Is there more information in Best Worst choice data?

Using the variance-covariance matrix to consider consumer heterogeneity

Dr Simone Mueller
Research Fellow Wine Marketing
Ehrenberg-Bass Institute for Marketing Science, University of South Australia
GPO Box 2471 Adelaide, South Australia 5001
Telephone: (61 8) 8302 0906; Facsimile: (61 8) 8302 0042
Email: simone.mueller@unisa.edu.au

Dr Cam Rungie
Senior Lecturer
Ehrenberg-Bass Institute for Marketing Science, University of South Australia
GPO Box 2471 Adelaide, South Australia 5001
Telephone: (61 8) 8302 9410; Facsimile: (61 8) 8302 0042
Email: cam.rungie@unisa.edu.au

Dr Steve Goodman
Senior Lecturer
The University of Adelaide Business School
Adelaide, South Australia 5005
Telephone: (61 8) 8303 4650; Facsimile: (61 8) 8223 4782
Email: steve.goodman@adelaide.edu.au

Professor Larry Lockshin
Director - Wine Marketing Group
Ehrenberg-Bass Institute of Marketing Science, University of South Australia
GPO Box 2471 Adelaide, South Australia 5001
Telephone: (61 8) 8302 0261; Facsimile: (61 8) 8302 0042
Email: larry.lockshin@unisa.edu.au

Dr Eli Cohen
Senior Lecturer
Guilford Glazer School of Business and Management
Department of Hotel and Tourism Management
P.O. Box 653, Beer Sheva 84105, Israel
Telephone: (972 8) 6472 776; Facsimile: (972 8) 6472 920
Email: elico@bgu.ac.il
Adjunct Senior Lecturer
School of Marketing, University of South Australia
GPO Box 2471 Adelaide, South Australia 5001
Email: eli.cohen@unisa.edu.au

Is there more information in Best Worst choice data?

Using the variance-covariance matrix to consider respondent heterogeneity

Abstract

Best Worst Scaling (BWS) has been shown to be a powerful method for preference and attribute importance measurement and is already widely established in wine marketing to analyse what drives wine consumers' purchase decisions. Most prior BWS studies utilised Best Worst scores on an aggregated level only to measure relative attribute importance for the total sample or used sociodemographic or wine behaviour related variables to describe a-priori segments. Not many studies considered individual differences in the Best Worst scores to find post hoc segments based on revealed differences of attribute importance. Is there more information contained in Best-Worst data than has been considered to approach this problem? We exemplify for data of British on-premise wine purchase behaviour how considering the attribute variance-covariance matrix allows valuable insights into what drives consumer heterogeneity. Attributes with high variance signal respondents' disagreement on their importance and indicate the existence of distinctive consumer segments. Attributes jointly driving those segments can be identified by a high covariance. Based on the variance-covariance matrix we identify five dimensions of utility with a principal component analysis which show to be a very efficient tool for an effective interpretation of behavioural drivers of clusters derived by latent clustering. Our analysis opens new doors for marketing research to a more insightful interpretation of BWS data. This information gives marketing managers powerful advice on which attributes they have to focus to target different consumer segments.

Keywords: Best Worst Scaling, discrete choice, heterogeneity, segmentation, visualisation, Latent Class Clustering, wine, on-premise, UK

Introduction

There has been a quiet revolution in consumer preference measurement with the advent of Best Worst scaling, which is derived from the method of discrete choice (Finn & Louviere, 1992; Marley & Louviere, 2005). Best-Worst Scaling uses consumer choices of the best and worst or most and least important items in a set, which are usually concepts or written attributes, in a designed study to create a ratio-based scale. Best Worst Scaling overcomes several biases resulting from scores or ratings such as their inherent assumption of interval scales with absolute differences between scale points and their inferior discriminatory power (Cohen & Neira, 2003).

Best Worst scaling has now been widely used in social sciences and marketing research (Auger, Devinney, & Louviere, 2007; Cohen & Orme, 2004; Goodman, Lockshin, & Cohen, 2006; Louviere & Islam, 2007). Especially in wine marketing Best Worst scaling has proven its strength for a cross cultural study (Goodman, Lockshin, & Cohen, 2007) involving more than ten international wine markets to compare wine attribute importance on one identical scale and thereby eliminating any bias potentially caused by different scale usage in different cultures.

The majority of Best Worst studies focused on attribute importance on an aggregated level only (Finn & Louviere, 1992; Goodman, Lockshin, & Cohen, 2005; Louviere & Islam, 2007) or formed a-priori segments based on sociodemographic and wine behaviour related variables

(Goodman et al., 2006). Lockshin, Spawton & Macintosh (1997) give an overview of segmentation in wine marketing and point out the necessity to consider wine consumer heterogeneity when drawing valid conclusions. There are generally two classes of segmentation methods: a-priori segmentation based on prior known groups (e.g. gender, age) and post hoc segmentation utilising results of prior data analysis like attitude measures or other important constructs to identify distinct clusters (Wedel & Kamakura, 1999). Wedel & Kamakura (1999)suggest the superiority of post hoc segmentation using revealed attribute utilities which resulted in more stable and time consistent clusters than a-priori clustering variables. Especially sociodemographic variables have shown to be only weakly related to differences in purchase behaviour (Lockshin, Spawton & Macintosh, 1997; Aurifeille, Quester, Lockshin & Spawton, 2002).

Different segmentation approaches based on Best Worst results have been used in other disciplines than wine marketing to take respondent heterogeneity into account. Auger et al.(2007) applied the Ward Clustering method to individual Best Worst scores to find consistent patterns in ethical beliefs across several countries. Cohen & Neira (2003) used Latent Class Modelling to find clusters, which grouped similar utility components concerned with drinking coffee. But no Best Worst study has analysed attribute importance heterogeneity based on post hoc individual Best Worst scores. We apply a very simple but powerful analysis of the variance-covariance matrix of individual Best Worst scores to detect which attributes are determining utility components and drive distinct consumer segments. Our detailed explanation and visualisation aims to help understand the underlying principles usually hidden in such more advanced procedures. The simplicity and ease of use of our method will help practitioners adopt it in their market analysis.

We use Best Worst data of the attribute importance of British wine consumers when purchasing wine on-premise (in a bar, café, or restaurant). We first describe the data sample in the next section. Afterwards we introduce a crude method to derive the variance-covariance matrix. In the result section we show how this information allows us to include consumer heterogeneity and attribute relationships in our BW analysis and how this information can further be used to interpret consumer segments. Thereby we include in our explanation how our method allows marketing managers a more thorough understanding of what drives their customers and provides insights in how to target different consumer segments. We finish with a conclusion and outline further research to advance this area.

Data Collection

The data were collected using an online survey instrument in February 2007. Respondents were invited to respond from a panel of consumers registered for online survey completion. Respondents were paid for completing the questionnaire and a quota system was used to ensure a proportionate response in line with English population profiles for age and gender. A number of areas were monitored to allow a gauge of the reliability of the results, from the time taken to complete through to drop-off rates. All measures were normal for this type of research.

