Revealed Preference Attribute Modelling using Repeated Purchases

Cam Rungie, Gilles Laurent, Nadhem Mtimet, Wade Jarvis

Multinomial logit (MNL) models are routinely used in marketing to estimate the impact of attribute levels on the potential success of new alternatives in a market. In the paper the MNL model is extended, using the Dirichlet Multinomial Distribution (DMD), to analyse the attributes involved in the patterns in repeated purchase revealed preference data. The result is a model for which MNL is a special case and which estimates the impact of attribute levels on market share and loyalty. The success of a new alternative can now be estimated in terms of market share and loyalty simultaneously. The paper uses a basic model which demonstrates the process but is too simple to generate accurate or practical estimates. This is a limitation of the specific model and not the process.

Keywords: Multinomial Logit, Dirichlet,

DMD for Brands

Consider a period of time, such as a year, a population of shoppers and a product category. For each shopper the count of the purchases of each brand in the category is recorded. This is the repeat purchase pattern for the shopper. Over the population of shoppers this pattern has a statistical distribution. The Dirichlet Multinomial Distribution (DMD) has been successfully fitted empirically to this distribution (Jeuland, Bass & Wright 1980; Goodhardt, Ehrenberg & Chatfield 1984; Ehrenberg 1988; Uncles, Ehrenberg & Hammond 1995; Ehrenberg, Uncles & Goodhardt 2004).

The DMD has one parameters for each brand. If there are h brands then there are h parameters $\alpha_1, \alpha_2, \ldots, \alpha_h$ (Johnson, Kotz & Balakrishnan 1997).

Let the mean proportion of purchases for brand j be μ_j. Under some models this is taken an indication of the Market Share for the brand.

Let $S = \alpha_1 + \alpha_2 + \ldots + \alpha_h$. The properties of the DMD include $\mu_j = \alpha_j/S$.

Generalized Linear Model

In this paper, the DMD is extended to model the distributions of brands and of several attributes in the category, simultaneously. Covariates are introduced to the DMD creating a General Linear Model which (1) has multinomial logit as a special case and (2) identifies the location of utility.

The DMD is a conditional distribution. It accounts for the distribution of preferences for the brands conditional on the category purchase rate. Thus, in extending the DMD, the relative preference for brands and attributes are being modelled. The utility of the category is not being modelled.

The DMD can be converted to a general linear model (GLM) (McCullagh & Nelder 1989)through the use of the exponential linear link function (Rungie, Laurent & Stern 2003; Rungie & Laurent 2004). Let there be covariates $X_1, X_2, \ldots, X_m$.

Let: $\alpha_j = \exp(\beta_0 + \beta_{j,1}X_1 + \ldots + \beta_{j,m}X_m)$.
Thus: $\mu_j = \exp(\beta_0 + \beta_{j,1}X_1 + \ldots + \beta_{j,m}X_m)/\Sigma(\exp(\beta_0 + \beta_{i,1}X_1 + \ldots + \beta_{i,m}X_m))$.

which is the logit function (McFadden 1974; Ben-Akiva & Lerman 1984; Louviere, Hensher & Swait 1999).

Table 1. Effects codes for the attribute levels and the estimates of the codes for the DMD model

