How do the relationships between variables impact behavioural loyalty? The Qualitative Multinomial Distribution (QMD) applied to wine consumption

Armando Maria Corsi[a], Leonardo Casini[a], Cam Rungie[b]

[a] University of Florence, Italy

armando.corsi@unifi.it; leonardo.casini@unifi.it

[b] Ehrenberg-Bass Institute for Marketing Science, University of South Australia, Adelaide

cam.rungie@unisa.edu.au

Abstract

Until nowadays, the studies on consumers' behavioural loyalty allowed the analysis of a brand, category or product attribute at a time. However, it has never been possible to observe and measure the relationships between brands/attributes from a loyalty perspective. Therefore, the paper aims at filling this gap presenting an innovative methodology, available through a new statistical distribution – called *Qualitative Multinomial Distribution* (QMD) – which detects the presence and the impact in terms of loyalty of interaction and correlation effects between brands categories and/or product attributes in both a partial and a comprehensive way. In this study, it has been shown that a comprehensive analysis offers a better representation of the connections between product attributes (or brand categories) than partial analyses. However, the QMD allows observing both approaches before deciding what the most suitable one is, thanks to a series of powerful diagnostic tools.

Keywords: Behavioural Loyalty, Qualitative Multinomial Distribution, Dirichlet Multinomial Distribution, Mixed Logit Models, Interaction Effects.

Topic area: Consumer buying and choice behaviour

1. Introduction and Literature Review

The studies of Lancaster (1966) have largely revolutionised the approach toward demand functions and the total utility consumers attach to a product or service. According to the Lancasterian approach, the demand of a good is defined as the sum of the values of the characteristics constituting the product/service, rather than to the product as a whole (Chan, 2005; Kym *et al.*, 2007). This definition lies on the assumption that the elements that form a good are perfectly substitutable between each other, thus making the estimated function additive. This approach has been applied to study several concepts. For example, Xu *et al.* (2009) observed that customer satisfaction can be explicitly measured as the overall satisfaction with respect to a combination of product attributes. Steele *et al.* (2008) stated that users' perceived quality is the result of the combination of product attributes that provide the greatest satisfaction to a specified user. Manaktola and Jauhari (2007) affirmed that consumers make choices according to the combination of product attributes that best meet the needs according to a series of values costs and prior satisfaction.

Assuming an additive model, however, implies that significant relationships between attributes do not exist (Caprihan and Wadwha, 2005). This assumption is not often true in reality, but it is generally difficult to confute it (Jaccard *et al.*, 1990). However, it has been shown that if model estimates are far from observed or expected values the additive model is invalid and, thus, alternative models should be taken into consideration. In particular, literature observed that one should look at models able to manage two kinds of relationships: interactions and correlations. Interaction occurs when the association between two variables varies according to the level of one or more covariates (Greenberg, 1985), hence showing the extent at which a factor impacts on the value of another. Conversely, correlation does not prove evidence of any causal relation between the variables, but it measures the strength of a linear or nonlinear relationship between them (Bharati and Chaudhury, 2004). In addition, it is important to stress that in case a study involves more than two variables, it is not sufficient to measure all the possible partial relations between them (two-way, three-way, etc.), but one should aim at finding a way to evaluate them all in one comprehensive analysis (Bai *et al.*, 2007), otherwise one always risks to make comments on incomplete results.

As shown, these concepts are already acquainted in the study of utility, satisfaction, perceived quality, etc., but they have never been applied to the study of customer behavioural loyalty. So far, in fact, it has only been possible to analyse loyalty levels of just one brand, category or product attribute (and their respective constituting elements) at a time, drawing from them conclusions about the overall behaviour of consumers towards goods or services (Ehrenberg *et al.*, 2004; Jarvis *et al.*, 2007a; b; Casini *et al.*, 2009). However, it has never been possible to observe and measure the relationships between brands/attributes from a loyalty perspective. Only East *et al.* (2005) discussed about the concepts of additivity and interaction in relation to loyalty. However, the meaning of these terms in that study did not refer to the relationships between variables, but on whether loyalty is singular; that is (i) only behavioural or attitudinal, or (ii) additive where total loyalty is a sum of the two components, or (iii) interactive with a kind of combination of the two.

Therefore, the paper aims at filling this gap presenting an innovative methodology, available through a new statistical distribution – called *Qualitative Multinomial Distribution* (QMD) – which allows analysing how interaction and correlation effects, if present, between brands categories and/or product attributes impact customer loyalty in both a partial and a comprehensive way. The

distribution will be applied on the purchases of wine[1] made in the three-year period 2003-2005 by a representative sample of the Italian population in the retail sector as reported in the AC Nielsen consumer panel. More specifically, given that it has been widely demonstrated that in a fragmented market, as that of wine, product attributes influence loyalty more than brands (Jarvis and Goodman, 2005), the analysis will be focused on three product attributes: formats, prices and quality designations.