The on-premise UK study is part of a larger cross cultural study to analyse purchase behaviour in 11 international wine markets (Goodman et al., 2007, Goodman, Lockshin, & Cohen, 2008 and www.winepreferences.com/influences/influences.html). The thirteen attributes (see Table 1) were chosen to represent a wide variety of on-premise wine choice drivers in all markets involved in the overall study. Lockshin and Hall (2003) reviewed current consumer behaviour research in wine and Lockshin et al. (2006) updated that review, but no specific articles focused on consumer choice criteria on-premise; all the articles focused on retail stores. The choice set, therefore, was developed after a review of the literature, using discussion with industry participants, consumers and then trialled in a pilot study. Every respondent answered a complete Youden square design of thirteen choice sets with choice set size of four attributes where every attribute appeared four times and pair frequency equalled one. Three hundred three completed questionnaires were used for data analysis.

The sample can be assumed to be representative for British wine consumers and has equal proportions of male and female respondents. A quota scheme ensured that age groups (18-24, 25-40, 41-50 and >50 years) were equally represented by 25% each. The distribution of respondents' household income is typical for the UK. The sample contains frequent wine consumers (33% more than once week, 51% once a week or less) and less frequent wine consumers (16% only at special occasions).

Research Method

The process of deriving aggregated Best-Worst scores from individual choices has already been extensively described in various publications (Flynn, Louviere, Peters, & Coast, 2007; Goodman et al., 2005; Mueller, Francis, & Lockshin, 2007) and is not our focus here. Best Worst Scaling produces an interval scaled utility score which is unbiased by individual scale usage (Marley & Louviere, 2007).

There exist two alternative approaches to derive the variance-covariance matrix, which contains attribute importance heterogeneity (variance) and the co-relation of attributes (covariance), a 'crude' method based on aggregated choices and a more precise method based on individual choices (Rungie, 2008).

We apply the 'crude' method, which calculates the variance-covariance matrix from individual BW scores which represent aggregated choices of best and worst over every respondent and attribute. The Best-Worst of attribute *i* (BW_i) was calculated by subtracting the number of times of that attribute was chosen least important for individual *i* from the frequency it was chosen most important for the same individual.

We use a principal component analysis of the BW scores to derive five distinct utility components, which drive consumers' choice behaviour. To model consumer heterogeneity we use Latent Class Clustering based on individual scores for each of the BWS attributes and then compare the derived clusters across the cognitive utility dimensions. This demonstrates the usefulness of analysing heterogeneity and linking it with a cluster analysis.

Results

1) Attribute importance

The attribute 'I have had the wine before and liked it' was most often chosen as most important (best) and least often chosen as least important (worst), accordingly its aggregated Best-Worst score is highest (see Table 1). The mean B-W represents the average B-W per respondent and is derived by dividing the overall B-W by sample size. The relative importance between attributes can be more easily interpreted when standardising the BW score to a probabilistic ratio scale.

This ratio scale can be derived by transforming the square root of Best divided by worst to a 0 to 100 scale (Mueller et al., 2007). The Sqrt(B/W) for all attributes is scaled by a factor such that the most important attribute with the highest Sqrt(B/W) becomes 100. All attributes can then be compared to each other by their relative ratio, e.g. ‘I matched it to my food’ is 0.54 times (approximately half) as important to the overall sample as ‘I had the wine before and liked it’. Similar, to try something different is twice as important as availability by the glass.

Overall, alcohol level and availability in small units such as by the glass or in half bottle are not very important for British on-premise wine consumers in average. In the middle there is a number of attributes with rather similar importance such as region, waiter recommendation, menu suggestion and varietal.

Table 1: Attribute importance on aggregated level (n=304)

Attribute	**Best**	**Worst**	**B-W**	**Mean (B-W)**	**Sqrt (B/W)**	**Sqrt stand**
I have had the wine before and liked it	790	70	720	2.37	3.36	100
I Matched it to my food	521	156	365	1.20	1.83	54
Suggested by another at the table	434	161	273	0.90	1.64	49
Try something different	333	174	159	0.52	1.38	41
Region	354	277	77	0.25	1.13	34
I had read about it, but never tasted	265	202	63	0.21	1.15	34
Waiter recommended	196	274	-78	-0.26	0.85	25
Suggestion on the menu	197	279	-82	-0.27	0.84	25
Varietal	164	275	-111	-0.37	0.77	23
Available by the glass	209	453	-244	-0.80	0.68	20
Promotion card on the table	213	508	-295	-0.97	0.65	19
Available in Half Bottle (375ml)	165	500	-335	-1.10	0.57	17
Alcohol level below 13%	111	623	-512	-1.68	0.42	13

All the importance measures B-W, mean(B-W) and standardised Sqrt(B/W) result in the same attribute order. For the remainder of this paper we will use the average B-W to measure attribute importance as it most closely related to the variance-covariance matrix. The average B-W score is visualised in Figure 1. As every attribute appeared four times in the choice design the maximum it could be chosen as most (best) and least (worst) important is 4., similarly the minimum of B-W is -4. Bars of items which were more often chosen as best than as worst are on the right hand side (B-W>0) whereas items more often chosen as worst than as best (B-W<0) are drawn to the left side of the vertical axis. For example the item ‘I liked the wine before’ was in average net 2.37 out of 4 appearances chosen as most important. Likewise ‘Alc < 13%’ was in average net 1.7 out of 4 possible times chosen as least important. Items in the middle where either not often chosen as best or worst or were chosen as best the same number of times as worst.

