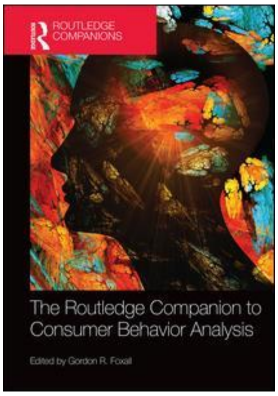

The Routledge Companion to Consumer Behavior Analysis

Gordon R. Foxall

Triple jeopardy in a behavioral perspective

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315850696.ch10
Andrew Rogers, Peter Morgan, Gordon R. Foxall
Published online on: 08 Sep 2015

10

Triple jeopardy in a behavioral perspective

A Bayesian hierarchical model

Andrew Rogers, Peter Morgan, and Gordon R. Foxall

Introduction

Estimating the volume demand of fast-moving consumer goods (FMCGs) is well established and the evidence of a price elasticity of demand is widely accepted by marketing scholars and practitioners alike (Foxall et al., 2013). The purpose of this chapter is to explore other aspects of marketing and psychological variables which may influence consumer behavior above and beyond that of price elasticity. The "law-like" marketing phenomenon of double jeopardy and psychologically based aspects of the Behavioral Perspective Model are combined with the concept of price elasticity to better understand consumer behavior choices.

Brand equity

A brand is defined as "the name, term, sign, symbol or design or a combination of them intended to identify the goods and services of one seller or group of sellers and to differentiate them from those of competition" (Kotler et al., 1999, p. 571). The brand is an "enduring and profitable asset" to a manufacturer (Dyson et al., 1996, p. 9), its formal recognition coming in 1988 when the wealth of the brand was included on the balance sheet of organizations (Allen, 1990). This occurrence was largely due to the willingness of organizations to pay a premium to influence how consumers associate with the brand both on a tangible and non-tangible basis (Dyson et al., 1996) and it is often seen as offering a bundle of benefits to the consumer (Webster, 1994). Some examples are Nestle paying eight times the market value for Rowntree, Grand Metropolitan paying $800 million for Henblein and Yahoo buying Tumblr for $1.1 billion despite Tumblr revenue for 2012 being just $13 million (Fleury, 2013; Kotler et al., 1999).

Given the importance of a strong brand, marketers have developed a need to measure this brand relevance and strength through what has become widely known as brand equity (Kotler et al., 1999). Equity is defined by the Marketing Science Institute as "the set of associations and behavior on the part of a brand's customers, channel members and parent corporation that permits the brand to earn greater volume or greater margins than it could without the brand name" (Chaudhuri, 1995, p. 27). The *Oxford English Dictionary* defines equity as "the commercial value that derives from consumer perception of the brand name of a particular product or service, rather than from the product or service itself'. Aaker (1991, p.15) defines it as "a set of brand assets and liabilities linked to a brand, name and symbol that add or subtract from the value provided by a product or service to a firm and/or to that firm's customers". The relationship of equity to the value and the strength of a brand has been discussed in many ways in the literature. Wood (2000, p. 663) says this determines "brand strength or the degree of brand loyalty", which are antecedents of value. Taylor et al. (2004) show loyalty is a result of equity and value, and Kotler et al. (1999) agree that equity creates brand loyalty. Whereas the benefit of equity and loyalty seems to be clear, there is some disagreement on the order of causality. However, despite this wide acceptance of their importance, the notion was challenged with work uncovering a double jeopardy of marketing effect, based on findings by sociologist William McPhee.

The double jeopardy phenomenon

In 1963, William McPhee observed that comic strips which were read by fewer people were also liked less by those fewer people. Having identified the same pattern among radio presenters, he concluded that smaller brands suffered in two ways: they attracted fewer buyers and were less popular among those fewer buyers. He called this "double jeopardy" or DJ (Ehrenberg et al., 1990). Extensive research shows a similar pattern more widely across behavioral categories and geography including media ratings, newspapers, automobiles, oil companies and many consumer packaged goods (Colombo & Morrison, 1989; Ehrenberg et al., 1990; Ehrenberg & Goodhardt, 2002; Wright & Sharp, 2001).

The reasoning behind double jeopardy

Ehrenberg and Goodhardt (2002) note that the relatively simple logic on which the DJ effect rests means that it can be modeled mathematically. On the basis of the account given by McPhee (1963), these authors explain the logic as follows. Consider a widely known restaurant "W" and a more obscure restaurant "O" which are equal in every way apart from W being better known. When a consumer is asked to name their favorite restaurant, most of the people frequenting W will say restaurant W (as they know not of the more obscure O), whereas of those frequenting O, half will say their favorite is O and half W (as they are equal in every way). Hence O suffers disproportionally: fewer diners and less loyalty among the fewer diners. This is how the DJ phenomenon comes about.

DJ challenges the fundamental importance of brand loyalty, since it implies that penetration is vital to brand growth and that loyalty will follow from this (Ehrenberg et al. 1990). However, Baldinger and Rubinson (1997) argue that brand loyalty and brand volume are highly correlated and loyalty is critical to marketing strategy, while Baldinger et al. (2002, p. 10) state that even though penetration appears to be more important than loyalty, for higher penetration brands, loyalty is "relatively more important than for smaller penetration brands". However, the absence of elasticity coefficients makes it difficult to establish the predictive magnitudes of these relationships. Furthermore, the correlations of loyalty are all less than the correlation of penetration (0.87 vs. 0.81 for small brands and 0.84 vs. 0.62 for smaller brands) which would confirm the importance of penetration over loyalty. Also the correlation of mid-share brands is 0.5, which does not lie between the 0.62 and 0.81 as may be expected given the logic presented.

However, Ehrenberg et al. (1990) do not suggest loyalty is not required, only that larger-share brands have a disproportional effect on loyalty towards those brands. Therefore, increasing share is more efficiently achieved through a focus on penetration which in turn will bring higher loyalty. These authors argue that marketing practitioners need to be aware of the effect of DJ on loyalty measures as they need to expect smaller brands' loyalty measures to be smaller than larger brands and not to overreact when this is the case. Indeed Ehrenberg and Goodhardt (2002, p. 2) state that "marketing people not knowing about [DJ] on customer loyalty is like rocket scientists not knowing that the earth is round". Furthermore, there is research evidence that new product launches attain a level of loyalty almost instantly, after which changes in loyalty are almost wholly accounted for by the DJ effect (Ehrenberg & Goodhardt 2000; Wright & Sharp, 2001). The implication is that brands are not strong or weak in equity: they are simply large or small in size (Ehrenberg & Goodhardt, 2000; Wright & Sharp, 2001).

The NBD-Dirichlet model has been extensively used to describe how FMCGs are purchased, including the DJ phenomenon (Goodhardt et al., 1984; Sharp et al., 2012). However, the NBD-Dirichlet does not incorporate marketing mix variables, since it assumes that the marketplace is essentially stationary (showing a small trend in sales in the medium term) and, on this basis, these variables can be assumed to have been taken into consideration (Bassi, 2011). This is because the marketing mix by and large determines the size of the brand and differences in loyalty are systematic (Ehrenberg et al., 1990). The model does not, however, assume the marketing mix variables are absent, but proposes that, within a stationary market, brand volume is unaffected by changes in the marketing mix (Ehrenberg, 1972). However, consumers are still making choices, usually among a repertoire of brands, and certain factors will influence which brand they choose at any one time. The NBD-Dirichlet states that these, on average, will form the predictive nature of the stationary market rather than on individual purchases of the consumer within that marketplace. Hence, the NBD-Dirichlet describes the pattern of purchase rather than the reason for the individual purchase, i.e. "why one person (or household) generally consumes more toothpaste or soup than others, or somewhat prefers brand j to k or vice versa, is not accounted for by the model and is in fact at this stage still largely unknown" (Goodhardt et al., 1984, p. 638). Research aimed at understanding this consumer behavior has been increasing over recent decades, spanning cross-disciplinary fields (Jobber, 2004; Kotler et al., 1999; Miller, 1995).

One model which incorporates different factors in explaining consumer choice is the Behavioral Perspective Model (BPM).

