Working Paper
n°: 2013-61-02

Integrating consumer characteristics into the stochastic modelling of purchase loyalty

Cam Rungie [a], Mark Uncles [b], Gilles Laurent [c]

a University of South Australia, GPO Box 2471, Adelaide SA 5001, Australia
b University of New South Wales, Sydney NSW 2052, Australia
c INSEEC Business School, 27 avenue Claude Vellefaux 75010 Paris, France

January 2013

An ulterior version of this article appeared in European Journal of Marketing, vol 47, Issue 10, October 2013, ISSN 0309-0566

Structured Abstract

Purpose

Many marketing datasets record repeated purchase/consumption by consumers. Measures for analyzing such data are in widespread use, including penetration/reach, average purchase frequency, sole buying, share of category requirements, and repeat rate – collectively these are known as Brand Performance Measures (BPMs). The measures are used to identify levels of behavioral loyalty, concentration, consistency over time and switching in a market. Most datasets also record substantial information on the characteristics of consumers, including demographics, geodemographics, psychographics and much more. These consumer characteristics are often used to identify consumer segments for targeting and resource allocation purposes. The paradox is that there has been no general method for analyzing both types of information at the same time.

Approach

This paper describes a model for predicting BPMs, covariates are then introduced into the model, with discussion of model specification, model estimation, overall model assessment, and the derivation of generalized theoretical BPMs. The outcome is a practical procedure for behavioral loyalty segmentation.

Findings

The implications for strategy and management in applying covariates to the BPMs are considerable. Where there are concentrations of consumers with high repeated purchase/consumption then many aspects of the marketing mix will be affected. An investigation of the role of covariates in understanding BPMs in the laundry detergent market is presented as an example and ways for market analysts to display results are demonstrated.

Originality/Value

Despite the fact that BPMs are the best operationalization of the concept of behavioral loyalty, until now there has not been a model to evaluate the impact of consumer characteristics as covariates on these BPMs. This paper's original contribution starts with the mathematical specifications for the BPMs. A model is created which fits covariates to the BPMs. New statistical and graphical methods are described.

Key words: behavioral loyalty; revealed preference; repeat purchase; repeat consumption; repeat choice; brand performance measures; covariates; segmentation; stochastic models; Generalized Dirichlet model; marketing research

1. Introduction

Over the years, marketing managers have developed many Brand Performance Measures (BPMs) based on panel data collected by the market research industry. BPMs include penetration/reach, average purchase frequency, sole buying, share of category requirements, and repeat rate. BPMs permit important strategy and management diagnoses. For example, typically strategic growth for a brand can be achieved by attracting new customers rather than selling more to existing customers - building penetration rather than average purchase frequency. Or, the degree of concentration of brand sales on a limited number of consumers might influence the selection of media and communications strategies.

Traditionally, each BPM is computed separately, on the basis of raw panel data. Our first contribution is to show that a dozen of the most popular BPMs used by managers can be interrelated through a single, well-validated statistical model, the Dirichlet. Thus, to evaluate the complete set of BPMs, there is only a need to estimate the model parameters, which is feasible through a practical procedure. At little cost, managers can build a comprehensive scorecard of BPMs, rather than limiting themselves to just one or two of them.

Even more importantly, we show how to take advantage of the considerable background information recorded by panel operators on consumer characteristics such as demographics, geodemographics, attitudinal responses, psychographics, stated preference discrete choice, etc. We show how to estimate the impact of those consumer characteristics on the Dirichlet model parameters, and therefore their impact on each BPM.

Our procedure, and the managerial implications, should be applicable in any category where data are used to record repeated purchase/consumption by consumers (e.g., packaged goods, grocery stores, fast foods, media consumption, transportation, telecommunications, blogging and online purchasing).

We first review BPMs, giving a precise definition of each. We then describe the well-known Dirichlet model, and show how each BPM can be estimated on the basis of the parameters of the Dirichlet. The next step is to introduce the Generalized Dirichlet model, in which covariates (here, consumer characteristics) impact those parameters. We then discuss in detail an empirical application to the buying of laundry detergent. Before concluding, we position our new model against its precursors.

2. Brand Performance Measures (BPMs)

2.1 Constructs

Managers often analyze markets with two aggregate constructs for each brand: total sales and market share. By contrast, BPMs draw upon data recording each consumer's repeated choices, permitting a deeper and more intuitive understanding of market phenomena. This enables answers to be provided for questions such as: what proportion of the population are buyers, what are their purchase frequencies, how much of each consumer's share of category requirement is met by a brand? All these constructs relate to repeated choice and are collectively known as behavioral loyalty because they reflect the extent to which purchases are concentrated in segments of loyal and consistent consumers (Jacoby and Chestnut, 1978; Ehrenberg, 1988).

2.2 Measures

For each BPM, we distinguish the observed sample statistic (calculated directly from a dataset) and the underlying population parameter (which is not directly observable, but is a component of the statistical distributions for the population). In the simple case of the mean of a random variable X, the sample statistic $\bar{x}$ is the observed average of the sample, while the population parameter μ is the unobserved expected value $E[X]$ of X.

The formula for calculating each BPM sample statistic from data is easily determined from the definitions and is computed in practice by suppliers such as MarketingScan, Nielsen, GfK and TNS. In contrast, the precise functional form for the underlying parameter of each BPM, and most importantly the relationships between the functional forms of the different BPMs, have not been established in the marketing literature. The first contribution of this paper is to identify these functional forms and relationships. This is important for two reasons: first, it allows managers to see that there is structural unity in what may seem at first sight to be a collection of independent measures; second, it makes possible the statistical analysis of the impact of consumer characteristics on the parameters of the functional forms, and therefore on the BPMs. The incomplete specification of the functional forms for the parameters has held back the development of statistical models for the BPMs.

To proceed we need precise notation. There are h brands in the product category. We analyze a fixed period of time, such as a year. Over the population of consumers, let there be

h random variables, $R_1, R_2, \ldots, R_h$, which are the purchases, by each consumer, of each brand. Over the same population of consumers, the total purchases of the category by each consumer is another random variable, the sum of purchases over all brands ($K = \sum_{j=i}^{h} R_j$). All these random variables, R and K, are counts and so are non-negative integers.

We can now specify, for each BPM, its precise definition, including the definition of its population parameter.

Consumers (households, shoppers, panelists, respondents, decision makers, etc.)

This is the unit for which the purchase data is recorded. We consider all consumers who are potential buyers (e.g., all panel members), regardless of whether or not they buy. A consumer can have an individual purchase rate of zero for the product category.

Category Buyer and Brand Buyer

A category buyer is a consumer who makes, over the period considered, at least one purchase from the product category. A buyer of brand j is a consumer who makes, over the period considered, at least one purchase of brand j. By definition, a buyer of a brand is a buyer of the category. A buyer of a brand may or may not be a buyer of another brand.

Purchase Rate

The number of purchases by the consumer over the specified time period. Each consumer has a separate purchase rate for each brand. The consumer's category purchase rate is the sum of her/his brand purchase rates.

For consumer i

brand purchase rates are $r_{i,1}, r_{i,2}, \ldots, r_{i,h}$

category purchase rate is k_i

Over the population of consumers r and k are random variables

brand purchase rates are $R_1, R_2, \ldots, R_h$

category purchase rate is K

where $K = \sum_{j=1}^{h} R_j$

Average Purchase Rate for the Category

The purchase rate of the category averaged over all consumers.

Population parameter E[K]

Penetration/Reach for the Category

The proportion of consumers who are buyers of the category.

Population parameter Pr{K>0}

Average Purchase Frequency for the Category

The average purchase rate for the category among buyers of the category.

Population parameter $E[K] / \Pr\{K>0\}$

Average Purchase Rate for Brand j

The purchase rate of the brand averaged over all consumers.

Population parameter $E[R_j]$

Where $$E[K] = \sum_{j=1}^{h} E[R_j]$$

Market Share for Brand j

Total purchases of the brand as a proportion of the total purchases of the category.

Population parameter $E[R_j] / E[K]$

Penetration/Reach for Brand j

The proportion of consumers who are buyers of the brand.