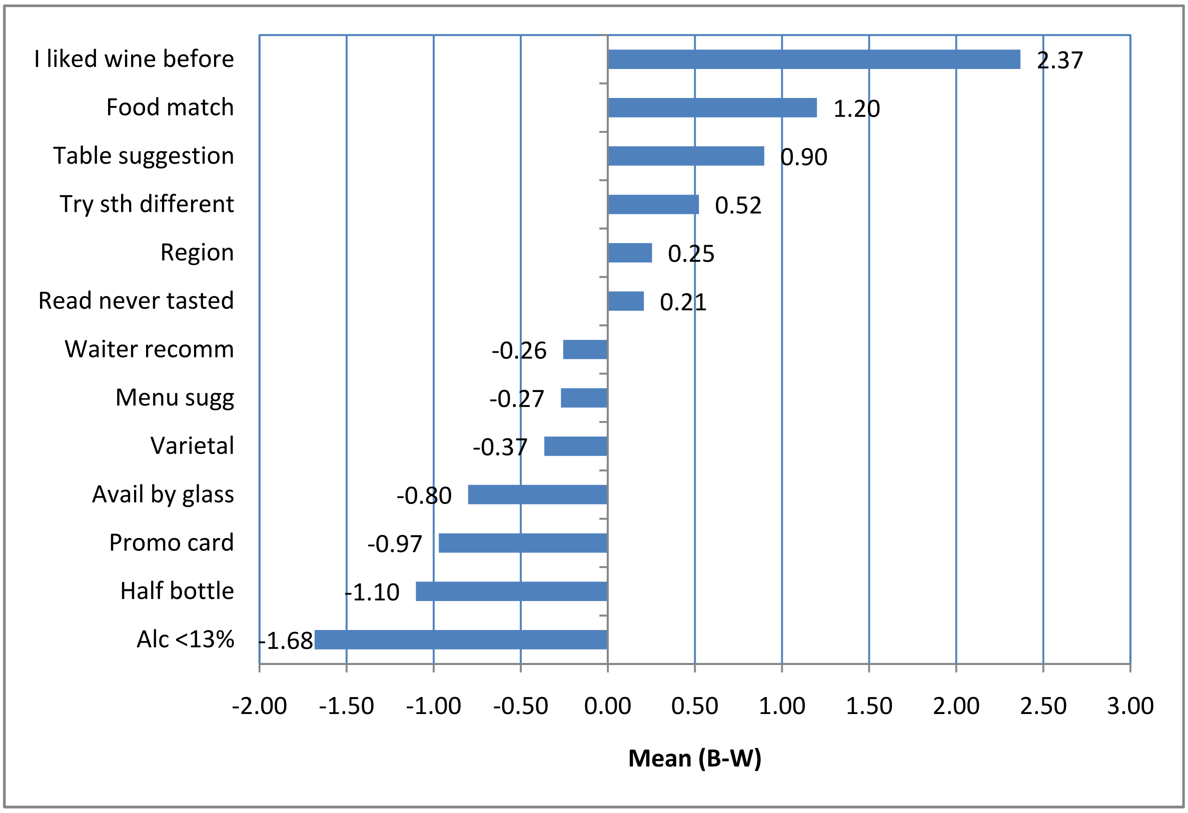

Figure 1: Average B-W score over total sample (n=304)

2) Importance heterogeneity

From the average B-W score we do not yet know if an attribute was similarly important to all consumers. The intermediate average BW score of an attribute such as region or waiter recommendation can either be caused by all respondents perceiving it as medium important or can be a result of averaging out respondents for whom it is very important with respondents for whom it is not very unimportant. The later case of consumer heterogeneity means marketing managers should respond very differently by targeting those consumers with high attribute importance with different products, channels or communication than consumers with low importance. The average alone does not yet give them any guidance related to this problem.

The degree of attribute importance heterogeneity is expressed by the variance or standard deviation of BW scores.

Table 2 shows the variance and standard deviation of the on-premise wine purchase attributes for our UK study. Just as the B-W score has a range of -4 to +4 for the design used in this data then it can be shown that the standard deviation is similarly bounded. Under extreme conditions of heterogeneity, which in practice will never occur, one attribute could have a standard deviation of 4 (half the respondents select the item as best at every opportunity and half select it as worst at every opportunity) and the others attributes would all have smaller standard deviations. In Table 2 all attributes have a standard deviation above one, which is a signal of existing consumer heterogeneity for all of them. There are some attributes, which show relatively higher agreement of their relative importance (e.g. menu suggestion, try something different and liked before). Other attributes such as region, availability by the glass, promotion card and matching with food have a higher standard deviation indicating respondents' disagreement on their relative importance.

Table 2: Variance and standard deviation of attribute importance (n=304)

Attribute	Mean B-W	Var(B-W)	Stdev(B-W)
I have had the wine before and liked it	2.37	2.68	1.64
I Matched it to my food	1.20	3.56	1.89
Suggested by another at the table	0.90	3.49	1.87
Try something different	0.52	2.47	1.57
Region	0.25	4.65	2.16
I had read about it, but never tasted	0.21	2.58	1.61
Waiter recommended	-0.26	3.22	1.79
Suggestion on the menu	-0.27	2.05	1.43
Varietal	-0.37	2.81	1.68
Available by the glass	-0.80	4.01	2.00
Promotion card on the table	-0.97	3.57	1.89
Available in Half Bottle (375ml)	-1.10	3.45	1.86
Alcohol level below 13%	-1.68	3.06	1.75

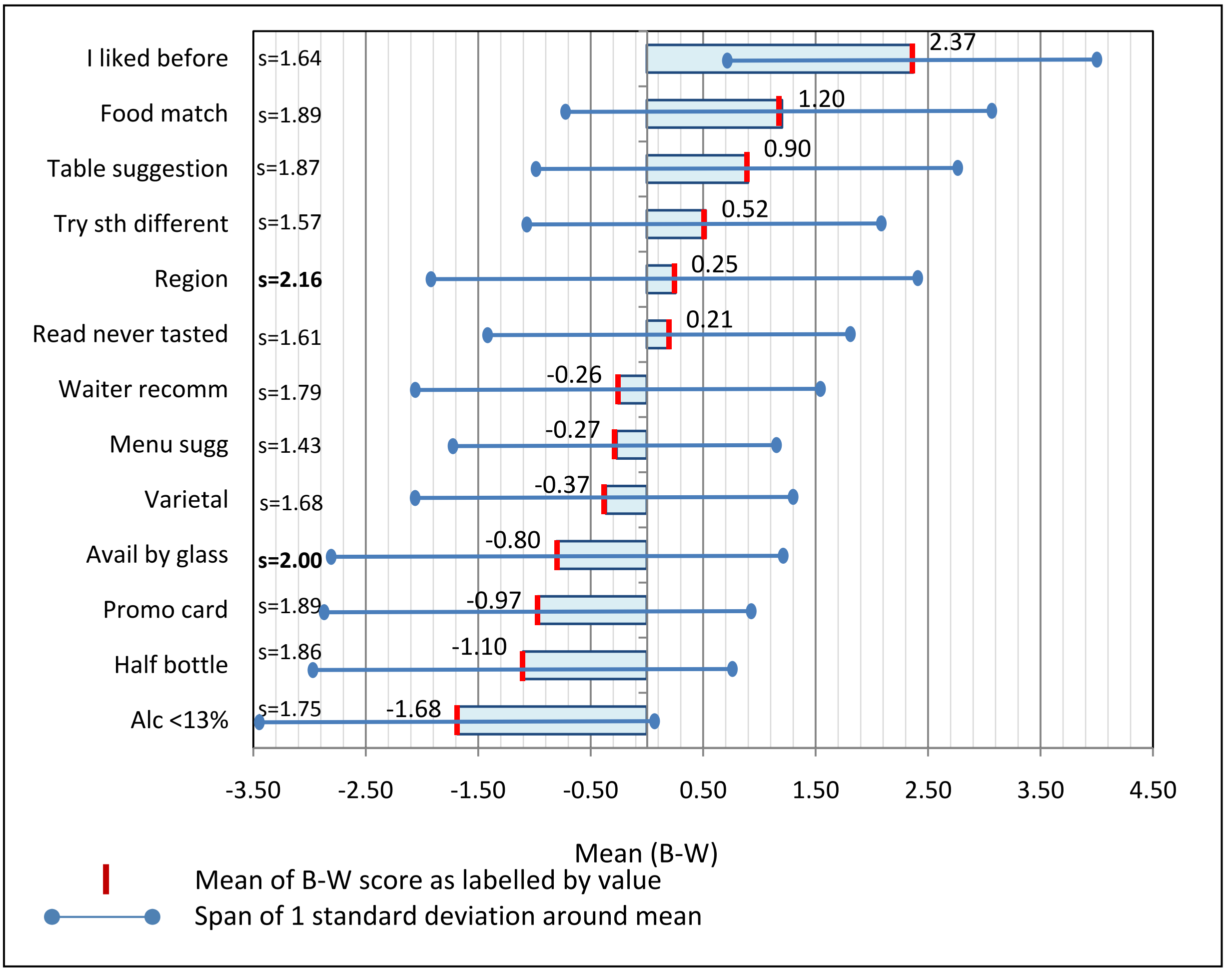

Figure 2: Attribute importance and standard deviation (n=304)

