J. R. Statist. Soc. A (1984),
147, *Part* 5, *pp.* 621–655

The Dirichlet: A Comprehensive Model of Buying Behaviour

By G. J. GOODHARDT, A. S. C. EHRENBERG† and C. CHATFIELD

City University Business School, London Business School and Bath University, UK

[*Read before the* Royal Statistical Society *on Wednesday, June* 27th, 1984, *the President* Dr W. F. Bodmer *in the Chair*]

SUMMARY

The Dirichlet is a stochastic model of purchase incidence and brand choice which parsimoniously integrates a wide range of already well-established empirical regularities.

Keywords: BUYER BEHAVIOUR; CONSUMER PURCHASING; STOCHASTIC MODEL; PURCHASE INCIDENCE; BRAND CHOICE; GAMMA DISTRIBUTION; POISSON DISTRIBUTION; MULTINOMIAL DISTRIBUTION; DIRICHLET DISTRIBUTION; MULTIVARIATE BETA DISTRIBUTION; BETA BINOMIAL DISTRIBUTION; NEGATIVE BINOMIAL DISTRIBUTION

The Dirichlet model describes how frequently-bought branded consumer products like instant coffee or toothpaste are purchased when the market is stationary and unsegmented. This is the common situation where, over the time-periods analysed,

(A) the sales of each brand show little variation,
(B) the different brands show no special groupings.

We introduce the empirical data in Section 1 and the model itself in Section 2. The model assumes a mixture of distributions at four levels (Section 2.1):

(i) Purchasing of the product-class takes the form of a Poisson process for each consumer,
(ii) The purchasing rates of different consumers follow a Gamma distribution,
(iii) Each consumer's choices among the available brands follow a multinomial distribution, and
(iv) These choice probabilities follow a multivariate Beta or "Dirichlet" distribution across different consumers.

In Section 2.2 we give justifications for these assumptions.

More importantly, the model has successfully described the patterns observed in more than 40 product-fields (Section 3) and therefore provides interpretative norms (Section 3.1). As input the model only requires the sales level of each brand and two parameters. These can be identified as two aspects of consumer diversity, namely how much people differ from each other in (a) their purchasing rates and (b) their brand-choice preferences (Section 3.2).

The model encompasses earlier, more limited formulations which often remain easier to use, as noted in Section 4.1. We briefly review the literature on other models in Section 4.2 and comment on aspects of model-building and on some future areas of work in Section 5.

1. THE DATA ANALYSED

Table 1 sets out an artificial example of purchases of three brands, X, Y and Z, over 12 weeks for a notional sample of 100 households. Such data usually come from large consumer panels, as operated for example by AGB Research in the UK (e.g. Buck, 1982) and MRCA in the USA. Our analysis unit here is the *buying occasion* or the nearest equivalent for a given brand. (The amount bought or price paid on each occasion would be analysed separately.)

† *Present address*: London Business School, Sussex Place, Regent's Park, London NW1 4SA.

 0035–9238/84/147621 $2.00

Buying behaviour tends to be fairly regular in the aggregate (e.g. 6 or 7 consumers out of the 100 in Table 1 bought brand X in each 4 weeks). But it is more irregular at the individual level. We take the view that individual purchases occur in an *as if random* manner (e.g. Ehrenberg, 1982,

TABLE 1

A simplified buying pattern for three brands over 12 weeks (a notional sample of 100 households; three brands X, Y, Z)

	Purchases in week												*12-week totals (Brands and Product)*
	1	2	3	4	5	6	7	8	9	10	11	12	
Household 1	X	X	Y	X	X	X	Y	X	.	Y	X	Y	$7X + 4Y = 11$
Household 2	.	.	X	.	Z	X	.	.	.	Z	X	X	$4X + 2Z = 6$
Household 3	.	.	Z	.	.	Z	.	X	X	.	Z	.	$2X + 3Z = 5$
Household 4	.	.	.	.	Y	.	Y	Y	.	X	.	.	$X + 3Y = 4$
Household 5	Y	X	.	.	X	.	.	.	.	.	.	Y	$2X + 2Y = 4$
Household 6	X	.	.	.	.	.	.	Z	Z	Y	.	.	$X + Y + 2Z = 4$
Household 7	.	Z	X	.	.	.	.	.	.	.	.	X	$2X + Z = 3$
Household 8	.	Y	.	Y	.	.	.	.	X	.	.	.	$X + 2Y = 3$
Household 9	Z	.	.	.	.	.	X	.	.	.	Y	.	$X + Y + Z = 3$
Household 10	.	.	.	X	.	Y	.	.	.	.	.	.	$X + Y = 2$
Household 11	.	.	.	Y	.	.	X	.	.	.	.	.	$X + Y = 2$
Household 12	.	.	.	.	.	.	.	.	.	X	.	.	$X = 1$
Households 13–100	.	.			(No purchases at all)						.	.	0
4-week totals													
No. of buyers	6 of X, 4 of Y, 3 of Z: 10*				6 of X, 3 of Y, 3 of Z: 9*				7 of X, 4 of Y, 3 of Z: 10*				12 of X. 8 of Y, 5 of Z: 12*
No. of purchases	$8X + 5Y + 3Z = 16$*				$8X + 5Y + 3Z = 16$*				$8X + 5Y + 3Z = 16$*				$24X + 15Y + 9Z = 48$*

[* Of *any* brand.]

p. 75). This does not mean that we think a housewife times her purchases and chooses her brands literally at random (e.g. by tossing mental pennies). Instead, people generally have "verifiable reasons" for making a purchase (e.g. "We'd nearly run out", or "My mother-in-law was coming"). But taken *en masse*, purchases are capable of being modelled probabilistically at the individual level, even though true randomness lies only in the model.

1.1. *Summary Measures*

We now introduce some regular features of such purchasing data which any buying behaviour model typically has to be able to predict. Five refer to a single brand and five to relations between different brands.

(i) *Penetration, b.* The proportion of the population buying a brand at least once in a given period is usually called its penetration, denoted by b. It increases with T, the length of the analysis period, though less than *pro rata*. Thus for brand X in Table 1, b increases from 2 per cent in Week 1 to 12 per cent in Weeks 1-12.

(ii) *Repeat-buying.* Of the 6 households that bought brand X in Weeks 1-4, 3 (or 50 per cent) bought it again in Weeks 5-8. These 3 households are called repeat-buyers for the time-periods in question and explain the less-than-pro-rata increase in b.

(iii) *Purchase frequency per buyer, w.* The average purchase frequency per buyer is denoted by w. It also increases with T, but generally more slowly than the penetration b, i.e. from 1.0 for X in 1 week to 2.0 in 12 weeks.

(iv) *Sales of brand X.* The *per capita* (or per household) sales level of a brand is denoted by m. It increases *pro rata* to T in a stationary market. In terms of b and w, this comes about through the "Sales Equation" that $m = b \times w$. In 12 weeks, penetration of brand X is 6 times that in 1 week and purchase frequency doubles, and so cumulative sales of X increase 12-fold. The increase is mainly in b, with more light buyers of X being added as T increases, at least for low values of b.

(v) *The distribution of brand purchases.* The distribution of purchases of a brand by individual consumers is reversed-J-shaped, as illustrated by the 12-week totals in Table 1.

Turning to the differences between the brands, we have

(vi) *Differing brand penetrations.* The different sales volumes of brands X, Y and Z—24, 15 and 9—are roughly proportional to their 12-week penetrations—12, 8 and 5.

(vii) *$w \doteqdot$ constant.* In contrast, buyers of the different brands buy them at much the same average rate of about 2 purchases in 12 weeks. But the precise rates decrease slightly with market-share:

$$\begin{aligned} &\text{Brand } X\text{:} \quad w = 24/12 = 2.0,\\ &\text{Brand } Y\text{:} \quad w = 15/8 \;\; = 1.9,\\ &\text{Brand } Z\text{:} \quad w = 9/5 \;\;\; = 1.8. \end{aligned}$$

This is known as a "Double Jeopardy" pattern (McPhee, 1963, Chapter 3): Z not only has fewer buyers than X, but they also buy it (slightly) less often.

(viii) *Total product purchases.* The buyers of brand X bought the product-class as a whole (i.e. any brand) on average 4 times in the 12 weeks (48/12 = 4.0). This is much higher than w, i.e. buyers of brand X tend to buy the *other* brands quite often. This average rate of product purchasing, denoted by w_P, also varies little from brand to brand but *increases* slightly with decreasing market-share (a trend in the opposite direction to the w's):

$$\begin{aligned} &\text{Brand } X\text{:} \quad w_P = 48/12 = 4.0,\\ &\text{Brand } Y\text{:} \quad w_P = 33/8 \;\; = 4.1,\\ &\text{Brand } Z\text{:} \quad w_P = 21/5 \;\; = 4.2. \end{aligned}$$

(ix) *Sole buyers.* The consumers who buy brand X and no other brand in a given period are called sole buyers. The proportion of such consumers is higher in short periods: 3 out of 6 buyers of X in the first 4 weeks, only 1 out of 12 in the 12 weeks.

(x) *Duplicated buyers.* The proportion of buyers of X in a period who in the same period also buy another specific brand, Y say, varies with the second brand's penetration. (Table 1 is somewhat atypical in this respect because the example is oversimplified.)

As we will illustrate in Section 3, the patterns in the small artificial example above have been found to apply in large samples for a wide range of different product-classes and brands. The task for any model here is therefore not to discover or model new results, but to summarize and integrate regularities that are already known, and to help highlight any discrepancies.

1.2. *A Stationary Market*

Aggregate measures of buying behaviour are often approximately steady over successive periods of time, i.e. "stationary", at least in the medium-term. For example, in Table 1 the total sales of brands X, Y and Z are 8, 5 and 3 in *each* of the three 4-week periods. (In real life they are not quite so steady, especially not with a small sample.) Again, repeat-buying of brand X is 3/6 or 50 per cent not only from Weeks 1-4 to Weeks 5-8, but also from Weeks 5-8 to 9-12 and for the *non*-consecutive periods 1-4 to 9-12.

The relatively low observed repeat-buying rates are therefore not due to a brand continually losing its customers (the "Leaky Bucket Theory"), but because many buyers buy infrequently. People who buy a brand only twice in a quarter (like Household 3 for brand X in Table 1) cannot be buying the brand in each 4-week period, let alone every week. However, the evidence is that in the medium-term they would tend to go on buying X at about the same average rate (see also Ehrenberg, 1972, Table 3.9; Ehrenberg and Goodhardt, 1979).

This lack of erosion in the level of repeat-buying leads to the view that such consumer goods markets tend to be stationary in the medium term not only for aggregate measures such as sales, but even approximately for the individual consumer. This in turn suggests formulating a stochastic model at the individual level with fixed probabilities of purchasing incidence and brand-choice. Most exceptions to approximately stationary markets are sharp but short-lived up-and-down variations in the sales of particular brands due to promotional activities.

2. THE DIRICHLET MODEL

We consider a population of N consumers making purchases in a product-class of g brands. The Dirichlet model specifies probabilistically how many purchases each consumer makes in a time-period and which brand is bought on each occasion. It combines both purchase incidence and brand-choice aspects of buyer behaviour into one model. The model should therefore allow us to predict the various summary measures introduced in Section 1.1.

We give the basic assumptions and formulation of the model in Section 2.1 and empirical and theoretical justifications of the assumptions in Section 2.2.

A crucial property of the model is that different brands can simply be combined into a "super-brand" (Section 2.3). This is used technically in estimating one of the parameters of the model (Section 2.4) and in evolving theoretical predictions of specific summary measures (Section 2.5).

2.1. *The Basic Assumptions*

Our formulation of the model is to specify the probability vector of the ith consumer making any specific combination $\{r_j\}$ of purchases of the $j = 1$ to g brands in an analysis-period of any chosen length T, that is r_1 purchases of brand 1, plus r_2 purchases of brand 2, etc., where $r \geqslant 0$. Summing over the $j = 1$ to g brands, $\Sigma r_j = n_i$ is the total number of purchases of the product-class made by the ith consumer in that period.

We arrive at the formulation of the particular model by making five assumptions. The first two concern brand-choice:

(A1) The ith individual's brand choices over a succession of purchases are as if random, with a probability $(p_j)_i$ of choosing brand j from $j = 1, \ldots, g$ brands. These probabilities are fixed over time and brand-choices at successive purchases are assumed independent. The number of purchases of each brand that individual i makes in a sequence of n_i purchases can therefore be modelled by a multinomial with parameters $n_i, (p_1)_i, \ldots, (p_g)_i$ (Wilks, 1962, p. 139).

(A2) The probabilities $(p_j)_i$ vary among individuals according to a Dirichlet distribution (e.g. Wilks, 1962, p. 177). This is a multivariate Beta-distribution given by the joint density function

$$Cp_1^{\alpha_1 - 1} \ldots p_g^{\alpha_g - 1},$$

where $p_j \geqslant 0$, $\Sigma p_j = 1$, $C = \Gamma(S)/\Pi(\Gamma \alpha_j)$, $S = \Sigma \alpha_j$ and $\alpha_j > 0$.The probability of choosing brand j then has the jth marginal distribution. This is the simple Beta-distribution

$$Cp_j^{\alpha_j - 1}\ (1 - p_j)^{(S - \alpha_j - 1)},$$

with mean α_j/S (Wilks, 1962, p. 173), which is the brand's market share. (In the traditional notation of the Beta-distribution with parameters α and β, S would equal $(\alpha + \beta)$.)

Assumptions A1 and A2 therefore say that the joint distribution of purchases of different brands across all consumers is given by a mixture of multinomials with a Dirichlet distribution. (When $g = 2$, this reduces to the well-known Beta-Binomial distribution, as noted in Section 2.5.)

We next make two assumptions regarding purchase incidence in the product-class.

(B1) Successive purchases of the ith individual behave as if random, and are assumed to be independent with a constant mean rate μ_i in some chosen "unit" length time-period (longer than the minimum inter-purchase time, which is usually a week for grocery products). The number of purchases n_i made in each of a succession of equal non-overlapping periods of relative length T then follows a Poisson distribution with mean $\mu_i T$.

(B2) The mean purchasing rates vary between individuals according to a Gamma-distribution with density function

$$\frac{e^{-\mu K/M}\ \mu^{K-1}}{(M/K)^K \Gamma(K)}.$$

From assumptions B1 and B2 it follows that the number of purchases of the product made by all individuals in a time-period of length T follows a Negative Binomial distribution (abbreviated NBD), with mean MT and exponent K. (We use capital letters for product-class parameters, and lower case letters for those relating to a specific brand.)

Our final assumption concerns the relationship between the (A) and (B) assumptions, namely:

(C) The brand-choice probabilities and the average purchase-frequencies of different consumers are distributed independently over the population.

Assumptions (A) to (C) are sufficient to specify a single model which we have called the *NBD-Dirichlet* model, or *Dirichlet* for short. In the notation of mixtures or compound distributions (e.g. Gurland, 1957), the number of purchases an individual (or household) makes of each of the g brands in a period of length T is given by a g-variate discrete random variable with the joint frequency distribution

$$[\mathscr{M}(\mathbf{r} \mid \mathbf{p}, n) \underset{\mathbf{p}}{\wedge} \mathscr{D}\ (\mathbf{p} \mid \boldsymbol{\alpha})] \underset{n}{\wedge} [\mathscr{P}(n \mid \mu) \underset{\mu}{\wedge} \mathscr{G}\ (\mu \mid MT, K)],$$

where $\mathscr{M}$, $\mathscr{D}$, $\mathscr{P}$ and $\mathscr{G}$ denote the Multinomial, Dirichlet, Poisson and Gamma distributions.