Population parameter $\Pr\{R_j>0\}$

Relative Penetration/Reach for Brand j

Penetration of the brand as a proportion of the penetration for the category.

Population parameter $\Pr\{R_j>0\} / \Pr\{K>0\}$

Average Purchase Frequency for Brand j

The average purchase rate of the brand among buyers of the brand.

Population parameter $E[R_j] / \Pr\{R_j >0\}$

Sole Buyers (100% Loyals) for Brand j

The proportion of buyers of the brand who buy only that brand (including consumers who only buy once).

Population parameter $\Pr\{R_j = K \mid K>0\}$

Share of Category Requirements (SCR) for Brand j

Defined as:

$$\frac{\text{Total Purchases of brand } j}{\text{Total Purchases of the category by buyers of brand } j}$$

Population parameter $E[R_j] / E[K \mid R_j >0]$

Average Portfolio Size for the Category

Average number of brands purchased by buyers of the category.

Population parameter $$\frac{\sum_{j=1}^{h} \Pr\{R_j > 0\}}{\Pr\{K > 0\}}$$

Repeat Rate for Brand j

The proportion of buyers of a brand at the last purchase occasion who repurchase the same brand at the next purchase occasion. (A purchase occasion is when the consumer makes a purchase from the category.)

Population parameter $\Pr\{R_j = 2 \mid K=2\} / \Pr\{ R_j = 1 \mid K=1\}$

This list of BPMs is comprehensive, but it does not purport to be exhaustive (see Ehrenberg, 1988 and Ehrenberg et al., 2004 for additional BPMs).

2.3 Distributions

In the marketing literature there is a long history of applying probability distributions to the random variables we have just discussed. Early work applied the negative binomial distribution to the number of purchases of a brand or of the category (Ehrenberg, 1959; Schmittlein et al., 1985). The beta binomial distribution was applied to the binary choice, on each purchase occasion, between one brand and all others (Chatfield et al., 1966; Chatfield and Goodhardt, 1970) and to similar choices in the domain of audience measurement (Sabavala and Morrison, 1977). Regular patterns were observed in the choices between brands in a single product category (Bass et al., 1976; Ehrenberg, 1988). Models were then developed which combined probability distributions for the number of purchases in the category with probability distributions for the choice among brands, conditional on purchases being made. One approach applied an Erlang distribution to the number of purchases and a beta binomial distribution to the choice among brands (Jeuland et al., 1980). The Dirichlet multinomial distribution, a multivariate version of the beta binomial distribution, was used to extend the results from the beta binomial distribution and to capture the previously observed regular patterns in the choices among brands. The Dirichlet model combined the negative binomial distribution for the number K of purchases in the category with the Dirichlet multinomial distribution for the choices $R_1, R_2, \ldots, R_h$ among brands, conditional on K, leading to a single multivariate distribution for $R_1, R_2, \ldots, R_h$ (Goodhardt et al., 1984; Uncles et al., 1995; Ehrenberg et al., 2004; Wright, Sharp and Sharp, 2002). We discuss this model in the next section.

3. The Dirichlet Model

The Dirichlet model has been fully described previously and we do not repeat here its derivation (Goodhardt et al., 1984) and assumptions (Dunn et al. 1983; Schmittlein et al.1985). All we do is provide enough information for the differences between the Dirichlet and the Generalized Dirichlet to be appreciated.

(a) Model Specification. The Dirichlet model specifies a probability density function for the distribution of purchases by a population of consumers of each of the brands, within a product category, over a specified period of time. The category might be, for example, laundry detergents; the period of time a year; the population all consumers (or households) in

France; and the brands the set of about twenty brands of detergents available in French supermarkets. The data in this example would record for every household in the sample the purchases of each of the brands over the year. The data are multivariate, in that there are several brands, and discrete, integer and nonnegative because we count purchases.

The model combines two probability density functions. The category purchase rate K is assumed to have a negative binomial distribution (NBD). Purchases of individual brands $R_1, R_2, \ldots, R_h$ are assumed to have a Dirichlet multinomial distribution (DMD), conditional on the category purchase rate, K, having a specific value k. The two distributions, the NBD and the DMD, are assumed to be otherwise independent. Most importantly, the model then specifies the theoretical values of the BPMs from the combined NBD/DMD distribution, which have been shown to fit well with the values observed from the data. The model was originally developed to reflect empirical generalizations that were identified in the patterns of the observed BPMs (e.g., double jeopardy, Ehrenberg et al., 1990).

(b) Probability Density Functions. As indicated above, the Dirichlet model combines a negative binomial distribution (NBD) for the category purchase rate, K, and a Dirichlet multinomial distribution (DMD) for the brand purchase rates, $R_1, R_2, \ldots, R_h$, conditional on K.

The NBD has two parameters (both positive): the shape parameter, γ, and the scale parameter (which influences the shape), β . The shape parameter is also a heterogeneity parameter which describes the variation in purchases of the category across consumers. The probability density function for the NBD is (Johnson et al., 1993):

(1) $$f_{\gamma,\beta}(k) = \frac{\Gamma(\gamma+k)}{\Gamma(\gamma)k!}\frac{\beta^k}{(1+\beta)^{(\gamma+k)}} \quad \text{for } k = 0,1,2,...$$

The DMD has h parameters, one for each brand. These parameters are $\alpha_1, \alpha_2, \ldots, \alpha_h$ where each is positive. The probability density function for the DMD is (Johnson et al., 1997):

(2) $$f_{\alpha_1,\alpha_2,\ldots,\alpha_h}(r_1,r_2,\ldots,r_h \mid r_1+r_2+\ldots+r_h=k) = \frac{\Gamma\left(\sum_{j=1}^{h}\alpha_j\right)k!}{\Gamma\left(\sum_{j=1}^{h}\alpha_j+k\right)}\prod_{j=1}^{h}\frac{\Gamma(\alpha_j+r_j)}{r_j!\Gamma(\alpha_j)}$$

The sum of the α parameters $S = \Sigma\alpha$ is a heterogeneity parameter which describes the variations in the preference for brands across consumers.

The Dirichlet model brings together these two probability density functions creating a single probability density function which specifies the purchases of all brands in a product category over a period of time, this is given by:

$$f_{\gamma,\beta,\alpha_1,\alpha_2,\ldots,\alpha_h}(r_1, r_2, \ldots, r_h) = f_{\gamma,\beta}(k) f_{\alpha_1,\alpha_2,\ldots,\alpha_h}(r_1, r_2, \ldots, r_h \mid r_1 + r_2 + \ldots + r_h = k) \quad (3)$$

There are $h + 1$ parameters: γ, β, α_1, α_2, ... , α_h. The number of brands in a product category, h, is usually of the order of 20 to 40, but can be as low as two or as high as 150. Empirical generalizations have been identified in the range of values for the parameters of the Dirichlet model. It is rare for the individual α for any brand to be greater than one and values are often in the range of 0.7 down to 0.01 and less. Together with market share, the sum of the α parameters over all brands – the S parameter – determines repeat purchase rates (Fader and Schmittlein, 1993; Sharp, Wright and Goodhardt, 2002). The value of the shape parameter, γ, is often of the order of 0.5 to 1, but can be higher or lower. It is higher for staple product categories which are purchased by most consumers and where relatively few rates are close to zero (e.g., as expected for laundry detergent). The value of the scale parameter, β, is proportional to the length of the time period over which the data is collected and can be of the order of 0.1 to 50 (Driesener, 2005).

(c) Formulae for the Theoretical BPMs. Based on the assumptions of the NBD and DMD in the Dirichlet model, we can derive the population parameters for the BPMs; i.e. the theoretical estimates of the BPMs as functions of γ, β, α_1, α_2, ... , α_h.

