

The Number of Levels Effect:

A Proposed Solution

In this paper, we first offer an analytic approach that confirms the existence of the algorithmic component to the number of levels effect for rank-order data. We then describe a second study which demonstrates a practical solution to the number of levels effect, regardless of the source of the effect.

Author: Paul Richard "Dick" McCullough

1999 Sawtooth Software Conference Proceedings,
February 1999, San Diego, CA, USA

The Number of Levels Effect:

A Proposed Solution

Abstract

The existence of the number of levels effect (NOL) in conjoint models has been widely reported since 1981 (Currim et al.). Currim et al. demonstrated that the effect is, for rank-order data, at least partially mathematical or algorithmic. Green and Srinivasan (1990) have argued that another source of this bias may be behavioral.

In this paper, we first offer an analytic approach that confirms the existence of the algorithmic component to the number of levels effect for rank-order data. We then describe a second study which demonstrates a practical solution to the number of levels effect, regardless of the source of the effect.

Introduction

The existence of the number of levels effect in conjoint models has been widely reported since 1981 (Currim et al.). The effect occurs when one attribute has more or fewer levels than other attributes. For example, if price were included in a study and defined to have five levels, price would appear more important than if price were defined to have two levels. This effect is independent of attribute range, which also can dramatically affect attribute relative importance.

NOL was originally observed for rank-order preferences but has since been shown to occur with virtually all types of conjoint data (Wittink et al. 1989). Currim et al. demonstrated, for rank-order data, that the effect is at least partially mathematical or algorithmic. Green and Srinivasan (1990) have argued that a source of this bias may also be behavioral. That is, attributes with higher numbers of levels may be given more attention by respondents than attributes with fewer levels. If true, this might cause respondents to rate attributes with a greater number of levels higher than attributes with fewer levels. Steenkamp and Wittink (1994) have argued that the effect is, at least partially, due to non-metric quality responses, which computationally causes ratings data to behave similarly to rank-order data.

The number of levels effect has been widely reported. It is generally agreed that the effect is a serious problem that can and often does significantly distort attribute relative importance scores, utility estimates and market simulation results. And largely due to the fact that the only known

practical method for removing this effect has been to hold the number of levels constant across attributes, it has often been ignored in commercial studies.

In this paper, we confirm the existence of an algorithmic component to the number of levels effect for rank-order data and offer a solution to remove the levels effects bias from full-profile conjoint models, both rank-order and ratings. The solution is demonstrated using ratings data.

Two separate studies are reviewed in this paper.

Methodology - First Study

In the first study, we examine a rank-order conjoint data set. The data come from a trade-off study of a new high-technology product. The trade-off data are derived from a rank-order card sort data collection exercise. The study design consists of 21 cards and three attributes. Attribute Price has 3 levels and is a vector attribute. Attribute Brand has 5 levels and is a partworth attribute. Attribute V has 2 levels and can be considered either vector or partworth. The combinations have been selected so the experimental design is reasonably orthogonal and balanced.

The data set is prepared two ways: 1) the data set is kept in its original format and 2) the data set is altered, i.e., degraded, so that one attribute at a time is redefined to have fewer levels. This degradation is achieved by simply removing any cards from the rank-order which include the omitted attribute levels. The degraded form of the data set for attribute Price has 14 cards. The degraded form of the data set for attribute Brand has six or eight cards, depending on which levels were exterior for a given respondent. Attribute V only has two levels so no degradation of the data set is necessary.

Conjoint utilities are estimated for each version of the data set. We estimate the existence and magnitude of the levels effect for each version of the data set using a slight variation of the regression model approach used by Steenkamp and Wittink (1994). In that approach, the relative importance scores for each respondent for a fixed attribute are regressed against the number of levels for that attribute. In the Steenkamp and Wittink study, the sample was split so that half of the sample saw attributes with one set of levels and half saw attributes with another set of levels. In our study, the two data sets, i.e., the complete data set and the degraded data set for a given attribute, are merged to provide variance in the number of levels. Because respondents are exposed to exactly the same stimuli in all versions of the data set (original and degraded), no behavioral component of NOL is possible. Any NOL effect detected will necessarily be due entirely to an algorithmic component.

Analysis - First Study

The regression models results (Table 1) show a levels effect for two different attributes. For the well-ordered, i.e., vector, attribute Price, the magnitude of the effect is approximately the same as cited by Steenkamp and Wittink.

For the non-well ordered, i.e., partworth, attribute Brand, the effect, although significant, is substantially less in magnitude.

Table 1: Regression of Attribute Importance on Number of Levels

Attribute	Beta ^a	R2	ANOVA
Price ^b	.077*	.03	.001* (F=10.32)
Brand ^c	.026*	.02	.004* (F=8.226)

^a Levels effects for each attribute were estimated by regressing the merged attribute importance scores of the original attribute and the degraded attribute on their respective number of levels.

^b Price is a well-ordered or vector attribute.

^c Brand is a non-well-ordered or partworth attribute. However, interior levels were excluded on a per-respondent basis to ensure no exterior levels were excluded in the two-level case.