A graphical representation of attribute importance heterogeneity can be seen in Figure 2. As in Figure 1 the bars represent the net average of how often an item was chosen as best (positive value) or as worst (negative value). For better visibility the bars' ends are marked with a heavy solid line. The whiskers around the average score represent one standard deviation (*s*) on each side, two *s* in total. Thus, attributes with a higher standard deviation have longer whiskers, implying respondent heterogeneity. The length of the whiskers can be interpreted as the share of respondents who have a lower or higher individual (B-W) for this attribute than the aggregated mean.

For the most important attribute 'I liked the wine before' the maximum possible mean(B-W) of +4 lies within one standard deviation, implying that a considerable portion of respondents chose this item always as best whenever it appeared in their choice set. Comparing two items with a total B-W average around zero such as 'region' and 'I read about the wine but never tasted before' it becomes clear that 'read but never tasted' was mostly neither chosen as best nor worst, whereas for region the heterogeneity in choices of best and worst cancelled each other out. For a marketing manager this means that there are some consumers who care about the wine's region, which can be specifically targeted, whereas having read about a wine is more unanimously considered as medium important by most on-premise wine consumers.

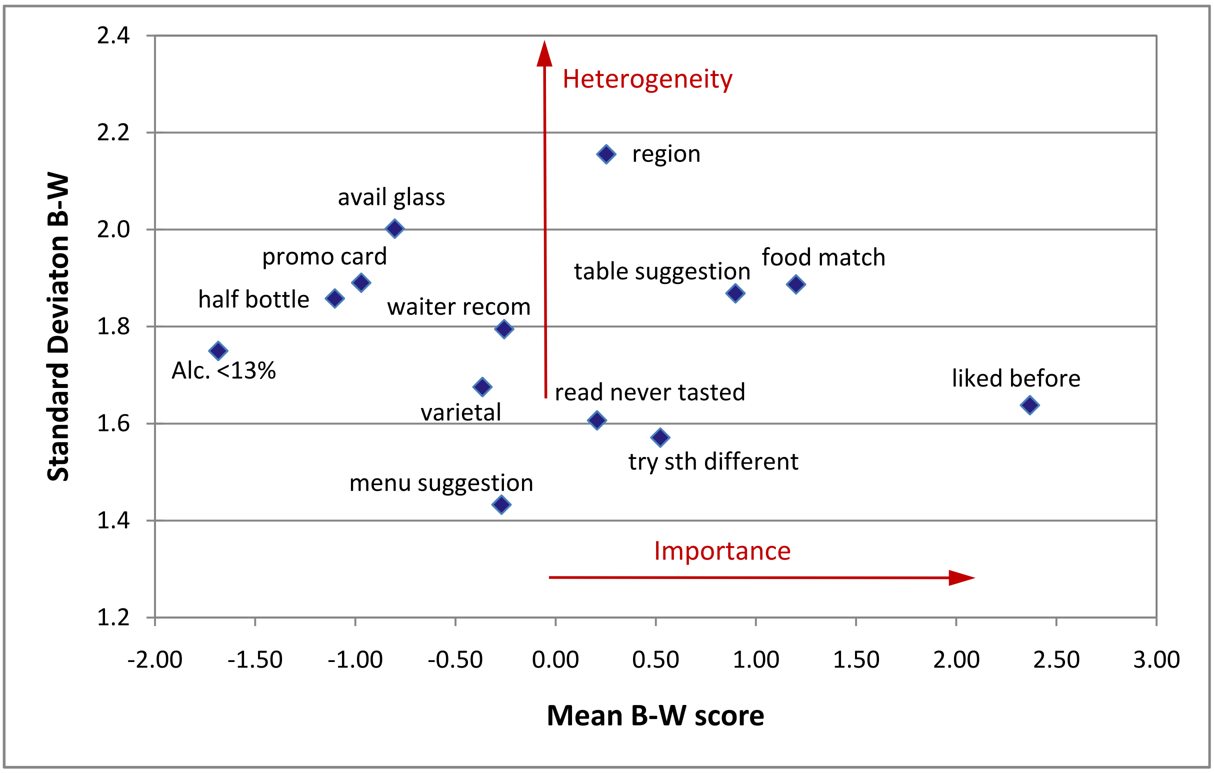

Figure 3: Attribute importance and heterogeneity

Both dimensions of attribute importance and heterogeneity are visualised in Figure 3 where the mean(B-W) and its standard deviation are graphed together. Companies should optimise those attributes with high importance. In addition companies should pay special attention to those attributes that show a high amount of heterogeneity and reasonable importance implying that they are very important to a subset of their customers, even though they may not be important to most consumers. Those attributes can be found in the right upper part of the coordinate system, such as region, food match and suggestion by someone else at the table. Attributes like available by the glass which have a low mean(B-W) but score high in heterogeneity are suitable for niche markets, if the company wants to develop a marketing mix for smaller numbers of customers.

3) Related drivers of heterogeneity

For wine marketers it would be interesting to know if important attributes with high heterogeneity (i.e. region, food match, and table suggestion) are distinct drivers of different consumer segments (see Appendix A) or if they are related and are jointly important for the same target group. The variance-covariance matrix show strongly every attribute pair varies together. If one of two attributes with a high positive covariance is above its expected value (mean(B-W)) than the other attribute also tends to score above its expected value. In other words, if one attribute is more important for an individual than for the average of all consumers than a high covariance implies a high probability that the other attribute is also more than of average importance. Thus attributes with a high positive covariance jointly drive the same segment. Similarly attributes with high negative covariance also drive the same segment, but in opposite directions.

A measure closely related to covariance is the correlation of two items, which is defined as their covariance divided by the standard deviation of every item. The correlation coefficient is often easier to interpret as it is limited to values between +1 and -1. The significance value of a correlation gives the probability that a correlation coefficient is significantly different from zero and can be a guide to finding strong attribute correlations (see Appendix B).