Average Purchase Rate for the Category

$$= \gamma\beta$$

Penetration/Reach for the Category

$$= 1 - \frac{1}{(1+\beta)^{\gamma}}$$

Purchase Frequency for the Category

$$= \frac{\text{Average Purchase Rate for the category}}{\text{Penetration for the category}} = \frac{\gamma\beta}{1 - \frac{1}{(1+\beta)^{\gamma}}}$$

Dirichlet S

$$S = \sum_{j=1}^{h} \alpha_j$$

Average Purchase Rate for Brand j

$$= \frac{\alpha_j \gamma\beta}{S}$$

Market Share for Brand j

$$\mu_j = \frac{\alpha_j}{S}$$

Penetration/Reach for Brand j

$$= \sum_{k=1}^{\infty} f_{\gamma,\beta}(k)\left(1 - \frac{\Gamma(S)\Gamma(S-\alpha_j+k)}{\Gamma(S+k)\Gamma(S-\alpha_j)}\right)$$

The expression $f_{\gamma,\beta}(k)$ is the probability density function for the NBD in Equation (1).

Purchase Frequency for Brand j

$$= \frac{\text{Average Purchase Rate for brand } j}{\text{Penetration for brand } j}$$

Sole Buyers (100% Loyals) for Brand j

$$= \frac{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k)\left(\frac{\Gamma(S)\Gamma(\alpha_j+k)}{\Gamma(S+k)\Gamma(\alpha_j)}\right)}{\text{Penetration for brand } j}$$

Share of Category Requirements (SCR) for Brand j

$$= \frac{\text{Average Purchase Rate for brand } j}{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k)k\left(1 - \frac{\Gamma(S)\Gamma(S-\alpha_j+k)}{\Gamma(S+k)\Gamma(S-\alpha_j)}\right)}$$

Average Portfolio Size for the Category

$$= \frac{\sum_{j=1}^{h} \text{Penetration for brand } j}{\text{Penetration for the category}}$$

Repeat Rate for Brand j

$$\rho_j = \frac{\alpha_j + 1}{S + 1}$$

Formulae for penetration/reach, sole buyers and share of category requirements specify infinite summations that can be numerically approximated with finite summations (Rungie and Goodhardt, 2004). For large values of k the expression within the summation is close to zero and gets even closer as k increases further. Thus, continuing the summation to higher values of k makes no change in the estimate of the BPM. We recommend starting the summation at k=1 and stopping when $\sum f_{\gamma,\beta}(k)$>.99999. Often this is achieved quickly.

(d) Model Estimation and the Likelihood Function. Let there be a sample of n consumers. Let $r_{1,i}$, $r_{2,i}$, ... ,$r_{h,i}$ be the purchases of consumer i and let k_i be the sum of these purchases over the h brands. Then, equations (1) and (2) each lead to the likelihood functions in Equations (4) and (5).

Log-likelihood function for the NBD:

$$LL_{\gamma,\beta} = \sum_{i=1}^{n} \Big\{ \ln(\Gamma(\gamma + k_i)) - \ln(\Gamma(\gamma)) - \ln(\Gamma(k_i + 1)) + k_i \ln(\beta) - (\gamma + k_i)\ln(1+\beta) \Big\} \tag{4}$$

Log-likelihood function for the DMD:

$$LL_{\alpha_1,\alpha_2,\ldots,\alpha_h} = \sum_{i=1}^{n} \Bigg\{ \ln\left(\Gamma\left(\sum_j \alpha_j\right)\right) + \ln(\Gamma(k_i+1)) - \ln\left(\Gamma\left(\sum_j \alpha_j + k_i\right)\right) + \sum_{j=1}^{h} \ln(\Gamma(\alpha_j + r_{i,j})) - \ln(\Gamma(r_{i,j}+1)) - \ln(\Gamma(\alpha_j)) \Bigg\} \tag{5}$$

As in (3) the joint probability density function is the product of the two separate density functions for the NBD and DMD. Similarly the joint log-likelihood function is the sum of the two separate log-likelihoods: $LL_{\gamma,\beta} + LL_{\alpha_1,\alpha_2,\ldots,\alpha_h}$. For a given data set the parameters estimates are those values which maximize the joint log-likelihood. However, as there are no parameters in common to both log-likelihood functions, each can be maximized separately. Thus, the distribution parameters γ, β can be estimated from equation (4) and α_j for $j = 1,2,\ldots h$ from equation (5). These log-likelihood functions can be maximized using the Nelder-Mead (1965) optimizer (Rungie, 2003) although Bayesian estimation could also be developed (Arora et al., 1998; Wong, 1998).

4 The Generalized Dirichlet Model

Unlike the Dirichlet model, our specification and use of the Generalized Dirichlet is new.

(a) Covariates. Most databases recording repeated choice also include considerable background information on the consumer: demographics, geo-demographics, psychographics, attitudes, stated preferences, as well as purchases of complimentary and substitute products, or consumption of a lifestyle nature such as usage of green products, computer technologies, and health and fitness products. Apart from the literature on distributions for repeated choice discussed above, over the last thirty years another literature has developed for the theory, models and analysis of discrete choice, notably random utility theory and multinomial logit models (McFadden, 1974; Ben-Akiva and Lerman, 1985; Louviere et al., 2000; Hensher et al., 2005). These models incorporate the characteristics of consumers as covariates. They focus on conditional choice – the selection if a choice is to be made – but not on the quantity of selections the consumer decides to make. Consequently they do not represent repeated choice in a manner which leads to estimates of the BPMs. Logit determines the relationship between many covariates and market share but cannot do the same for the BPMs which reflect quantity. The alternative approach is to extend the Dirichlet model to incorporate covariates, as we do here.

(b) Model Specification. We extend the Dirichlet model to include covariates by creating a generalized linear model (GLM) (Wrigley and Dunn, 1985; McCullagh and Nelder, 1989). Let there be a vector of covariates $\mathbf{x}$. We express each of the distribution parameters γ, β and α_j, for j = 1 to h, as a function of the covariates. Here, we use the linear exponential link function.

NBD component of the model:

(6) $$\gamma(\mathbf{x}) = \exp(\boldsymbol{\psi}'\mathbf{x}) \qquad \beta(\mathbf{x}) = \exp(\boldsymbol{\varphi}'\mathbf{x})$$

DMD component of the model:

(7) $$\alpha_j(\mathbf{x}) = \exp(\boldsymbol{\theta}'_j\mathbf{x}) \quad \text{for each brand } j$$

Thus, there are several vectors of model parameters $\boldsymbol{\psi}$, $\boldsymbol{\varphi}$, and $\boldsymbol{\theta}_j$ for j = 1 to h. The number of elements in each of these vectors is the same as the number of covariates in the vector **x.**

Equation (6) assumes that we use linear combinations of the covariates. Broadening the model to other link functions is not problematic as long as the distribution parameters γ, β and α_j, for j = 1 to h, all remain positive.

Substituting Equation (6) in Equation (1) gives the Generalized NBD:

$$(8) \qquad f_{\boldsymbol{\psi},\boldsymbol{\varphi}}(k \mid \mathbf{x}) = \frac{\Gamma(\exp(\boldsymbol{\psi}'\mathbf{x}) + k)}{\Gamma(\exp(\boldsymbol{\psi}'\mathbf{x}))k!} \frac{\exp(\boldsymbol{\varphi}'\mathbf{x})^{(k)}}{(1 + \exp(\boldsymbol{\varphi}'\mathbf{x}))^{(\exp(\boldsymbol{\psi}'\mathbf{x}) + k)}}$$

Substituting Equation (7) in Equations (2) gives the Generalized DMD:

$$(9) \qquad f_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_h}\left(\mathbf{r} \mid \mathbf{x} \text{ and } \sum_{j=1}^{h} r_j = k\right) = \frac{\Gamma\left(\sum_j \exp(\boldsymbol{\theta}'_j\mathbf{x})\right)k!}{\Gamma\left(\sum_j \exp(\boldsymbol{\theta}'_j\mathbf{x}) + k\right)} \prod_j \frac{\Gamma(\exp(\boldsymbol{\theta}'_j\mathbf{x}) + r_j)}{r_j!\,\Gamma(\exp(\boldsymbol{\theta}'_j\mathbf{x}))}$$

Combining Equations (8) and (9) gives the Generalized Dirichlet Model

$$(10) \qquad f_{\boldsymbol{\psi},\boldsymbol{\varphi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_h}(\mathbf{r} \mid \mathbf{x}) = f_{\boldsymbol{\psi},\boldsymbol{\varphi}}(k \mid \mathbf{x}) f_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_h}\left(\mathbf{r} \mid \mathbf{x} \text{ and} \sum_{j=1}^{h} r_j = k\right)$$

Equation (10) is the Generalized Dirichlet Model. We have replaced the distribution parameters, γ, β and α_j with vectors of model parameters, $\boldsymbol{\psi}$, $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}_j$ for j = 1 to h. We have introduced covariates into the model through **x**.