To activate the model we need to estimate the $(g + 2)$ quantities α_j, M and K (Section 2.4). Rather than enumerate numerically all the relevant probabilities in calculating theoretical values of any desired summary measure, we try to develop algebraic short-cuts (Section 2.5).

2.2. *Justifications of the Assumptions*

The main justification of the Dirichlet model is that in practice it fits many different aspects of buying behaviour under a wide range of conditions, as documented in numerous papers and reports and illustrated in Section 3. In addition there are reasons why the specific distributions of brand-choice and purchase incidence should be as just assumed and not something else:

A. *The Brand-choice distributions*

Assumption A1 of a multinomial distribution is in line with the evidence outlined in Section

1.2, that stochastic buying behaviour at the individual level tends to be irregular but stationary, at least in the medium-term and ignoring sharp but short promotional fluctuations.

For Assumption A2 there is a strong characterization result: given the unsegmented nature of the market, the mixing distribution of the brand-choice probabilities across different individuals *must* be of the Dirichlet form. Thus lack of segmentation means that choosing between the different brands should in some sense be independent. But the choice probabilities for each individual are constrained to add to 1 and cannot therefore be strictly independent.

This type of problem was noted by Karl Pearson (1897). More recently, Mosimann (1962) has introduced the idea of "independence except for the constraint", whereby for two brands j and k, p_j and $p_k/(1-p_j)$ should be independent rather than p_j and p_k. This theory has been developed by Connor and Mosimann (1969), Darroch and Ratcliffe (1971), Darroch and James (1974) and James (1975), and is linked to certain ideas of rational choice by Luce (1959). It expresses mathematically what we mean in marketing when we say a market is "not segmented": the proportion of purchases devoted to any particular brand is independent of the way the remaining purchases are distributed between the *other* brands.

The crucial point now is that independence except for the constraint $\Sigma p_j = 1$ is a characterization of the Dirichlet distribution (Mosimann, 1962, 1984). The use of this distribution is therefore the quantitative analogue of our purely qualitative marketing criterion of non-segmentation.

Aitchison (1982, p. 142) has recently said that the Dirichlet distribution seldom if ever provides an adequate description of compositional data because of its strong independence structure. But we have found that it is precisely this independence structure which empirically fits so well here. In a strictly unsegmented market where the multinomial choice probabilities for each individual are fixed over time, the Dirichlet distribution is the only possible model.

B. *The purchase incidence distributions*

Assumption B1 of a Poisson process with mean μ_i for the ith individual's purchases of the product-class over time rests on the basic observation that purchase incidence tends to be effectively independent of the incidence of previous purchases (for periods greater than some minimum like a week) and so irregular that it can be regarded *as if random* (see Section 1). The Poisson assumption and possible alternatives, like some Erlang distribution for interpurchase times, have been extensively discussed, especially for *brand* purchases (e.g. Ehrenberg, 1959, 1972; Herniter, 1971; Chatfield and Goodhardt, 1973; Dunn *et al.*, 1983). The Poisson process remains a workable approximation.

Assumption B2 of a Gamma mixing distribution for the values of μ_i can probably be justified as follows. If for different product-classes P, Q, R, S, etc. (like toothpaste, breakfast cereals, canned soup, etc.)

(1) the average purchase rate of P is independent of the rates for other products $Q, R, S, \ldots$, and

(2) P's proportion of a consumer's total purchases, namely $P/(P+Q+R+S+\ldots)$, is independent of her total rate of purchasing all the products,

then it can be shown, following a similar characterization for *brands* (e.g. Goodhardt and Chatfield, 1973; Chatfield, 1975), that the distribution of the mean rates of purchase of P must be Gamma. These independence conditions are likely to be approximately fulfilled in practice. Thus heavy buyers of toothpaste are not necessarily heavy buyers of canned soup, nor do heavy buyers of the whole range of products necessarily devote an above average proportion of their consumer goods expenditure to toothpaste. In both cases, any correlation—e.g. due to household size—is likely to be fairly low. No direct empirical analysis for different product-classes have yet been made, but much suitable data exists.

Assumption C is in line with general experience across a wide range of product-fields (e.g. Ehrenberg, 1972), including the specific cases referred to in this paper (e.g. Table 5). Shoemaker *et al.* (1977) have noted that in three product-fields some of the brands had varying shares among light, medium and heavy buyers, but the differences appear to have been small and not consistent.

2.3. *The Additivity Property*

An important property of the model is that any two brands j and k with means α_j/S and α_k/S can be combined into a super-brand with mean $(\alpha_j + \alpha_k)/S$. Nothing else in the specification of the Dirichlet model is affected (Wilks, 1962). This feature is not common to other such models (e.g. Herniter, 1971).

This additivity property helps to explain why minor brands can be grouped into an "all other brands" category and, more generally, why the model can cope when a brand is made up of different pack-sizes or flavours, or is bought from different retail outlets. It is also of marked computational help in estimating the model in Section 2.4 and dominates the algebraic short-cuts we use in calculating values of summary measures in Section 2.5.

2.4. *Fitting the Model*

In order to apply the Dirichlet model to a g-brand market, we must first estimate the values $\alpha_1, \alpha_2, \ldots, \alpha_g$ of Assumption A2 and the M and K of Assumption B2. In principle we assume that detailed data as in Table 1 are available for some base period (say 12 weeks) which we will take to be of unit length. In practice, we use summary statistics, namely penetrations and average purchase rates to fit the model.

Using the g observed *per capita* (or per household) average purchasing rates M_j as input variable, we can calculate the product-class purchasing rate M as $M = \Sigma m_j$, and equate the theoretical and observed market-shares (the means of the marginal Beta-distributions)

$$\alpha_j/\Sigma\alpha_j = m_j/\Sigma m_j.$$

As the brand shares must add up to one, there are only $(g - 1)$ independent equations of this form. To solve the equations we need to estimate $\Sigma\alpha_j$, which we call S—one of the two structural parameters of the model.

We believe that the simplest way to estimate S is to equate the observed proportion $1 - b_j$ of the population not buying brand j in the chosen time-period to the corresponding theoretical probability, solve for S, and then form a weighted average of the separate values of S across all g brands. The theoretical Dirichlet formula for $(1 - b_j)$ is however not in closed form for S, and hence needs an iterative solution. It also involves an infinite series for the NBD probabilities of buying the product-class and this calls for an *ad hoc* truncation rule. The computational details of what we do are set out in the Technical Appendix.

The parameter K is calculated by fitting an NBD to the distribution of purchases for the whole product class. If the distribution is reverse J-shaped, we fit by "mean and zeros", i.e. by solving $1 - B = (1 + M/K)^{-K}$ for K. This then has high efficiency (e.g. Anscombe, 1950; Ehrenberg, 1959, 1972; Chatfield, 1969). If the distribution is not reverse J-shaped, we use the method of moments. The theoretical NBD of product-class purchases can then be generated by

$$P_n = \left(1 + \frac{M}{K}\right)^{-K} \frac{\Gamma(K + n)}{n!\,\Gamma(K)} \left(\frac{M}{M + K}\right)^n$$

for $n = 0, 1, \ldots$ purchases and compared with the observed distribution in order to assess goodness-of-fit.

The Empirical-Dirichlet model

For some products the distribution of product-class purchases shows systematic deviations from an NBD. Toothpaste, which we use illustratively in Section 3, is an example. Table 2 shows that there is a small short-fall of once-only buyers. We believe that such discrepancies occur for "saturated" markets, i.e. products which generally have no direct substitute like toothpaste, and where the additivity type of property then does not apply at the product-class level. But more extensive empirical work is needed.

TABLE 2

Purchases of toothpaste in 12 weeks (fitted by mean and zeros)

	0	1	2	3	4	5	6+	*Av. no. per buyer*
Observed (%)	44†	19	14	9	6	3	4	2.6†
NBD (%)		22	13	8	5	3	5	

† Used in fitting.

The departure from an NBD here is not large enough to make us discard the NBD part of the model. But where it is, we can use the rest of the Dirichlet model by using the observed distribution as input instead of the fitted NBD. Thus in Section 2.5 we replace the theoretical NBD probability P_n in the various formulae by the *observed* proportion of consumers making n product purchases. The model is then referred to as the *Empirical-Dirichlet* model. It tends to give a better fit in the base period, but is less elegant and we are no longer able to make predictions for time-periods of a different length T. Thus we normally prefer to use the full NBD-Dirichlet model except when the goodness-of-fit is rather poorer than in Table 2.

2.5. *Applying the Model*

Having estimated the parameters, we can calculate the theoretical value of any specific aspect of buying behaviour, such as the summary measures outlined in Section 1.1. We can do this in the base-period and also for any period of length T ($T \gtrless 1$). The calculations are straightforward in principle though computationally tedious, even when using algebraic short-cuts.

Our main simplification is to try to reduce the calculations down to those of a Beta-Binomial distribution (e.g. Chatfield and Goodhardt, 1970). Thus in calculations for a specific brand j with brand-share α_j/S, we combine all the other brands into a single superbrand with brand-share $(S-\alpha_j)/S$ (see Section 2.3). This means that the probability of making r_j purchases of brand j, conditional on n purchases of the product-class having been made ($r_j \leqslant n$), is given by the Beta-Binomial distribution

$$p(r_j \mid n) = \binom{n}{r_j} B(\alpha_j + r_j, S-\alpha_j + n - r_j)/B(\alpha_j, S-\alpha_j),$$

where B here denotes the Beta function.

The proportion of consumers buying the product-class n times and buying brand j r_j times is then given by the product of the equations for P_n and $p(r_j \mid n)$

$$p(r_j, n) = P_n\, p(r_j \mid n).$$

(A capital P stands for probabilities of buying the product-class, a lower case p for those of buying a brand.) By summing $p(r_j, n)$ over appropriate values of n and r_j, we can calculate any statistic of interest for the brand. Summations over the values of n need to be truncated in the way described in the Technical Appendix for estimating S. In these computations both P_n and $p(r \mid n)$ can be calculated using appropriate recurrence relations.

Estimates for brand j

Examples for brand j (dropping the subscript j) are:

The penetration b of brand j is estimated as $1-p(0)$, the proportion not buying the brand, where

$$p(0) = \sum_{n=0} \{P_n\, p(0 \mid n)\}$$

and $p(0 \mid n)$ is the probability of making zero purchases of brand j given that n purchases of the product-class have been made in the analysis-period. The summation is again truncated.

Here $p(0 \mid n)$ is, from the Beta-Binomial, writing α for α_j.

$$p(0 \mid n) = \frac{(S-\alpha)(S-\alpha+1)\ldots(S-\alpha+n-1)}{S(S+1)\ldots(S+n-1)} \qquad \text{for } n \geqslant 1,$$

$$p(0 \mid 0) = 1.$$

The theoretical number of purchases of brand j per buyer of j is calculated as

$$w = \sum_{n=1} \left\{ P_n \sum_{r=1}^{n} rp(r \mid n) \right\} \Big/ [1 - p(0)],$$

and their average number of purchases of the product-class as

$$w_P = \sum_{n=1} \{nP_n [1 - p(0 \mid n)]\}/[1 - p(0)].$$

The proportion of the population who buy brand j only (the "sole" buyers) is given by

$$\sum_{n=1} \{P_n p(n \mid n)\},$$

since if they buy the product n times they must be buying the brand n times also. Their average purchase frequency per buyer is

$$\left\{ \sum_{n=1} [nP_n p(n \mid n)] \right\} \Big/ \left\{ \sum_{n=1} [P_n p(n \mid n)]. \right\}$$

In a period of length T, relative to the base period (with $T \gtrless 1$), all the above formulae are unchanged in the NBD-Dirichlet except that M in the NBD equation becomes MT.

To estimate the number of repeat buyers from one period of length T to another of equal length we enumerate b_{2T} for the double period and b_T for a single period and calculate $2b_T - b_{2T}$. But we have at this stage developed no simple way of calculating the theoretical average purchase frequency of such repeat-buyers in each period. Instead, we generally use the more convenient NBD/LSD theory in this area (see Section 4).

Duplicated buyers of brands j and k

Theoretical estimates of purchase combinations of two or more specific brands, j, k, etc. would generally require more number-crunching than we have yet attempted. (Another approach would be to use the model to construct theoretical purchases of a "sample" panel of 10 000 simulated households and then tabulate the relevant summary measures, as we routinely do with observed data.)

A simple method has however been developed for estimating the numbers of duplicated buyers of any particular pair of brands j and k, which has long been an important feature of the study of buyer behaviour (e.g. Section 1.1 (x), Tables 6 and 8 and Section 4). Here we use the additivity property of Section 2.3 in a different way. We form a composite brand $(j+k)$ and estimate its penetration $b_{(j+k)}$. This then gives us b_{jk}, the theoretical proportion of the population buying both brands at least once as

$$b_{jk} = b_j + b_k - b_{(j+k)},$$

and hence also "conditional" proportions like $b_{j/k} = b_{jk}/b_k$.

3. THE FIT OF THE MODEL

The Dirichlet model directly or indirectly describes the buying patterns that have been found for many different products, food and non-food, over many years, in the UK and USA, etc. (e.g. Ehrenberg and Goodhardt, 1979a, Table 5.3, and many other references). Thus in the relevant parameter range, the Dirichlet gives much the same results as our successful earlier sub-models (see Section 4.1). Direct analyses so far cover products like aviation fuel, petrol, toilet soap, soup, instant coffee, breakfast cereals and cosmetics.

Here we illustrate the fit of the model with results for toothpaste in the UK (Aske Research, 1975a). The input is in effect that in the January to March quarter of 1973, 56 per cent of the AGB panel of 5240 continuously-reporting households bought toothpaste on average 2.6 times (Table 2) which gives $M = 1.5$ and $K = 0.78$, and that the brand penetrations averaged 9 per cent to give $S = 1.2$. Table 3 shows how given the market-shares of the then eight leading brands, the model recovers the individual brand penetrations b, and purchase frequencies w and w_{p} for the quarter and, extrapolating, for the year as a whole (i.e. putting $T = 4$).

TABLE 3

Penetration and average purchase frequencies of toothpaste brands
(observed "O" and Dirichlet "D": fitted to quarterly data)

Leading brands (and market-shares)		Quarterly % Buying		Quarterly Av. purchases of Brand		Quarterly Av. purchases of Product		Annual % Buying		Annual Av. purchases of Brand		Annual Av. purchases of Product	
		O	D	O	D	O	D	O	D	O	D	O	D
Colgate DC	(25%)†	20	20	1.8	1.8	3.1	3.2	34	37	3.7	3.8	10	8.7
Macleans	(19%)†	17	17	1.7	1.7	3.0	3.3	32	32	3.2	3.6	10	8.9
Close Up	(10%)†	9	8	1.6	1.7	3.5	3.3	15	17	3.0	3.2	10	9.2
Signal	(10%)†	8	9	1.9	1.7	3.5	3.3	17	18	3.4	3.3	11	9.2
Ultrabrite	(9%)†	8	8	1.7	1.7	3.2	3.3	17	17	2.9	3.2	10	9.2
Gibbs SR	(8%)†	7	7	1.7	1.7	3.2	3.3	17	14	2.8	3.2	10	9.3
Boots Priv. Label	(3%)†	3	2	1.4	1.7	2.6	3.4	6	5	2.4	3.0	9	9.4
Sainsbury Priv. Lab.	(2%)†	2	2	1.5	1.6	3.1	3.4	3	4	3.2	3.0	11	9.5
Average		9†	9	1.7	1.7	3.2	3.3	18	18	3.1	3.3	10	9.2

[† Used in fitting]

For example, for Beecham's Macleans with a market-share of 19 per cent the model predicts:

(a) That in the quarter, $b = 17$ per cent of households bought it on average $w = 1.7$ times and made a total of $w_P = 3.3$ purchases of any toothpaste.