In our case the correlation matrix in Appendix B shows a moderately strong positive relationship between availability by the glass and availability in half bottles (r=0.38), implying a similar importance for this target segment. Promotion card is rather strongly negatively correlated with food match (r=-0.42) and region (r=-0.36), implying that those consumers influenced by a promotion card on the table did not select the wine according to its region of origin nor to match their food. According to Cohen and Cohen (1983) correlations below .35 are considered rather low, while those above .45 are considered moderate to high. Because the aim of the study was to cover the most important drivers for on-premise wine choice, a series of very strong correlations would indicate that mainly redundant characteristics would have been selected, risking that other important ones were omitted. Therefore for the purpose of this study correlations approaching .35 are also considered as relevant. With this in mind, region of origin and food match show a moderate correlation (r=0.21), suggesting that both attributes are of above average importance for similar consumers. Other significant attribute correlations are promotion card on the table with menu suggestion (r=0.17) and region with varietal (r=0.17). On the other hand, respondents for whom extrinsic attributes as region and grape variety are important perceive promotion cards and availability by the glass as less important for their purchase decision (negative correlation) but do consider if the wine matches their food.

4) Utility structure

The information contained in the variance-covariance matrix can be further condensed by a principal component analysis, which reports latent factors influencing consumers' purchase behaviour. These factors can be interpreted as cognitive utility dimensions, which determine individuals' behaviour (Luce, 1959). Therefore, these factors do not yet in themselves represent different consumer groups but consumer segments can be described by the utility dimensions, which dominate their behaviour.

A principal component analysis with varimax rotation and Kaiser Normalisation resulted in five factors, which explain 61% of variance. This score is strong evidence that the thirteen attributes used in the study are not independent and represent about five independent underlying dimensions of preference. While individual pair-wise correlations in the data are not overly large, the overall correlation structure for the best-worst scores for the thirteen attributes is highly correlated.

Each utility dimension is defined by those attributes, which load highest and lowest on each factor. Thus, each independent dimension is defined by its two end points, the attributes with the highest positive and negative factor loadings. Table 3 shows the factor loadings of all 13 attributes, the highest positive (bold and gray under laid) and negative values (bold) for every factor highlighted.

Table 3: Choice attribute factor loadings for principal component analysis

Utility factors	**1**	**2**	**3**	**4**	**5**
Variance explained by factor	17.5%	14.3%	11.7%	8.9%	8.5%
Factor name	Ease of trial	New experience	Restaurant advice	Low risk food matching	Cognitive chooser
Promotion card on the table	**0.78**	-0.07	0.03	0.07	0.12
Suggestion on the menu	**0.38**	-0.14	**0.57**	**0.29**	0.18
Available by the glass	**0.35**	**-0.40**	**-0.59**	-0.17	0.11
Available in Half Bottle (375ml)	**0.25**	-0.27	**-0.49**	-0.41	0.01
Waiter recommended	0.11	-0.15	**0.73**	-0.23	-0.17
Try something different	0.09	**0.77**	-0.14	-0.07	0.11
I have had the wine before, liked it	0.03	-0.16	-0.10	**0.72**	-0.35
Alcohol level below 13%	0.01	-0.16	-0.10	**-0.70**	-0.10
I had read about it, but never tasted	0.00	**0.68**	0.09	0.11	-0.06
Suggested by another at the table	-0.11	0.14	0.15	0.16	**-0.77**
Varietal	-0.37	0.26	0.04	0.11	**0.65**
Region	**-0.66**	0.00	0.07	-0.04	**0.22**
I matched it to my food	**-0.70**	-0.30	-0.10	**0.30**	0.04

The *first factor* is defined by the difference between attributes simplifying the decision and reducing the risk of choice and attributes implying a very cognitive decision such as matching the wine to food or choosing by region. We therefore label this factor ‘ease of trial’ as a high loading implies that respondents value suggestions by promotion cards on the table and on the menu and availability by the glass and in half bottles as very important for their on-premise wine choice. A negative loading on this utility dimension implies that consumers prefer to choose by region and by matching the wine to their food.

The *second factor* can be characterised as looking for a new experiences with high loadings on try something different and read about the wine but never tasted it. For respondents scoring high on this factor availability by the glass or half bottle are rather unimportant. Suggestions on the menu and waiter recommendation load high on the *third factor*, which captures the utility dimension of advice from the restaurant. The opposite of this utility dimension is availability by the glass and in half bottles. These people expect to buy a standard bottle of wine. The most important attribute loading on the *fourth factor* is prior experience and liking, which implies low risk decision-making. Also, the desire to match the wine to food and to follow suggestions on the menu indicate a traditionalist wine approach. Finally, the last factor shows high importance of varietal and region, both extrinsic wine attributes which require a certain cognitive understanding of wine. To follow suggestions by others at the table forms the most negative aspect of the fifth utility dimension, thus we label it with cognitive chooser.

These five utility dimensions of on-premise choice behaviour derived from the covariance of attribute importance help us to understand consumers' cognitive networks and underlying behavioural choice processes. In the next step we will show how those utility dimensions can be utilised to understand and visualise behavioural differences between different consumer clusters.

5) Preference clusters and their utility structure

Whereas the utility dimensions span a five dimensional utility space, which is universal for all consumers, each consumer differs regarding the relative importance of each factor for his or her position in the utility space. Every consumer can be located and therefore characterised by his or her spatial utility location, defined by five coordinates, the regression factor scores of his attribute importance of all thirteen attributes on the five factors. Thereby a consumer with an individual high positive factor loading is driven by those attributes, which load high on this factor. Opposing attributes with negative loadings are unimportant for on-premise wine choice.

We found four distinct consumer clusters with a Latent Class Clustering Model (Vermunt & Magidson, 2005) by using the thirteen Best Worst attributes as dependent variables. A detailed description of the four clusters with their average attribute scores is shown in Appendix D. Respondents in each cluster are characterised by a similar wine choice behaviour driven by similar utility structures. That means they are located in similar regions of the utility space, characterised by their factor score coordinates.

Table 4: Cluster means of individual utility dimension factor scores for Latent Class 4-Cluster solution

	C1		**C2**		**C3**		**C4**			
n	135		74		55		40			
	44%		24%		18%		13%		ANOVA	
Cluster means of individual utility factor scores									F value	p
Factor 1: Ease of trial	0.59	a	-0.90	b	-0.65	b	0.58	a	89.16	0.00
Factor 2: New experience	-0.26	a	0.02	a	0.69	b	-0.09	a	13.50	0.00
Factor 3: Restaurant advice	-0.10	a	-0.78	b	0.90	c	0.56	c	51.50	0.00
Factor 4: Low risk food matching	-0.56	a	0.32	b	0.36	b	0.81	c	38.42	0.00
Factor 5: Cognitive chooser	0.14	a	0.16	a	0.00	a	-0.77	b	10.25	0.00

Latent Cluster solution, LL = -7463.25, Classification R^2=0.89, Classification error=0.048
Different superscript letters: significantly different at p=0.05, post-hoc Tukey-test.

Table 4 shows the ex-post cluster averages of individual factor loadings of the five utility dimensions (factors). The ANOVA F-values indicate that the four consumer clusters are highly significantly different in their location on every utility dimension. A consumer segment is driven by those utility dimensions on which it scores highly positive or negative. Factor scores are normalised between +1 and -1 and therefore allow a very easy comparison of the relative importance of each utility dimension between the four clusters. This advantage becomes especially clear if one tries to find behavioural drivers out of the differences in attribute importances in Appendix C. There, the value for every cluster is a confound of average attribute importance and relative differences between the clusters. Instead, utilising utility factor scores avoids this confound of utility drivers with average attribute importance and allows a distillation of the main underlying behavioural differences between different consumer clusters.