(c) Model Estimation and the Likelihood Functions. Let $\mathbf{x}_i$ be the observed values for the vector of covariates for consumer i. Equations (8) and **(9)** each lead to the log-likelihood functions in Equations (11) and (12).

Log-likelihood function for the Generalized NBD:

(11)
$$LL_{\boldsymbol{\psi},\boldsymbol{\varphi}} = \sum_{i=1}^{n} \Big\{ \ln(\Gamma(\exp(\boldsymbol{\psi}'\mathbf{x}_i) + k_i)) - \ln(\Gamma(\exp(\boldsymbol{\psi}'\mathbf{x}_i))) - \ln(\Gamma(k_i + 1)) + k_i \ln(\exp(\boldsymbol{\varphi}'\mathbf{x}_i)) - (\exp(\boldsymbol{\psi}'\mathbf{x}_i) + k_i)\ln(1 + \exp(\boldsymbol{\varphi}'\mathbf{x}_i)) \Big\}$$

Log-likelihood function for the Generalized DMD:

(12)
$$LL_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_h} = \sum_{i=1}^{n} \Bigg\{ \ln\left(\Gamma\left(\sum_j \exp(\boldsymbol{\theta}'_j\mathbf{x}_i)\right)\right) + \ln(\Gamma(k_i + 1)) - \ln\left(\Gamma\left(\sum_j \exp(\boldsymbol{\theta}'_j\mathbf{x}_i) + k_i\right)\right) + \sum_{j=1}^{h} \ln(\Gamma(\exp(\boldsymbol{\theta}'_j\mathbf{x}_i) + r_{i,j})) - \ln(\Gamma(r_{i,j} + 1)) - \ln(\Gamma(\exp(\boldsymbol{\theta}'_j\mathbf{x}_i))) \Bigg\}$$

The log-likelihood function for the full model is the sum of equations (11) and (12). Again, as there are no parameters in common to both log-likelihood functions, each can be maximized separately. Thus, the vectors of model parameters $\boldsymbol{\psi},\boldsymbol{\varphi}$ can be estimated from equation (11) and $\boldsymbol{\theta}_j$ for $j = 1,2,...h$ from equation (12). The estimates ate the values that maximize the log-likelihood function. For models of this nature the shape of the function can be such that in a local area there is a maximum, less than the global maximum, which may misdirect the optimizing algorithms leading to inappropriate estimates. To minimize this risk, the log-likelihood functions should be maximized more than once using different start points (i.e. different initial values).

(d) Overall Model Assessment. In the usual manner with likelihood theory we can generate likelihood-ratio tests for each of the covariates (Edwards, 1976; Eliason, 1993). There are effectively two models within the Generalized Dirichlet Model, the Generalized NBD and the Generalized DMD. Thus, there are two likelihood-ratio tests for each covariate. The null hypothesis is that the covariate has no impact on the model (i.e., the model parameter for the covariate is zero). If a covariate is significant for one component there is no necessity it will be significant for the other component; covariates significant in the Generalized NBD relate to category purchase rates, whereas covariates significant in the Generalized DMD relate to brand preference. For instance, household size might be significant in terms of category purchase rates of laundry detergent, but not at all significant with respect to choice of a specific brand such as Ariel or Omo. Conceivably, if covariates are to be included then demographic factors may influence the category purchase/NBD side more (Ehrenberg et al.,

2004) and attitudinal brand preference may influence the brand choice/DMD side more (Horsky et al., 2006).

(e) Covariates for the BPMs. Substituting Equation (6) into the formulae for the theoretical BPMs gives rise to generalized functional forms in which, in each case, the BPM is expressed as a function of the covariates.

(f) Tabulation and Graphing of the Effects of Covariates. We suggest the following procedures for examining the impact of the covariates. From the several covariates in the model, select one at a time as the focus for the analysis of the outputs. Let it vary and examine how the model's estimates of the BPMs change. We emphasize that this is after the model parameters have been estimated. The procedures described here all use the one set of estimates for the unconstrained model parameters. We are varying the covariates and watching the changes in the estimates of the BPMs.

Two calculation procedures are used: (i) marginal effects (i.e., as the covariate of interest varies, the other covariates are held constant, at their average value), and (ii) goodness-of-fit estimates (i.e., as the covariate of interest varies, the other covariates also vary exactly matching the patterns and collinearity in the data – this is achieved using the data themselves: the model is applied to the observed values of the covariates for each respondent separately and the estimates of the BPMs averaged over the sample of respondents).

Let us emphasize the difference between the two procedures. The goodness-of-fit estimates are a check. They capture and reflect collinearity in the covariates and should closely estimate the observed BPMs if the model fits the data well. By contrast, the marginal effects remove the indirect influences of the other covariates in the data generated by the collinearity and show the underlying direct impact of the covariate.

In these two procedures the remaining issue is to determine how much the covariate of interest should vary. Options are: (i) by one standard deviation, (ii) by a unit value of one (useful for dummy variables which take on a value of zero and one), or (iii) over the range in the data. These options allow the impact of the covariate to be examined under similar but slightly different conditions.

In the example below results are presented in both tabular and graphical form. The tabular results show the marginal effect, where the covariate of interest has been varied by one standard deviation (from the mean to the mean + one standard deviation). The graphs show three results: (i) the marginal effect over the range of the covariate, (ii) the observed BPM for each segment of the covariate, and (iii) the goodness-of-fit estimate for each segment of the covariate.

5. An Application: The Purchasing of Brands of Laundry Detergent

5.1 Goals and Data

The modeling procedure described above opens up a substantial area of marketing analysis. In broad terms, a product might create its sales through a large number of consumers choosing it occasionally or through a small concentration of consumers choosing it very frequently and consistently. Total sales and market shares will be the same in these two cases, but the structure of the market will be very different and many aspects of strategy will be influenced. A brand manager in this situation will need to distinguish between promotional activities that attract more customers and activities that result in selling more to existing customers. The analysis here compares three BPMs that capture these contrasting options: penetration/reach, average purchase frequency, and share of category requirements.

Data were supplied by MarketingScan and GfK. Purchases of all brands in the laundry detergent category were recorded for 3439 households for one year in France. The information is typical of the revealed preference data described earlier and that is now routinely available to analysts, manufacturers and retailers. For illustrative purposes, we focus on the top five national brands (with all remaining brands grouped into a super-brand labeled 'Others') and particular attention is given to the case of Le Chat.

Market shares and the three BPMs of interest are shown in Table 1, together with Dirichlet theoreticals (without covariates). Brands are arranged by market share enabling the analyst to see that penetrations are lower for brands with smaller market shares, as are SCRs (in keeping with the double jeopardy effect).

[Insert Table 1 about here]

5.2 Covariates

Four covariates are considered: (i) household-size (number of persons in the household), (ii) expenditure (average spend in supermarkets by the household each week), (iii) visits (average number of visits to supermarkets by members of the household each week), and (iv) loyalty (loyalty to favorite supermarket – SCR of favorite supermarket over a year). Typical values for each covariate are shown in Table 2.

[Insert Table 2 about here]

Potentially, all four covariates might influence the selection of promotional activities. When consumers spend more in supermarkets it may be because they have high disposable income or because they have a larger household and a greater requirement for staples like laundry detergent. Generally, the former group can be expected to be less price elastic and the latter group more elastic and more likely to react to price promotions. The model combines category purchases and the selection of brands when a purchase occurs and so elasticity can be relative to total purchases and to conditional switching. Certainly with a staple product category like detergents the quantity purchased is more likely to vary with household size than with income but the preference for brands may follow a more complex pattern; some brands might be preferred by higher income households and some by larger households.