(b) That in the year as a whole, $b = 32$ per cent, $w = 3.6$ and $w_P = 8.9$.

The fit is close, e.g. to within about ± 1 for the b's across all eight brands. The only bias concerns w_P in the year, with $\bar{O} = 10$ and $\bar{D} = 9$. This seems due to the short-fall of once-only buyers in Table 2. Fitting the *Empirical Dirichlet* to the annual data gives $\bar{w}_P = 10$, as observed.

The predictions also reflect the small trends in w and w_P noted in Section 1.1(vii) and (viii), namely that

(c) The w's generally decrease with market-share. (The observed values jump around somewhat with subsample sizes down to 100. But the four largest brands average $\bar{O} = 3.3$ and $\bar{D} = 3.5$ in the year, and the four smallest $\bar{O} = 2.8$ and $\bar{D} = 3.1$.)

(d) There is a contrary small *upward* trend in the theoretical values of w_P, though not in the *observed* ones here. (It has however been observed in almost all the other cases analysed, e.g. Ehrenberg, 1972, Chapters 9 and 10; Ehrenberg and Goodhardt, 1976; Ehrenberg and Goodhardt, 1979a; Wrigley, 1980; Kau, 1981, and numerous Aske Research reports). It occurs as a statistical selection effect despite Assumption C and the independence structure of the model.

The Dirichlet model also successfully predicts the other measures of buying behaviour, such as

(e) The frequency distributions of purchases of the individual brands, as illustrated in Table 4.

TABLE 4

Distributions of annual purchases of toothpaste (two typical brands)

48 weeks		Number of purchases 0	1	2	3	4	5	6+
Macleans	Obs.	67	13	6	4	2	2	5
	Theo.	68	12	6	4	3	2	5
Sainsburys	Obs.	94	3	1	1	0.5	0.3	0.4
	Theo.	94	3	1	1	0.4	0.2	0.6
Average†	Obs.	81	8	4	2	1	1	3
	Theo.	81	8	4	2	1	1	3

† The eight leading brands.

(f) The penetration and average purchase frequency among frequent or infrequent buyers of the product-class, as shown in part in Table 5 (a rather direct test of the independence Assumption C).

TABLE 5

Penetration and purchase frequency among infrequent (1–6) *and very frequent* (13+) *buyers of toothpaste*

	Annual purchases of toothpaste							
	1–6 purchases				13 + purchases			
	b		w		b		w	
48 weeks	Obs. %	Theo. %	Obs.	Theo.	Obs. %	Theo. %	Obs.	Theo.
Colgate DC	35	37	1.8	1.8	65	68	6.3	6.2
Macleans	33	32	1.7	1.8	64	61	5.4	5.8
Close Up	24	24	1.7	1.7	54	44	4.7	5.2
Signal	14	14	1.8	1.7	44	40	5.3	4.9
Ultrabrite	17	16	1.6	1.6	35	35	4.7	4.7
Gibbs SR	17	16	1.8	1.6	34	34	4.4	4.7
Boots Priv. Label	6	5	1.9	1.6	11	12	3.0†	4.1
Sainsbury Priv. Label	2	3	1.9	1.5	8	8	4.5	4.1
Average	19	19	1.8	1.7	39	38	4.8	5.0

[† Sub-sample base: 81 households]

(g) The incidence and average purchase frequencies of 100 per cent-loyal or "sole" buyers of each brand, to within a few per cent both in short and long periods (not shown here in detail).

(h) The proportion of the quarterly buyers of Macleans who also buy a specific *other* brand at least once in the quarter, as shown in Table 6. The annual predictions here show a 10–20 per cent bias in line with that for w_P in Table 3 and also reflect some discrepancies in the trend with T of $b_{j/k}/b_j$ (Aske Research, 1975a gives details.) A related problem is that the parameter S in the model should be constant irrespective of T, but it varied from 1.0 when estimated from 4-week data to 2.2 for the 12-week data, to 1.8 in the year. This needs further study. Nonetheless, the predictions illustrated in (a) to (h) generally work well,

TABLE 6

Quarterly duplication of Macleans and the other brands (% buyers of Macleans also buying other brand)

	Colg. DC	*Macl.*	*Close Up*	*Signal*	*Ultra-brite*	*Gibbs SR*	*Boots PL*	*Sains. PL*
Observed (%)	28	(100)	10	9	9	10	2	1
Theoretical (%)	24	(100)	11	10	10	8	4	2

both here and in many other product-groups, as illustrated further in Table 8 in the next section.

3.1. *Uses of the Model*

The Dirichlet model summarizes many aspects of buying behaviour. Practical applications include providing interpretative norms, both for stationary and non-stationary markets.

Providing norms for stationary markets

Suppose some data show that 53 per cent of the buyers of Macleans toothpaste in a given quarter bought it again in the next quarter. The question is whether this is low (*only* 53 per cent repeat-buyers), or high (as *many* as 53 per cent), or what? The model's prediction of 51 per cent then tells us that the observed rate is neither particularly low nor high, but merely just about normal for any toothpaste brand with a 19 per cent market-share. (Similarly, had the observed figure been 65 per cent, we could judge it *high* only by having an interpretative norm.)

Such interpretative norms are needed whenever we deal with a new data set. For example, some years ago we were faced for the first time with a range of data from the USA. But we did not have to start from scratch: our earlier NBD model of repeat-buying had already been well established in the UK and was then found to apply to the US data as well. This could be summarized succinctly (Ehrenberg and Goodhardt, 1968), as "American and British repeat-buying habits are the same", a finding which has since been extensively confirmed (e.g. Aske Research, 1969d, 1974, 1975b).

Recent extensions to consumers' *store-choice* (e.g. Wrigley, 1980; Kau and Ehrenberg, 1984) could similarly be summarized as "Store-choice is like brand-choice". Table 7 illustrates in detail

TABLE 7

Penetration growth of instant coffee at different store-groups (four leading store-groups)

		Analysis-period (in weeks)			
		1 week	*6 weeks*	*12 weeks†*	*24 weeks*
Any instant coffee	Obs. (%):	17.0	54	65	76
	Theo. (%):	16.0	50	65†	77
Coop	Obs. (%):	3.8	16	20	28
	Theo. (%):	3.7	14	20†	27
Tesco	Obs. (%):	2.2	9	14	21
	Theo. (%):	2.0	9	14†	20
Independents	Obs. (%):	0.9	5	7	11
	Theo. (%):	1.0	4	7†	10
Fine Fare	Obs. (%):	0.4	3	3	5
	Theo. (%):	0.5	2	3†	5
Av. store-group‡	Obs. (%):	2.1	9	13	18
	Theo. (%):	2.1	8	13†	17

[‡ Eight leading groups.] [† Used in fitting.]

how the penetration growth of instant coffee at different store-groups can be closely predicted by the same Dirichlet or NBD models as used for brands (Kau, 1981).

Table 8 similarly shows the general fit of the model for half-yearly duplication of purchasing between the different store-groups. The model predicts that the duplication in each column should be about the same, effectively proportional to the brand penetrations (as in the old Duplication of Purchase Law—see Section 4.1), but with a very small upward trend with decreasing market-share, as already noted empirically (e.g. Ehrenberg, 1972). The fit for the individual duplications in Table 8 is to within a mean deviation of 4 percentage points, representing a 0.91 correlation between the observed and predicted values.

TABLE 8
Duplication of purchase of different store-groups (instant coffee)

		% also buying at							
24 weeks		*Misc.*	*Coop*	*Kwik*	*Tesco*	*Asda*	*indep*	*Symb*	*Fine*
Buyers at									
Miscellaneous	%	–	32	25	29	13	15	16	8
Coop	%	38	–	21	22	19	13	12	4
Kwiksave	%	37	26	–	22	9	11	9	7
Tesco	%	45	29	23	–	17	13	14	7
Asda	%	31	38	14	24	–	9	14	3
Independent	%	52	34	23	24	12	–	18	5
Symbol	%	52	32	20	30	20	19	–	8
Fine Fare	%	50	23	29	27	9	11	16	–
Average	%	43	31	22	25	14	13	14	6
1.2 X Penetration†	%	40	34	26	25	17	13	12	6
Penetration	%	33	28	22	21	14	11	10	5

†Approximate Dirichlet prediction.

A further illustration of the use of interpretative norms is that 10 years ago it was found that almost two-thirds of the customers of the London Business School's executive courses had sent only a single participant in 2 years. Should this have been a cause for concern? The fit of the NBD (Charlton and Ehrenberg, 1976a; Powell and Westwood, 1978; Sichel, 1982) showed that the pattern was little or no different from that for brands of breakfast cereals, detergents, toothpaste, petrol, or whatever. Similar findings arise with airline contracts for aviation fuel with oil companies across different airports (Ehrenberg, 1975), and with a consulting firm's clients (Coggill and Simpson, 1984).

More general conclusions include that advertising is not needed to create the normal levels of brand loyalty (Charlton and Ehrenberg, 1976b) and that buyer behaviour can be normal again *immediately* after a major market upset, with no lagged effects (Ehrenberg and Goodhardt, 1969; Ehrenberg, 1984b).

Interpreting change

Theoretical norms for a stationary market also provide a base-line for interpreting *change* (i.e. non-stationary situations), without having to match the results against an empirical "control sample". Thus just by comparing non-stationary data with the stationary norms we can assess whether an increase in sales came from attracting more buyers or from existing buyers buying more, and if the latter, whether it was heavy or light buyers doing so. Analyses of real life cases have been discussed elsewhere (e.g. Ehrenberg, 1972; Ehrenberg and Goodhardt, 1979a; Goodhardt and Ehrenberg, 1967, and D. Wellan, 1984) as well as examples involving deliberate experimentation (e.g. Charlton and Ehrenberg, 1976b; Ehrenberg, 1981, 1984b; Ehrenberg and Charlton, 1973a, b; Motes and Woodside, 1984; Motes *et al.*, 1984; Castleberry, 1983, and S. Castleberry *et al*, 1984).

Prescriptive uses

The model can also help us in decision-making. Table 3 implies that a brand like Ultrabrite could double its sales by getting more buyers, since other brands have higher b's. It might also seem that they could buy more Ultrabrite without having to use more toothpaste, since Ultrabrite's customers mostly buy other brands as well ($w_P = 10$ and $w = 3$ in the year). But the model tells us that this will not happen and is supported by the facts: w hardly differs for the different brands, whether observed or predicted. A marketing plan which aims at increasing sales by getting existing buyers to buy much more of the brand would therefore be aiming at something altogether unusual or unlikely, like making pigs fly. There is a law of nature against it, or so it would appear.

Understanding the nature of markets

The generalized empirical knowledge summarized by the model also helps us to understand better the nature of consumer goods markets. Examples are:

The number of customers which a brand (or retail store) has varies dramatically, but predictably, with the length of the time-period analysed.

Brand loyalty exists but is low and not exclusive (i.e. consumers habitually or "loyally" buy more than one brand).

Different aspects of loyalty, like repeat-buying and multi-brand buying, are directly related.

Brands hardly differ from each other in the degree of loyalty each attracts.

In the medium term, competitive marketing inputs (pricing, advertising, etc.) only show up in the brands' market-shares.

Most branded goods markets are largely unsegmented.

The structure of buyer behaviour is the same for radically different kinds of product-classes (like breakfast cereals or detergents), for different advertised brands (like Kellogg's Corn Flakes and Nabisco's Shredded Wheat), and irrespective of "exogenous" variables like the interest rate, or that the toothpaste data in Section 3 arose 2 years after the 1973 oil crisis, and while Mr Ford was President of the USA.

3.2. *The Nature of the Input*

The input required by the Dirichlet model is not only very parsimonious—the g values α_j and the two structural parameters S and K—but is also readily interpretable.

The sales levels

As we have already noted in Section 2.4, the g values α_j reflect the *per capita* or per household purchasing levels m_j of the g brands in the chosen unit time-period. A possible measurement simplification is that these values could be obtained from *sales* data (e.g. retail audits or ex-factory shipments), without having to observe individual consumers' purchases at all.

In practice we reformulate the m_j as $M = \Sigma m_j$, the average rate of purchase of the total product-class, and the market-shares m_j/M (which amount to $(g-1)$ independent values). With this formulation only M varies with T, the length of the time-period analysed.

The different brands in a product-class generally differ and compete with each other in many ways (product-formulation, pricing, packaging, advertising and promotion, retail distribution, usage patterns, etc.). But the close fit of the model shows that in an unsegmented near-stationary market all these other variables have generally no net effect on the structure of buyer behaviour, but are subsumed by the brands' market-shares. Such a strong causal interpretation is possible because it is negative (see also Ehrenberg 1982, Chapter 20). We are saying that lack of correlation probably implies lack of causation.

Two measures of diversity

The two parameters, K and S, are characteristics of the product-class and can be interpreted as reflecting the heterogeneity or diversity of consumers. Thus from Assumptions B1 and B2 in

Section 2.1, the standard deviation of the resultant NBD is given by $\sqrt{\{M(1+M/K)\}}$. Hence K reflects how much peoples' individual product-purchases differ from the overall mean M.

In Assumptions A1 and A2, the variance of the individual probabilities $(p_j)_i$ in the marginal Beta-distribution for brand j is $\{m_j(M-m_j)\}/\{M^2(1+S)\}$. S therefore reflects the extent that people differ from each other in their propensities to buy each brand. Thus if S is very large, the variance is near-zero and everyone has much the same probability m_j/M of buying brand j (minimum diversity). If S is small, the individual $(p_j)_i$ differ more across i. Indeed, the distribution reduces to just two "spikes" if $S=0$: a proportion m_j/M of individuals that always buy brand j, and a proportion $(1-m_j/M)$ that never do. (This is maximum diversity; but no one switches between brands.)

Neither of these two kinds of consumer variability need however be measured directly. Instead, we can in principle, and often in practice, estimate K from the observed values of M (which could be got from *sales* data, as already noted) and $(1-B)$, the observed proportion of non-buyers of the product. The latter is a qualitative rather than a quantitative form of input. We can similarly estimate S without using any quantitative measurement of peoples' brand-choice behaviour. In effect we use only $(1-\bar{b})$, a weighted average of the proportions of the population who do *not* buy each brand. One therefore does not need the individual details of continuous consumer panel data to activate the model.

4. THE BACKGROUND

In this section we briefly refer to our previous more limited models of buyer behaviour in Section 4.1, and give a brief review of other models of buying behaviour in Section 4.2.