From the cluster differences in utility factor scores it becomes clear that the first and fourth clusters are both driven by the factor ‘ease of trial’. Within this utility dimension the *first cluster* is more influenced by the availability of wine by the glass and half bottles, whereas the relatively small *fourth cluster* is more attracted by promotion cards on the table and menu suggestions. This need for external assistance in their wine choice of the fourth cluster is also reflected in the high loading of the third factor ‘restaurant advice’ and the lowest loading on ‘cognitive chooser’. Fairly opposite to the low risk food matching fourth cluster is the *second cluster*, which is driven by the negative loading attributes of the utility factors ‘ease of trial’ and ‘restaurant advice’. At the same time it reveals the highest loading on the ‘cognitive chooser’ utility dimension, indicating that suggestions from the table are not important for those consumers enjoy the cognitive task of choosing a wine. This second cluster has the highest importance of all to the region and varietal in wine selection, is able to match the wine to food and is highly influenced by prior experience. Finally, the *third cluster* is most of all looking for restaurant advice in the form of menu suggestions and waiter recommendations (high loading of third factor). Furthermore consumers of this cluster seek new experiences when choosing wine in an on-premise environment. Just as for the second cluster ease of trial and availability in small volumes is not important for this cluster (high negative loading of first factor).

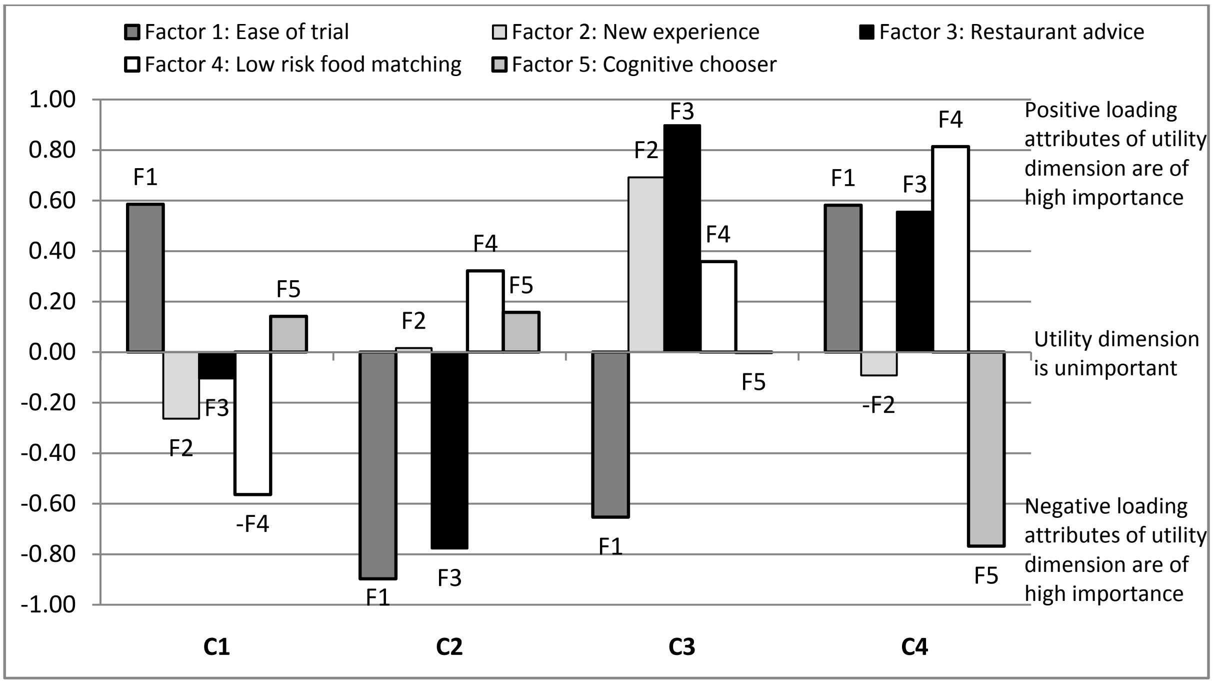

Figure 4: Comparison for loadings of utility components for preference clusters

Figure 4 visualises the differences in loadings of the five utility dimensions for all four clusters. Again it becomes clear that a visual representation of the five utility dimensions allows an easier interpretation and overview of behavioural differences between the clusters than an analysis of differences of all thirteen attributes (Appendix C), which are partly related and therefore contain repeated information (DeSarbo & Wu, 2001). Instead, utility factors as distilled underlying cognitive utility dimensions are much more powerful in explaining behavioural differences between the clusters.

To be able to locate and target these four on-premise wine consumer clusters we need to characterise them by wine and dining behaviour as well as by sociodemographic variables. Here we can only briefly outline major differences between the four clusters; the space restrictions of this conference paper deter us from a more detailed presentation of all distributions of the mostly categorical variables. Therefore, Table 5 only gives a broad overview on all wine and dining behaviour as well as sociodemographic variables which were found to be statistically different at p=0.05 for at least two of the four clusters.

Table 5: Overview of sociodemographic, wine and dine behaviour cluster differences

	C1	**C2**	**C3**	**C4**
	44%	24%	18%	13%
drink frequency	medium	medium	low	high
wine involvement	low	medium	high	low
dine involvement	low	low	high	high
café frequency	low	medium	medium	high
wine choice takes time	disagree	agree	agree	disagree
last wine purchased	more by glass	more by bottle	more by bottle	more by glass
price for wine at fine dining	lower	medium	higher	medium
gender	equal	more male	more male	more female
age	lower	higher	higher	lower

Cluster differences in respondents’ stated wine choice behaviour in Table 5 substantiate the findings of wine choice drivers derived from the Best Worst data and highly agree with the differences in utility dimensions and attribute importance in Table 4. The more convenience oriented first and fourth clusters are younger, have a lower wine involvement, disagree that their on-premise wine choice takes time and most often chose wine by the glass in their last on-premise wine purchase. Compared to the first cluster, consumers in the fourth cluster have a higher wine consumption frequency, a higher dining involvement, more often dine in café-style restaurants and show a higher willingness to pay for wine in fine dining situations. These consumers are more likely to be female.

The relative similarity of the second and third cluster in prior findings is again confirmed. Both clusters agree that their wine choice takes time, purchase wine rather by the bottle than by the glass, have a higher wine involvement and willingness to pay for wine. Consumers in these clusters are more likely to be older and male. Whereas wine consumers of the third cluster have the highest wine and dining involvement their wine drinking frequency is the lowest of all clusters.