The work is organized as follows. After this introduction and literature review, an explanation of the salient aspects of the methodology and data collection will be presented. Then, results will be analysed and discussed, while conclusions close up the paper.

2. Methodology

Given the innovativeness of the model, it is only possible to refer to a work of Rungie (2007) in order to explain the methodology applied in this research.

When one wants to study regularities and variations over a population of consumers for the analysis of consumer behaviour and discrete choices, a useful way to conduct these studies is to apply *ad hoc* statistical distributions. Two of them are *Multinomial Logit* (MNL) and the *Dirichlet Multinomial Distribution* (DMD), as they both present unique and useful properties. The former is able to accommodate variable choice sets and deconstruct choice into utilities and partworths. The latter, the DMD, is considered the multivariate extension of a *Beta Binomial Distribution* (BBD), which is applicable when consumers make repeated choices from the same binary choice set. The DMD conceptualizes each choice by each individual as a Bernoulli trial and his/her repeated choice as a multinomial trial based on a fixed latent conditional choice probability for the consumer. Over the population, these probabilities have a Dirichlet distribution. Hence, the DMD is a multinomial mixed by a Dirichlet. This feature is unique in choice modelling and the main outcome is that the properties of the repeated choice (DMD) and the latent conditional choice probabilities (Dirichlet) can be estimated from the data. This means that variances are known, thus the DMD allows identifying reliability, partitioning the variance and establishing behavioural loyalty levels. This distribution, however, presents two limitations. As implicit assumptions of the DMD (i) only brands (or any one attribute) can be evaluated and (ii) there is no variation in the underlying loyalty toward the levels assumed by the brands (or the levels of the attribute). This explains the necessity to apply the BBD to each single item *j*, so as to analyse the loyalty expressed as a binary choice between it and all the other items in the choice set.

Therefore, MNL and DMD have extraordinary properties, but each has some that the other has not. Hence, a new distribution is needed (a) to let the MNL identify reliability, separately measure the between and within consumer variance and analyse the impact of choice sets and attributes on loyalty and (b) to bring to the DMD variable choice sets and remove the undesirable implicit constraint on loyalty. In addition, this new distribution uses repeated choice for the identification of the structure of partworths in a way similar to structural equation modelling. This new distribution is called *Qualitative Multinomial Distribution* (QMD).

The main assumptions are summarised here: (i) repeated bivariate binary choice, (ii) the functional form which links variable partworths to latent conditional choice probabilities via variable choice sets is logit, (iii) the sequence of choices for each consumer is independent and the

[1] We refer to all the purchases of Italian and foreign wines with an EAN code made by the AC Nielsen consumer panel. These data do not include non-EAN wines and the following categories: champagne, marsala, sherry, port, grape must, wine based aromatized beverages, sangria, aromatized wines, natural sparkling wines, fortified wines, and the *spumante* category. Hence, the data represents for AC Nielsen, the 71% in volume and the 78% in value of the wine purchased in Italy (average 2005-2009) in the retail channel.

partworths stationary, (iv) the partworths have a Gaussian distribution and (v) the interaction partworths are independent. These assumptions are specifically derived for a bivariate binary choice, but it can easily be transformed in its multivariate counterpart.

Due to space limitations, it is not possible to go into the mathematical details of the QMD, but we suggest interested readers to look at Rungie (2007) for more information about it.

3. Data & Analysis

Data have been gathered from the AC Nielsen Italian consumer panel. The sample comprises 5299 Italian households, whose wine purchases in the retail sector have been registered for the three-year period 2003-2005. A further subsample has been extracted from them, in order to include only those households with somewhat regular purchase behaviour. The subsamples include the families who (1) bought wine on more than one occasion in each of the two three-year periods and (2) bought more than 10 units of wine.

QMD could be applied in a multivariate way, but, given the innovativeness of the methodology, researchers decided to apply the QMD in a bivariate binomial way, hence analysing two attributes with two levels at a time. Regarding prices, it has been decided to consider two price ranges – '<€3' and '>€3' – as the basic tier (< €3) accounts alone for a 68.2% of the market (IRI Infoscan, 2007). Segmentation according to denominations classified wine based on the presence/absence a quality designation. Thus, wines have catalogued as 'GI-DOC-DOCG' vs. 'NOT GI-DOC-DOCG'. In relation to formats wines in 'up to 0.75litres bottles' were grouped together. The other group has been organised in order to account for the sales of 1litre carton wines and larger formats, including 3litre bag-in-box wines.