There was considerable collinearity between the covariates as is shown in Table 3. This is to be expected. For example, the correlation between expenditure and household size was high (0.48). Some consumers who spend more do so because they come from larger households, and therefore any analysis of expenditure should also analyze household size. The existence of collinearity highlights the importance of including all four covariates simultaneously in the multivariate analysis and not analyzing each separately.

[Insert Table 3 about here]

5.3 BPMs – with Covariates

(a) Overall Model Assessment. We fitted the Generalized Dirichlet Model to the data. The iterative process for maximizing the log-likelihood functions described above gives a log-likelihood value for the unconstrained model, with all covariates, and then one log-likelihood value for each constrained model in which only one covariate is omitted. For each covariate this creates two likelihood-ratio tests, one examining its contribution to the NBD (via equation (11)) and the other to the DMD (via equation (12)). Recall, the NBD side of the model analyses the impact of the covariate on the category purchase rate, and the DMD side analyzes the impact on the probabilities for each of the brands being selected when a purchase is made.

From equation (6) in the NBD, there are two parameters for each covariate and so two degrees of freedom. From equation (7) in the DMD, there is one parameter per brand for each covariate and so the degrees of freedom equate to the number of brands; i.e. 6 (5 national brands, plus ‘Others’).

The results for the likelihood-ratio tests are summarized in Table 4. All covariates are significant in the NBD and all but one is significant in the DMD (Tables 4(a) and 4(b) respectively). It is not unexpected that covariates may be significant in one part of the model and not the other as a basic assumption of the model is that the brand choice is independent of the purchase incidence for the category (an assumption which in further research can be tested by including purchase incidence as a covariate in the DMD). Specifically, we see that one covariate, visits, is significant in the NBD but not the DMD, which indicates that those who shop more often (i.e., those who record a higher number of visits) also buy more of the category, but their brand choice probabilities do not differ from those who shop less often.

[Insert Table 4 about here]

The significance tests focus on the two components and two likelihood functions for the model. However, the results we discuss below combine these two components. We now examine the impact of the covariates on the BPMs.

(b) Tabulation of the Marginal Effects of Covariates. For each covariate we standardize the impact by examining how the BPMs change as the covariate varies from its mean value to the mean + one standard deviation. Table 5(a) shows the impact on penetration: depicted is the theoretical penetration from the standard Dirichlet model and then the variation as the covariate changes. Typically, a covariate will vary over the sample by plus or minus two standard deviations. Thus, the table gives a conservative indication of the impact of the covariate. Equivalent results for average purchase frequency and share of category requirements are shown in Tables 5(b) and 5(c).

[Insert Table 5 about here]

(c) Graphs of Marginal Effects of the Covariates. Plots of the marginal effect of each of the covariates on the BPMs for Le Chat are shown in Figure 1. These plots show that: as household size increases so penetration is constant and SCR falls; and as household weekly spend increases so penetration increases and SCR falls dramatically.

(d) Goodness-of-Fit Estimates. We also show in Figure 1 the average observed results, as 'Observed' dots, which are useful in assessing the goodness-of-fit of the model. This approach is only possible because we recoded all the covariates to be discrete. For each covariate, we divided the sample into sub-samples according to the value of the discrete covariate. For each sub-sample it was then possible to present two additional results: (i) the observed BPMs were calculated for each sub-sample and graphed as 'Observed' dots, and (ii) the model was fitted to each member of the sub-sample individually and the average fitted value for the theoretical BPM was calculated for the sub-sample. We graphed this as the 'Goodness-of-fit Estimate' line, as described earlier. The approach provides a graphical picture of the fit of the model. If the model fits the data well then the line 'Estimated' will be close to the 'Observed' dots. In the figures the line is close to the dots, indicating that the Generalized Dirichlet Model fitted the data well. For the higher levels of expenditure (800 and above) the 'Observed' does not match the goodness-of-fit estimates as well. This is due to sampling variation: the sample sizes are small (see Table 2).

[Insert Figure 1 about here]

(e) Comparison of Tables and Graphs. We presented the results in two formats; Table 5 equates to Figure 1. Both formats are useful, presenting and emphasizing different aspects of the output. The tables are more exact, as they present actual numerical results, and can provide a greater array of results. The graphs are useful in emphasizing individual trends and doing so in a visual way. We recommend the use of both formats.

[Insert Table 6 about here]

(f) Interpretation of Parameter Estimates. For each covariate the model estimates two parameters for the Generalized NBD and one parameter for each brand in the Generalized DMD, see Table 6. There is practical value in comparing the impact of each covariate on the BPMs as is discussed above but additional insights can be gained from the actual parameter estimates. Table 6 indicates the way in which covariates can have counter influences. At the brand level in the Generalized DMD the influence of household size on Skip is slightly negative whilst still being significant. As household size increases the preference for Skip decreases. However, household size also has a positive influence at the category level in the Generalized NBD. As household size increases so too does the category purchase rate. Household size reduces preference for Skip but increases preference for the category. So there are contrary influences on the penetration and purchase frequencies where the category influence is greater. While the overall impact is not shown here, the influence of household size on penetration and purchase frequencies for Skip is positive. On the one hand there is value in considering the impact of covariates on penetration and purchase frequency, and at the same time there is value in evaluating the actual parameters.

5.4 Discussion of the Application

For some covariates the marginal effect is not the same as the pattern evident in the raw data (i.e. in the 'Observed' dots). For example, as the number of visits increases the marginal effect on penetration reduces, whereas in the raw data penetration increases. In a classic sense for a regression model, this is due to collinearity. This is the strength of analyzing several covariates at the same time.

As a second example, in the raw data, penetration for Le Chat increases with size of household; it appears that bigger households are more likely to be buyers of Le Chat. However, in the model which has multiple covariates, this effect is not evident. In the marginal analysis there is no effect of household size on penetration because of collinearity with the other covariates and, in particular, with weekly spend. Thus, the fact that in the raw data bigger households were more likely to be buyers of Le Chat arises because bigger households also spend more. Household size itself is not a relevant covariate of penetration of Le Chat. The apparent effect is misleading. In the marketing of Le Chat it is more important to focus on variations due to spend rather than household size.

Figure 1(a), and Table 5(a), can be used to segment on the basis of penetration. It is possible to identify those households where Le Chat has greater penetration (households which spend more) and where it did not. Thus the management of the brand can be adjusted accordingly, either to further strengthen it in areas where it already has strength or to tackle areas where it has a weakness. The Generalized Dirichlet Model provides a tool to identify the segments based on penetration. Similarly Figure 1(c), combined with Table 5(c), can be used to segment on the basis of SCR. Le Chat has higher SCR amongst households who spend less, are smaller, visit less often and are more loyal to their favorite supermarket.

So, taking all this information into account, how should the brand manager of Le Chat react? What promotional activities should be planned? Firstly, household size should not be given too much attention. For example, in targeted advertising the emphasis would be on creative concepts communicating household spend rather than size. Consumers with higher expenditure should be targeted using promotions designed to sell more to existing customers (volume/loyalty rewards, etc.). Conversely, consumers with low expenditure should be targeted with promotions designed to attract new customers (free samples, etc.). Amongst those consumers who are more loyal to a single store, Le Chat has lower penetration but higher share of category requirements. Promotion should look for ways of using the loyalty to the store to tap into attracting new customers to Le Chat.

The Dirichlet model also makes a broader theoretical contribution as its assumptions provide insight into the process of repeated choice. Some of the Dirichlet model's applications focus on the BPMs but others focus on the nature of the model, its parameter and their values. The *S* parameter, which is the sum of the α parameters, is a measure of category loyalty and the *γ* parameter is a measure of category concentration. The parameters closely represent key aspects in the assumptions of the model. Similarly, examining the parameters of the generalized model, as in Table 6, gives direct insight into the impact of the covariates on the processes of repeated choice. We see greater variation between brands resulting from the covariates in Table 6, which examines parameters, than the earlier tables examining the BPMs. (And the generalized model provides the additional output of standard errors for the parameter estimates.) The parameters deconstruct choice while the BPMs aggregate it into a range of dimensions. The applications of the generalized model differ accordingly; some will be developmental, examining choice processes and theory, and some will be applied and will assist marketing decisions.