The Behavioral Perspective Model

The BPM has been used as a theoretical and methodological behavioral framework to explain consumer choice (Foxall & James, 2003; Foxall & Schrezenmaier, 2003; Foxall et al., 2004; Foxall et al., 2006; Oliveira-Castro et al., 2005; Romero et al., 2006). The model (Figure 10.1), an extension of the Skinnerian three-term contingency, proposes that behavior can be viewed as a function of a consumer's learning history within a specific temporal setting together with the benefits (or disbenefits) to be gained from the action (Foxall, 1990; Foxall, 2004).

The BPM states that consumption behavior is followed by a combination of utilitarian and informational reinforcement, and that this *pattern of reinforcement* influences the rate of subsequent behavior of a similar kind (Foxall, 2005). Consumer behavior settings range from relatively open (e.g. browsing supermarket shelves where a variety of alternative behaviors are available) to relatively closed (e.g. standing in line in an airport security queue, where a rather inflexible sequence of behaviors is enjoined upon the consumer). Hence, the freedom (in the sense of the number of behavioral options available) the consumer enjoys varies along this open–closed continuum (Foxall et al., 2013). The consumer behavior setting includes physical surroundings, including temporal constraints, and social surroundings, which include verbal rules (Foxall, 2007). Discriminative stimuli, that comprise the consumer behavior setting, include marketing mix variables. Hence, brand and product characteristics are discriminative stimuli that set

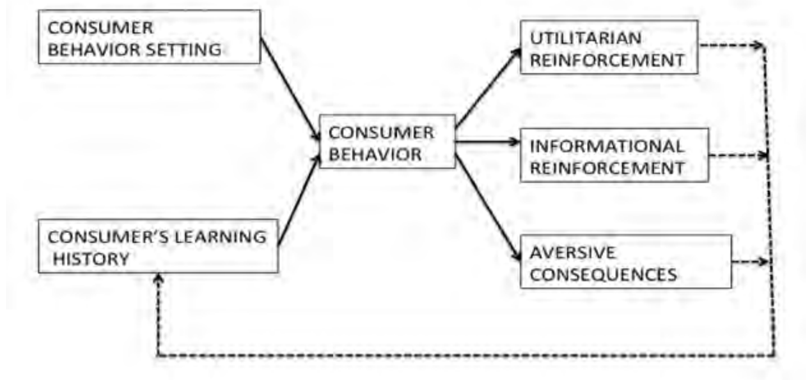

Figure 10.1 The Behavioral Perspective Model

Source: After Foxall (2010)

the occasion for reinforcement conditional on the consumer's enacting specific purchase and consumption responses (Foxall, 1987). The BPM also embraces the principles of behavioral economics within the purchase decision (Foxall, 2001). Choice is not assumed to be an internal psychological process but a consequence of reinforcements within a situational setting (Foxall, 1986a; Foxall 1986b).

Triple jeopardy

There is evidence of further effects beyond that of DJ on consumer choice. Bhat and Fox (1996) note that, compared to larger stores, smaller stores have fewer customers who visit less regularly (i.e. DJ) *and also* spend less money during their visits, a phenomenon to which they refer as "triple jeopardy" (TJ). However, presumably this would be due to the distribution of items rather than a systematic TJ effect. Recalling McPhee's explanation of DJ among restaurants "O" and "W", one might be led to agree with Sharp and Riebe (2005) that we should not expect the smaller-share restaurant "O" to also suffer further by people eating/spending less per visit than in restaurant "W". Presumably a nationwide fast food outlet enjoys higher penetration and more regular visits than say a Michelin star restaurant, though doesn't benefit further by higher average revenue per visit. In fact research into the TJ claim at the level of the retail chain (accounting somewhat for distribution limits) is dismissive of this notion of TJ (Sharp & Riebe, 2005).

At the level of the individual consumer, however, there has emerged some evidence that a further effect may indeed be present. Bandyopadhyay et al. (2005) observe that lower-volume brands from smaller consumer repertoires systematically score better on attitudinal measures than lower-volume brands in large consumer repertoires (albeit both less than larger brands as would be expected under DJ). This may suggest a further effect above and beyond that of DJ. Fader and Schmittle (1993) add weight to this argument via their finding instances when the NBD-Dirichlet model could not explain the market share of excessively high or low-volume brands, a strong indicator that there are other factors in play. Furthermore, research by Chaudhuri (1995) indicates that brand loyalty is a mediating variable in the creation of brand equity which supports both the equity and the DJ concepts.

We must conclude at this point, however, that there is no clear evidence on whether a TJ may exist, though some research seems to signal the possibility at the consumer level.

The research we describe in this chapter seeks to understand whether a TJ effect may exist on a psychological level which can be described within the variables that comprise the BPM. To establish the existence of this TJ, the model needs to account for the well-established elasticity of demand (Foxall et al., 2013). This can be obtained using the following equation (Equation 1). These coefficients are compatible with economic theory and consistent over time (Oliveira-Castro et al., 2006).

Equation 1. Elasticity of demand:

$$LogQ_i = \alpha + \beta Log\mathrm{Price}_\mathrm{i} + \varepsilon_\mathrm{i} \tag{1}$$

The DJ phenomenon is also extensive in the literature and hence these will be included in the model to account for its existence. A suitable model structure is required.

Models for evaluating the DJ effect

The NBD-Dirichlet model is widely acknowledged to explain the DJ effect and, as we have said, the marketing mix variables are assumed within this model to be accounted for by the nature of the marketplace. However, the model also assumes the market is not segmented (Goodhardt et al., 1984). Within this study, the data are segmented by biscuit type and by supermarket store, which imply a segmented market. The specific aim is to understand the effects of the utilitarian and informational measures rather than the overall structure of the marketplace and hence the Dirichlet model is less relevant to the purpose of this study.

These informational and utilitarian elements of the BPM were not available at the time of development of the Dirichlet and a predictive model is, therefore, required to test these while also taking into account established factors affecting the size of a brand. Evaluating loyalty within the literature is not restricted to the Dirichlet model, for example Labeaga et al. (2007) build a discrete choice model where positive and significant coefficients for the loyalty measures demonstrate smaller brands suffer from less loyalty, hence showing the DJ effect. Chaudhuri and Holbrook (2001) use the LISREL structural equation model to assess the role of loyalty while Bandyopadhyay et al. (2005) also use a structural equation modeling approach. The implication is that other models outside of the Dirichlet have successfully been developed which incorporate or demonstrate the DJ effect and a relevant model is required in this case. Before discussing model selection, a discussion of the available data and a preliminary analysis is conducted.

Preliminary analysis of the data

The data relate to a panel sample of 1,594 households and 75,563 purchases of the UK biscuit (sweet and savory) market from the week ending 17 July 2004 to 9 July 2005. The data are assembled at stock keeping unit (SKU) level, whereby each descriptor contains a string relating to the brand and the number of items within the pack. The SKU element is not structured in any systematic way, hence to extract information about the SKU, the 2,783 SKUs are individually analyzed and the relevant information is consistently extracted and coded. This information relates to the brand name, the weight and the number of items per pack.

Some records appear to have an extremely low price per SKU (as low as 1 pence per item) and a decision is required as to how these observations are treated. The lowest-value biscuit ranges are classed as supermarket own label or value brands. There seems to be a minimum price of 20p per pack. Hence a minimum price of 20p is used as a minimum acceptable price for a

Table 10.1 Biscuit categories

	Count	*% Count*	*Volume*	*% Volume*
Blsc choc countlines	17,293	28.3%	5,089,771	28.3%
Blsc choc fully coated	1,715	2.8%	468,712	2.6%
Blsc choc semi ctd/latticed	7,033	11.5%	2,736,751	15.2%
Blsc coconut	397	0.6%	119,320	0.7%
Blsc cream/jam filled inc sandwich	3,381	5.5%	986,035	5.5%
Blsc digestives exc choc	526	0.9%	176,500	1.0%
Blsc fruit filled	1,910	3.1%	552,825	3.1%
Blsc ginger	1,124	1.8%	362,650	2.0%
Blsc marshmallow fully choc ctd	821	1.3%	191,293	1.1%
Blsc marshmlls choc semi/uncoated	76	0.1%	17,484	0.1%
Blsc sav crispbrd/rice cakes	5,332	8.7%	913,185	5.1%
Blsc sav extruded crckrs/waterbiscs	2,723	4.5%	775,180	4.3%
Blsc sav remaining	6,650	10.9%	1,543,272	8.6%
Blsc shortbread	1,295	2.1%	443,455	2.5%
Blsc shortcakes	1,295	2.1%	381,622	2.1%
Blsc sweet rem types	6,197	10.1%	2,001,843	11.1%
Blsc sweet/semi sweet assortmnt	913	1.5%	601,675	3.3%
Blsc tea & coffee	1,062	1.7%	324,066	1.8%
Blsc wafers	1,344	2.2%	285,418	1.6%

packet of biscuits. In the same manner, there are very low price per 100g values. Likewise, an analysis of the supermarket value range suggests a cut-off point of 15p per 100g is appropriate and hence this is used as a cut-off for all brands. This leaves a base sample of 61,087 records to analyze (80.8% of the original sample). As well as the SKU name, there is a product description field. Table 10.1 shows the distribution of data within this.