* Significant at the 95% confidence level

A Proposed Solution to the Number of Levels Effect

The above analysis suggests a possible solution to eliminate entirely the number of levels effect regardless of its source.

Methodology - Second Study

The subject of this study was high-end ice hockey skates. The study was designed to be a full-profile metric conjoint study that had these attributes:

- Brand (4 levels)
- Price (4 levels)
- Visual Design (3 levels)
- Weight (3 levels)
- Psychological Price Point (2 levels)

Brand and Visual Design were partworth attributes. Price and weight were vector attributes. Psychological price point was a metric attribute with \$0 and \$1 as its two levels. Respondents were shown a product price that was the sum of the values for the price attribute and the psychological price point attribute. For example, if price were \$399 and psychological price point were \$1, respondents would see a price of \$470. If price were \$399 and psychological price point were \$0, respondents would see a price of \$399.

Respondents participated in a two-stage conjoint exercise. The first conjoint exercise had only two levels for each attribute. The levels used were the exterior levels, i.e., those levels that had maximum and minimum utility for each individual respondent. For the partworth attributes, exterior levels, that is, the most preferred and least preferred levels, had to be identified for each respondent prior to the first conjoint exercise. This was done by direct questioning. For the vector attributes, the numeric maximum and minimum values were assumed to be exterior for all

respondents. There were 18 different versions of this two-level design (6 different Brand pairs times three different Visual Design pairs). Respondents in this section of the study rated 12 different hockey skates for purchase interest.

The second conjoint exercise was a full-profile metric conjoint exercise utilizing all levels of all attributes. For this exercise, respondents rated 18 cards.

As was the case in the earlier study, both of these experimental designs were reasonably orthogonal and balanced.

The general concept is to identify attribute relative importance scores from the first stage conjoint exercise (exterior levels only). Utility estimates from this stage should exhibit no number of levels effect since all attributes have the same number of levels. The second stage conjoint exercise should establish the relative preference of levels within attribute.

The full-level utility estimates can then be linearly scaled into the two-level estimates. The resulting utilities will exhibit the correct attribute relative importance and also maintain the relative positions of levels within each attribute.

Data for this second study were collected in late December, 1998, via a Web-based survey, which allowed greater design flexibility and experimental control than paper-and-pencil data collection.

Analysis - Second Study

Prior to analysis, the data sets were edited so that respondents with individual-level conjoint models that were not significant at least at the 75% confidence level were excluded from further analysis. Approximately one-third of the sample was discarded at this stage.

Additionally, respondents were excluded who did not provide consistent claimed and derived exterior levels, i.e., exterior levels from the direct questioning which were the same as the exterior levels computed from the full-levels conjoint. This second criterion caused a dramatic reduction in sample size. Approximately two-thirds of the remaining sample was discarded at this stage.

The initial sample size was 425. The final sample size was 79.

Upon reviewing possible sources for this high percentage of inconsistent respondents, it was concluded that the wording of the exterior levels direct questions were confusing and misleading. Other sources of this inconsistency might have been model instability, irrational respondents or poor data quality due to the Web-based collection method.

However, the exterior levels questions were redesigned for a subsequent paper-and-pencil study which employed the same study design. Results from that study, while improved, were still disappointing. Approximately half of the sample did not provide consistent claimed exterior levels when compared to derived exterior levels. If either question wording or quality of Web-based data were the primary source of this inconsistency, the paper-and-pencil study should have shown much greater improvement.

Further, we would expect most unstable models and irrational respondents to be excluded by discarding all models which were not significant at least at the 75% confidence level.

Additional possible explanations include respondent indifference to alternative levels, respondent fatigue, confusion- or fatigue-motivated simplification where the respondent would focus on one attribute that was important to him or her and ignore the others. Interaction effects, i.e., respondents may impute certain properties to certain levels that are not inherent in those levels, may also distort the claimed exterior levels identified by direct questioning. For example, a respondent may assume that a specific brand is expensive, heavy and/or traditional looking during the direct questioning (thus coloring his or her responses to those questions) but may change that opinion when shown an alternative that lists that brand with the attribute levels low price, light weight and stylish.

Additional research needs to be conducted to explore possible reasons for the high degree of inconsistency in respondent data between claimed and derived exterior levels.

Results - Second Study

Table 2 shows the attribute relative importance scores for the full-levels stage and for the two-levels stage. There are differences in relative importance, particularly for psychological price point. It is suspected that differences in attribute relative importance scores might have been more dramatic if the variance in number of levels across attributes had been greater.

Table 2:
Attribute Relative Importance: Complete and Two-level

<i>Mean Relative Importance</i>		
Attribute	Full-levels n=79	Two-levels n=79
Price	13%	14%
Brand	53%	51%
Psychological Price Point	6%	10%
Visual Design	17%	14%
Weight	11%	11%

However, the two-level design does not provide information about all of the attribute levels that may be of interest to management.