6. Precursors of the Generalized Dirichlet Model

The Generalized Dirichlet Model is an example of a set of techniques designed to incorporate covariates into analyses of category purchase incidence and brand choice (e.g., Vilcassim and Jain, 1991; Chintagunta, 1993). There are several important precursors.

First, there is the Generalized NBD, also known as count regression, and various forms of Poisson regression (Wagner and Taudes, 1986; Bawa and Ghosh, 1990; Winkelmann, 1997; Cameron and Trivedi, 1998). There are two parameters for the NBD. The usual approach to generalizing the NBD is to include covariates in one of the distribution parameters (usually the mean = $\gamma\beta$) and then to express the other distribution parameter (now the standard deviation or variance) as a multiple of the first. However, the form of Generalized NBD given in our paper is even more generalized. Covariates are fitted to each of the distribution parameters γ and β separately. This is because the two parameters have separate influences on the BPMs.

Second, covariates have been incorporated through multinomial logit model specifications – discussed in an exploratory way by Jones and Zufryden (1980) and then empirically demonstrated (Heckman and Willis, 1977; Wrigley and Dunn, 1985; Uncles, 1987; Vilcassim and Jain, 1991; Bucklin and Gupta, 1992; Fader, 1993; Fader and Lattin, 1993; Russell and Kamakura, 1994; Elrod and Keane, 1995). Related are a variety of latent class specifications (Fader and Hardie, 1996; Yim and Kannan, 1999; Kannan and Yim, 2001; Danaher et al., 2003). These earlier approaches are extremely important for the development of the work described here, but they tended to look at the potential to improve model fit rather than explicitly explore the properties of the BPMs themselves.

Third, covariates have been added to the Dirichlet distributions leading to forms of regression where the dependent variables are multivariate proportions (Hijazi, 2003; Gueorguieva et. al., 2008). The component of the generalized Dirichlet model developed here relating to brand choice uses the same generalized linear model approach but is applied to the Dirichlet multinomial distributions and to data reporting counts rather than proportions.

Fourth, in a nonparametric (or semi-parametric) tradition, case-specific 'mass points' have been used to model consumer heterogeneity rather than assuming the Dirichlet's smooth Gamma and Beta distributions (Dunn et al., 1987; Colombo and Morrison, 1989; Chintagunta et al. 1991). In these models the effect of covariates is incorporated into the estimation of the mass points; however, this comes at the expense of computational efficiency and these approaches have not reported the same range of BPMs as is now commonly expected in Dirichlet modeling.

The contribution to marketing of the Generalized Dirichlet model is of a very different nature. We have seen that the generalized model enables analysis of relationships between the covariates and the BPMs. It is a method for segmentation based on market share along with any other of the BPMs, extending the usual models for segmentation to incorporate behavioral loyalty. Furthermore, there is scope for significant future development in that the generalized model not only includes covariates which are characteristics of consumers, but it has the capacity to include covariates for the marketing mix of the brands and the purchase situation in a multi-attribute framework.

6. Summary

We generalize the Dirichlet model so it can incorporate the characteristics of consumers as covariates. In so doing the model is substantially expanded to provide output on how the characteristics of consumers are related to BPMs and loyalty. The Generalized Dirichlet model developed here has the capacity to continue to capture all the patterns identified previously for BPMs (e.g., double jeopardy). It achieves this by simultaneously modeling every BPM for every brand in the product category. In addition, like logit and regression models, it can incorporate multiple covariates and evaluate the impact of each. It provides far greater power and validity in evaluating the impact of covariates on the BPMs, and it provides formal testing procedures. Statistically, it does for the Dirichlet model what has been available for many years in regression and multinomial logit. But the outcome is entirely new. It provides a method for rigorously and fully evaluating the impact of characteristics on the BPMs and in particular on consumer loyalty. This is potentially very useful for marketing analysts.

References

Arora, N., G.M. Allenby and J.L. Ginter, (1998), "A hierarchical Bayes model of primary and secondary demand", Marketing Science, Vol. 17, No. 1, pp. 29-44.

Bass, F.M., A. Jeuland and G.P. Wright, (1976), "Equilibrium stochastic choice and market penetration theories: Derivations and comparisons", Management Science, Vol. 22 (10), pp. 1051-1063.

Bawa, K. and A. Ghosh, (1990), "The covariates of regularities in purchase timing", Marketing Letters, Vol. 2, No. 2, pp. 147-157.

Ben-Akiva, M. and S.R. Lerman, (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press., Boston.

Bucklin, R.E. and S. Gupta, (1992), "Brand choice, purchase incidence, and segmentation: An integrated modelling approach", Journal of Marketing Research, Vol. 29, No. 2, pp. 201-215.

Cameron, A.C. and P.K. Trivedi, (1998), *Regression Analysis of Count Data*, Cambridge University Press, New York, NY.

Chatfield, C., A.S.C. Ehrenberg and G.J. Goodhardt, (1966), "Progress on a simplified model of stationary purchasing behaviour", Journal of the Royal Statistical Society, Series A (General), Vol. 129, No. 3, pp. 317-367.

Chatfield, C. and G.J. Goodhardt, (1970), "The beta-binomial model for consumer purchasing behaviour", Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol. 19, No. 3, pp. 240-250.

Chintagunta, P., D.C. Jain and N.J. Vilcassim, (1991), "Investigating heterogeneity in brand preference in logit models for panel data", Journal of Marketing Reseach, Vol. 28, pp. 417-428.

Chintagunta, P.K., (1993), "Investigating purchase timing incidence, brand choice and purchase quantity decisions of households", Marketing Science, Vol. 12, No. 2, pp. 184-208.

Colombo, R.A. and D.G. Morrison, (1989), "A brand-switching model with implications for marketing strategy", Marketing Science, Vol. 8, pp. 89-99.

Danaher, P.J., I.W. Wilson and R.A. Davis, (2003), "A comparison of online and offline consumer brand loyalty", Marketing Science, Vol. 22, No. 4, pp. 461-476.

Driesener, C, (2005), “Empirical generalisations in the parameters of the Dirichlet model: an examination across 50 categories”, PhD Thesis, University of South Australia.

Dunn, R., S. Reader and N. Wrigley, (1983), "An investigation of the assumptions of the NBD model as applied to purchasing at individual stores", Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 32, No. 3, pp. 249-259.

Dunn, R., S. Reader and N. Wrigley, (1987), "A nonparametric approach to the incorporation of heterogeneity into repeated polytomous choice models of urban shopping behaviour", Transportation Research A, Vol. 21, No. 4/5, pp. 327-343.

Edwards, A.W.F., (1976), *Likelihood - An account of the Statistical Concepts of Likelihood and its Application to Scientific Inference*, Cambridge University Press, Bath, UK.

Ehrenberg, A.S.C., (1959), "The pattern of consumer purchases", Applied Statistics, Vol. 8, No. 1, pp. 26-41.

Ehrenberg, A.S.C., (1988), *Repeat-Buying, Facts, Theory and Applications*, Oxford University Press, New York, NY.

Ehrenberg, A.S.C., G.J. Goodhardt and T.P. Barwise, (1990), "Double jeopardy revisited", Journal of Marketing, Vol. 54, No. 3, pp. 82-91.

Ehrenberg, A.S.C., M.D. Uncles and G.G. Goodhardt, (2004), "Understanding brand performance measures: Using Dirichlet benchmarks", Journal of Business Research, Vol. 57, No. 12, pp. 1307-1325.

Eliason, S.R., (1993), *Maximum Likelihood Estimation - Logic and Practice*, Sage Publications Inc., California.

Elrod, T. and M.P. Keane, (1995), "A factor-analytic probit model for representing the market structure in panel data", Journal of Marketing Research, Vol. 32, pp. 1-16.

Fader, P. and B.G.S. Hardie, (1996), "Modelling consumer choice among SKU's", Jounal of Marketing Research, Vol. 23, pp. 442 - 452.

Fader, P.S., (1993), "Integrating the Dirichlet-multinomial and multinomial logit models of brand choice", Marketing Letters, Vol. 4, No. 2, pp. 99-112.