Some categories have low counts and hence are not suitable for modeling. A method is required to group the data. Chang (2007, p. 107) suggests a five-band classification yielding the groups shown in Table 10.2.

From the SKU field, it is possible to identify the number of items per pack. For each SKU, the number of items per pack is extracted manually. The number ranges from as few as one biscuit per pack to 48. There are also other larger formats such as drums, bags and barrels which do not contain the actual number of items but all imply larger packs. A sensible structure for modeling purposes is required. Hence the biscuit pack sizes are grouped as per Table 10.2 based on both the distribution of records within groups and logical groupings. Note that for packets of biscuits which contain many standard biscuits (such as digestive), the figure relates to the number of packets within the SKU; in this case, one. Where biscuits are individually wrapped, biscuits tend to be single-serve portions rather than multiple serve. For example, a single packet of six Kit-Kat biscuits will be classed as "6".

Initial analysis of utilitarian and informational reinforcement

Before a model is built, the analysis seeks to explore the potential relationship between the size of brands and the informational and/or utilitarian variables. Within the data, the informational score is obtained by averaging two scales and hence a score can be obtained for each brand. The

Table 10.2 Redefined biscuit categories

Subcategory	*Definition*
Chocolate countlines	Individually wrapped chocolate-covered biscuit bars which can be sold in multipacks, including Penguin, Club, Breakaway, Classic, Kit-Kat, Twix etc., which are marketed and packaged both as confectionary and biscuits.
Plain sweet biscuits	Plain sweet biscuits are uncoated, untopped or unfilled but can be flavored, for example coconut or chocolate, including chocolate chips, digestives, sweet assortment, shortbread, shortcakes, wafers, coconut, tea/coffee biscuits and ginger.
Chocolate-coated biscuits	Plain sweet biscuits coated partially, topped or completely with chocolate.
Filled biscuits	Sweet cookies which can either be filled, topped or sandwiched between plain biscuits.
Non-sweet biscuits	Plain savory biscuits, savory crackers and bread-like savory biscuits. Often flavored or topped with salt, cheese or other savory products.

	Count	*% Count*	*Volume*	*% Volume*
Countlines	17,293	15.2%	5,089,771	14.9%
Plain sweet	14,153	12.4%	4,696,549	13.7%
Chocolate coat	9,645	8.5%	3,414,240	10.0%
Filled	5,921	4.6%	1,538,860	4.5%
Non sweet	14,705	12.9%	3,231,637	9.4%

utilitarian scores are two categories and it will be more difficult for the data to show patterns which may correlate with the size of the brand. However, category 2 is seen to be a higher level of utilitarian benefit than group 1, therefore the utilitarian element is used as an ordinal/categorical variable and analysis techniques will reflect this.

With regards to the informational variable, the correlation with brand volume can be assessed. A Pearson's correlation test is considered. The informational variable is robustly normally distributed and hence no transformation is undertaken (see Figure 10.2).

The biscuit volume data are not normally distributed and hence a natural logarithmic transformation is undertaken with the resulting data being robustly normally distributed (recognizing longer tails). See Figures 10.3 and 10.4 respectively.

Table 10.3 Biscuit pack size distribution

	Count	*% Count*	*Volume*	*% Volume*
1	33743	55.2%	9,354,867	52.1%
2–5	3922	6.4%	1,195,547	6.7%
6–7	6880	11.3%	1,665,044	9.3%
8–11	7349	12.0%	2,056,553	11.4%
12+	6771	11.1%	2,597,211	14.5%
pack	2422	4.0%	1,101,835	6.1%

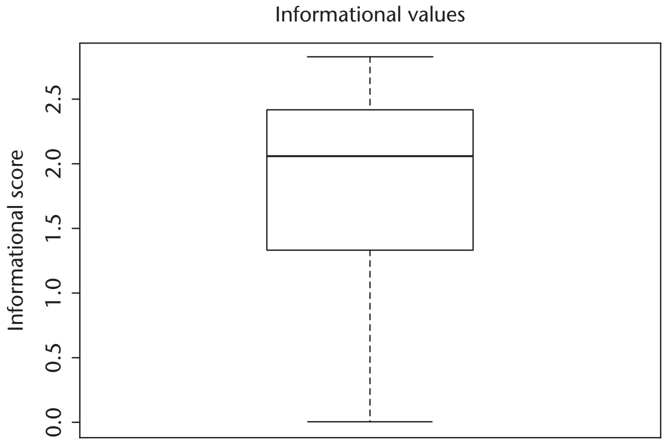

Figure 10.2 Boxplot of informational reinforcement

For DJ to be tested, the volume of each brand needs to be summed across the database. If DJ exists this will be correlated to the average informational score attributed to that brand. The correlation is set up by assuming the null hypothesis of no correlation between the informational and (naturally logged) volume. The two-tailed Pearson's correlation coefficient is .17, which is statistically significant and hence this null hypothesis is rejected, suggesting a positive correlation between the levels of informational scores and the volume size of each brand. Hence it seems the DJ phenomenon is present as larger brands have a statistically significant higher level of perceived informational benefit than smaller brands.