For the purposes of the study, authors first analysed the partial relationships, from now onwards called '2x2 analysis', between the three attributes and levels (priceXquality designation, priceXformat, quality designationXformat), then the comprehensive analysis, from now onwards called '2x2x2 analysis' (priceXquality designationXformat) was run.

The QMD model estimates the parameters of four variants of the QMD. The 'base model' only shows the means and the variance of the attributes, but it does not include interaction or correlation effects. Conversely, the 'full model' takes into account both effects, hence it presents 7 parameters – means and variance of the two attributes, means and variance of the interaction effect and the parameter relative to the correlation effect.

The second step is looking at the means (μ_1, μ_2, μ_3) – which represent the partworth utility consumers show when moving from one level of the attribute to the other – and the standard deviations (σ_1, σ_2, σ_3) – which (a) express the extent at which consumers differ in the utility they attach to each attribute and (b) tell us what attribute drives more loyalty (Rungie, 2007) – of the 2x2 and the 2x2x2 analysis.

The means (μ_4, μ_5, μ_6) and the standard deviations (σ_4, σ_5, σ_6) of the interaction section and the correlation coefficient (ρ) tell us if there is a positive or negative interaction/correlation effect when the attributes are combined together.

Once defined the nature of the relationship between attributes and levels, it is then possible to look at the loyalty levels that singular attributes and levels – in case the model suggest the absence of interaction or correlation effects – or combinations of them – in case the QMD detects some kinds of relationships – stimulate on consumers.

4. Results

The preliminary analysis of the results shows a general presence of interaction effects among attribute levels in determining the overall level of consumers' loyalty (table 1). The observation of the log likelihood tests for each of the three 2x2 analyses shows that the model able to explain at best what kind of relationship exists among attributes levels is the one which includes the presence of interaction effects, but excludes the presence of correlation. This means that for any of the 2x2 combinations, the levels which determine an increase in consumers utility (which also means an increase in customer loyalty) should be present together in order to generate the highest loyalty levels. In particular, as one can observe in table 2 (μ_4, μ_5, μ_6 - σ_4, σ_5, σ_6), price and format turned up to show the strongest interaction (1.5771-0.8061 and 0.8759-0.6540), followed by price and quality designation (1.0842-0.7192 and 0.8081-0.1984), and format and quality designation (0.9319-0.9223 and 0.4017-0.5884).

Tab. 1: Log-likelihood values of the four variants of the QMD

ATTRIBUTES	ALTERNATIVES	MODEL	NO. OF PARAMETERS	LOG LIKELIHOOD
PRICE x FORMAT	Base	No Correlations; No interaction	4	237005.1560
	Correlation	Main correlation, No interaction	5	235023.5508
	Interaction	No correlation; Interaction	6	195920.2753
	Full	Main correlation, Interaction	7	195915.5720
PRICE x DESIGN.	Base	No Correlations; No interaction	4	259292.1317
	Correlation	Main correlation, No interaction	5	258067.2106
	Interaction	No correlation; Interaction	6	234671.7133
	Full	Main correlation, Interaction	7	234630.5650
DESIGN. x FORMAT	Base	No Correlations; No interaction	4	275402.0889
	Correlation	Main correlation, No interaction	5	274406.2504
	Interaction	No correlation; Interaction	6	244204.0983
	Full	Main correlation, Interaction	7	244202.5184

If we now look at the means (μ_1, μ_2, μ_3) of both 2x2 and 2x2x2 analyses (table 2), we observe that consumers show a positive utility by moving from a higher to a lower price when price is modelled together with format (1.4104), quality designation (1.2792) and in the 2x2x2 analysis (1.6726). Similarly, panel members attach higher utility to GI-DOC-DOCG wines when quality designation interacts with price (0.6201), format (0.1030) and in the 2x2x2 analysis (0.6736).

So far, these results seem to demonstrate that there is no point in conducting a 2x2x2 analysis, as already with 2x2 analyses, the results are all consistent. However, authors would like to focus on formats. One notes a decrease in utility passing from a smaller to a bigger format (-0.6303), when the latter is analyzed together with price. At the same time, one observes an increase in utility (0.3281) when the latter is analyzed together with denomination. These results appear to be in contradiction. However, when one looks at the mean value of format in the 2x2x2 analysis, one sees a value of -0.0096, which is approximately equal to 0. In such a situation the 2x2x2 gives more information than those we obtain by studying format with a 2x2 analysis. The 2x2x2 suggests that the two levels really divide the population in two segments, one loyal to smaller formats and the other loyal to bigger ones. The 2x2 analysis gives some clue on it, by showing two opposing results, but due to some latent component (that the 2x2 is not able to catch), we do not exactly know what to conclude from these results. Only the 2x2x2 allows having a clear picture of the preferences of consumers in respect to the format chosen from a loyalty perspective.