4.1. *Our Previous Models*

The Dirichlet model describes stochastically when a purchase of the product-class is made and which brand is chosen. In contrast, our previous modelling started with when an individual brand was bought, using the Poisson-Gamma NBD model as in Assumptions B1 and B2, but applied to the individual brand, not the product-class (e.g. Ehrenberg, 1959). Finding various generalizable empirical regularities and developing theoretical formulations then generally went hand in hand. Next, we linked the results for different brands, for example that b and w for different brands j and k could be related as $w_j(1-b_j) \doteqdot w_k(1-b_k) = \text{constant}$. Finally, there were results concerning the extent to which any one consumer bought more than one brand over time, along the lines of (vi) to (x) of Section 1.1. Included here were results showing that the proportion of buyers of brand j who only buy j in the analysis-period should, on an independence assumption, equal $(1-B)/(1-b_j)$, and that brand duplication between brands j and k follows the Duplication of Purchase Law $b_{j/k} = Db_j$, where D is constant for all pairs of brands.

The derivation and generally close fit of such sub-models of specific aspects of buying behaviour have been described elsewhere in some detail (e.g. Chatfield *et al.*, 1966, Ehrenberg, 1972, Ehrenberg and Goodhardt, 1979a, and some 100 other papers and reports). The Duplication of Purchase Law and the level of w_P were, however, two empirically-established regularities for which no predictive theoretical models had been developed, but which are now also accounted for by the Dirichlet. Further extensions and applications of some of these earlier models were also developed by others (e.g. Rothman (SRS, 1965); Grahn, 1969; Morrison, 1969; Charlton *et al.*, 1972, Jephcott, 1972; Charlton and Ehrenberg, 1976a; Paull, 1978; Easton, 1979; Frisbie, 1980; Wrigley, 1980; Greene, 1982; Sichel, 1982).

The characterization of the Gamma-distribution in the NBD model for brands (Goodhardt and Chatfield, 1973) then led to the development of a special case of the Dirichlet model with complete independence between different brands as a direct generalization of the NBD model for specific brands (Aske Research, 1973; Chatfield and Goodhardt, 1975; Chatfield, 1975). This in turn led to the more general but mathematically different Dirichlet model (e.g. Aske Research, 1974, 1975b; Ehrenberg and Goodhardt, 1976). In the present paper we are bringing together these and more recent developments.

Our earlier models give numerically close predictions to those of the Dirichlet model, in the usual parameter range of the observed data. However, they are mathematically different. For example, the distribution of purchases of a given brand in the Dirichlet model is not an NBD (except in the case of independence, when $S = K$). Nonetheless, the two tend to agree closely as illustrated in Table 9; in general they agree if anything more closely with each other than either does with the observed data. The older sub-models like the NBD still tend to be used in practice because predictions for a single brand are computationally simpler and are not affected by non-stationarities for the *other* brands.

TABLE 9

The frequency distribution of brand purchases (observed, Empirical-Dirichlet and NBD)

		Number of purchases						
48 weeks		0	1	2	3	4	5	6+
Macleans	Obs. (%)	67	13	6	4	2	2	5
	Dir. (%)	68	12	6	4	3	2	5
	NBD (%)	(67)†	13	6	4	3	2	5
Average brand	Obs. (%)	81	8	4	2	1	1	3
	Dir. (%)	81	8	4	2	1	1	3
	NBD (%)	(81)†	8	4	2	1	1	3

† Used in fitting.

4.2. *Other Models*

Other published work on mathematical models of buyer behaviour has mainly been concerned with how markets change rather than with first describing the steady-state case (see Massy *et al.*, 1970; Montgomery and Ryan, 1974; Lilien and Kotler, 1983, for reviews). The first-order Markov model is the best-known example. It assumes that the probabilities of repeat-buying and of switching from brand j to k for successive purchases are invariant characteristics of each brand, i.e. remain constant even when the market-shares of the brands change. But empirically the reverse is the case. Repeat-buying and switching propensities tend to be independent of the specific brands and vary instead with their market-shares. The known facts and the Markov concept could not differ more (see also Ehrenberg, 1965) and Markov has slowly dropped out of favour.

A formulation which may be dealing with stationary markets is the Hendry model (Butler, 1966; Herniter, 1973; Kalwani and Morrison, 1977; Ehrenberg and Goodhardt, 1979b). It centres on the equivalent of the Duplication of Purchase Law $b_{j/k} = Db_j$, but only for pairs of successive purchases (where D becomes $S/(S + 1)$ in terms of the Dirichlet parameter). This ignores purchase incidence and population heterogeneity (as does the Markov model). In addition, the model assumes that D can be estimated from a rather incoherent "entropy" assumption (e.g. Ehrenberg and Goodhardt, 1973), which does not in itself lead to a good empirical fit (e.g. Ehrenberg and Goodhardt, 1974).

A more substantial approach to stationary markets is the work by Bass and his colleagues (e.g. Bass *et al.*, 1976; Jeuland *et al.*, 1980) who formulated models similar to the ones here, probably starting in part from Chatfield and Goodhardt (1975). But the work has lacked systematic evidence of empirical regularities to be fitted by the model. It also did not give characterizations of the distributions as in Section 2.2, nor an adequate estimation procedure for the crucial parameter S. More generally, the Dirichlet has been gaining increasing currency in recent discussions of consumer behaviour models as a convenient extension of the univariate beta-distribution. An early mention was Pyatt (1969).

5. DISCUSSION

The Dirichlet model successfully and parsimoniously describes many different aspects of buying behaviour in approximately stationary non-segmented markets. It makes explicit that there are simple, general and rather precise regularities in a substantial area of human behaviour where this has not always been expected.

In Section 5.1 we note that the main variations in the data are nonetheless predictable without the model's probabilistic formulation. In Section 5.2 we briefly contrast the model-building approach we have followed here with the regression approach which is so prevalent in modern statistics. In Section 5.3 we outline areas for further work.

5.1. *Predicting the Aggregate Patterns*

The predictability of the observed patterns does not depend on the Dirichlet model as such. We illustrate this by considering the values of b, w and w_P for buyers of a brand like Macleans toothpaste.

Table 10 gives the values of w for the average brands of a variety of product-classes and shows how we know from direct observation that w_P is generally much bigger than w (e.g. Aske Research, 1969–1981). We also know empirically that neither w nor w_P vary much from brand to brand. Hence we can predict from the toothpaste values in Table 10 that $w = 3$ and $w_P = 10$ for Macleans. This agrees with the observed Macleans result shown in Table 3 ($w = 3.2$ and $w_P = 9.6$, rounded to 10 there). We also know that there are small trends in w and w_P which can be modelled by heuristic curve-fitting or low-level theories (e.g. Ehrenberg, 1972) to improve the prediction marginally.

TABLE 10

Annual rates of buying (per buyer of a brand: w, $w_P - w$ and w_P)

	Average number of purchases of		
48 weeks	*The brand* (w)	*Other brands* ($w_P - w$)	*The product* (w_P)
Per buyer of the av. brand of			
Cigarettes†	50	100	150
Petrol	13	65	78
Biscuits	3	70	73
Chocolate Bars†	4	46	50
RTE Cereals (USA)	4	42	46
RTE Cereals (UK)	5	38	43
Margarine†	10	30	40
Detergents	7	32	39
Soup	7	23	30
Frozen Fish	4	20	24
Refrigerated Dough	4	17	21
Take-home Beer†	5	8	13
Tinned Spaghetti	4	8	12
Toothpaste	3	7	10
Nail Polish	1.5	1.4	2.9
Average	8	34	42

† Estimated from 12- or 24-week data.

The variation of w and w_P with T, the length of the analysis periods, is also broadly predictable without any underlying model. In Table 3, we used as an input to the model the observation that quarterly buyers of toothpaste bought the product on average 2.6 times. From this much follows. Thus simple statistical insight and empirical experience suggest that w_P for any specific brand like Macleans will be somewhat higher (the observed w_P is 3.2). Experience of the less than *pro rata* increases from a quarter to a year also suggests that the *yearly* w and w_P will be somewhat less than four times the quarterly ones. (Suitable formulae are $(w_T - 1) \doteqdot T^{0.82}(w - 1)$, and $(w_{PT} - 1) \doteqdot T(w_P - 1)$; e.g. Ehrenberg, 1972.) In a stationary market the annual value of b follows from the Sales Equation of Section 1.1(v).

Ad hoc systematizing of regular empirical patterns can therefore lead to much the same resul as the predictions in Table 3, which at first sight seemed perhaps rather remarkable: Merely fro Macleans' quarterly market-share of 19 per cent one was successfully led to predict that abo 32 per cent of the sample bought the brand in the year on average 3.2 times, and also bough other brands about 7 times.

Even the consumption habits and brand preferences of individual consumers appear to be fairl stable in the short- to medium-term. Indeed, that is what the underlying specification of th Dirichlet model assumes, and the resulting predictions tend to fit closely. But why one person (o household) generally consumes more toothpaste or soup than others, or somewhat prefers bran j to k or vice versa, is not accounted for by the model and is in fact at this stage still largel unknown. (Obvious demographic factors like household size or income only account for a portio of the variation.) Nevertheless, we believe that this will in due course turn out to be largel explainable.

In any event, a parsimonious stochastic model like the Dirichlet is a very convenient analysi tool (e.g. Section 3). It models the "as if random" variability of precise purchase incidence an specific brand-choice by simple "independent" probability processes. It also makes explicit tha the different aspects of buyer behaviour under stationary conditions are all interrelated and tha they can be very largely accounted for by the specified inputs of the model.

5.2. *The Approach to Model-Building*

The type of model-building in this paper has little in common with the fitting of regressio equations which accounts for much of the modelling found in the statistical literature. Insteac of the typical *ad hoc* equation which is in some sense "best", though only for the particula sample data it is fitted to, the approach followed here leads to results of potentially lasting value We believe it does so because we started with empirical knowledge of generalizable patterns laboriously identified over some 30 years and based on many different sets of data (for differen brands, products, time-periods, countries, etc.). The modelling now merely seeks to synthesize this prior knowledge (e.g. Ehrenberg, 1984a). This is quite unlike the analytic use of regressior to instantly discover and simultaneously describe a previously unknown relationship.

The model described here is special in that:

(i) Most of its predictive power lies in the shape of the distributions involved rather than just in the values of fitted parameters.

(ii) Nevertheless, the shapes of these distributions are the necessary results of quite simple, empirically verifiable "independences" (as proved by the characterizations of Section 2.2).

(iii) The two structural parameters K and S are not just *ad hoc* numerical coefficients, but measures of identifiable properties of specific aspects of the behaviour described.

(iv) From a single "independent" variable for each brand, namely its sales level, a wide variety of "dependent" variables such as $b, w, w_P, b_{j/k}$, etc. are successfully predicted, always using the same two parameters, K and S.

(v) These "dependent" variables can be literally predicted rather than merely "fitted"; i.e. it is not necessary to observe the dependent variables in any given situation before fitting the appropriate model.

5.3. *Further Work*

Although the Dirichlet model now provides us with a good tool for handling stationary purchasing data, a good deal of further work still needs to be done. Tables 3–8 show some discrepancies from the model. They are mostly small and may be of little marketing importance. But are they generalizable and, if so, can they then be related to other variables, like pricing or advertising or future market trends?

Other discrepancies concern the lack of constancy of the parameter S noted in Section 3, and similar problems for the trend of the duplication coefficient D with T, and the cases where the

NBD does not fully fit (the "Empirical-Dirichlet" situation of Table 2). There is also one product, cigarettes, for which repeat-buying appears to differ from the NBD/Dirichlet predictions (Aske Research, 1971).

Analytically, developing closed-form approximations to the Dirichlet would be helpful, especially where NBD/LSD approximations do not already exist (e.g. for the duplication coefficient D or for w_P). We also need to explore the numerical nature of the model more fully.

An empirical question is whether the two basic "diversity" parameters K and S are product-class characteristics which are the same for different demographic sub-groups, different store-groups, different countries, and so on.

Extensions to other market conditions, types of economies and types of products beyond frequently-bought branded consumer goods are also needed. Some have already been started, as noted in Section 3. But more can be done, for example with phenomena where the NBD, Duplication Law or related types of patterns are already known to occur. This includes people's viewing of television programmes (e.g. Goodhardt *et al.*, 1975), the readership of print media (e.g. Hyett, 1958; Leckenby and Kishi, 1984), the incidence of human accidents (e.g. Irwin, 1964) and the original NBD studies on the abundance of species (e.g. Fisher *et al.*, 1943).

We can also increase the number of choice dimensions from one—i.e. which brand of a product is chosen, or at which store the product is bought—to cover both brand and store choice (Kau and Ehrenberg, 1984), or brand and flavour, brand and pack-size, etc. The choices might be hierarchical, e.g. first brand then flavour, or vice versa (Goodhardt in discussion on Aitchison, 1982).

A major extension will be to modelling non-stationary markets. But here we first need many empirical studies to see what if any generalizable patterns exist (e.g. Ehrenberg, 1982, 1984b, and D. Wellan, 1984).

More generally still, there is also scope for using the model to help our understanding of competition, of marketing inputs like advertising, price, product-quality and retail distribution, and of the psychologically "low-involvement" nature of consumers' brand-choice (e.g. Barwise, 1984).

ACKNOWLEDGEMENTS

We are indebted to AGB Research for the supply of data tapes and to Beecham Products for permission to quote extensively from an Aske Research Report prepared for them. Preparation of this paper has in part been supported by AGB, British American Cosmetics, Lever Brothers, London Weekend Television, Kellogg's, Mars, RHM Foods, Thames TV, J. Walter Thompson, and United Biscuits. Some of our earlier work on the Dirichlet arose from analyses for Unilever and the Market Research Corporation of America.

G. J. Goodhardt is Sir John E. Cohen Professor of Consumer Studies at City University Business School, and A. S. C. Ehrenberg is WCM Professor of Marketing and Communication at the London Business School.

TECHNICAL APPENDIX

Estimating the Dirichlet Parameter S

This Appendix describes how we have in practice calculated the structural parameter $S = \Sigma\alpha_j$. As noted in Section 2.4, we form an estimate of S iteratively for each brand separately. We then combine these as a weighted average. (Some improvement in the procedures could probably now be made.)

The simplest estimating statistic to use appears to be the number of non-buyers of a brand. By the additivity property of the Dirichlet (Section 2.3), the distribution of purchases of brand j reduces to a Beta-Binomial distribution for "brand j" or "not-brand j". For consumers buying the product-class n times in the analysis-period, the conditional probability of *not* buying brand j is therefore

$$p(0 \mid n) = \frac{(S-\alpha_j)(S-\alpha_j+1)\ldots(S-\alpha_j+n-1)}{S(S+1)\ldots(S+n-1)}, \tag{A1}$$

where $\alpha_j = Sm_j/M$ from Section 2.4.

Multiplying this by the probability P_n of making n purchases (estimated from the fitted NBD a in Section 2.4) gives the overall probability of not buying brand j:

$$p(0) = \sum_{n=1}^{\infty} \{P_n\, p(0 \mid n)\}. \qquad \text{(A2}$$

This can then be equated to the observed proportion of non-buyers of j, $1 - b_j$, and solved for S We deal with the infinite summation in equation (A2) and the non-closed form of (A1) for S a follows.