Fader, P.S. and J.M. Lattin, (1993), "Accounting for heterogeneity and nonstationarity in a cross-sectional model of consumer purchase behavior", Marketing Science, Vol. 12, No. 3, pp. 304-317.

Fader, P.S. and D.C. Schmittlein, (1993), “Excess behavioral loyalty for high-share brands: deviations from the Dirichlet model for repeat purchasing”, JMR, Vol. 30, No. 4, pp 478-493.

Goodhardt, G.J., A.S.C. Ehrenberg and C. Chatfield, (1984), "The Dirichlet: a comprehensive model of buying behaviour", Journal of the Royal Statistical Society, Series A (General), Vol. 147, No. Part 5, pp. 621-655.

Gueorguieva R., R. Rosenheck and D. Zelterman, (2008), “Dirichlet component regression and its applications to psychiatric data”, Computational Statistics and Data Analysis, Vol. 52, Issue 12, pp 5344-5355.

Hijazi, R.H., (2003), “Analysis of compositional data using Dirichlet covariate models”, Ph.D. Dissertation, The American University, Washington, DC.

Heckman, J.J. and R.J. Willis, (1977), "A beta-logistic model for the analysis of sequential labour force participation be married women", Journal of Political Economy, Vol. 85, No. 1, pp. 27-58.

Hensher, D.A., J.M. Rose and W.H. Greene., (2005), *Applied Choice Analysis: A Primer*, Cambridge University Press, Cambridge.

Horsky, D., S. Misra and P. Nelson, (2006), "Observed and unobserved preference heterogeneity in brand-choice models", Marketing Science, Vol. 25, No. 4, pp. 322-335.

Jacoby, J. and R.W. Chestnut, (1978), *Brand Loyalty Measurement and Management*, John Wiley & Sons, New York, NY.

Jeuland, A.P., F.M. Bass and G.P. Wright, (1980), "A multibrand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes", Operations Research, Vol. 28, No. 2, pp. 255-277.

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

Johnson, N.L., S. Kotz and A.W. Kemp, (1993), *Univariate Discrete Distributions*, John Wiley & Sons, Inc., New York, NY.

Jones, M.J. and F.S. Zufryden, (1980), "Adding explanatory variables to a consumer purchase behavior model: An exploratory study", Journal of Marketing Research, Vol. 17, pp. 323-334.

Kannan, P.K. and C.K. Yim, (2001), "An investigations of promotions on across-submarket competition", Journal of Business Research, Vol. 53, No. 3, pp. 137-149.

Louviere, J.J., D.A. Hensher and J. Swait, (2000), *Stated Choice Methods Analysis and Application*, Cambridge University Press, Cambridge.

McCullagh, P. and J.A. Nelder, (1989), *Generalized Linear Models*, Chapman & Hall/CRC, London.

McFadden, D., (1974), "Conditional logit analysis of qualitative choice behavior", Zarembka, P. (Ed.), *Frontiers in Econometrics*, Academic Press, New York, NY.

Nelder, J.A. and R. Mead, (1965), "A simplex method for function minimization", The Computer Journal, Vol. 7, pp. 303-313.

Rungie, C., (2003), "How to estimate the parameters of the Dirichlet model using likelihood theory in Excel", Marketing Bulletin, Vol. 14.

Rungie, C. and G. Goodhardt, (2004), "Research note: Calculation of theoretical brand performance measures from the parameters of the Dirichlet model", Marketing Bulletin, Vol. 15.

Russell, G.J. and W.A. Kamakura, (1994), "Understanding brand competition using micro and macro scanner data", Journal of Marketing Research, Vol. 31, pp. 289-303.
Sabavala, D.J. and D.G. Morrison, (1977), "A model of TV show loyalty", Journal of Advertising Research, Vol. 17, No. 6, pp. 35-43.
Schmittlein, D.C., A.C. Bemmaor and D.G. Morrison, (1985), "Why does the NBD model work? Robustness in representing product purchases, brand purchases and imperfectly recorded purchases", Marketing Science, Vol. 4, pp. 255-266.
Sharp, B., (2010), *How Brands Grow*, Oxford University Press, South Melbourne.
Sharp, B., M. Wright and G. Goodhardt, (2002), "Purchase loyalty is polarised into either repertoire or subscription patterns", Australasian Marketing Journal, Vol. 10, Issue 3, pp. 7-20.
Uncles, M., (1987), "A beta-logistic model of mode choice: Goodness of fit and intertemporal dependence ", Transportation Research Part B: Methodological, Vol. 21, No. 3, pp. 195-205.
Uncles, M.D., A.S.C. Ehrenberg and K. Hammond, (1995), "Patterns of buyer behavior: Regularities, models and extensions", Marketing Science, Vol. 14, No. 3, 2 of 2, pp. G71-G78.
Vilcassim, N.J. and D.C. Jain, (1991), "Modelling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity", Journal of Marketing Research, Vol. 28, No. 1, pp. 29-41.
Wagner, U. and A. Taudes, (1986), "A multivariate polya model of brand choice and purchase incidence", Marketing Science, Vol. 5, No. 3, pp. 219-44.
Winkelmann, R., (1997), *Econometric Analysis of Count Data*, Springer, Berlin.
Wong, T.T., (1998), "Generalized Dirichlet distribution in Bayesian analysis ", Applied Mathematics and Computation, Vol. 97, No. 2/3, pp. 165-181.
Wright, M., A. Sharp and B. Sharp, (2002), "Market statistics for the Dirichlet model: Using the Juster scale to replace panel data", International Journal of Research in Marketing, Vol. 19, pp. 81-90.
Wrigley, N. and R. Dunn, (1985), "Stochastic panel-data models of urban shopping behavior: 4. Incorporating independent variables into the NBD and Dirichlet Models", Environment and Planning A, Vol. 17, No. 3, pp. 319-331.
Yim, C.K. and P.K. Kannan, (1999), "Consumer behavior loyalty: A segmentation model and analysis", Journal of Business Research, Vol. 44, No. 2, pp. 75-92.

Table 1 Laundry Detergent, France:

Penetrations, Average Purchase Frequency and Share of Category Requirements over a Year

Brand	Market Share	Penetration		Average Purchase Frequency		SCR	
		Obser-ved	*Theor-etical*	Obser-ved	*Theor-etical*	Obser-ved	*Theor-etical*
		%	*%*			%	*%*
Category	100	92	*90*	7.0	*7.1*	100	*100*
Others	43	62	*62*	4.5	*4.4*	63	*62*
Skip	16	31	*31*	3.3	*3.4*	49	*48*
Ariel	13	24	*25*	3.5	*3.3*	53	*46*
Le Chat	10	20	*20*	3.2	*3.2*	44	*45*
Omo	9	21	*20*	2.8	*3.2*	40	*45*
X-Tra	9	19	*18*	3.0	*3.1*	39	*44*
MAD			*0.2*		*0.18*		*3.5*

Table 2 Laundry Detergent, France: Covariates

Household-Size

Range	Value	Frequency
1	1	777
2	2	1143
3	3	592
4	4	553
5 or more	5	374
	Total	*3439*

Visits

Range	Value	Frequency
0.75 or less	0.5	749
0.75 to 1.25	1	1159
1.25 to 1.75	1.5	707
1.75 to 2.25	2	372
2.25 to 2.75	2.5	199
2.75 to 3.25	3	106
3.25 to 3.75	3.5	50
3.75 or more	4	97
	Total	*3439*

Expenditure

Range	Value	Frequency
150 or less	100	515
150 to 250	200	682
250 to 350	300	653
350 to 450	400	526
450 to 550	500	380
550 to 650	600	278
650 to 750	700	167
750 to 850	800	119
850 or more	900	119
	Total	*3439*

Loyalty

Range	Value	Frequency
.35 or less	0.3	77
.35 to .45	0.4	247
.45 to .55	0.5	410
.55 to .65	0.6	369
.65 to .75	0.7	456
.75 to .85	0.8	470
.85 to .95	0.9	636
.95 or more	1	774
	Total	*3439*

Table 3 Laundry Detergent, France: Correlations between Covariates

	Expenditure	Visits	Loyalty
Household-Size	0.48	0.12	-0.07
Expenditure		0.47	0.08
Visits			-0.07