Moreover, the values of standard deviations (σ_1, σ_2, σ_3) for the 2x2 analyses show a situation where format (1.48) drives loyalty more than quality designation (1.02), price (1.24) dominates over

quality designation (0.95), and format (1.34) is stronger than price (1.07). According to a transitive property, we could be lead to affirm that the order of attribute "strength" in driving loyalty is format>price>quality designation. However, the results of the 2x2x2 analysis (σ_4, σ_5, σ_6) show a different ranking, with price (1.18) leading over format (1.06) and quality designation (0.84). This once again proves that only the 2x2x2 analysis is able in our case to describe how product attributes are really related between each others.

Tab. 2: Parameters estimates[2]

TABLE SUMMARY		ATTRIBUTES						INTERACTION					
		PRICE		FORMAT		DESIGNATION		PRICE		FORMAT		DESIGNATION	
		μ1	σ1	μ2	σ2	μ3	σ3	μ4	σ4	μ5	σ4	μ6	σ6
2 x 2	**PRICE** HIGH to LOW			1.4104	1.0754	1.2792	1.2429			1.5771	0.8061	1.0842	0.7192
	FORMAT SMALL TO BIG	-0.6303	1.3489			0.3281	1.4848	1.5771	0.8061			0.9319	0.9223
	DESIGNATION TABLE TO GI-DOC-DOCG	0.6201	0.9512	0.1030	1.0272			1.0842	0.7192	0.9319	0.9223		
2 x 2 x 2	**PRICE** HIGH to LOW			1.6726	1.1876	1.6726	1.1876			0.8759	0.6540	0.8081	0.1984
	FORMAT SMALL TO BIG	-0.0096	1.0662			-0.0096	1.0662	0.8759	0.6540			0.4017	0.5884
	DESIGNATION TABLE TO GI-DOC-DOCG	0.6736	0.8432	0.6736	0.8432			0.8081	0.1984	0.4017	0.5884		

Finally, once shown that the 2x2x2 analysis is able to describe at best the nature of the relationships between attributes and levels in determining consumers' loyalty, we can now look at the loyalty values that the eight product combinations, are able to generate. These values should be interpreted exactly as those of the polarization index (φ_c), when applied to single brand categories or product attributes (Rungie, 2007). Hence, a value of 0 means total disloyalty, while the closer the index is to 1 the higher is the loyalty. This clarification helps us understanding (table 3) that two combinations are able to stimulate the highest loyalty values. On the one side we find (#1) more expansive wines (≥€3), with a quality designation (GI, DOC or DOCG) sold in 0.75litres bottles (φ_c=0.5204); on the other (#2), cheaper (≤€3) table wines sold in bigger formats, which totalize a φ_c of 0.5304. However, it is also interesting that alternatives #5, #6 and #7 reaches a loyalty value on average higher than #2, #3 and #4. This could be probably related to the strength of price, in particular of lower price (≤€3), in driving loyalty more than the other two attributes.

Tab. 3: Loyalty values of the eight product combinations

	Alternative	Description	Polarization
			φ_j
2 x 2 x 2	1	≥€3 - GI/DOC/DOCG - ≤0.75l	0.5204
	2	≥€3 - GI/DOC/DOCG - >0.75l	0.2574
	3	≥€3 - NOT GI/DOC/DOCG - ≤0.75l	0.1936
	4	≥€3 - NOT GI/DOC/DOCG - >0.75l	0.1048
	5	<€3 - GI/DOC/DOCG - ≤0.75l	0.3339
	6	<€3 - GI/DOC/DOCG - >0.75l	0.4019
	7	<€3 - NOT GI/DOC/DOCG - ≤0.75l	0.3668
	8	<€3 - NOT GI/DOC/DOCG - >0.75l	0.5305

[2] The correlation section is missing, as results suggest taking into account a model, which only includes interaction.

It is important to remember that it is not always the case that a comprehensive (2x2x2) analysis offers a better representation of the connections between product attributes (or brand categories) than partial analyses (2x2s). However, the QMD allows observing both approaches before deciding what the most suitable one is, thanks to a series of powerful diagnostic tools.

5. Conclusions

The present research showed a new statistical distribution, called *Qualitative Multinomial Distribution* (QMD), which is able to analyse the relationships (none, interaction, correlation or a mix of the two) between product attributes, in respect to the loyalty levels that the latter are able to generate.