Attribute	Attribute Level	Effects Code	Parameter Estimate DMD Model	Effects Code Estimate DMD Model
Constant		β_0	-4.4963	-4.4963
Market	Top 10	$-(\beta_{1,2}+\beta_{1,3}+\beta_{1,4})$		0.2814
Share	11 to 30	$\beta_{1,2}$	0.1213	0.1213
of Brand	31 to 50	$\beta_{1,3}$	-0.5369	-0.5369
	51 and over	$\beta_{1,4}$	0.1341	0.1341
Price	$7.49 and under	$\beta_{2,1}$	-1.4326	-1.4326
Group	$7.50 to $12.49	$-(\beta_{2,1}+\beta_{2,3}+\beta_{2,4})$		0.2997
	$12.50 to $17.49	$\beta_{2,3}$	0.5307	0.5307
	$17.50 and over	$\beta_{2,4}$	0.6022	0.6022
Region of	Australia or State Only	$-(\beta_{3,2}+\beta_{3,3}+\beta_{3,4})$		0.2238
Origin	Australian specific, high aware	$\beta_{3,2}$	0.7462	0.7462
	Australian specific, low aware	$\beta_{3,3}$	0.0339	0.0339
	Foreign	$\beta_{3,4}$	-1.0039	-1.0039
Grape	Cabernet	$\beta_{4,1}$	-0.1703	-0.1703
Variety	Shiraz	$\beta_{4,2}$	0.1265	0.1265
	Cabernet Shiraz blend	$-(\beta_{4,1}+\beta_{4,2}+\beta_{4,4})$		0.0135
	Other	$\beta_{4,4}$	0.0304	0.0304

The contrast between the multinomial logit (MNL) and the DMD is substantial when repeated purchase data is available. For each shopper MNL provides one probability for each purchase and these probabilities are combined as if they are from different shoppers. By comparison the DMD provides one probability for the shopper's complete set of purchase counts for all brands. This creates a new functional form for the likelihood function. If there is only one purchase per shopper, then the functional forms for DMD and MNL are identical. However, where there are multiple purchases for each shopper, then the likelihood based on the DMD has its own functional form which has one very special property. The constant is now far more identified than for MNL. An example is given below.

Example

The category was red wine. The period of time was one year. Number of shoppers = 2036. Shoppers were only selected if they had a purchase rate of 10 or more. Total purchases = 85902 bottles (min=10, Q1=16, M=26, Q3=49 max=1066).

The many SKUs in the category were recoded into 4 attributes (1) size of market share for the brand (2) price (3) region of origin and (4) grape variety. Each attribute had four levels. Thus there were 4x4x4x4 = 256 concepts in the universal choice set.

Contrasting dummy variables were used for each attribute level, see

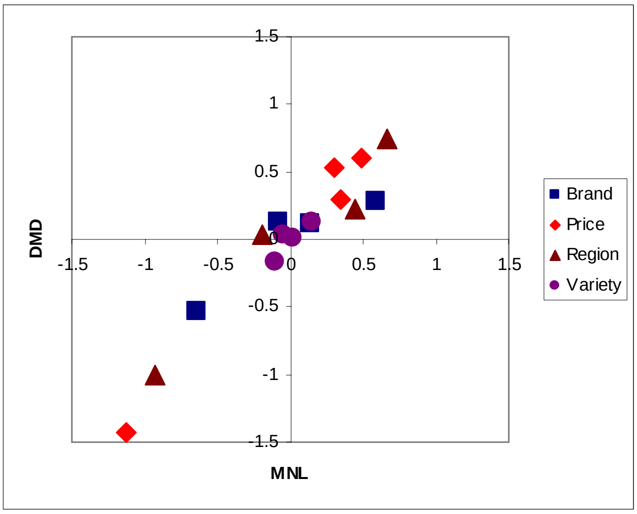
Table 1. The levels were contrasted against the concept with the greatest market share, which was, a top ten brand, cabernet shiraz blend, in the $7.50 to $12.49 price bracket and labelled as being from a region of origin which was Australia or State specific only.

Two models were fitted (1) traditional MNL and (2) DMD. Figure 1 shows the plots of the effects codes from MNL and DMD. (Also the estimates for the DMD model are given in

Table 1.) It can be seen from the figure that the two models generated very similar estimates. The conclusion is that replacing MNL with DMD did not lead to radically different results.

Figure 2 shows the plot of the loglikelihood function for the DMD model as the constant was varied. This is a standard procedure for examining the potential for a likelihood function to achieve a global maximum. All parameters are held constant and one parameter, in this case the constant, is allowed to vary. As there is only one parameter changing it is possible to plot the result with the parameter on the horizontal axis and the loglikelihood function on the vertical axis. In a model with many parameters a global maximum in such a plot (where only one parameter varies) is not conclusive evidence of a global maximum over all parameters but it is useful evidence. Thus, while Figure 2 is not 'proof' that the constant is identified it is a good demonstration.