A truncation procedure

To equate (A2) with $(1 - b_j)$, we must approximate (A2) by a finite sum. This we do by firs truncating the NBD probabilities at a value n^*, where those in the tail add to less than 0.1 per cen (say). Then we find a value n' such that the truncated distribution up to n^* plus a lump at n reproduces the mean:

$$\sum_{n=1}^{n^*} (nP_n) + n' \left(1 - \sum_{n=0}^{n^*} P_n\right) = M.$$

Since n' is generally non-integral, we separate out this lump without bias at $[n']$, the integral valu of n', and at $([n'] + 1)$, by determining two probabilities $P_{[n']}$ and $P_{[n']+1}$ such that

$$P_{[n']} + P_{[n']+1} = 1 - \sum_{n=0}^{n^*} P_n$$

and

$$[n']\, P_{[n']} + ([n'] + 1) P_{[n']+1} = n' \left(1 - \sum_{n=0}^{n^*} P_n\right).$$

The two probabilities are now inserted into (A2) to give an approximate finite sum so that we can equate

$$\Sigma \{P_n\, p(0 \mid n)\} = 1 - b_j \qquad \text{(A3)}$$

where the summation includes terms up to n^* and then at $[n']$ and $([n'] + 1)$.

Solving for S

Equation (A3) involves (A1) and cannot be solved analytically for S, and so we proceed by iteration. We start with some reasonable value for S, say S', This could be derived from previous knowledge or taken as unity. We then calculate the corresponding value of the left-hand side of (A3) and compare it with the observed value of $(1 - b_j)$. If it is greater, we increase the trial value of S by trying $S'' = 2S'$; if it is smaller we try $S'' = S'/2$. At the next iteration we interpolate or extrapolate linearly from S' and S'' to produce a new trial value S''', and so on. We continue until the observed and predicted penetrations differ by less than 0.0001, say. Three or four iterations are usually sufficient.

Combining the estimates for all brands

We estimate S in this way for each brand, producing a series of g trial estimates $S_1, S_2, \ldots, S_g$. One advantage is that we can check that the values of S are mostly much the same for the different brands. If one or two values are very different, we leave these values out and proceed with the rest as being more or less homogeneous. (One can still apply the model to all the brands and see where it fits and where it fails to fit.)

Given values of S_j for g^* ($\leqslant g$) brands, we find a weighted average to produce a single overall estimate by calculating

$$\hat{S} = (\sum_{g^*} S_j m_j/M)/(\sum_{g^*} m_j/M). \tag{A4}$$

This formula weights the S_j according to brand-share and therefore gives more importance to "large" brands. The brand-share is approximately proportional to brand penetration and hence to the sample size buying each brand. When all g brands are included in (A4), the denominator is unity and $\hat{S}$ is just the sum of the g α's estimated separately for each brand.

Although the penetrations of each brand were used in the process of estimating the values of $\{S_j\}$, the use of $\hat{S}$ from (A4) is an overall one and the Dirichlet penetration norms are not constrained to equal the observed penetrations.

REFERENCES

Aitchison, J. (1982) The statistical analysis of compositional data. *J. R. Statist. Soc.* B, **44**, 139–177.

Anscombe, F. J. (1950) Sampling theory of negative binomial and logarithmic series distributions, *Biometrika*, **37**, 358–382.

Aske Research (1969–1981) *Dentifrice* (1969); *Ready-to-Eat Cereals* (1969); *The Structure of the Detergent Market* (1969); *Petrol* (1972); *The Structure of the Biscuit Market* (1976); *Colour Cosmetics* (1981); London: Aske Research Ltd.†

——— (1969d, 1974, 1975b) *Refrigerated Dough: An Initial Report* (1969d);, *Ready-to-Eat Cereals* (1974); *The Dirichlet Model* (1975b). New York: Aske Research Inc.†

——— (1971, 1973, 1975a) *Cigarettes: Spring 1969* (1971); *Results Concerning Brand-Choice* (1973); *The Structure of the Toothpaste Market* (1975a). London: Aske Research Ltd.†

Barwise, T. P. (1984) Mass attitudes and routine choice behaviour. Ph.D. Thesis, London University.

Bass, F. N., Jeuland, A. and Wright, G. P. (1976) Equilibrium stochastic choice and market penetration theories. *Management Sci.*, **22**, 1051–1063.

Buck, S. F. (1982) Consumer panels in the UK: past, present and future. In *Proceedings of the 25th Annual Conference*. London: The Market Research Society.

Butler, B. F. (1966) *Hendrodynamics: Fundamental Laws of Consumer Dynamics*. Croton-on-Hudson: The Hendry Corporation.

Castleberry, S. (1983) Longitudinal field-experiment of repeat-purchase behaviour: Effects of varying intrinsic quality, price, and brand availability. Ph.D. Thesis, University of Alabama.

Castleberry, S., Ehrenberg, A. S. C. and Motes, W. H. (1984) *An Extended Sales Test of Product Quality*. Univ. of Georgia: Working Paper.

Charlton, P. and Ehrenberg, A. S. C. (1976a) Customers of the LEP. *Appl. Statist.*, **25**, 26–30.

——— (1976b) An experiment in brand choice. *J. Marketing Res.*, **13**, 152–160.

Charlton, P., Ehrenberg, A. S. C. and Pymont, B. (1972) Buyer behaviour under mini-test conditions. *J. Market Res. Soc.*, **14**, 171–183.

Chatfield, C. (1969) On estimating the parameters of the Logarithmic and Negative Binomial distributions. *Biometrika*, **56**, 411–414.

——— (1975) A marketing application of a characterisation theorem. In *Statistical Distributions in Scientific Work*, Vol. 2. (G. P. Patil, S. Kotz and J. K. Ord, eds). Dordrecht: D. Reidel.

Chatfield, C., Ehrenberg, A. S. C. and Goodhardt, G. J. (1966) Progress on a simplified model of stationary purchasing behaviour. *J. R. Statist. Soc.* A, **129**, 316–367.

Chatfield, C. and Goodhardt, G. J. (1970) The Beta-Binomial model for consumer purchasing behaviour. *Appl. Statist.*, **19**, 240–250.

——— (1973) A consumer purchasing model with Erlang inter-purchase times. *J. Amer. Statist. Ass.*, **68**, 828–835.

——— (1975) Results concerning brand-choice. *J. Marketing Res.*, **12**, 110–113.

Coggill, C. and Simpson, S. (1984) *Analysis of Cambridge Division Sales*. Royston: PA Technology.

Connor, R. J. and Mosimann, J. E. (1969) Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Amer. Statist. Ass.*, **64**, 194–206.

Darroch, J. N. and James, I. R. (1974) *F*-independence and null correlation of continuous, bounded-sum, positive variables. *J. R. Statist. Soc.* B, **36**, 467–483.

Darroch, J. N. and Ratcliffe, D. (1971) A characterization of the Dirichlet distribution. *J. Amer. Statist. Ass.*, **66**, 641–643.

† Access can be obtained through the authors.

Dunn, R., Reader, S. and Wrigley, N. (1983) An investigation of the assumptions of the NBD model as applied to purchasing at individual stores. *Appl. Statist.*, **32**, 249–259.

Easton, G. (1979) Stochastic models of industrial buying behaviour. *Omega*, **8**, 63–69.

Ehrenberg, A. S. C. (1959) The pattern of consumer purchases, *Appl. Statist.*, **8**, 26–4.

——— (1965) An appraisal of Markov brand-switching models. *J. Marketing Res.*, **2**, 347–362.

——— (1972) *Repeat-Buying: Theory and Applications*. Amsterdam and New York: North Holland and Elsevier.

——— (1975) The structure of an industrial market: the case of aviation fuel. *Industrial Marketing Management*, **4**, 273–285.

——— (1981) Marketing experiments. In *Proceedings of the 24th Annual Conference*. London: The Market Research Society.

——— (1982) *A Primer in Data Reduction*. London and New York: Wiley.

——— (1984a) Data analysis with prior knowledge, In *Proceedings of the 50th Anniversary Conference of the ISU Statlab*. Ames: Iowa University Press (in press).

——— (1984b) The 1982/3 LBS marketing experiment. Working Paper, London Business School.

Ehrenberg, A. S. C. and Charlton, P. (1973a) McConnell's experimental brand-choice data. *J. Marketing Res.*, **10**, 302–307.

——— (1973b) The analysis of simulated brand-choice. *J. Advert. Res.*, **13**, 21–33.

Ehrenberg, A. S. C. and Goodhardt, G. J. (1968) A comparison of American and British repeat-buying habits. *J. Marketing Res.*, **5**, 15–18.

——— (1969) The evaluation of a consumer panel. *Admap*, **5**, 388–393.

——— (1973) Setting budgets and allocating advertising effects. In *Proceedings of the 19th Annual Conference*. New York: The Advertising Research Foundation.

——— (1974) The Hendry brand-switching coefficient. *Admap*, **10**, 232–238.

——— (1976) *Decision Models and Descriptive Models in Marketing*. Cambridge: Marketing Science Institute.

——— (1979a) *Understanding Buyer Behavior*. New York: J. Walter Thompson and MRCA.

——— (1979b) The switching constant. *Management Sci.*, **25**, 703–705.

Fisher, R. A., Corbett, R. S. and Williams, C. E. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Animal Biol.*, **12**, 42–58.

Frisbie, G. A. (1980) Ehrenberg's Negative Binomial model applied to grocery store trips. *J. Marketing Res.*, **17**, 385–390.

Goodhardt, G. J. and Chatfield, C. (1973) Gamma-distribution in consumer purchasing. *Nature*, **244**, No. 5414, 316.

Goodhardt, G. J. and Ehrenberg, A. S. C. (1967) Conditional trend analysis: a breakdown by initial purchase level. *J. Marketing Res.*, **4**, 155–161.

Goodhardt, G. J., Ehrenberg, A. S. C. and Collins, M. A. (1975) *The Television Audience: Patterns of Viewing*. Aldershot: Gower Press.

Grahn, G. L. (1969) The Negative Binomial distribution model of repeat-purchase loyalty: an empirical investigation. *J. Marketing Res.*, **6**, 72–78.

Greene, J. (1982) *Consumer Behavior Models for Non-Statisticians*. New York: Praeger.

Gurland, J. (1957) Some interrelations among compound and generalised distributions. *Biometrika*, **44**, 265–268.

Herniter, J. (1971) A probabilistic market model of purchase timing and brand selection. *Management Sci.*, **18**, 102–113.

——— (1973) An entropy model of brand purchase behavior. *J. Advert. Res.*, **10**, 361–375.

Hyett, G. P. (1958) *The Measurement of Readership*. Seminar paper at the London School of Economics.†

Irwin, J. O. (1964) The personal factor in accidents – a review article. *J. R. Statist. Soc.* A, **129**, 438–451.

James, I. R. (1975) Multivariate distributions which have Beta conditional distributions. *J. Amer. Statist. Ass.*, **70**, 681–684.

Jephcott, J. St. G. (1972) Consumer loyalty – a fresh look. In *Proceedings of the 15th Annual Conference*, London: The Market Research Society.

Jeuland, A. P., Bass, F. and Wright, G. P. (1980) A multibrand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes. *Operations Res.*, **28**, 255–277.

Kalwani, M. U. and Morrison, D. G. (1977) A parsimonious description of the Hendry system. *Management Sci.*, **27**, 467–477.

Kau, Ah Keng (1981) Patterns of store choice. Ph.D. Thesis, University of London.

Kau, A. K. and Ehrenberg, A. S. C. (1984) Patterns of store-choice. *J. Marketing Res.*, **21** (in press).

Leckenby, J. D. and Kishi, S. (1984) The Dirichlet multinomial distribution as a magazine exposure model. *J. Marketing Res.*, **21**, 100–106.

Lilien, G. and Kotler P. (1983) *Marketing Decision-Making: A Model-Building Approach*. New York: Harper & Row.

Luce, R. D. (1959) *Individual Choice Behaviour: A Theoretical Analysis*. New York: Wiley.

Massy, W. F., Morrison, D. G. and Montgomery, D. B. (1970) *Stochastic Models of Buyer Behavior*. Boston: MIT Press

McPhee, W. N. (1963) *Formal Theories of Mass Behavior*. New York: Free Press.

Montgomery, D. B. and Ryan, A. B. (1974) Stochastic models of consumer choice behavior. In *Consumer Behavior: Theoretical Foundations* (T. S. Robertson and J. Scott Ward, eds).

Morrison, D. G. (1969) Conditional trend analysis: a model that allows for non-users. *J. Marketing Res.*, **6**, 342–346.

Mosimann, J. E. (1962) On the compound multinomial distribution, the multivariate *B*-distribution, and correlations among proportions. *Biometrika*, **49**, 65–82.

——— (1984) Size and shape analysis. In *Encyclopaedia of the Statistical Sciences* (S. Kotz and N. L. Johnson eds), New York: Wiley (in press).

Motes, W., Castleberry, S. B. and Motes, S. G. (1984) A longitudinal test of price effects on brand choice behavior. *J. Bus. Res.*, (in press).

Motes, W. and Woodside, A. G. (1984) Field test of package advertising effects on brand choice behavior. *J. Advert. Res.*, **24**, 39–45.

Paull, A. E. (1978) A generalised Poisson model for consumer purchase panel data analysis. *J. Amer. Statist. Ass.*, **73**, 706–713.

Pearson, K. (1897) Mathematical contributions to the theory of evolution: on a form of spurious correlation which may arise when indices are used in the measurements of organs. *Proc. Roy. Soc.*, **60**, 489–498.

Powell, N. and Westwood, J. (1978) Buyer-behaviour in management education. *Appl. Statist.*, **27**, 69–72.

Pyatt, G. F. (1969) A model of brand-loyalties. In *Proceedings of the 1968 CEIR/SCICON Symposium on Model-Building in Business and Government* (M. G. Kendall, ed.), London: Griffin.

Shoemaker, R. W., Staelin, R., Kadane, J. B. and Shoaf, F. R. (1977) Relation of brand-choice to purchase frequency. *J. Marketing Res.*, **14**, 458–468.

Sichel, H. S. (1982) Repeat-buying and the Poisson-generalised inverse Gaussian distribution. *Appl. Statist.*, **31**, 193–204.

SRS (1965) *The SRS Motorist Panel*. London: Sales Research Services.

Wellan, D. (1984) *Shield's Launch*. London Business School: Working Paper.

Wilks, S. S. (1962) *Mathematical Statistics*. New York: Wiley.

Wrigley, N. (1980) An approach to the modelling of shop-choice patterns. In *Geography and the Urban Environment*, Vol. III. (D. T. Herbert and R. J. Johnston, eds). London and New York: Wiley.

DISCUSSION OF THE PAPER BY PROFESSOR GOODHARDT, PROFESSOR EHRENBERG AND DR CHATFIELD

Professor D. J. Bartholomew (London School of Economics): This paper provides a welcome opportunity for the Society to discuss the stochastic modelling of human behaviour – whether it is possible and how it should be done. The possibility should not be taken for granted. I was recently told by an experienced social statistician that he did not believe that models had any place in social statistics. Such a view is difficult to maintain in the face of empirical evidence of the kind given in the paper but such beliefs are strongly held. Given the possibility there are two broad approaches. One, using multiple regression or time series methods, builds on the correlation structure of the data. In Stochastic Models for Social Processes (Bartholomew, 1983, p. 4) I called these "black box" models because they merely relate the explanatory variables to the dependent variable without attempting to model the mechanism by which one is turned into the other. Like the authors, I think that this approach is very much the second best. The second course, followed here, is to identify the basic elements of the choice process at the individual level, specify the probability distributions associated with each and then link them together to give the aggregate behaviour. Not only does such a model give greater insight but we can extrapolate its behaviour with more confidence since its component parts can be separately validated.