If, for each respondent, his/her utility weights for an attribute with three or more levels are linearly scaled into his/her utility weights for the same attribute with two levels, then attribute relative importance is maintained as well as level importance within attribute.

Table 3 shows the utility weights for attribute levels for the second conjoint exercise (full-levels) and attribute levels for the second conjoint exercise rescaled to have the same attribute relative importance scores as the attributes from the two-levels conjoint. The relationship between levels within attribute from the full-levels stage (stage 2) are preserved while the attribute relative importance scores from the two-levels stage (stage 1) are also preserved.

Table 3: Original and Rescaled Attribute Level Utilities

Attribute	Original Full-levels n=79	Rescaled Two-levels n=79
Price	-.00287	-.00274
Brand		
A	-.461	-.479
B	.657	.724
C	-.221	-.383
D	.024	.186
Psych. Pt.	-.088	-.026
Design		
A	-.009	.036
B	-.06	-.091
C	.069	.125
Weight	-.067	-.002

Alternative Solutions to Respondent Inconsistency

To avoid the problems of respondent inconsistency, there are at least three possible alternatives. For aggregate models, one could conduct a full-levels conjoint exercise, calculate utility weights, identify exterior levels among aggregate, mean utilities, then conduct a subsequent two-level exercise with a fresh sample. It may be the case, however, that the problems of heterogeneity normally associated with aggregate models could affect the accuracy and usefulness of this approach.

For disaggregate models, one could conduct a full-levels conjoint exercise, calculate utility weights during the interview, identify exterior levels for each respondent, create an appropriate questionnaire (in real time), then conduct a subsequent two-level exercise with the same respondent. This approach is necessarily adaptive and would require some form of computer-assisted interviewing. It also assumes that the psychological component of the number of levels effect is extremely short-term. This assumption would need to be tested before this alternative could be accepted.

Another alternative for disaggregate models would be to rescale the full-levels utilities into the two-levels utilities, regardless of whether or not the two-levels utilities are exterior. It is not clear that the resulting attribute relative importance scores would or would not accurately reflect the true two-levels importance scores, i.e., the attribute relative importance scores that would have been computed had all respondents been shown exterior levels in the two-levels exercise.

Summary

The existence of both psychological and algorithmic components to the number of levels effect has been demonstrated in prior studies.

Here, we have demonstrated a potential solution to eliminate the number of levels effect regardless of its source. Given an appropriate data collection methodology, such as Web-based surveys, and a two trade-off study design, conjoint utilities can be estimated for all attributes in their original specifications as well as for all attributes redefined to the two level case. The original utility weights can be linearly scaled into the two-level utility weights to remove the number of levels effect and more accurately reflect attribute relative importance.

More work must be done, however, to increase the consistency between claimed exterior levels and derived exterior levels or to find an alternative way to identify exterior levels.

The author wishes to thank Rich Johnson, Dick Wittink, Jayme Plunkett and Jamin Brazil for their invaluable assistance with this paper.

References

Currim, I.S., C.B. Weinberg, D.R. Wittink (1981), "The Design of Subscription Programs for a Performing Arts Series," *Journal of Consumer Research*, 8 (June), 67-75.

Green, P.E., and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*, 54 (October), 3-19.

Steenkamp, J.E.M., and D.R. Wittink (1994), "The Metric Quality of Full-Profile Judgments and the Number-of-Levels Effect in Conjoint Analysis," *International Journal of Research in Marketing*, Vol. 11, Num. 3 (June), 275-286.

Wittink, D. R., (1990), "Attribute Level Effects in Conjoint Results: The Problem and Possible Solutions," 1990 Advanced Research Techniques Forum Proceedings, American Marketing Association.

Wittink, D. R., J. C. Huber, J. A. Fiedler, and R. L. Miller (1992), "The Magnitude of and an Explanation for the Number of Levels Effect in Conjoint Analysis," working paper, Cornell University (December).

Wittink, D. R., J. C. Huber, P. Zandan, R. M. Johnson (1992), "The Number of Levels Effect in Conjoint: Where Does It Come From and Can It Be Eliminated?," 1992 Sawtooth Software Conference Proceedings, 355-364.

Wittink, D.R., L. Krishnamurthi, and D.J. Reibstein (1989), "The Effects of Differences in the Number of Attribute Levels on Conjoint Results," *Marketing Letters*, 1, 113-23.

© 1999 / MACRO Consulting, Inc.

Published in the 1999 SAWTOOTH SOFTWARE CONFERENCE PROCEEDINGS, February 1999, San Diego, CA, USA.

We are an independent **marketing research consulting firm** dedicated to helping you make the most informed, insightful marketing decisions possible. We specialize in technology, consumer, and new product research, and are well recognized for our **State-of-the-Art Research** techniques.

Ultimately, we provide more than just technical expertise. We focus on developing **pragmatic solutions** that will have a positive impact on the profitability of our clients.

CONTACT US:

Telephone: 650-823-3042

General Inquiries:
info@macroinc.com

Advanced Analysis Inquiries:
analysis@macroinc.com

richard@macroinc.com

www.macroinc.com