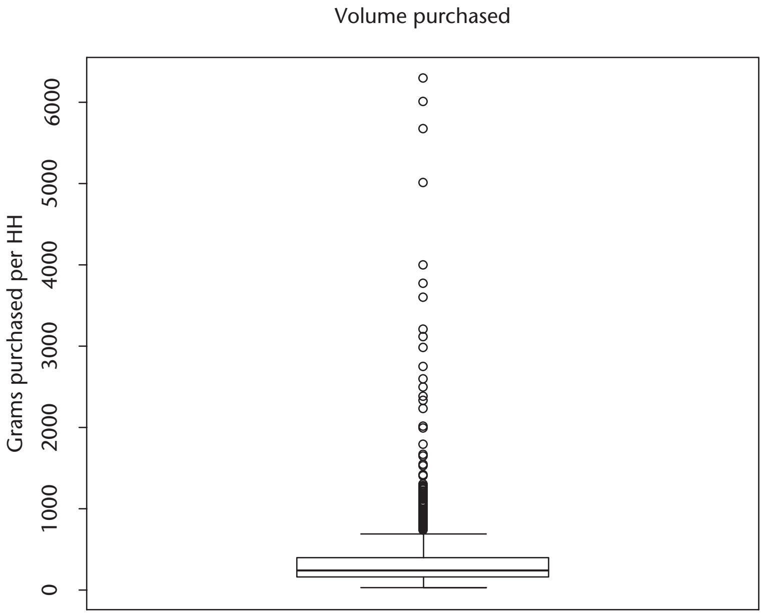

Figure 10.3 Boxplot of biscuit volume

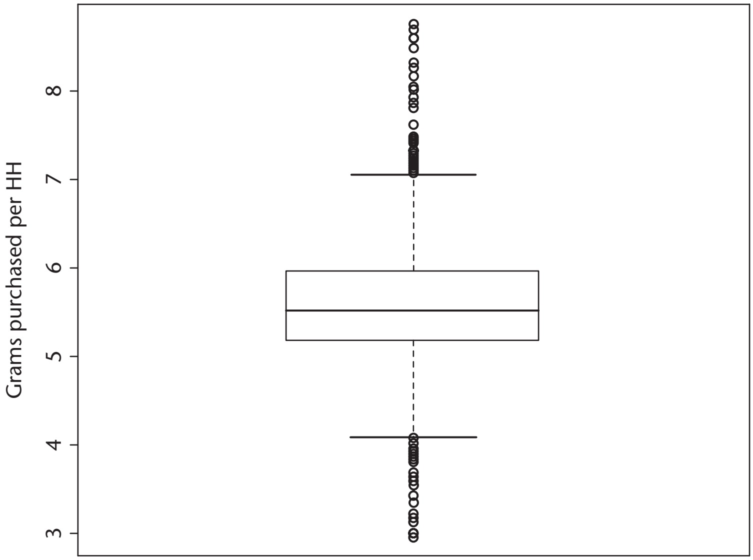

Figure 10.4 Boxplot of logged biscuit volume

Next, the analysis seeks to understand whether the higher utilitarian group has a higher average volume than the lower group. This would again inform the presence of DJ within the utilitarian element of the BPM (at least for the category in question). The mean LN (volume) values of the lower and upper utilitarian groups are 5.39 and 5.57 with 95% confidence intervals of (5.38, 5.40) and (5.57, 5.58) respectively. The lack of overlap of the confidence intervals suggests these values are significantly different to each other. A formal test is set up. The utilitarian variable is treated as categorical and hence an analysis of variance (ANOVA) approach is employed on the LN (volume) variable, given the Gaussian assumptions of the test. A two-tailed ANOVA is constructed with the null hypothesis of no difference between the means of the two groups. The ANOVA yields an F-ratio of 1,262, hence strong evidence to reject the null hypothesis and suggest the mean difference is statistically significant. This means that it is evident that the DJ phenomenon is present within the utilitarian group as higher utilitarian group brands have a higher mean volume.

Let us reflect on the analysis thus far. It has been shown that on a univariate basis the DJ of marketing seems to be present within the utilitarian and informational elements of the BPM when looking at the biscuit data set. It is important to note, however, these have been based on univariate analysis. Past studies within this category have shown other factors have implications for market share. Chang (2007) shows that the price elasticity of the biscuit category is circa –0.5 and hence changes in price will influence the volume of brands. The UK biscuit category is a stationary market and therefore Sharp et al. (2002) claim the DJ effect will be present in this category. Therefore it is wise to take these into consideration before conclusions are drawn.

Model build

As stated, the aim is to isolate the effect of the utilitarian and informational variables and test whether the DJ effect is present above and beyond those variables which would naturally be expected to predict the volume of the biscuit category. Hence a discussion as to the variables to be included in the model is presented.

Price. Recent studies have indicated the biscuit category data have a negative elasticity of demand and this value is close to –0.5 (Chang, 2007; Oliveira-Castro et al., 2006) and in general food products tend to be inelastic (Driel et al., 1997). To gain comparability across brand size, the price per 100g is used. The natural logarithm of this variable ensures the data are robustly normally distributed and the coefficient can be interpreted as the elasticity given a log-log model.

Supermarket own brand price. An offset relating to the prevalence of a supermarket own brand is included to account for systematic differences in supermarket elasticity.

Behavioral DJ effects. Traditional DJ studies show that larger brands have more purchasers and more loyalty among them. To account for the first of these measures, a variable is constructed that indicates how many unique purchasers (i.e. households) have purchased the product over the 52-week period. This is referred to as "*Penetration*". To represent these purchasers' loyalty, a variable is constructed to represent the share of how many times each brand was purchased within each individual's purchase history. This matches the share of category requirements (Goodhardt et al., 1984). This is called "*Loyalty*" and is the brand's share of each household's purchases. The larger this share, the larger the loyalty towards that brand. The combination of these two variables will allow for the identification of DJ within the purchases of a category, i.e. do smaller brands suffer twice by having a smaller set of purchasers (*penetration*) and less loyalty among this smaller set (*loyalty*)? If both these constructed variables have a statistically significant and positive coefficient then the DJ effect is evident.

Effects of utilitarian and informational reinforcement. The informational variable is included as scalar and utilitarian as dichotomous. The base value will relate to the lower utilitarian group of brands and the offset will allow the calculation of the higher utilitarian group. If the coefficient of the offset is positive and significant then this will indicate the presence of DJ within the utilitarian variables.

Including brand attributes as independent variables

Foxall et al. (2013) suggest that the characteristics of the brand may contribute to the purchase quantity of that brand. From a model perspective, the data being analyzed span many SKUs. Modeling all SKUs separately is demanding in terms of the degrees of freedom required (Fader & Hardie, 1996). Conversely, modeling all SKUs in a pooled functional form can ignore differences between SKU sizes and effects. An alternative structure proposed by Fader and Hardie (1996) sets to use the brand attributes as variables in themselves, arguing that consumers rarely choose brands but select an item based on their attributes. Singh et al. (2005, p. 196) and Webster (1994) agree and say consumers perceive a product as a "bundle of benefits" based on these attributes. Also Foxall (1987) says these characteristics are discriminative stimuli with regards to purchase response. The utilitarian variables from the earlier analysis seemed to have a positive effect on biscuit purchase volume, though product characteristics may be the cause of this and so they are included (Foxall et al., 2013).

Therefore, before firm conclusions can be made, the analysis proceeds by incorporating these terms into a multivariate model to test whether the utilitarian and informational DJ characteristics are present when simultaneously accounting for these other known factors. Equation 2 represents this model.

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1 LN(\Pr ice_j) \\
&+ \beta_2 Utilitarian_1 \star Informational_j + \beta_3 Utilitarian_2 \star Informational_j \\
&+ \beta_4 Supermarket \star LN(\Pr ice_j) \\
&+ \beta_5 Penetration_j \\
&+ \beta_6 Loyalty_j \\
&+ \sum_{i=1}^{5} \beta_{i+6} Biscuit\,type_i \\
&+ \sum_{i=1}^{6} \beta_{i+11} Pack\,type_i \\
&+ \varepsilon_j
\end{aligned}
$$

Equation 2: Non-hierarchical model:

$$where\ \varepsilon_j \sim N(0, \sigma^2) \qquad j = 1,2,\ldots,n \qquad (2)$$

Fixed and random effects

The Homescan panel data are constructed of multiple purchases within households. It is likely that purchases within households are less likely to be independent and a model to accommodate this is desired. This data structure lends itself to a hierarchical framework, simultaneously allowing different households to take different values at each level (Duncan et al., 1996). This is an advantage over conventional regression models which ignore any hierarchy in the data (Coker et al., 2013; Merino & Vargas, 2013) and can be subject to generalizing relationships, and hence "explain everything in general and nothing in particular". Multilevel models allow relationships to be calculated both across levels and also at specific levels (Duncan et al., 1996, p. 819). The multilevel model allows the cross-level estimation of the parameters, which may partially explain the variable that it is nested within (Merino & Vargas, 2013). Hence the assumption of data independence is relaxed (Field et al., 2012). Not accounting for the hierarchical structure may result in underestimated regression standard errors which may produce causal relationships when in fact the occurrence is through chance (Browne & Rabat, 2004).

Within multilevel models, variables can be modeled as fixed or random effects. A fixed effect is one where all possible variations of the population are contained in the variable. If the variable contains only a sample of the wider population values, it is deemed a random effect (Field et al., 2012). Within the Homescan database, the product attributes represent the available products within the population of the stores (or at least of the stores within the study). It is suggested these are treated as fixed effects. However, since the Homescan database is a representative sample of the UK population, it is logical these households are included as a random effect within the model.

Hence a random element is added to the current fixed effects model above. The random part of the model represents the purchaser and *panel id* is used to represent this. The functional form of the model is updated and shown in Equation 3.

$$
\begin{aligned}
LN(Volume_j) = \beta_{0j}Purchaser_{ij} \\
&\beta_{0j} = \beta_0 + \nu_{0j} \\
&+ \beta_1 LN(\Pr ice_j) \\
&+ \beta_2 Utilitarian_1 \mid Informational_j + \beta_3 Utilitarian_2 \mid Informational_j \\
&+ \beta_4 Supermarket \mid LN(\Pr ice_j) \\
&+ \beta_5 Penetration_j \\
&+ \beta_6 Loyalty_j \\
&+ \sum_{i=1}^{5} \beta_{i+6} Biscuit\,type_i \\
&+ \sum_{i=1}^{6} \beta_{i+11} Biscuit\,type_i \\
&+ \varepsilon_j
\end{aligned}
$$

Equation 3: Hierarchical model:

$$
where\ \varepsilon_j \sim N(0,\ \sigma^2) \qquad \nu_{0j} \sim N(0,\ \sigma^2_{\nu 0}) \qquad j = 1,2,\ldots,n \tag{3}
$$

Bayesian estimation of the parameters

Traditionally, statistics has been dominated by what is known as a frequentist approach (Poirier, 2006), so-called as probability is defined as the long-run frequency of an event (Koop et al., 2007). A statistical paradigm, which has come to be known as Bayesian statistics, was first published posthumously in 1763 in a work by the Reverend John Bayes titled "An essay towards solving a problem in the doctrine of chances".