The authors express the qualms that many have felt about the propriety of treating people as if they made decisions by tossing "mental pennies". I, myself, would not rule out the possibility of a purely random element in impulsive behaviour but the justification for using probabilities does not depend on this. Many purely deterministic processes display random aspects. The generation of pseudo-random numbers in computers is a good example. A short run from such a sequence betrays no sign of its deterministic origins. Human behaviour is somewhat similar. There is simply not enough information in the few overt acts to identify the complex web of motivation which lies behind them. Adding to this the high degree of independence between individuals it becomes almost inevitable that collective action will show random characteristics.

The paper exemplifies several interesting aspects of modelling and it may be useful to relate them to a wider social context. Variation between individuals is universal and a successful model must reckon with the fact. The authors introduce such variation at two points; the rate at which individuals make purchases and their relative preferences as expressed by the choice probabilities $\{p_j\}$. Variation of the former kind is commonly incorporated into occupational mobility models and the use of the Dirichlet distribution for a set of proportions, which goes back at least to Martin (1967), was used in a social context by Bartholomew (1975).

The particular choice of mixing distribution is usually dictated by mathematical convenience but one of the most elegant features of the present paper is that it shows that the choice need not be arbitrary. By the use of characterization theorems the authors are able to show that certain empirical features which the model should reproduce virtually determine the form of the mixing distribution (gamma for the rate of purchasing and Dirichlet for the choice vector). The link here with Mosimann's work on size and shape is indicative of other potential applications.

Even a relatively complex model, like the present one, can be criticized on the grounds of over-simplification. To do so is to overlook another important function which social models serve. The authors refer to their use as norms and this is very similar to the idea of a "base-line" model as the term is used among sociologists (see Mayhew (1984)). A base-line model embodies the simplest possible stochastic assumptions. Its aim is not so much to describe the system accurately as to provide a bench mark in relation to which what actually occurs can be interpreted. In a sense it "takes out" of the observed pattern what can be explained on minimal assumptions and so focusses on what remains to be explained. This approach has proved very useful in practice.

Thus far my remarks have picked out some of the many admirable features of the paper but tradition demands that congratulation should be tempered with criticism. It seems to me that an opportunity may have been missed of improving the summary measures used in marketing. The various quantities listed in 1.1 are mainly functions of time. Functions in their entirety are not easy to assimilate and so their values at fixed points, 12 weeks say, are used instead. This is somewhat arbitrary and liable to conceal as much as it reveals. Should we not be encouraging market researchers to describe their data in terms, for example, of the alphas, K and S? The authors show in 5.2(iv) that straightforward meanings can be given to them. In other areas of social science (e.g. labour turnover and social mobility) models have thrown considerable light on how to measure such complex phenomena. It appears that there are similar opportunities in market research.

There are many other matters which I should like to raise on methods of estimation and on goodness of fit – and whether it matters – but these must be left for another contributor or another day. It remains only to congratulate the authors and on behalf of the Society to propose a warm vote of thanks.

Dr A. W. Kemp (University of St. Andrews): I would like to begin by congratulating the authors on the clear exposition of their assumptions and their estimation procedure. Their demonstration of the evident practical utility of this work is particularly impressive.

I would also like to thank the authors for provoking several hours of fruitful argument concerning the exact nature of the subject of the paper, that is to say the Dirichlet model. Statistical modelling has two stages. Firstly there is the abstraction from a practical situation to a scientific model which makes hopefully realistic assumptions and preserves the essential features of the situation. Secondly there is the formulation of this scientific model in mathematical terms, thus providing the statistical model. Statisticians are sometimes alleged to be better at mathematical formulation than at scientific understanding. Here possibly the reverse is true. I will therefore centre my remarks on the mathematical aspects of the model and their ramifications.

The authors' formulation of the NBD – Dirichlet model, or Dirichlet for short, is

$$[\mathscr{M}(\mathbf{r} \mid n, \mathbf{p}) \underset{\mathbf{p}}{\wedge} \mathscr{D}(\mathbf{p} \mid \boldsymbol{\alpha})] \underset{n}{\wedge} [\mathscr{P}(n \mid \mu) \underset{\mu}{\wedge} \mathscr{G}(\mu \mid K, K/MT)]$$

where $\mathscr{M}$, $\mathscr{D}$, $\mathscr{P}$ and $\mathscr{G}$ denote the Multinomial, Dirichlet, Poisson and Gamma distributions. Note that I have adjusted the parameterizations of the component distributions to accord with standard practice. Now the multinomial distribution has probability generating function (p.g.f.) $(p_1 z_1 + \ldots + p_{g-1} z_{g-1} + p_g)^n$ where $\Sigma p_j = 1$, $j = 1, \ldots, g$, whereas the singular multinomial distribution has p.g.f. $(p_1 z_1 + \ldots + p_{g-1} z_{g-1} + p_g z_g)^n$ where $\Sigma p_j = 1$, $j = 1, \ldots, g$. The difference is non-trivial in the present context. Only the singular multinomial distribution will yield the joint distribution for all g brands – without the use of the generating variable z_g information on the number of purchases of the gth brand will be suppressed. The same difficulty arises if one works with the multinomial probabilities rather than the non-singular p.g.f.

The second source of argument has been mixing *vs* generalizing. For me the more natural formulation of the model is to say that total product purchases have the distribution

$$\mathscr{P}(n \mid \mu) \underset{\mu}{\wedge} \mathscr{G}(\mu \mid K, K/MT)$$

whilst the distribution between brands is

$$\mathscr{M}(\mathbf{r} \mid n, \mathbf{p}) \underset{\mathbf{p}}{\wedge} \mathscr{D}(\mathbf{p} \mid \boldsymbol{\alpha}),$$

and consequently the model is

$$[\mathscr{P}(n \mid \mu) \underset{\mu}{\wedge} \mathscr{G}(\mu \mid K, K/MT] \;\vee\; [\mathscr{M}(\mathbf{r} \mid n, \mathbf{p}) \underset{\mathbf{p}}{\wedge} \mathscr{D}(\mathbf{p} \mid \boldsymbol{\alpha})]$$

using Gurland's (1957) notation, where $F_1 \vee F_2$ symbolizes the p.g.f. $g_1(g_2(z_1,\ldots,z_g \mid \phi, \boldsymbol{\alpha}))$. Now if

$$\{g_2(z_1, \ldots, z_g \mid \phi, \boldsymbol{\alpha})\}^n = g_2(z_1, \ldots, z_g \mid \phi n, \alpha)$$

then

$$\begin{aligned} F_1 \vee F_2 &= \sum_n P_n \{g_2(z_1, \ldots, z_g \mid \phi, \boldsymbol{\alpha})\}^n \\ &= \sum_n P_n\, g_2(z_1, \ldots, z_g \mid \phi n, \boldsymbol{\alpha}) = F_2 \underset{n}{\wedge} F_1, \end{aligned}$$

(sometimes known as Gurland's theorem). This condition does not here hold immediately – careful examination of the interchangeability of the various integrations and summation is needed to convince one that

$$[\mathscr{P} \underset{\mu}{\wedge} \mathscr{G}] \vee [\mathscr{M} \underset{\mathbf{p}}{\wedge} \mathscr{D}] \sim [\mathscr{M} \underset{\mathbf{p}}{\wedge} \mathscr{D}] \underset{n}{\wedge} [\mathscr{P} \underset{\mu}{\wedge} \mathscr{G}].$$

Once this has been established many other mathematically equivalent formulations can be made, such as

$$\begin{aligned} &[\mathscr{P} \vee \text{Logarithmic}] \vee [\mathscr{M} \underset{\mathbf{p}}{\wedge} \mathscr{D}] \sim [\text{Neg. bin.}] \vee [\mathscr{M} \underset{\mathbf{p}}{\wedge} \mathscr{D}] \\ &\sim \left[\mathscr{P} \vee \mathscr{M}\right] \begin{matrix} \underset{\mathbf{p}}{\wedge}\mathscr{D} \\ \underset{\mu}{\wedge}\mathscr{G} \end{matrix} \sim \left[\text{Multivariate Poisson}\right] \begin{matrix} \underset{\mathbf{p}}{\wedge}\mathscr{D} \\ \underset{\mu}{\wedge}\mathscr{G} \end{matrix} \\ &\sim \{[\mathscr{P} \underset{\mu}{\wedge} \mathscr{G}] \vee \mathscr{M}\} \underset{\mathbf{p}}{\wedge} \mathscr{D} \sim [\text{Neg. sing. multinomial}] \underset{\mathbf{p}}{\wedge} \mathscr{D}. \end{aligned}$$

These are not arid mathematical results, but have important physical meaning – each formulation corresponds to a different scientific model but leads to the same multivariate statistical distribution. These various scientific models can arise in many fields besides marketing, a point touched upon by the authors only very lightly in Section 5.3.

Consider for instance [Neg. sing. multinomial $\underset{\mathbf{p}}{\wedge}$ $\mathscr{D}$] for which the p.g.f. is

$$\int_0^1 \cdots \int_0^1 \left\{1 + \frac{MT}{K} - \frac{MT}{K}(p_1 z_1 + \ldots + p_g z_g)\right\}^{-K} \frac{\Gamma(\Sigma_1^g \alpha_j)\, \Pi_1^g\, p_j^{\alpha_j - 1}}{\Pi_1^g\, \Gamma(\alpha_j)}\, dp_1 \ldots dp_{g-1}.$$

This is not quite the same as Mosimann's (1963) compound negative multinomial with p.g.f.

$$\int_0^1 \cdots \int_0^1 p_g^K \{1 - p_1 z_1 - \ldots - p_{g-1} z_{g-1}\}^{-K} \frac{\Gamma(\Sigma_1^g \alpha_j)\, \Pi_1^g\, p_j^{\alpha_j - 1}}{\Pi_1^g\, \Gamma(\alpha_j)}\, dp_1 \ldots dp_{g-1},$$

but is a case of Mosimann's more general distribution considered in the same paper with p.g.f.

$$\int_0^1 \cdots \int_0^1 \left\{ \frac{1}{u_0} - \left(\frac{u_1 z_1}{u_0} + \ldots + \frac{u_h z_h}{u_0} \right) \right\}^{-K} f(u_0, \ldots, u_h)\, du_1 \ldots du_h$$

where $1/u_0 = 1 + MT/K$ and $u_i/u_0 = MTp_{i/K}$, $\Sigma_{i=0}^{h} u_i = 1$. All of Mosimann's results for his general compound negative multinomial are applicable here.

The physical situation giving rise to Mosimann's [Neg. multinomial $\underset{\mathbf{p}}{\wedge}$ $\mathscr{D}$] was as follows. Pollen grains at a particular depth in a pollen-bearing sediment were sampled by noting how many grains of four different kinds were encountered when inverse sampling for 100 grains of a fifth kind, i.e. $K = 100$ and $g = 4$. However, were one to sample a fixed amount of sediment and count the number of grains of each of the $g = 5$ kinds, then Goodhardt, Ehrenberg and Chatfield's distribution would be appropriate, see also Taillie, Ord, Mosimann and Patil (1979).

Returning to mathematical considerations, I am unimpressed by the author's estimation procedure for α_i, $i = 1, \ldots, g$. Their estimation of α_i/S, where $S = \Sigma_1^g \alpha_i$, from the market shares is straightforward. However their estimation procedure for S necessitates obtaining for each brand an estimate of S by iteration from the marginal probability $P(X_j = 0)$, and then combining these estimates into an overall estimate using a weighted average.

However from a knowledge of the model's unconditional probabilities $P(X_1 = x_1, \ldots, X_g = x_g)$

$$= \frac{\Gamma(K + \Sigma_1^g x_j)\, \Gamma(\Sigma_1^g \alpha_j)\, \Pi_1^g \Gamma(x_j + \alpha_j)}{\Gamma(K)\, \Pi_1^g x_j!\, \Pi_1^g \Gamma(\alpha_j)\, \Gamma(\Sigma_1^g (x_j + \alpha_j))} \left(\frac{MT}{K + MT} \right)^{\Sigma_1^g x_j} \left(\frac{K}{K + MT} \right)^K$$

(a formula which should surely have been given, at least in the technical appendix), it is straightforward, though algebraically tedious, to obtain the marginal distributions and their moments. Each $\hat{\alpha}_j$ can then be obtained as the same simple function of the appropriate moments. This involves no iteration and no combination of possibly discordant estimates. Each marginal distribution will also yield an estimate of S and these can be examined against $\Sigma_1^g \hat{\alpha}_j$ for homogeneity. Alternatively, given the above formula one could consider maximum likelihood or minimum chi-square estimation.

Mr. President, we are grateful to the authors for presenting so admirably their long experience with the successful application of this distribution to the purchase patterns for so many product-fields. The ramifications of their Dirichlet model, however, extend beyond the confines of marketing and merit the attention of all concerned with multivariate modelling. I have great pleasure in seconding this vote of thanks.

The vote of thanks was passed by acclamation.

Mr T. Sharot (Audits of Great Britain Ltd): I would like to make two observations on this paper.

First, the model presented enables us for the first time to make theoretical statements about the sampling error of certain estimators, notably the brand share of each brand, and changes in these over time. This is valuable for three reasons.

One, in many markets the brand share is more important to manufacturers than are actual purchases as an indicator of success, since, for instance, increased sales are little solace if accompanied by a reduction in share. Second, there is a positive correlation between observed purchases of each brand and purchases of the product-class, which means that the brand share will have a lower coefficient of variation than brand purchases. Third, though measurement errors such as under-recording are always present to some extent in panel data, brand share estimates tend to be little affected in comparison to estimates of the absolute purchasing-level.

Correspondingly, I have derived the formulae for the standard error of an observed brand share, for a change in brand share, and for the difference between two brands' shares.

All three results are fairly simple and certainly elegant and will be made available to any interested parties.

My second point concerns the property of independence implicit in the Dirichlet distribution. I would argue that the model fits many aspects of consumer buying behaviour very well, not because of this independence, but despite departures from it.

I would say that virtually every market is in fact segmented, either by differences in product formulation, or retail distribution, by price, by quality, by usage occasion, or in one of several other ways. This partitioning in turn affects consumer buying behaviour; as a result this model cannot fit satisfactorily the observed variations in brand buyer duplication or brand switching between different pairs of brand. This can be seen for the two Private Label brands in Table 6 of paper. It would be very useful to extend the model to accommodate these variations, particularly to provide norms for brand switching in the stationary case, against which to judge the effect of promotions or new brand launches. Aitchison's work on compositional data could well form the basis of this.

Notwithstanding this limitation, the present model represents a considerable advancement in the analysis of buying behaviour and for this I too would like to thank "G, E & C Ltd."

Mr D. Bloom (London SW3): I address myself only to the marketing implications, particularly Sections 3 and 5.

It is asserted in Section 3, under the heading *Prescriptive Uses*, that –

> "increasing sales by getting existing buyers to buy much more of the brand would therefore be aiming at something altogether unusual or unlikely, like making pigs fly. There is a law of nature against it, or so it would appear."

While that may generally be the case, there are exceptions. I would point to one spectacular one to demonstrate my point. This concerns Arm and Hammer – a brand of bicarbonate of soda in the United States.