In the specific case, the model able to explain at best the relationships between the three attributes under analysis – price, format and quality designation – is the one which includes the presence of interaction, but excludes that of correlation. Moreover, it was also noted that the three partial (2x2s) analyses were not able to give a clear picture of the relationships between the attributes, although they offered some information and clues on them. As a consequence, a comprehensive analysis (2x2x2) was run on the product attributes. In this way, it was possible to have a clearer picture of the combinations, which drive loyalty at most.

Future researches will repeat the same analysis to the data relative to the three-year period 2006-2008, in order to see how the relationships between product attributes changed over time. Secondly, it is interesting to apply the QMD to the study of loyalty towards combination of brands and brand categories. Finally, researchers will try to extend the bivariate-binomial analysis to its multivariate counterpart, so as to have a more sensible representation of reality.

6. References

Bai, J., Wahl, T. I., and Wandschneider, P. R. (2007), "Valuing Attributes of Fluid Milk Using Choice-Based Conjoint Experimental Design", Paper presented at the American Agricultural Economics Association Annual Meeting, Portland, Oregon (USA).

Bharati, P., and Chaudhury (2004), "An empirical investigation of decision-making satisfaction in web-based decision support systems", *Decision Support Systems*, Vol. 37 No. 2, pp. 187-197.

Caprihan, R., and Wadhwa, S. (2005), "Scheduling of FMSs with information delays: a simulation study", *International Journal of Flexible Manufacturing Systems*, Vol. 17 No. 1, pp. 39-65.

Casini, L., Rungie, C., and Corsi, A. M. (2009), "How loyal are Italian consumers to wine attributes?", *Journal of Wine Research*, In press.

Chan, T. Y. (2006), "Estimating a continuous hedonic-choice model with an application to demand for soft drinks", *Journal of Economics*, Vol. 37 No. 2, pp. 466-482.

East, R., Gendall, P., Hammond, K., and Lomax, W. (2005), "Consumer loyalty: singular, additive or interactive?", *Australasian Marketing Journal*, Vol. 13 No. 2, pp. 10-26.

Ehrenberg, A. S. C., Uncles, M. D., and Goodhardt, G. (2004). "Understanding brand performance: using Dirichlet benchmarks", *Journal of Business Research*, Vol. 57, pp. 1307-1325.

Greenberg, R. S., and Kleinbawn, D. G. (1985), "Mathematical Modeling Strategies for the Analysis of Epidemiologic Research", *Annual Review Public Health*, pp. 223-245.

Jaccard, J., Wan, C. K., and Turrisi, R. (1990), "The detection and interpretation of interaction effects between continuous variables in multiple regression", *Multivariate Behavioural Research*, Vol. 25 No. 4, pp. 467-478.

Jarvis, W., and Goodman, S. (2005), "Effective marketing of small brands: niche positioning, attribute loyalty and direct marketing", *Journal of Product & Brand Management*, Vol. 14 No. 5, pp. 292-299.

Jarvis, W., Rungie, C., and Lockshin, L. (2007a), "Revealed preference analysis of red wine attributes using polarization", *International Journal of Wine Business Research*, Vol. 19 No. 2, pp. 127-138.

Jarvis, W., Rungie, C., and Lockshin, L. (2007b), "The polarization method for merging data files and analysing loyalty to product attributes, process and brand in revealed preference", *International Journal of Market Research*, Vol. 49 No. 4, pp. 489-513.

Kim, J., Allenby, G. M., and Rossi, P. E. (2007), "Product attributes and models of multiple discreteness", *Journal of Econometrics*, Vol. 138, pp. 208-230.

Lancaster, K. J. (1966), "A new approach to consumer theory", *Journal of Political Economy*, Vol. 74 No. 2, pp. 132-157.

Manaktola, K., and Jauhari, V. (2007), "Exploring consumer attitude and behaviour towards green practices in the lodging industry in India", *International Journal of Contemporary Hospitality Management*, Vol. 19 No. 5, pp. 364-377.

Rungie, C. (2007), Repeated Choice, Under Review.

Steele, C., Jones, R., and Leaney, P. (2008), "Improved tennis ball design: incorporating mechanical and psychological influences", *Journal of Engineering Design*, Vol. 19 No. 3, pp. 269–284

Xu, Q., Jiao, R. J., Yang, X., Helander, M., Halimahtun, M. K., and Opperud, A. (2009), "An analytical Kano model for customer need analysis", *Design Studies*, Vol. 30 No. 1, pp. 87-110.