Managerial Implications

Together with the differences in utility dimension these insights allow powerful implications for on-premise managers. For the majority of wine consumers (first and fourth cluster) the ease of trial is very important for their wine choice. Both clusters are either attracted by being able to purchase small volumes of wine or by being offered help in their choice in the form of (impersonal) suggestions in the menu or on the table. Only a third of consumers (third and fourth cluster) are looking for personal advice in the form of waiter recommendations. Restaurants targeting high-involved wine consumers should offer an interesting wine selection by the bottle, which offers the consumer a new experience. A clear statement of the varietal and region on the wine menu is important for these highly involved wine consumers who are also willing to pay higher wine prices. Promotion cards and offering wine in smaller volumes are less suited to target this group. Almost a fifth of UK on-premise wine consumers are highly involved and actively seeking personal waiter advice, which requires special trained and qualified staff to meet the needs of this consumer segment.

Conclusion

We have shown for Best Worst choice data how the variance-covariance matrix can provide important insights into what drives the behaviour of different consumers by finding underlying utility dimensions. Analysing the variance of Best Worst scores can distinguish those choice attributes, which are of similar important to all respondents (low variance) from those which vary in their importance between consumers (high variance). Attributes which show a rather high average score and a high variance are most important to focus on by marketing managers as they are major drivers of purchase behaviour for different consumer segments. Those attributes, which are strongly correlated jointly drive similar segments and load high on the same utility dimension. The distillation of utility dimensions out of the variance-covariance matrix allows a simpler and easier analysis of behavioural differences between consumer clusters than the analysis of differences in the choice attributes.

Best Worst Scaling has proven to be a reliable preference and attribute importance measurement method with high discriminative power. As Best Worst Scaling is a very simple method it is especially attractive for marketing practitioners. Our approach will allow marketing managers and researchers to retrieve more information from Best Worst Scaling data to take consumer heterogeneity into account. The visualisation of differences in underlying preference dimensions is a powerful tool for comparing differences segments and drawing managerial relevant implications.

Further Research

Future research should focus on two major avenues, the validation of the stability of the utility dimensions and a refinement of the method to derive the variance-covariance matrix.

We found five utility dimensions for on-premise wine choice for consumers in the UK and characterised behavioural differences between four consumer segments. Future research in different markets should validate the stability of these preference drivers.

The second focus of future research is to develop a more precise method to derive the variance-covariance matrix, which does not use aggregated but single choices and which can be applied to choices of concepts which combine several attributes and levels. Our analysis has focussed first on the variances for the scores for each attribute and has shown that these variances are both numerically large and of particular management relevance. The analysis has then focussed on the correlations between attributes and has shown that over all 13 attributes there is a correlation structure which also is numerically large and of managerial relevance. The method used here for calculating variances and correlations is well suited to this analysis and to derive the conclusions presented here, but it has limitations. It is a two-step process in which the Best Worst scores for each respondent are calculated and then the variances and correlations are calculated.

These two steps can be combined into a single process known as Qualitative Multinominal Distribution (QMD). This more 'accurate' method directly derives this matrix from individual choices, thus avoids one aggregation step. Further research should be undertaken demonstrating the application of the QMD to this data (Rungie, 2008). The first benefit of the QMD is greater statistical efficiency and accuracy. The second benefit relates to the design of the Best Worst experiment. In the data used here all respondents saw the same balanced 13 choice sets. If different respondents had been presented with different choice sets than the method used here would be less appropriate as individual Best Worst scores would reflect the choice set more than the respondents' preferences. In such cases QMD is still valid. It allows for variable choice sets; however this needs to be tested across various data sets.

References

Auger, P., Devinney, T., & Louviere, J. (2007). Using Best–Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries. *Journal of Business Ethics, 70*, 299-326.

Aurifeille, J.-M., Quester, P., Lockshin, L., & Spawton, T., (2002), "Global versus International Involvement Based Segmentation: A Cross National Exploratory Study", *International Marketing Review*, Vol 19 (4), 369-386.

Cohen, J., Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates, Inc.

Cohen, S., & Orme, B. (2004). What's your preference? *Marketing Research, 16*(2), 32-37.

Cohen, S., & Neira, L. (2003). *Measuring preference for product benefits across countries: Overcoming scale usage bias with Maximum Difference Scaling.* Paper presented at the ESOMAR 2003 Latin America Conference Proceedings, Uruguay.

DeSarbo, W. & Wu, J. (2001), The joint spatial representation of of multiple variable batteries collected in marketing research, *Journal of Marketing Research*, 38, 244-253.

Finn, A., & Louviere, J. J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. *Journal of Public Policy & Marketing, 11*(2), 12-25.

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics, 26*(1), 171-189.

Goodman, S., Lockshin, L., & Cohen, E. (2005). *Best-worst scaling : a simple method to determine drinks and wine style preferences*. Paper presented at the Second Annual International Wine Marketing Symposium, Sonoma State University, July 8-9, 2005.

Goodman, S., Lockshin, L., & Cohen, E. (2006, 6-8 July). *Using the Best-Worst method to examine market segments and identify different influences of consumer choice.* Paper presented at the 3rd International Wine Business and Marketing Research Conference, Montpellier.

Goodman, S., Lockshin, L., & Cohen, E. (2007). influencers of consumer choice in a retail setting - more international comparisons. *Wine Industry Journal, 22*(6), 42-48.

Goodman, S., Lockshin, L., & Cohen, E. (2008). Influencers of consumer choice in the on-premis environment: more international comparisons. *The Australian & New Zealand Grapegrower & Winemaker*(February 2008), 76-78.

Lockshin, Larry and John Hall. (2003). "Consumer Purchasing Behaviour for Wine: What We Know and Where We are Going", International Wine Marketing Colloquium, Adelaide, July, CD-ROM.

Lockshin, L., Jarvis, W., D'Hauteville, F. & Perrouty, J. (2006,) 'Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice', *Food Quality and Preference*, 17, 166-178.

Lockshin, L.S., Spawton, A.L. & Macintosh, G. (1997), "Using Product, Brand, and Purchasing Involvement for Retail Segmentation," *Journal of Retailing and Consumer Services*, Vol. 4 (3), 171-183.

Louviere, J., & Islam, T. (2007). A comparison of importance weights/measures derived from choice-based conjoint, constant sum scales and best-worst scaling. *Journal of Business Research, forthcoming.*

Luce, D. R. (1959), *Individual choice behaviour: a theoretical analysis*, Dover Publ., Mineola, New York.

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

Marley, A. A. J., & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology, 49*(6), 464-480.

Mueller, S., Francis, L., & Lockshin, L. (2007). Comparison of Best-Worst and Hedonic Scaling for the Measurement of Consumer Wine Preferences. Working paper Ehrenberg-Bass Institute for Marketing Science, University of South Australia.

Rungie, C. (2008). Repeated choice. School of Marketing, University of South Australia.

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

Wedel, M., & Kamakura, W. A. (1999). *Market segmentation : conceptual and methodological foundations* (2nd ed.). Boston: Kluwer Academic.