Despite the Bayesian theorem being prevalent throughout statistics literature, applying the paradigm in the building of predictive models has been limited until relatively recently. The main reason is that the nature of the model requires integration of the posterior function, which for non-trivial functional forms is extremely complicated (Lunn et al., 2000). However, the development of a Markov Chain Monte Carlo (MCMC) algorithm in 1995 has led to strong growth within the Bayesian discipline (Poirier, 2006). This MCMC allows the posterior to be constructed by generation of a Monte Carlo style method which bypasses the need for integration of the function (Lunn et al., 2000). Instead, the posterior distribution is derived through a large number of iterations. This, paired with increased computational power during the same period (which directly facilitates this MCMC methodology), has led to Bayesian models being applied to a broad range of disciplines (Lunn et al., 2000). So much so that the Bayesian framework is now seen as a "well established alternative to classical inference" (O'Hagan, 1994, p. 1).

The Bayes theorem states the conditional probability of a parameter (θ) given the observed data (X_i) is proportional to the probability of the data given the parameter, multiplied by the probability of the parameter (Congdon, 2003). Or mathematically,

$$
P(\theta \mid X_i) \propto P(X_i \mid \theta)P(\theta)
$$

The left-hand part of the equation is known as the posterior probability. The right-hand side terms are known as the likelihood and prior, respectively. A major difference of Bayesian statistics over frequentist statistics is the embracing of the prior knowledge about an event or parameter. This prior is the initial belief of a parameter or event before any (new) data are considered. It can come from past studies, expert opinion or what may be considered as common sense (Hansen et al., 2004; O'Hagan, 1994). The likelihood is the addition of new data to be evaluated. The posterior probability is the blend of the prior and the likelihood, and the result is seen as the updated view of the estimate of the parameter (or event). Bernardo (1999) argues this concept closely matches what is seen in everyday life and Bernardo and Smith (2000, p. 4) say it illustrates how beliefs "fit together in the light of changing evidence".

However, risk assessment work by Viscusi (1985) states that a person's prior knowledge can be systematically biased and, although not criticizing Bayesian learning, points out the challenges by citing work by Lichtenstein (1978). This work demonstrates individuals over-assessing small risks and under-assessing larger risks. Indeed the frequentist perspective argues any knowledge should be free from any interpretation bias; however, Leamer (1992) says that in practice the researcher must have some prior incline and would reject any absurd parameters anyway. Koop (2006) argues that more information is better than less and allows for uncertainty to be embraced and quantified, not ignored in the decision-making process (Aspinall, 2010), while O'Hagan (1998, p. 21) says developing realistic prior estimates is preferable to relying on "ignorance".

Another difference in the paradigms relates to the parameter distribution itself. The frequentist views a parameter of a model as unknown but fixed (Abelard, 2013). This means the parameter has a definite value and the analysis is the probability of observing the data given the estimated parameter value (Abelard, 2013), i.e. $P(X_i | \theta)$. The Bayes theorem turns this on its head and calculates the probability of the parameter, given the data, i.e. $P(\theta | X_i)$. While no criticism of the frequentist methodology is offered by the author, it seems clear why O'Hagan (1994) says the Bayesian interpretation is more intuitive to management and allows more transparent means of embracing the uncertainty of a parameter. The debate of the Bayesian vs. frequentist approach has been prevalent in statistical literature and this text is not intended to input into this discussion. The author acknowledges the benefit of Bayesian methodology has been favored in this short text, though his wider philosophical view is very much in line with Efron (2005), who states that the 21st century will involve problems being resolved using a combination of different analytical tools and approaches from both paradigms. The purpose of this study is not to evaluate the paradigms, but to use their benefits and contextualize the results within a management decision framework. Hansen et al. (2004) make use of relevant frequentist diagnostics when evaluating the relevance of the Bayesian model and parameter estimates.

WinBUGS

To calculate these Bayesian functions, specialized software is required. The BUGS Bayesian software stands for Bayesian inference Using Gibbs Sampler (Lunn et al., 2000). WinBUGS is the Windows version of the software. Both use the MCMC method to analyze Bayesian full probability models (Lunn et al., 2000). The Gibbs sampler samples from the full conditional probability of each parameter in turn, keeping the others constant, and convergence is monitored and achieved after a number of iterations (Spiegelhalter et al., 2003).

An overview only of the WinBUGS process is offered here. The model is specified within the WinBUGS environment. The software first checks the syntax of the model. Next the data are loaded and WinBUGS checks they are of the correct format. The model is compiled, which creates the data structure for the Gibbs sampler. Finally, initial values are loaded or generated from the prior distributions of each parameter (Spiegelhalter et al., 2002). The parameters need to be monitored to ensure convergence and also to obtain the posterior statistics when convergence is achieved (Spiegelhalter et al., 2002).

Prior discussion of informative model

The Bayesian model requires the inclusion of prior estimates of the parameter. Congdon (2003) says it may be difficult to do and sometimes better to resort to non-informative priors. Therefore a model is initially constructed where the prior information is considered non-informative or vague and is defined as normally distributed with mean zero and small precision (0.0001). The term precision is used when referring to the prior distribution within a Bayesian modeling framework and is the inverse of the variance (Congdon, 2003). This model will be known as model "vague".

A second mixed-effect model is constructed where the move is to informed priors, embracing knowledge available from past studies or from common sense. Refer to this as model "informative". Informed priors are now discussed.

Price. It is natural to assume that price elasticity will be negative given the non-luxury nature of the category. Also, as discussed earlier, other studies of the same data suggest this measure is of small magnitude. However, given the different model structure, the author does not wish to impose too rigid a prior distribution and hence a truncated normal distribution is applied where the upper boundary is controlled to be a maximum of 0. The small precision allows the model to produce any estimate on the negative real number line $\Re^-$.

DJ variables. The traditional DJ variables would both be positive given the prevalence of the phenomenon, but the author poses no more influence other than this, as these values have not been run in similar studies and hence the data should have a greater influence over and above this truncation. Hence a lower limit of 0 is offered. The small precision allows the data to take any range within the positive real numbers $\Re^+$.

The informational and utilitarian variables would both be positive if DJ exists and hence a lower boundary of zero and small precision will infer a coefficient estimate within $\Re^+$. Other coefficients offer no prior logic of scale or sign and hence these are retained with vague prior distributions.

Model diagnostics

The coefficients of the three models are shown in Table 10.4. The R-squared (adjusted) values for the hierarchical models are far larger than the non-hierarchical model, suggesting that allowing for within-household effects is positively contributing to the fit of the model. Table 10.4 below shows the mean estimate of the parameter and their standard error.

As well as the higher R-squared (adj) value, the deviance information criteria (DIC) measurement for the hierarchical model is less than the non-hierarchical (45,543 vs. 71,677, respectively), suggesting the inclusion of the random element better predicts a replicated data set of the same structure (Spiegelhalter et al., 2002). DIC is preferred in this case to the Bayesian information criteria (BIC) as the random effects are of interest (Spiegelhalter et al., 2002). Furthermore the non-hierarchical model will assume all observations are independent; however, the percentage of the random variance component versus the total model residual is:

$$\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} = 32.7\%$$

demonstrating the importance of the household random effect in estimating the model. The parameters are visualized graphically in Figure 10.5, demonstrating the differences between the hierarchical and non-hierarchical estimates.