All four covariates were correlated (P<0.0002)

Table 4: Laundry Detergents, France: Likelihood-Ratio Tests

(a) Likelihood-Ratio Tests for the NBD side of the Model

	Log Likelihood	Likelihood Ratio	# Para-meters	Degrees of freedom	P
Full Model	9405		10		
Covariate Removed					
Constant	9557	305	8	2	0.0000
Visits	9416	23	8	2	0.0000
Household-Size	9421	31	8	2	0.0000
Expenditure	9709	607	8	2	0.0000
Loyalty	9421	31	8	2	0.0000
All covariates are significant cf the Full Model					
Null Model (Constant only)	10032		2		

(b) Likelihood-Ratio Tests for the DMD side of the Model

	Log Likelihood	Likelihood Ratio	# Para-meters	Degrees of freedom	P
Full Model	13904		30		
Covariate Removed					
Constant	13980	152	24	6	0.0000
Visits	13908	6	24	6	0.3747
Household-Size	13922	34	24	6	0.0000
Expenditure	13916	22	24	6	0.0010
Loyalty	13923	37	24	6	0.0000
All covariates are significant cf the Full Model except visits					
Null Model (Constant only)	13952		6		

Table 5 Laundry Detergents, France: The Impact of Covariates

(a) Impact of the Covariates on Penetration

Brand	Theoretical Penetration %	Absolute change in Penetration when covariate increases by one standard deviation %			
		Visits	Household Size	Expenditure	Loyalty
Others	62	1.1	1.2	6.7	-1.5
Skip	31	0.6	-0.3	8.4	-1.1
Ariel	25	0.2	-2.3	6.8	0.6
Le Chat	20	-0.6	0.6	7.5	-0.8
Omo	20	-0.9	1.1	5.6	-0.5
X-Tra	18	0.7	4.7	2.4	-0.9

(b) Impact of the Covariates on Average Purchase Frequency

Brand	Theoretical Purchase Frequency	Absolute change in Average Purchase Frequency when covariate increases by one standard deviation			
		Visits	Household Size	Expenditure	Loyalty
Others	4.4	0.21	0.30	1.49	0.36
Skip	3.4	0.10	0.17	1.07	0.32
Ariel	3.3	0.09	0.13	0.98	0.35
Le Chat	3.2	0.06	0.17	0.96	0.31
Omo	3.2	0.06	0.18	0.92	0.32
X-Tra	3.1	0.09	0.24	0.84	0.31

(c) Impact of the Covariates on Share of Category Requirements

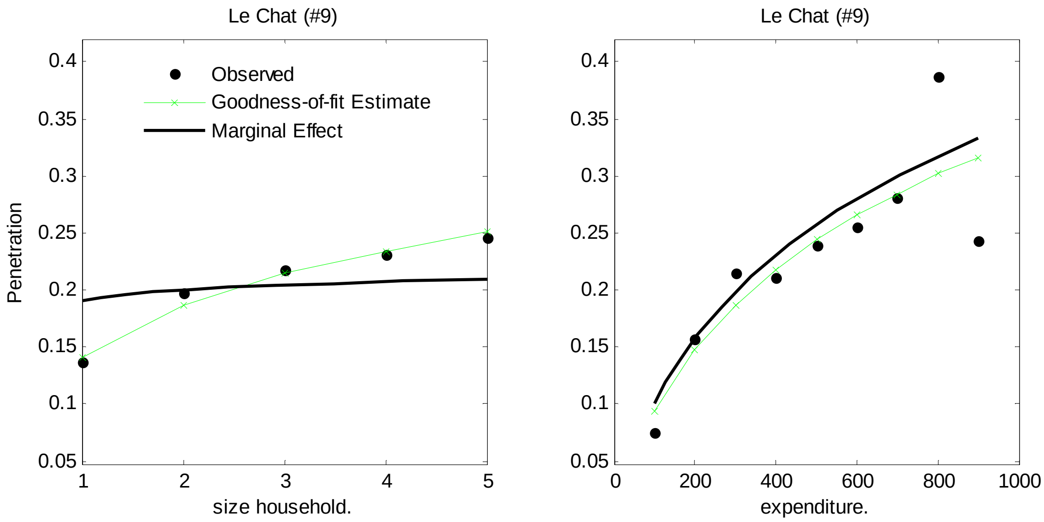
Brand	Theoretical SCR %	Absolute change in SCR when covariate increases by one standard deviation %			
		Visits	Household Size	Expenditure	Loyalty
Others	62	-0.1	-0.8	-6.6	0.7
Skip	48	-0.9	-1.4	-5.7	1.3
Ariel	46	-1.1	-1.9	-6.1	1.7
Le Chat	45	-1.3	-1.1	-5.8	1.3
Omo	45	-1.4	-1.0	-6.2	1.4
X-Tra	44	-1.0	-0.2	-6.9	1.2

Table 6 Parameter Estimates
The coefficients for the covariates

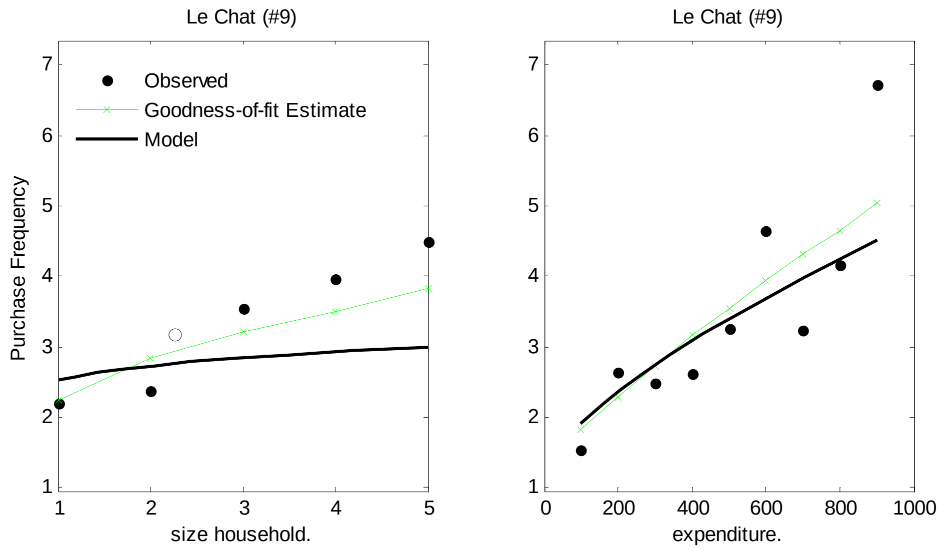
	Constant	Visits	log of Household size	log of Expenditure	log of Loyalty
NBD Regression Coefficients					
gamma (shape) parameter	-1.88(.03)***	-0.16(.04)***	-0.08(.06)	0.48(.02)***	-0.30(.13)*
beta (scale) parameter	-0.78(.16)***	0.21(.05)***	0.24(.06)***	0.26(.03)***	0.53(.13)***
DMD Regression coefficients for alpha parameter for each brand					
Others	-0.35(.04)***	0.07(.12)	0.00(.07)	-0.10(.03)***	-0.51(.03) ***
Skip	-2.31(.02)***	0.05(.05)	-0.09(.05)*	0.09(.01)***	-0.51(.07) ***
Ariel	-2.20(.03)***	0.04(.08)	-0.26(.05)***	0.07(.03)***	-0.29(.12) ***
Le Cha	-3.45(.03)***	-0.02(.11)	-0.01(.04)	0.20(.03)***	-0.53(.06) ***
Omo	-2.55(.03)***	-0.04(.04)	0.03(.07)	0.06(.01)***	-0.45(.07) ***
X-Tra	-1.97(.04)***	0.07(.11)	0.39(.11)***	-0.15(.01)***	-0.54(.09) ***

Figure 1 Laundry Detergent, France: The Impact of Covariates

(a) The Impact of Covariates on the Penetration for Le Chat



(b) The Impact of Covariates on the Average Purchase Frequency for Le Chat



(c) The Impact of Covariates on the Share of Category Requirements for Le Chat