Figure 1 . While changing from the traditional MNL to the DMD model similar results are still generated for the estimates of the effects of each attribute level.

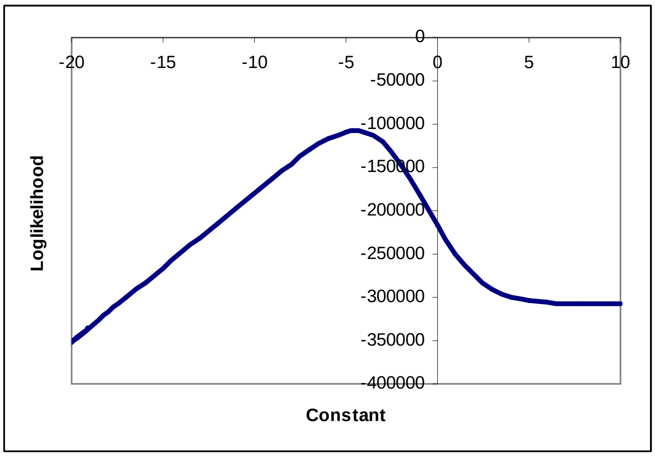


Repeat Rate

The DMD also identifies loyalty. Consider the original DMD model where there were h brands. Let the Repeat Rate for brand j be ρ_j. The repeat rate is the probability of a randomly selected shopper choosing the brand at the next purchase occasion given the shopper chose the brand at the last purchase occasion. A purchase occasion is an event where the shopper purchases from the product category. The Repeat Rate is a measure of loyalty. It shows the extent to which the brand holds onto its buyers (Rungie & Laurent 2003).

The properties of the DMD include $\rho_j= (1+\alpha_j)/(1+S)$. Thus, the repeat rate is identified.

Figure 2. The peak in the log likelihood function demonstrates that the constant in the model is identified.



In the general linear model, because the constant is identified, then the repeat rates are also identified. For example consider the concept:

Attribute	Level	Description	Coefficients
		Constant	-4.5
Brand	1	The group of top 10 brands	0.28
Price	2	\$7.50 to \$12.49	0.30
Region	1	Australia or State Only	0.22
Variety	3	Cabernet Shiraz blend	0.01
		Total: Metric Utility	-3.68
		Alpha: exp(Metric Utility)	0.025

Over all 256 concepts the sum of the alphas was $S = 4.57$. The estimated market share for this concept was $\mu = 0.025/4.57 =$ 0.55%. The estimated repeat rate was $\rho = (1+0.025)/(1+4.57) =$ 18.4%.

It has been demonstrated that the model can be used to estimate the mean and the repeat rates for specific product concepts. The model simultaneously estimates the impact of attribute levels on market share and loyalty. This potentially valuable contribution to marketing is not possible with the traditional MNL model because the constant (-4.5) is not identified. With the DMD model the constant is identified.

Marketing Outcome

A very simple demonstration of fitting the DMD to four product attributes has been given. In modelling terms the outcome has been that the constant is identified; i.e. the location of utility is identified. The practical outcome is that the model now reports the impact of attributes on the repat rate. The MNL model traditionally has reported the impact of attributes on market share. The DMD model also reports the impact on market shares and loyalty.

Discussion

This has been an overly simple model created to demonstrate the process and the outcomes.

The design, with four attributes (factors) and four levels per attribute, created a universal choice set of 256 concepts (alternatives). The model assumed the choice set was always all 256 concepts. For the process demonstrated here this assumption was not necessary. As further research, an evaluation of competing models with less than 256 alternatives in the choice set is being undertaken.

The MNL and DMD models presented here examine main effects only. Interaction effects should be examined for, and certainly can be expected in this wine data. The DMD model presented above had a poor fit for the market shares and the repeat rates. The correlations between the observed and estimated values for the market shares was 38% and for the repeat rates was 25%. Some of this misfit is very likely to be due to the absence of interaction effects from the model. Further analysis of the data is being undertaken.