With the trend to prepared meals the conventional market was in sharp decline. The proprietors hit on the device of persuading American housewives to put whole packs of the stuff in their refrigerators in order to dry and deodorize them. After this triumph they went on to suggest further new uses. The net result was a huge increase in the quantity sold per buyer: not impossible after all.

I would suggest to the authors, who I am sure are familiar with this criticism, that it is of the nature of their model that it cannot explain change and so not give much guidance about what the marketing man wants to do. If that guidance could be given, of course, their model would be immensely popular, and I think the level of understanding would increase by leaps and bounds.

Also, at the more practical level, there are problems in observing change – because change does come about, brands do change in their relative popularity, [brand] shares vary over time. The problem is simply that the commercially available panels are not very large. Divergences from norms within the data yielded by these panels can plausibly be represented as within margins of error that would be expected, given the size of those panels. But quite small divergences, if systematic, and if they persisted over fairly short periods of time, could bring about appreciable change. We are really unable to observe whether or not these occur.

That being so, I suggest that it would be sensible to develop extremely large-scale simulations to see what changes in the behaviour of those very large numbers of artificial consumers would be compatible with what is observed empirically with the relatively small panels commercially available, and yet not abort the laws that the authors have described so admirably in this paper.

Mr J St G. Jephcott (Audits of Great Britain Ltd): I congratulate the authors, on providing people who, like me, work with consumer panel data, with a powerful and practical tool for analysing such data. As the authors suggest it seems likely that the Dirichlet model will have applications to consumer behaviour beyond the types of buying behaviour illustrated in their paper. I will be interested for example, to see how well it "fits" off-air TV viewing behaviour in environments where there are a large number of channels on offer.

Returning to the Buying Behaviour model, the authors' justify its validity primarily by showing that in its aggregate form it successfully describes observed patterns of behaviour "in more than 40 product fields". The empirical approach is valuable in the context in which the model will normally be used, but, as the authors are themselves aware, it does not necessarily validate the assumptions on which the theoretical model is constructed.

Indeed, the assumptions made at the individual consumer level are not supported by empirical studies. Their comment on the assumption (B1) of a Poisson process describing an individual's purchases, specifically that incidence of previous purchases are so irregular that it can be regarded "as if random" providing an appropriate length of time is used, is perfectly reasonable and does not exclude the possibility of rational models of individual buying behaviour. On the other hand their assumption (A1) that each consumer's choices among the available brands follow a multinomial distribution is less adequately dealt with—their justification that it is in line with their observations on stationarity, is clearly insufficient.

As it happens there have been a number of studies of brand buying sequences which have shown that, on average, individuals' buying patterns provide fewer i.e. longer sequences or runs on a particular brand that would be predicted under this multinomial assumption. My own studies, reported in 1972, showed that in the product fields I looked at, that there was a small but significant serial correlation in brand choice, and, for obvious reasons, a more substantial effect when store choice was considered. My comments in no way detract from the practical value of the aggregate model, but do, I hope, introduce an appropriate degree of caution into any "insight" on consumer behaviour at the micro-level that some may claim for their model.

Professor C. D. Kemp (University of St Andrews): I would like to make three points. The first concerns the logical structure of the model. Essentially we are presented with two quite distinct models: one of brand choice (assumptions A) and one of purchasing frequency (assumptions B). Assumption C links them together and is chosen to make this linking easy. However, assumptions A1 and B1 are both *intra*-individual, whilst A2 and B2 are both *inter*-individual. It would therefore seem more logical to write the model as

$$(\mathscr{M} \underset{n}{\wedge} \mathscr{P}) \underset{\mathbf{p},\mu}{\wedge} \mathscr{D}-\mathscr{G}$$

where $\mathscr{D}-\mathscr{G}$ denotes a multivariate distribution whose marginal distributions are Dirichlet and gamma. Assumption C simplifies $\mathscr{D}-\mathscr{G}$ by making its density function merely the product of its marginal density functions, and enables us to write

$$(\mathscr{M} \wedge \mathscr{P}) \wedge \mathscr{D}-\mathscr{G} \equiv ((\mathscr{M} \wedge \mathscr{P}) \wedge \mathscr{D} \wedge \mathscr{G} \equiv ((\mathscr{M} \wedge \mathscr{P}) \wedge \mathscr{G}) \wedge \mathscr{D}.$$

C is an independence assumption more of convenience than necessity, unlike the assumption on $\{p_i\}$ which characterizes the Dirichlet distribution. Relaxing C might make for a more realistic model.

Secondly, there is a complete absence of information in the paper about standard errors of parameter estimates. Whilst I am not one who believes that any estimate which does not carry a standard error (asymptotic or otherwise) is necessarily useless, I do not think this matter can be entirely neglected. Further, if a fairly large number of parameters is to be estimated, covariances can be important so it would be useful to have some kind of estimated variance–covariance matrix: this, of course, should be obtainable given the expressions for the probabilities provided by Dr Kemp.

Finally, it is interesting to compare what appears to be a very successful application of mixed distributions with the application area in which, from about 1920, the concept of mixing began to be seriously developed *viz* accident proneness. I have suggested elsewhere (Kemp, 1970) that whilst the accident work contributed much to the development of discrete distribution theory, it had relatively little success in achieving its fundamental objectives (accident reduction and accident prevention). However, these two application areas seem to differ in at least two notable respects. Firstly, as far as the average individual is concerned, accidents are, fortunately, rare events whereas purchases are not. But perhaps more importantly, the accident work was primarily concerned with individual behaviour whereas this successful marketing application seems to be predominantly concerned with aggregate behaviour.

Professor D. R. Cox (Imperial College, London): Mention has been made of characterization theorems. Are there corresponding stability theorems available? That is, how much do the conclusions of the characterization theorem change when the initial assumptions are perturbed?

The following contributions were received in writing, after the meeting.

Professor F. Bass (University of Texas at Dallas): Stochastic brand choice behaviour of consumers can be represented and measured in two equivalent ways. One of these ways is *purchase incidence*, fully discussed by Goodhardt, Ehrenberg, and Chatfield in their "comprehensive" review of the Dirichlet distribution. Another representation of stochastic choice behaviour, however, is a measure of *brand switching* of consumers. The brand switching measure indicates the fraction of consumers switching from one brand to another on adjacent purchase occasions. The choice between these two representations is largely arbitrary, depending upon taste and inclination. Bass, Jeuland, and Wright (1976) have shown how brand switching measures and purchase incidence measures may be derived from the same set of assumptions. In their discussion of the Dirichlet, Goodhardt, Ehrenberg, and Chatfield have elected to focus only upon the purchase incidence measure.

Dr S. F. Buck (Audits of Great Britain Research plc): Practitioners dealing with consumer panel data, both in research agencies and in client companies, will benefit considerably from this paper. It improves the understanding of purchasing behaviour and includes at the same time a useful resumé of the basic definitions relating to the considerable material produced over the years by the authors in this area.

The authors themselves point out a number of items associated with the model which require further investigation and I would urge them to direct their attention to an examination of cross product analysis in order to better understand the loyalty to brand names and to individual retailers. In this context I do not entirely agree with the authors that weight of purchasing has a low correlation with demographics like size of household. In addition more work is required to understand the systematic deviations that occur from the model for some product fields if they are to be accepted as the exception that proves the rule!

I know that my colleagues are eager to examine the fit of the model against more recent data given the considerable changes that have occurred in the grocery market in recent years. The enormous importance of a small number of retail chains and the move to more heavily segmented markets should provide a good background for further application of this interesting and useful system.

Dr Peter J. Diggle (University of Newcastle upon Tyne): The authors rightly emphasize the interpretation of a fitted model, and place some emphasis on the interpretation of the negative binomial parameter K as "reflecting the heterogeneity or diversity of consumers". The negative binomial follows from assumptions (i) and (ii), but I did not find the arguments in favour of (i) compelling. Intuition suggests that the sequence of purchase times for an individual consumer might be more regular than Poisson, albeit still stochastic. Let us assume that an individual consumer's number of purchases per unit time has a distribution with mean μ and variance $\alpha\mu$ for some $0 < \alpha \leqslant 1$, and that μ is sampled from a gamma distribution with mean M and exponent K. It follows that the number of purchases per unit time by a randomly sampled customer has mean M, variance $M(\alpha + MK^{-1})$ and variance-to-mean ratio $\alpha + MK^{-1}$. The Poisson assumption (i) corresponds to $\alpha = 1$. As far as second-order properties are concerned, a reduction in α can be balanced by a corresponding reduction in K. This suggests to me that unless abundant data are available, the negative binomial might appear to fit well in cases when $\alpha < 1$. Put another way, if the model with the additional parameter α is not ruled out *a priori*, there may be some ambiguity of interpretation between the degree of regularity in an individual consumer's behaviour and the degree of heterogeneity between consumers.

Professor A. Jeuland (University of Chicago): In this short note it is impossible to touch on the numerous issues researchers must recognize when modelling purchasing behaviour. I will thus limit my comments to two directions for future research.

First, we need to go beyond simply describing observed behaviour with ever more sophisticated statistical models. Having recognized some patterns, we now need to develop theories for the observed phenomena. We need to answer the question: Why is this the case? As an example, Goodhardt and his co-authors have observed that the "... average rate of product purchasing, denoted by w_p, also varies little from brand to brand but increases slightly with decreasing market share (a trend in the opposite direction to the w's)."

A simple explanation based on the law of diminishing marginal utility and the law of consumer heterogeneity may be the answer: If consumers who buy marginal brands are doing so by self-selection because they have high utility for the product class, we do expect a negative correlation between w_p and market share.

My second point deals with using stationary models for interpreting change. I believe that a superior approach is to incorporate the dynamics into the stationary model itself so as to have estimates of the changes. These estimates then have known statistical properties. However, incorporating dynamics into a model of the type described by Goodhardt and his co-authors certainly is a major undertaking!

Dr F. Phillips (MRCA. TX): This impressive and useful paper is the culmination and synthesis of the authors' many years of observation, theory building and publication of partial results. Because of its completeness, this paper will be a valuable reference in my own work in interpreting consumer panel data, and it will be cited frequently in marketing research journals.

The authors' rigorous justification of the model's distributional assumptions is unmatched in the American marketing literature, and is welcome. The relation of the Dirichlet assumption to market segmentation (Section 2.2.A) is an excellent contribution. Similar mathematical structures appearing in the American literature have been largely *ad hoc* modelling efforts. This paper goes beyond these efforts by claiming to characterize a process in equilibrium.

The authors imply they built a theory of stationary markets by "ignoring sharp but short promotional fluctuations" (Section 2.2.A). These fluctuations are the most interesting aspect of the market to many managers; indeed these managers may be paid to create fluctuations. The authors later indicate uses of their model for evaluating exceptions. In this Section, though, the authors gloss over some important questions. What do these "sharp but short" fluctuations typically look like, in terms of raw data and impact on the model parameters? By what statistical method may they be "ignored" (i.e., how were the data smoothed)? One imagines this would be difficult in the several categories, like soft drinks or beer, that are promoted nearly continuously in the US.

The "means and zeros" *vs* method of moments discussion gives a helpful criterion for choosing a fitting method. In American companies the choice is still made on the basis of tradition or available software, and frustration results.

The remarks on advertising, two paragraphs below Table 8, are very provocative, and an expanded discussion would be in order.

As I have remarked, professional marketers devote their careers to destroying market equilibria. Technological change, successful new products, and changing consumer tastes translate to non-stationary purchase probabilities, and show evolution rather than equilibrium is the rule for markets. The evolution of a market may be a sequence of temporary equilibria. Given the long period over which the authors have been collecting data and validating their model, one would suspect that the Dirichlet model can describe these successive stationary conditions, by changing parameter values and the definition of the product category.

Even if so, this would dilute the authors' statements about baselines and norms in Section 3.1 *et seq*. These operating guidelines have less value if one is not sure whether a fluctuation will damp out or accelerate and lead to a new equilibrium. For this reason our own research has been toward tests for stationarity, which will enhance the important role of consumer panels in monitoring for market change.

Dr Robert W. Shoemaker (New York University): This is an interesting paper that makes important contributions in several areas. First it provides a comprehensive summary of the NBD–Dirichlet (N–D) model and its fit to observed buying patterns. Second, the authors note several systematic departures from the model predictions. These observations should prove helpful for developing improved models and predictions.

The authors are to be complimented for outlining numerous ways in which this research could be extended. One other possible extension is to use a common set of data and to compare the predictive accuracy of the N–D model with that of the CNBL model (Zufryden, 1977, 1978). This comparison would be particularly interesting since it appears that both models can be used for predictions on brand penetration, repeat purchasing and other measures.

One minor clarification concerns the paper's description of the Markov model (Section 4.2). The paper refers to the probabilities of switching from brand j to k as being invariant characteristics of each brand. As I understand the current N–D model, the purchase probabilities here are also invariant over time.

The authors may be a bit harsh on other approaches to analysis (Section 5.2). In particular, it appears that several contributions toward the development of the current N–D model have been made by other researchers with different objectives and different research procedures. Both are helpful.

In considering the total paper the authors are to be commended for being able to predict a wide range of buyer behaviour on the basis of a small number of parameters developed from more aggregate data. The paper provides valuable new evidence on purchase regularities across brands and product classes.

Drs Neil Wrigley and Richard Dunn (University of Bristol): We would like to take this opportunity to make two comments on this paper which arise from recent work at the Department of Geography, University of Bristol.

The first is that the Dirichlet model of buying behaviour appears also to hold for purchasing patterns at *individual* stores, with one important proviso. Using data for a number of product fields taken from a specially collected consumer panel survey in Cardiff, U.K., we have found

(i) that the same sorts of empirical regularities noted for brand-choice also occur for store-choice, and

(ii) that the full range of theoretical predictions of the Dirichlet model fit our observed data extremely well.

These results, and further information, are reported in Wrigley and Dunn (1984, 1985).

The important proviso in the case of purchasing at individual stores is that the Dirichlet model must be calibrated on an appropriate sample – a geographically defined sample which does not include large numbers of individuals who will never use certain stores because of constraints of distance or the existence of intervening opportunities. This is an example of the "relevant population" argument of Massy *et al.* (1970, p. 337), which we have also found to be important when the NBD model is applied to purchasing at individual stores (Dunn *et al.* 1983).

These results at the level of individual stores add further weight to the authors' comment that store-choice appears to be like brand-choice. We feel it is particularly important for geographers, planners and retailers that complex patterns of urban shopping behaviour appear to be accurately predicted by this simple and elegant model.

Our second point is that the store-choice (brand-choice) component of the model, the multinomial–Dirichlet mixture, has also been used as the basis for certain econometric models for the analysis of panel data, specifically the beta-logistic model of Heckman and Willis (1977). This regression–type model is used to ascertain whether the mean probabilities of choosing particular stores (brands), and the distribution of probabilities about those means, vary systematically between individuals as a function of individuals' characteristics. (See Dunn and Wrigley, 1985).

Although we concur with the authors' emphasis on generalizable patterns and replicable results, the analysis of *specific* samples using regression–type models may have an important complementary role. In the traditional brand–choice marketing applications it may be of interest to determine whether patterns of multibrand purchasing vary systematically with such factors as household income or household composition for a particular product field. Similarly, in the analysis of urban shopping patterns it may be important to establish how, for a particular panel in a city, the probabilities of using stores vary with distance-to-store, income or car-ownership.