Appendix A:

Variance Covariance Matrix of Attributes (n=304)

	1 Alc <13%	2 waiter recom	3 food match	4 liked before	5 menu suggest	6 table suggest	7 avail glass	8 try different	9 varietal	10 region	11 promo card	12 half bottle	13 read not tasted
1 Alc <13%	3.06	-0.24	-0.34	-0.77	-0.37	-0.53	0.43	-0.15	-0.49	-0.49	-0.02	0.50	-0.59
2 waiter recom	-0.24	3.22	-0.65	-0.13	0.34	0.12	-0.80	-0.45	-0.40	-0.52	0.01	-0.42	-0.09
3 food match	-0.34	-0.65	3.56	0.34	-0.43	0.00	-0.58	-0.65	0.53	0.87	-1.50	-0.75	-0.40
4 liked before	-0.77	-0.13	0.34	2.68	0.07	0.37	-0.43	-0.36	-0.51	-0.44	-0.13	-0.59	-0.10
5 menu suggest	-0.37	0.34	-0.43	0.07	2.05	0.04	-0.49	-0.20	-0.21	-0.58	0.47	-0.58	-0.09
6 table sugg	-0.53	0.12	0.00	0.37	0.04	3.49	-0.90	-0.14	-0.71	-0.49	-0.48	-0.80	0.02
7 avail glass	0.43	-0.80	-0.58	-0.43	-0.49	-0.90	4.01	-0.57	-0.66	-1.02	0.57	1.40	-0.96
8 try different	-0.15	-0.45	-0.65	-0.36	-0.20	-0.14	-0.57	2.47	0.39	-0.22	-0.12	-0.43	0.43
9 varietal	-0.49	-0.40	0.53	-0.51	-0.21	-0.71	-0.66	0.39	2.81	0.50	-0.80	-0.59	0.14
10 region	-0.49	-0.52	0.87	-0.44	-0.58	-0.49	-1.02	-0.22	0.50	4.65	-1.46	-0.75	-0.05
11 promo card	-0.02	0.01	-1.50	-0.13	0.47	-0.48	0.57	-0.12	-0.80	-1.46	3.57	0.16	-0.28
12 half bottle	0.50	-0.42	-0.75	-0.59	-0.58	-0.80	1.40	-0.43	-0.59	-0.75	0.16	3.45	-0.61
13_read never tasted	-0.59	-0.09	-0.40	-0.10	-0.09	0.02	-0.96	0.43	0.14	-0.05	-0.28	-0.61	2.58

Appendix B:

Pearson Correlation Matrix of Attributes (n=304)

	1 Alc <13%	2 waiter recom	3 food match	4 liked before	5 menu suggest	6 table sugg.	7 avail glass	8 try different	9 varietal	10 region	11 promo card	12 half bottle	13 read not tasted
1 Alc <13%	1	-0.08	-0.10	-0.27 **	-0.15 *	-0.16 **	0.12 *	-0.05	-0.17 **	-0.13 *	-0.01	0.15 **	-0.21 **
2 waiter recom	-0.08	1	-0.19 **	-0.04	0.13 *	0.04	-0.22 **	-0.16 **	-0.13 *	-0.14 *	0.00	-0.13 *	-0.03
3 food match	-0.10	-0.19 **	1	0.11 *	-0.16 **	0.00	-0.15 **	-0.22 **	0.17 **	0.21 **	-0.42 **	-0.21 **	-0.13 *
4 liked before	-0.27 **	-0.04	0.11	1	0.03	0.12 *	-0.13 *	-0.14 *	-0.18 **	-0.12 *	-0.04	-0.19 **	-0.04
5 menu suggest	-0.15 *	0.13 *	-0.16 *	0.03	1	0.01	-0.17 **	-0.09	-0.09	-0.19 **	0.17 **	-0.22 **	-0.04
6 table sugg	-0.16 **	0.04	0.00	0.12 *	0.01	1	-0.24 **	-0.05	-0.23 **	-0.12 *	-0.14 *	-0.23 **	0.01
7 avail glass	0.12 *	-0.22 **	-0.15 **	-0.13 *	-0.17 **	-0.24 **	1	-0.18 **	-0.20 **	-0.24 **	0.15 **	0.38 **	-0.30 **
8 try different	-0.05	-0.16 **	-0.22 **	-0.14 *	-0.09	-0.05	-0.18 **	1	0.15 **	-0.07	-0.04	-0.15 *	0.17 **
9 varietal	-0.17 **	-0.13 *	0.17 **	-0.18 **	-0.09	-0.23 **	-0.20 **	0.15 **	1	0.14 *	-0.25 **	-0.19 **	0.05
10 region	-0.13 *	-0.14 *	0.21 **	-0.12 **	-0.19 *	-0.12 *	-0.24 **	-0.07	0.14 *	1	-0.36 **	-0.19 **	-0.02
11 promo card	-0.01	0.00	-0.42 **	-0.04	0.17 **	-0.14 *	0.15 *	-0.04	-0.25 **	-0.36 **	1	0.05	-0.09
12 half bottle	0.15 *	-0.13 *	-0.21 **	-0.19 **	-0.22 **	-0.23 **	0.38 **	-0.15 *	-0.19 **	-0.19 **	0.05	1	-0.20 **
13 read not tasted	-0.21 **	-0.03	-0.13 *	-0.04	-0.04	0.01	-0.30 **	0.17 **	0.05	-0.02	-0.09	-0.20 **	1

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed).

Appendix C:

Cluster means of attribute importance for Latent Class 4-Cluster solution

	C1		C2		C3		C4			
n	135		74		55		40			
	44%		24%		18%		13%		ANOVA	
Cluster means BWS attributes									Fvalue	p
Liked the wine before	1.67	a	2.81	b	2.42	b	3.83	c	25.39	0.00
I matched it to my food	0.24	a	2.68	b	1.89	c	0.78	a	42.14	0.00
Suggested by another on table	0.40	a	0.72	a	1.55	b	2.03	b	11.58	0.00
Try something different	0.41	ab	0.55	ab	1.13	b	0.00	a	4.57	0.00
Region	-0.44	a	1.38	b	1.16	b	-0.75	a	20.96	0.00
Read about it, never tasted	-0.28	a	0.19	ab	1.11	c	0.65	bc	12.17	0.00
Waiter recommended	0.03	a	-1.70	b	0.42	a	0.53	a	28.30	0.00
Menu suggestion	-0.21	a	-1.09	b	0.13	a	0.50	c	15.45	0.00
Varietal	-0.75	a	0.42	b	0.40	b	-1.58	c	22.47	0.00
Available by the glass	0.27	a	-0.82	b	-3.04	c	-1.33	b	57.22	0.00
Promotion card on table	-0.14	a	-2.41	b	-1.82	b	0.05	a	43.17	0.00
Half bottle available	-0.18	a	-0.74	a	-3.00	b	-2.28	b	56.60	0.00
Alcohol below 13%	-1.04	a	-1.97	b	-2.35	b	-2.43	b	13.27	0.00

Latent Cluster solution, LL = -7463.25, Classification R^2=0.89, Classification error=0.048
Different superscript letters: significantly different at p=0.05, post-hoc Tukey-test.