,340 | 131,472 |
| | 9 | 132,978 | 131,860 | 131,379 | 131,187 | 131,158 | 131,325 |
| | 10 | 132,909 | 131,768 | 131,294 | 131,060 | 131,017 | 131,173 |
| | 11 | 132,828 | 131,690 | 131,231 | 131,015 | 130,981 | 131,147 |
| | 12 | 132,800 | 131,663 | 131,206 | 130,990 | 130,998 | 131,151 |
| | 13 | 132,817 | 131,672 | 131,213 | 130,997 | 130,993 | 131,193 |

## *REFERENCES*


Andrews, Rick L. and Imran S. Currim (2003), "A Comparison of Segment Retention Criteria for Finite Mixture Logit Models", *Journal of Marketing Research*, **40** (2), 235–243.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), "Maximum Likelihood from Incomplete Data via the EM-algorithm", *Journal of the Royal Statistical Society, Series B*, **39** (1), 1–38.

Gifi, Albert (1990), *Nonlinear Multivariate Analysis*, Wiley, Chichester, England.

Teicher, Henry (1963), "Identifiability of Finite Mixtures", The Annals of Mathematical Statistics, **34** (4), 1265-1269.

Yang, Chih-Chien and Chih-Chiang Yang (2007), "Separating Latent Classes by Information Criteria", *Journal of Classification*, **24** (2), 183–203.