Table 10.4 Model results

	Vague		*Informative*		*Non-hierarchical*	
	Beta (st err)		*Beta (st err)*		*Beta (st err)*	
Log price	–0.5537	(0.0036)	–0.5482	(0.00631)	–0.6778	(0.00387)
Informational	0.0206	(0.00271)	0.0207	(0.00301)	0.0165	(0.0032)
Informational (Util 2)	0.0305	(0.00329)	0.0302	(0.00331)	0.0541	(0.00379)
Log price (Sprmkt)	0.0838	(0.00322)	0.0818	(0.00373)	0.1351	(0.00373)
Penetration	0.0007	(0.00008)	0.0006	(0.00011)	0.0025	(0.00008)
Loyalty	0.1575	(0.00094)	0.1617	(0.00442)	0.0727	(0.00074)
Coatline	base		base		base	
Chocolate	0.1028	(0.00534)	0.1022	(0.00547)	0.1204	(0.00633)
Plain sweet	0.1100	(0.00744)	0.1097	(0.00738)	0.1581	(0.0087)
Filled	0.0045	(0.00663)	0.0055	(0.00702)	–0.0135	(0.00776)
Non sweet	0.0048	(0.0086)	0.0067	(0.00856)	0.0140	(0.00958)
Number 1	base		base		base	
Number 5	0.1500	(0.00627)	0.1482	(0.00649)	0.1930	(0.00758)
Number 7	0.0971	(0.00562)	0.0964	(0.00556)	0.1030	(0.00644)
Number 11	0.1749	(0.00604)	0.1733	(0.00657)	0.2119	(0.00733)
Number 12	0.2512	(0.00566)	0.2486	(0.00623)	0.3348	(0.0066)
Number pack	0.3653	(0.0079)	0.3584	(0.01033)	0.5030	(0.00909)
R-squared *(adj)*	71.0%		71.1%		53.2%	

The similarity (almost exact) of the vague and informative hierarchical models is due to the high level of agreement between the prior distribution and the likelihood. The informative prior distributions are relatively weak in terms of magnitude but tight in terms of signs and the likelihood is reinforcing this prior belief. From the coefficients, it can be seen that assuming independence between all observations and hence ignoring the within-household hierarchy leads to possible overestimation of the price elasticity, supermarket own brand effect on price elasticity, informational benefits in the higher utilitarian group and penetration, while underestimating the effect of informational impact and loyalty.

Given the similarity of the two hierarchical models, one can be discarded. The vague mixed model is used as the basis for the mixed effects model as the prior belief is not impacting the coefficients. The actual and predicted fit of the hierarchical model is shown in Figure 10.6, showing the relationship between the data and the included variables.

The residual plot is shown in Figure 10.7 and demonstrates a spread of residuals, with a few exceptions. While interesting to understand which characteristics may be driving these outliers, given the charted fits and the R-squared (adjusted) values, the discussion assumes a good model diagnostics and a discussion of the parameter values now follows.

Discussion

Price elasticity

Given the log-log model, the coefficient is the price elasticity. The non-hierarchical model shows a higher coefficient (slightly) than the mixed model, suggesting the ignorance of the

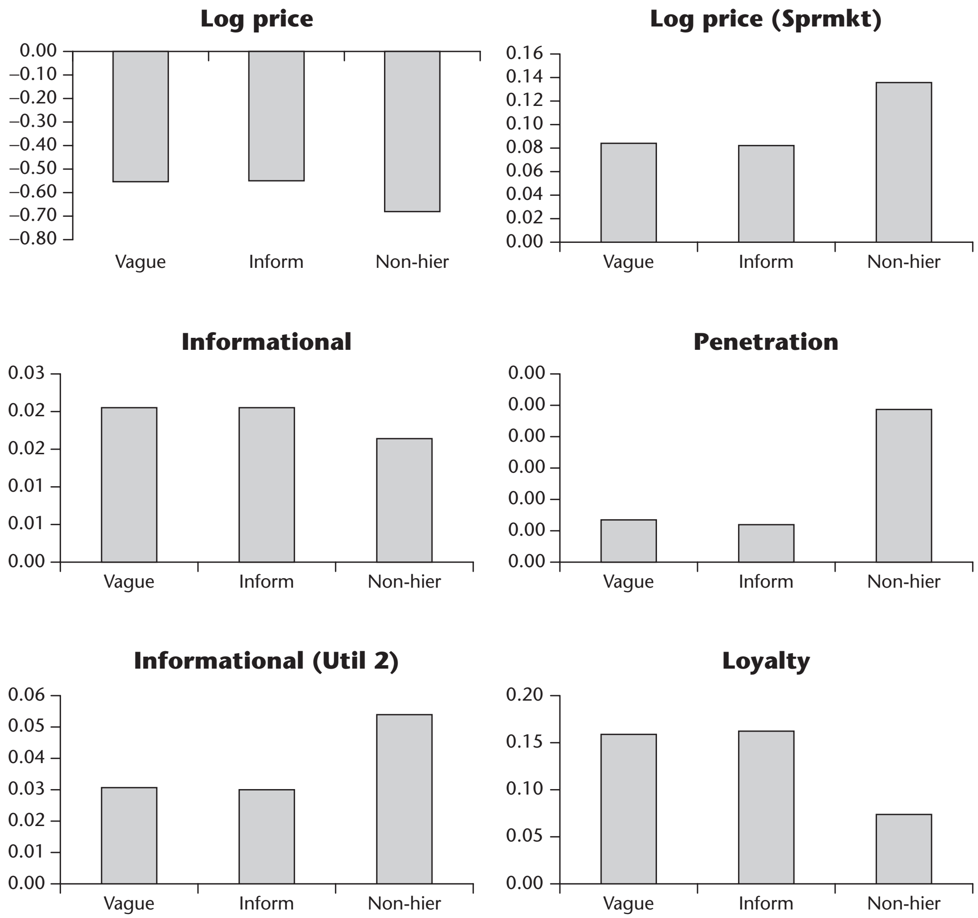

Figure 10.5 Plot of model coefficients

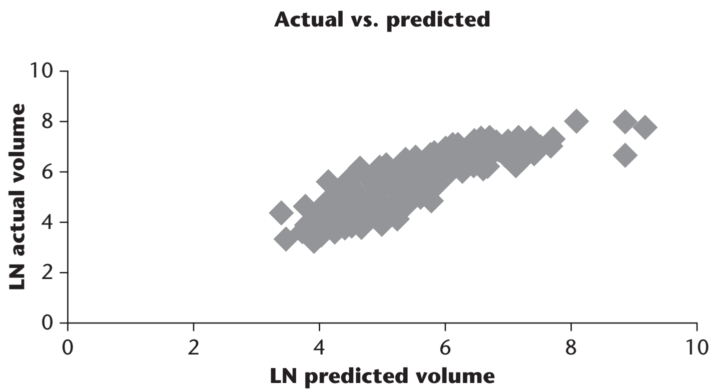

Figure 10.6 Actual vs. predicted

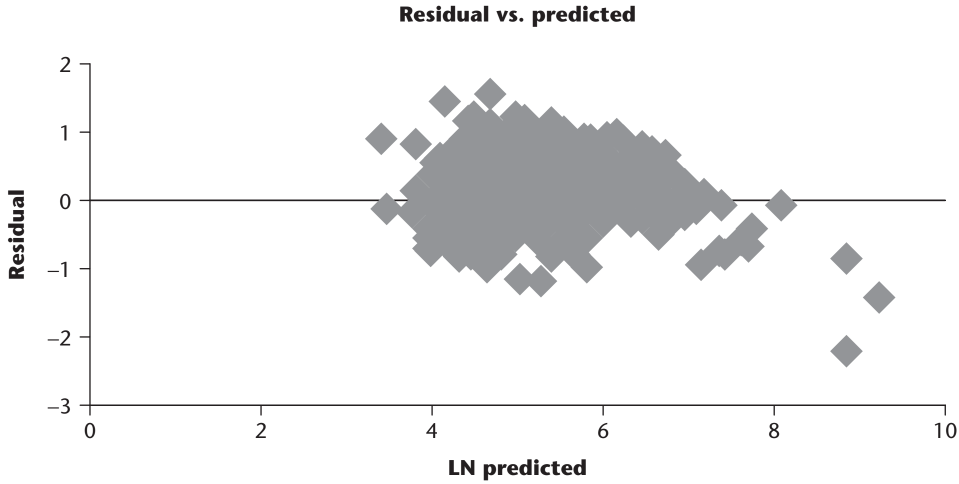

Figure 10.7 Residual plot

hierarchy of the data may produce a greater magnitude coefficient for price. Recall, the Bayesian estimation allows the probability distribution of the parameter to be estimated. The mixed-level model result is displayed in Figure 10.8.