The DMD is outstanding in modelling repeated purchases, in identifying the location of utility and in estimating the impact of attributes on repeat rates and loyalty. BUT, it is the wrong distribution for sets of product attributes. The DMD contains an overly restrictive assumption regarding covariance; the correlations between the utilities for pairs of attribute levels. This has the consequence of imposing one underlying level of loyalty on all attributes and all levels. As a result while repeat rates are identified they are always just a linear function of the market share. This is very likely to explain some of the misfit of the model in estimating market shares and repeat rates. The basic DMD model is too simplistic. A new less restrictive set of distributions is required. They are entirely achievable and are under development. The DMD has been useful in demonstrating the process for identifying the location of utilities and modelling loyalty. It has closed forms which are easy to discuss. The new distributions which replace the DMD will be more complex, particularly in that they will not have closed forms and will require numerical approximations. Generally, as with the DMD, identification of the effects codes will still be possible.

Conclusion

Repeated revealed preference data has been used to identify the location of utility. This leads the way to models which measure the impact of product attributes on repeat rates, which are a fundamental measure of loyalty. The traditional multinomial logit model, which estimate the impact of attributes on market share, has been expanded to also estimate the impact on loyalty. The basic DMD model presented here has been too simplistic to generate accurate results but it has demonstrated the potential of the process to contribute to discrete choice models and to marketing.

References

Ben-Akiva M & Lerman SR (1984). *Discrete Choice Analysis: Theory and Application to Travel Demand.* London, The MIT Press.

Ehrenberg ASC (1988). *Repeat-Buying, Facts, Theory and Applications.* New York, New York: Oxford University Press.

Ehrenberg, ASC, Uncles MD & Goodhardt GG (2004). Understanding Brand Performance Measures: Using Dirichlet Benchmarks. *Journal of Business Research* (forthcoming).

Goodhardt G J, Ehrenberg ASC & Chatfield C (1984). The Dirichlet: A Comprehensive Model of Buying Behaviour. *Journal of the Royal Statistical Society, Series A (General)* 147(Part 5), 621-655.

Jeuland AP, Bass FM & Wright GP(1980). *A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes. Operations Research* 28 (2).

Johnson NL, Kotz S & Balakrishnan N (1997). *Discrete Multivariate Distributions*. New York, John Wiley & Sons, Inc.

Louviere J J, Hensher DA & Swait J (1999). Stated Choice Methods Analysis and Application. Cambridge, Cambridge University Press.

McCullagh P &. Nelder JA (1989). Generalized Linear Models. London, Chapman & Hall/CRC.

McFadden D (1974). *Conditional Logit Analysis of Qualitative Choice Behaviour. Frontiers in Econometrics*. P. Zarembka. New York, Academic Press: 105-142.

Rungie CM & Laurent G (2003). *Repeated Binary Logit (RBL): Fitting Explanatory Variables to the Beta Binomial Distribution*. Adelaide, School of Marketing, University of South Australia.

Rungie CM & Laurent G (2004). *Research Note: Statistical Summary of the Generalized Dirichlet Model*. Adelaide, School of Marketing, University of South Australia.

Rungie CM, Laurent G & Stern P (2003). *Modeling Long-Run Dynamic Markets*. Adelaide, School of Marketing, University of South Australia.

Uncles MD, Ehrenberg ASC &. Hammond K (1995). Patterns of Buyer Behavior: Regularities, Models and Extensions. *Marketing Science* 14(3, 2 of 2): G71-G78.

Cam Rungie is a Senior Lecturer, School of Marketing, University of South Australia,

Gilles Laurent is the Carrefour Professor of Marketing, Marketing Department, HEC School of Management, 78350 Jouy-en-Josas, France,

Nadhem Mtimet is in the Unidad de Economia Agraria, Centro de Investigacion y Tecnologia Agroalimentaria de Aragon – CITA, and

Wade Jarvis is a Lecturer in the School of Marketing, University of South Australia