Professor Fred S. Zufryden (University of Southern California): The authors describe a model that includes a multinomial brand choice behaviour process with choice probabilities varying jointly over a population of consumers as a Dirichlet distribution. This model is combined with a Negative Binomial Distribution (NBD) characterizing product class purchase incidence behaviour.

Although the proposed multi-brand model provides the advantage, over a simpler Beta Binomial model, of considering estimates of buyers of brand combinations, this is not without computational and estimation difficulties. The computational aspects are, by the authors own admission, quite burdensome and necessitate lumping brands into an all other category and then essentially using a Beta Binomial model. Moreover, the proposed estimation procedure involves an iterative method that attempts to estimate the parameter S independently from the $p(0)$ of each brand and then reconcile any differences by, rather arbitrarily, weighting the derived S values to obtain the unique S parameter of the Dirichlet.

An alternative estimation method might involve first estimating the NBD parameters, inserting these and replacing α_j by Sm_j/M in the $p(0)$ equation of each brand and then minimizing the squares of the deviations between theoretical and observed $p(0)$'s, summed over each brand j, by choice of S. This least-squares method would still require truncation of the $p(0)$ terms and an iterative technique to search for a globally optimal S value. The individual α_j's could then be obtained from the brand market share equations.

The authors emphasize that their model applies to stationary markets and that efforts are currently under way to encompass more general market situations. This and the integration of marketing variables within stochastic models would seem to be worthwhile directions which concur with recent work in the USA (e.g., Zufryden, 1981; Jones and Zufryden, 1980).

The **authors** replied later, in writing, as follows.

We are grateful for the wide range of comments. They seem to fall under three broad headings—the assumptions of our model, its practical uses, and questions of fit and estimation. We comment accordingly.

As regards our five assumptions, Dr Diggle and Professor Shoemaker are not the first to question B1, the Poisson assumption. We would refer them to Chatfield and Goodhardt (1973) and Dunn *et al.* (1983) for extensive discussions. The evidence is that the departures from B1 are small and certainly not large enough to stop the overall model from "working". We were similarly encouraged rather than depressed by Mr Jephcott's 1972 investigation relating to A1, the Multinomial assumption, which showed only a small serial correlation in an isolated case.

Professor Kemp suggests it would be "more realistic" to relax assumption C, of the independence of the Dirichlet and Gamma distributions, by introducing a more general distribution. We would be interested if he could tell us about any candidates, and the number of parameters needed. Given the already close empirical fit of our "independent" model we are not inclined to invest in exploring his route ourselves.

The most doubts were expressed about the assumed lack of segmentation in A2. We have already reported cases and traces of segmentation, and there certainly are gradients in buying rates with household size, for instance, (e.g. Drs Buck, Wrigley and Dunn's comments) which lead to interesting but unresolved theoretical aggregation problems (e.g. Ehrenberg, 1959). But the effects are generally small compared with the vast *irregular* variation that we observe between individuals. The overall fit of the model in so many different ways and under so many different conditions is far stronger than any of these doubts. To a first (and close) degree of approximation, the question is not why the model fits so badly, but why it fits so well.

We therefore agree in principle with Professor Jeuland about the need to seek explanations. However, on the specific matter of the observed trends of the average purchasing rates w and w_P with market-shares, we feel we have already given an adequate explanation. It is that no special explanation is necessary! Professor Jeuland offers one based on the supposition that buyers of marginal brands have a higher utility for the product-class. But the value of our model, as we see it, is that it predicts the observed trends without recourse to any such supposition. Indeed, Assumption C of the model categorically denies Professor Jeuland's specific conjecture.

More generally, no specific hypotheses are needed to explain differences between one brand and another over and above the difference in their market-shares. Similarly, differences between one product-class and another need no explanation other than the two product-class parameters K and S.

We differ from Professor Bartholomew, however, when he suggests that practitioners ought to take these two time-invariant parameters on board, rather than the more directly measurable quantities b and w. That these increase non-linearly with T is a bit more complicated but highly meaningful—e.g. some $b = 400{,}000$ people buy instant coffee at Tesco in a week, and $b = 5$ million do so in a year (see Table 7)! Our message for the practitioner is simply that this increase in b is predictable. Hence b and w in a month tell essentially the same story as the numerically different b and w in a year. "The amount and complexity of the information that has to be considered is therefore greatly reduced" (Ehrenberg 1972, p. 36).

The only one of our assumptions which has escaped doubts and criticism from any discussant is B2, relating to the Gamma Distribution for heterogeneity of product-class purchasing. It is ironic that this is the assumption with the least empirical justification (although it is enormously strongly supported for individual *brands*). Indeed, its marginal failure for some product-classes (see Table 2) is one of the reasons necessitating the use of the Empirical–Dirichlet formulation in such cases. (We agree with Dr Buck about the attraction of cross-product field analyses, not least because it will provide evidence relating to B2; we have plans in hand.)

Perhaps the final word on assumptions A, B and C arises, as so often, from Professor Cox's comment. Are there, he asks, *stability* theorems available? As far as we know, the answer is "no". (Do they exist for other practical models?) Yet, less elegantly, we do know quite a lot about the sensitivity of the Dirichlet model to our precise assumptions. We know that none of the assumptions are quite true, yet empirically the model works pretty well.

As regards our restriction to "stationary" markets, Dr Phillips asks how we smooth away the fluctuations which occur, usually short but sharp, due for example to promotional activities. No special "smoothing" is required – we just fit the model when the broad pattern is approximately stationary over a relatively long period of time. In Table 7 for example, the results in the shorter periods are for the *average* period of that length (but the model tends to fit for most individual periods too, if there was no blip in b). In the longer periods, the short-term fluctuations become largely submerged. This is then justified pragmatically – it works. And *departures* from the model prediction can then often be explained by the market fluctuations (see Ehrenberg 1972, p. 35, for an example).

Many of the product-fields that have been analysed are ones where there is in fact almost continual promotional activity. But most of this is defensive, trying to keep each brand's share where it was, i.e. aiming to keep the market more or less stationary. Dr Phillips' view that marketing managers earn their living by trying to upset the equilibrium is widely held, but rather naively accepts such marketing managers' view of their own self-importance. Nonetheless, there are occasions where marked upsets do occur, as we discuss in Section 3.1.

We are grateful to Dr Kemp and also Professor Kemp for suggesting alternative formulations which lead to the same multivariate distribution in a single time period, though not necessarily to the same joint distribution over several time periods. A similar type of phenomenon arose in the studies of accident proneness, where several different models could explain the good fit of the negative binomial distribution to accident distributions. Dr Kemp's results will be of interest to researchers tackling other problems, while Professor Kemp's results are of direct relevance to consumer purchasing, as some results are best derived by varying the order of mixing.

However, we are confident that our formulation is the most appropriate one here, for two reasons. Firstly it is a straightforward representation of the assumptions underlying the model and clearly distinguishes the purchase incidence and the brand choice components. Secondly the first two "mixings" both yield discrete distribution, viz. $(P \wedge G) \sim$ Negative Binomial and $(M \wedge D) \sim$ Dirichlet-Multinomial (cf. Beta-Binomial). For practical estimation purposes this is an immense advantage.

The second main topic raised by the discussants is that of applications. Professor Bass rightly says that we have ignored the analysis of consecutive purchase occasions (what he calls "switching"). The Dirichlet model does of course cover this – as it does *any* aspect of buyer behaviour under stationary conditions. But we have not done much empirical work on consecutive purchases (and runs etc) because the extensive work done by others has not led to any generalizable patterns. (We have noted the reasons elsewhere. Consumers' heterogeneity of purchase rates means that different consumers' purchase occasions quickly get out of step, both with each other and with any happenings in the market place, like promotional activities, price changes, etc. Hence our "time-period" approach to the data.)

Professor Bartholomew believes that the norms provided by our stationary model are like the "base-line models" used by sociologists. But we believe that the two things differ since the base-lines are not in practice true (or *expected* to be true), but merely embody the "simplest possible stochastic assumptions".

Mr Bloom rightly comments on marketing implications. However, his Arm and Hammer example concerned the whole product-class, not an individual brand. It is certainly possible for purchase frequencies to change in a product-class – e.g. seasonally – but this does not occur (except perhaps momentarily) for just one brand.

We agree with Mr Bloom that our model does not "explain" change. Nobody knows much about change, so we cannot try to explain it, yet. We therefore feel it is premature to incorporate marketing variables in stochastic models, as suggested by Professor Zufryden, except as a "theoretical exercise". But our model successfully provides the background for studying change, as noted in Section 3.1. Here we believe in empirical observation and practical experimentation, rather than in just playing with one's assumptions, i.e. simulation, as suggested by Mr Bloom.

Finally, various questions of estimation and goodness of fit tests and the like have been raised.

Dr Kemp and Professor Zufryden for example are unhappy that in our estimation of S we use discrepant first-stage estimates. Our computational procedure certainly is not the last word. But it shows up possible discrepancies in the market. Furthermore, the model is such that any discrepant brand can then simply be left out, without causing any theoretical or practical problem – an extraordinarily powerful and, we believe, *unusual* form of robustness. Professor Bartholomew reckons that there are problems of goodness of fit and estimation but leaves these up in the air. Dr Buck says there are product-fields where the fit is not good, and we would like to know about them. Cases of big systematic departures from the model would give us much new insight.

Professor Kemp suggests that there is a great need for standard errors for our estimates. But we do not find much need for explicit theoretical calculations of this kind. Dealing with many different brands and many different time-periods for many different product-classes as we do, the empirical results such as in Tables 3 and 8, provide *observable* upper limits of sampling error (see Ehrenberg 1982, Section 9.5). Sampling errors for meaningful marketing indicators, like changes in brand share, are a different matter, and we are pleased Mr Sharot has found our model useful in this respect. (We think, however, he goes oddly wrong in saying brand share is more important than sales!) Of course, any sampling error calculations need to take account of the fact that the data came from complex, and not simple, random samples.

We are grateful to Dr Kemp for writing down the unconditional joint distribution of purchases and noting that it is possible to derive an expression for the marginal distributions. However it is only possible to express the latter in closed form by using the four-parameter hypergeometric function (Abramowitz and Stegun 1965, p. 556). *This would normally still have to be evaluated as an infinite summation*! For the record the formula for the marginal distribution of purchases is

$$P(r) = \frac{\Gamma(\alpha + r)\,\Gamma(K + r)\,\Gamma(S)}{\Gamma(\alpha)\,\Gamma(K)\,\Gamma(S + r)\,r!} \left(\frac{K}{MT + K}\right)^K \left(\frac{MT}{MT + K}\right)^r F(S - \alpha, K + r; S + r; MT/(MT + K))$$

for $r = 0, 1, \ldots$, where α is the α– parameter for the given brand.

Jeuland *et al.* (1980) similarly claim to have expressed brand penetration in closed form. But their formula also contains hypergeometric functions requiring doubly infinite summations! Thus for most purposes it is much easier to use an expression such as that in equation (A2) of the paper.

Dr Kemp and Professor Zufryden have both suggested new procedures for estimating the α-parameters. While accepting that our own procedure is pragmatic, we see no reason to prefer either of their two methods. One advantage of ours is that the observed and predicted brand-shares are constrained to be equal. This makes the fitted model more acceptable to the prime potential users, namely marketing men and market researchers. Unfortunately Professor Zufryden's method does not satisfy this requirement.

Furthermore, as the $p(0)$'s vary markedly in size, we see no reason why S should be chosen to minimize the sum of squared deviations between observed and predicted $p(0)$'s, as suggested by Professor Zufryden. Dr Kemp's proposed method involves using the first two moments of the marginal distributions. Unfortunately, the observed second moment is not usually reported in practical market reports. It would of course be possible in principle to calculate it from the raw data, but that may not be easily or cheaply accessible. Such is life. In addition, the available results on the efficiency of parameter estimates for the negative binomial distribution show that for reverse J-shaped distributions it is better to fit by mean and zeroes (as in our method) than by moments. In any case, Dr Kemp's suggestion still leaves one with the problem of producing a single overall S-value.

REFERENCES IN THE DISCUSSION

Abramowitz, M. and Stegun, I. A. (eds) (1965) *Handbook of Mathematical Functions*. New York: Dover.

Bartholomew, D. J. (1975) Errors of prediction in Markov Chain models. *J. R. Statist. Soc.* B, **37**, 444–456.

——— (1982) *Stochastic Models for Social Processes*, 3rd Edition. Chichester: Wiley.

Heckman, J. and Willis, R. (1977) A beta-logistic model for the analysis of sequential labor force participation of married women. *J. Polit. Econ.*, **85**, 27–58.

Jones, J. M. and Zufryden, F. S. (1980) Adding explanatory variables to a consumer purchase behavior model – an exploratory study. *J. Marketing Res.*, **17**(3), 323–334.

Kemp, C. D. (1970) 'Accident proneness' and discrete distribution theory. In *Random Counts in Scientific Work*, Vol. 2 (G. P. Patil, ed.), pp. 41–65. University Park: Pennsylvania State University Press.

Martin, J. J. (1967) *Bayesian Decision Problems and Markov Chains*. New York: Wiley.

Mayhew, B. H. (1984) Baseline models of sociological phenomena. *J. Math. Sociol.*, **9**, 259–281.

Mosimann, J. E. (1963) On the compound negative multinomial distributions and correlations among inversely sampled pollen counts. *Biometrika*, **50**, 47–54.
Taillie, C., Ord, J. K., Mosimann, J. E. and Patil, G. P. (1979) Chance mechanisms underlying multivariate distributions. In *Statistical Distributions in Ecological Work* (J. K. Ord *et al.*, eds). Maryland, USA: International Cooperative Publishing House.
Wrigley, N. and Dunn, R. (1984) Stochastic panel-data models of urban shopping behaviour: 2, Multistore purchasing patterns and the Dirichlet model. *Environment and Planning* A, **16**, 759–778.
——— (1985) Stochastic panel-data models of urban shopping behaviour: 4. Incorporating variables into the NBD and Dirichlet models. *Environment and Planning* A, **17**(3).
Zufryden, F. S. (1977) A composite heterogeneous model of brand choice and purchase timing behaviour. *Manag. Sci.*, **24**, No. 2, 121–136.
——— (1978) An empirical evaluation of a composite heterogeneous model of brand choice and purchase timing behavior. *Manag. Sci.*, **24**, No. 7, 761–773.
——— (1981) A logit Markovian model of consumer purchase behavior based on explanatory variables: empirical evaluation and implications for decision making. *Decision Sci.*, **12**, 645–659.

As a result of the ballot held during the meeting, the following were elected Fellows of the Society.

Honorary Fellows:

Jøreskøg, Professor K G, Dept of Statistics, University of Uppsala
Santalo, Mr L A, Academia Nacional de Ciencias Exactas, Fisicas y Naturales, Buenos Aires.

Ordinary Fellows:

Abu-Elzin, Fawzi M.
Calder, Patricia
Cox, Michael A. A.
Darvell, Brian W.
Davis, Mark H. A.
Donne, Kelvin E.
Edwards, David
Egan, Blaise F.
Fachel, Jandyra M.
Gradwell, Ann D.
Holloway, Roberta
Lesaffre, Emmanual E. H.
Murphy, David L.
Nagle, Sean C.
Rios-Insua, Professor Sixto
Scully, John A.
Sim, Jung-Wook
Beeyendeza, Justinus
Campbell, Rosemary
Dunne, Timothy T.
Shade, John S.
Smith, Helen K.
Tio, Ngee Ping
Walton, Jamie Edward