This can be interpreted by saying that given the prior information and the likelihood gained from the data, there is a 95% probability that the posterior estimate of the parameter is between −.560 and −.546, with −.554 being the mean estimate. Whereas frequentist methods would construct a hypothesis around the probability of observing the data, given the parameter estimate of −0.554, Bayesian inference instead calculates the probability of the value of the parameter. This posterior parameter is normally distributed, given the assumptions of the prior distribution. If the distribution of the parameter straddles zero then there is a probability the parameter may be zero (and hence redundant). However, there is a very small probability that the parameter is in fact greater than −0.543.

Figure 10.8 Price coefficient

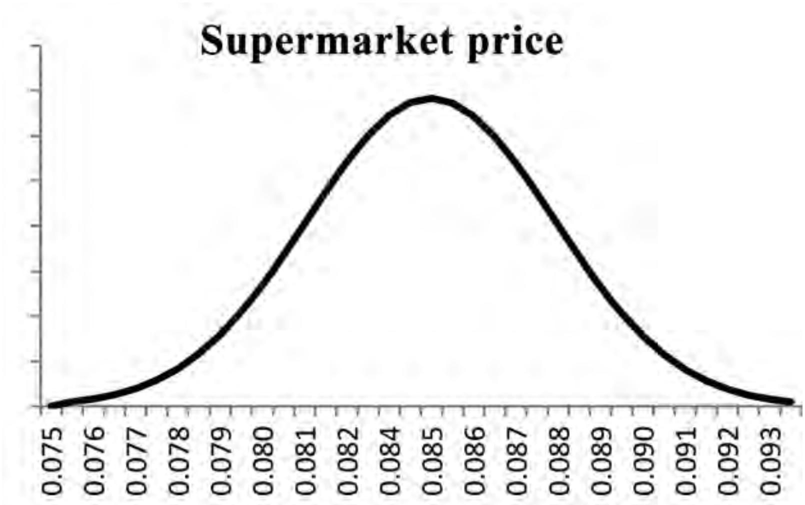

Figure 10.9 Supermarket own price offset coefficient

Supermarket own brand price elasticity offset

For both hierarchical and non-hierarchical models, the offset for supermarket own brands is positive and significant with values of .08 and .14, respectively. This means supermarket own brands are less responsive to average price changes. This may be due to the promotional nature of the category where branded names tend to benefit from price reductions over supermarket own brands. The hierarchical model's distribution of the posterior of this parameter is shown in Figure 10.9. Again, there is no reason to believe this parameter is redundant as there is very little probability that the distribution contains the value 0.

Traditional DJ variables

The penetration (or number of households purchasing) variable is linear and hence, given the dependent variable is naturally logged, the coefficient is raised to the exponential to interpret the value.

$$\text{linear coef} = e^{\beta} - 1$$

The linear coefficient is therefore 0.246% and 0.068% for the non-hierarchical and hierarchical models, respectively, so there is a factor of 10 difference between them. Figure 10.10 shows the hierarchical posterior distribution of the penetration parameter. The numbers are small but still there is just a very small chance that the parameter value is zero and hence redundant in the model.

What does this mean? There are 1592 unique households purchasing the category. If one extra household decides to purchase, then this is an increase of 1/1592 = 0.063% increase in the number of households buying. This is compared to a change of 0.068% change in the volume for the hierarchical model. Equating these gives a 0.068%/0.063% = 1.09. Hence every 1% increase in the number of households purchasing a biscuit brand will increase the volume by 1.09% on average. The equivalent number for the non-hierarchical model is 3.9%. This would

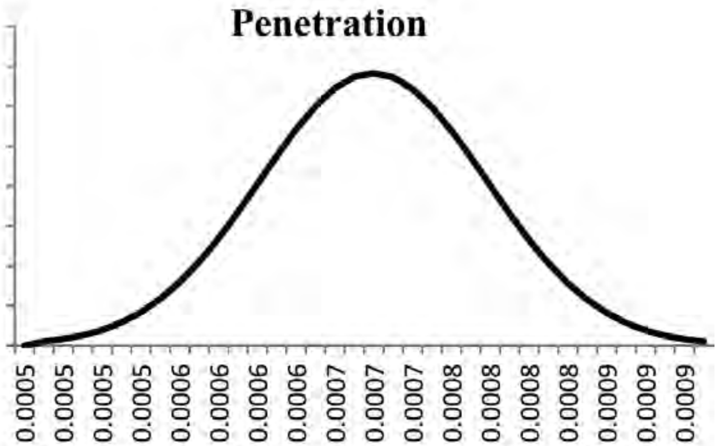

Figure 10.10 Penetration coefficient

imply that increasing the households by 1% increases volume by 3.9%. Logic would suggest that increasing the amount of households purchasing by 1% would increase the volume by a like 1% which is remarkably close to what the hierarchical coefficient suggests.

The loyalty measure is already in percentage terms (i.e. values fluctuate between a possible 0 and 100%). Hence the measures, when modeled against the logged volume, closely reflect the percentage change in the variable and therefore can be read as elasticity measures. Figure 10.11 shows the hierarchical model's posterior distribution for loyalty.

The elasticity measure for the hierarchical model is 0.16% and 0.07% for the non-hierarchical model so again there is some difference between the two.

We have noted that Ehrenberg et al. (1990) state that DJ commands smaller brands have fewer consumers and less loyalty among the smaller set of consumers. The same effect is evident here with the positive and statistically significant coefficient signifying the same observation.

Figure 10.11 Loyalty coefficient

Ehrenberg et al. (1990) state that penetration is more important than loyalty and this pattern is observed in this data. Hence the DJ effect is apparent within the data, above and beyond the effect of price.

Utilitarian and informational reinforcement variables

The informational variable is the base value and the informational variable for utilitarian group 2 (the higher group) is an offset, hence the base informational coefficient can be interpreted as the value for utilitarian group 1 (the lower utilitarian group). Adding the offset will give the value for utilitarian group 2.

The coefficients are transformed to linearity in the same way as the DJ variables:

$$\text{linear coef} = e^{\beta} - 1$$

Figure 10.12 shows the hierarchical posterior distribution estimates of the informational variable.

The values for the lower utilitarian groups are 0.021 and 0.017 for the hierarchical and non-hierarchical models, respectively, showing more similarity between the two models, though the hierarchical model in general is preferred for the reasons discussed. From Figure 10.12, again we see very little evidence to suggest this parameter is zero. Therefore we see that the nature of the positive coefficient suggests that larger (volume) brands within the lower utilitarian group are being perceived to have a higher informational benefit than smaller brands, over and above what can be accounted for by behavioral economic and traditional DJ effects.

Figure 10.13 shows the hierarchical posterior distribution for the offset informational variable for utilitarian group 2.

The offset values are positive and there is no evidence to suggest that the value of the parameter is non-positive given the distribution of the parameter in Figure 10.13. This suggests the higher utilitarian group is enjoying a higher perceived level of informational benefit over the lower utilitarian group. Combining the results of the two informational variables, it can be seen that, within the BPM structure, having taken account of known behavioral economic variables and traditional DJ effects, there seems to be a TJ effect, in that larger brands are being perceived

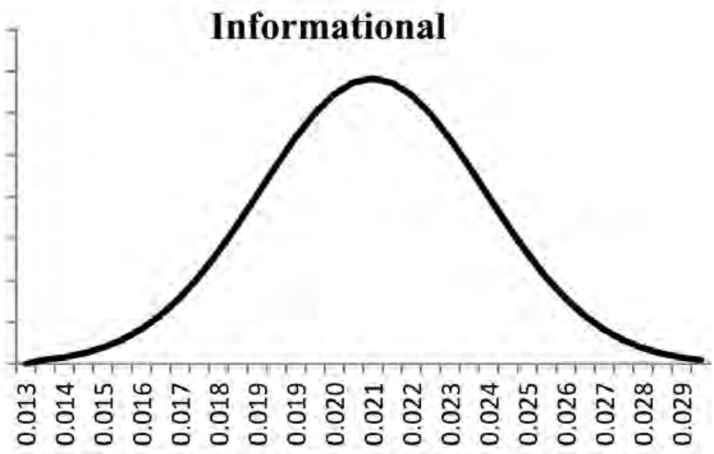

Figure 10.12 Informational coefficient

Figure 10.13 Informational coefficient offset for utilitarian group 2

to have a higher level of informational benefits. Furthermore, this effect is greater for brands which are perceived to have a greater utilitarian benefit. At least, this effect seems present within the UK biscuit category.

Conclusion

Within the biscuit category in the UK, a model has been built to describe consumer behavior. The model demonstrates that the volume of the brands within the category can be explained by a combination of elements from behavioral economics, marketing theory and psychological factors. Combining these complements the level of consumer behavior understanding which can be gained from this particular data set.

Price elasticity of demand in the UK biscuit category is negative and inelastic, in line with past studies (Chang, 2007; Driel et al., 1997; Oliveira-Castro et al., 2006) with a coefficient of Regular 0.554 (which decreases in magnitude to −0.47 for supermarket own brands). This change in magnitude can be explained through the promoted nature of the category where supermarket own label brands are not responding as well to promotional reductions as branded products.

The two constructed DJ variables are positive and hence show that smaller brands are suffering twice: less penetration and less loyalty among the smaller set of buyers. This is in line with marketing theory discussed mainly through the DJ studies of Ehrenberg. The elasticity of the penetration variable of 1.09% suggests an increase of 1% in the number of households would yield a 1.09% increase in volume, which is logical. Furthermore, the loyalty elasticity of 0.16% is much lower, again in line with DJ theory discussed earlier. This shows the importance of penetration in stimulating brands' growth. However, Ehrenberg et al. (1990) say that for DJ to be present, the brands need to be indistinguishable in a blind taste test. However, the DJ effect is apparent within this taste- and shape-diversified category which questions whether this assumption is required. Similarly, how a blind taste test can be applied to, for example, aviation fuel, comics or radio stations is questionable.

While the DJ effect is apparent, Ehrenberg and Goodhardt's (2002) claim that brands are not strong or weak in equity, simply large or small in size, is challenged by the existence of a further psychological effect demonstrated in the model. While much credibility is given to the claim, there is also an effect seen from the informational and utilitarian elements of the BPM. For the lower utilitarian group, a one-point increase in informational score would produce a 2.1% increase in volume (and a 3.1% increase in the higher utilitarian group). Therefore the perception of informational and utilitarian values is also exhibiting a DJ-type effect above and beyond what can be described by price elasticity or the (traditional) DJ variables. This TJ effect could start to explain the importance of psychological reinforcements to consumers, which may be construed as a form of equity. This may potentially start to explain why manufacturers are willing to pay a premium for brands eliciting stronger equity. By no means is the author stating this in any way replaces or supersedes the "law-like" DJ effect, but it is additional to it. Note the scale of the informational variable is a three-point scale. Hence the level of incremental volume which could be realized *purely* through the BPM informational and utilitarian reinforcements[1] would be a maximum of 6.2% (moving from level 1 to level 3 in the higher utilitarian group). Compare this to the DJ effect, where, according to this model, a 6% increase in penetration would render a ($6 \times 1.09 = 6.54\%$) increase in volume and also a ($6 \times 0.165 = 0.96\%$) further increase in volume through loyalty (so a total of 7.5%). Though this study is currently just for one category in one market, it does have implications for management.

Prior to this study, the recommended way in which to grow a brand would be through price changes/discounting and/or through increasing penetration (which would also increase loyalty a little). Focusing for the moment on the marketing theory and assuming constant pricing, increasing penetration for smaller or medium-sized brands clearly shows the benefit of the DJ theory in increasing volume. However, brands with high penetration may be reaching saturation point and hence the TJ effect presents a further opportunity to management to consider increasing brand size.

The study also offers contributions to the topic in the way the model is technically built. The hierarchical model structure better represents the data with a higher R-squared (adjusted) value than the non-hierarchical model and the random effects variance component accounts for 32.7% of the total variability of the model. The removal of the assumption of independence of purchases within households is arguably more logical than assuming purchases are independent. This allows the model to better understand purchase levels of brands given a specific household and hence allows the parameters to better represent the behavior. In some instances the difference between the hierarchical and non-hierarchical model is small (e.g. log price) and some are much higher (e.g. penetration coefficient is over twice the magnitude). The implications to management in ignoring the hierarchical structure of the data could lead to over or underestimation of the anticipated effect of marketing levers on consumer behavior. The random effects here are limited to the household intercept as these are a random sample of the population, whereas the other factors in the model are representative of the population (Field et al., 2012).

Bayesian estimation has been used in this instance to estimate the parameters. Bayesian estimation allows for prior information to be mathematically incorporated into the model. The informative model has produced estimates almost exactly in line with the non-informative model and hence the prior information is in strong agreement with the likelihood, resulting in aligned posterior estimates. The Bayesian estimation process also allows the direct observation of the posterior distribution of the parameters, given the data, and hence a confidence interval of the mean of the estimate can be directly observed. For each parameter, the posterior distribution does not straddle zero, indicating a high probability that each parameter is contributing to the model predictiveness.

Further research

The current research is restricted to one category within one market and hence further research is required to verify results and offer any general hypothesis. The coefficients of the DJ seem sensible, but further studies could confirm or contradict these findings. The price elasticity is very much in line with other studies and hence a higher degree of confidence is maintained. The "triple jeopardy" effects are new to this study and again need to be tested further in different studies. However, this model provides a good representation of the data; the assumptions and structure it offers make logical sense and the interpretation seems a sensible offering for management decisions.

Note

1 This is purely the informational and utilitarian effects of the model, not accounting for situational or learning history effects.

References

Aaker, D. (1991). *Managing brand equity*. New York: The Free Press.

Abelard (2013). *Cause, chance and Bayesian statistics*. Available at http://www.abelard.org/briefings/bayes.htm.

Allen, D. (1990). Creating value – The financial management of brands. *CIMA*. Available at http://hdl.handle.net/10068/507293.

Aspinall, W. (2010). A route to more tractable expert advice. *Nature* 463(21), 294–295.

Baldinger, A. L. and Rubinson, J. (1997). The jeopardy in Double Jeopardy. *Journal of Advertising Research* November, 37–49.

Baldinger, A. L., Blair, E. and Echambadi, R. (2002). Why brands grow. *Journal of Advertising Research*, 7–14.

Bandyopadhyay, S., Gupta, K. and Dube, L. (2005). Does brand loyalty influence double jeopardy? A theoretical and empirical study. *Journal of Product and Brand Management* 14, 414–423.

Bassi, F. (2011). The Dirichlet model: analysis of a market and comparison of estimation procedures. *Marketing Bulleting* 22, Technical Note 1.

Bernardo, J. M. (1999). Bayesian methods in the sciences. *Special Issue of Rev. Acad. Cien Madrid* 93. Available at http://www.uv.es/~bernardo/academia.html.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian theory*. Chichester: Wiley.

Bhat, S. and Fox, R. (1996). An investigation of jeopardy effects in store choice. *Journal of Retailing and Consumer Services* 3(3), 129–133.

Browne, W. and Rabat, J. (2004). *New developments in modeling, multilevel modeling*. In M. Hardy and A. Bryman eds. *Handbook of data analysis*. London: Sage.

Chang, S. W. (2007). *Behavioral economics: an empirical investigation of consumer buying behavior and price responsiveness in the British biscuit market*. PhD: Cardiff.

Chaudhuri, A. (1995). Brand equity or double jeopardy? *Journal of Product and Brand Management* 4, 26–32.

Chaudhuri, A. (1999). Does brand loyalty mediate brand equity outcomes? *Journal of Marketing Theory and Practice* 7, 136–146.

Chaudhuri, A. and Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty. *Journal of Marketing* 65(2), 81–93.

Coker, C., Altobello, S. A. and Balasubramanian, S. K. (2013). Message exposure with friends: the role of social context on attitudes toward prominently placed brands. *Journal of Consumer Behaviour* 12, 102–111.

Colombo, R. A. and Morrison, D. G. (1989). A brand switching model with implications for marketing strategies. *Marketing Science* 8, 89–99.

Congdon, P. (2003). *Applied Bayesian modeling*. London: Wiley.

Driel H. V., Nadall, V. and Zeelenberg, K. (1997). The demand for food in the United States and the Netherlands: a systems approach with the CBS model. *Journal of Applied Economics* 12, 509–532.

Duncan, C., Jones, K. and Moon, G. (1996). Health related behavior in context: a multilevel approach. *Social Science and Medicine* 42, 817–830.

Dyson, P., Farr, A. and Hollis, N. S. (1996). Understanding, measuring and using brand equity. *Journal of Advertising Research* November/December, 9–21.

Efron, B. (2005). Bayesians, frequentisits and scientists. *Journal of the American Statistical Association* 100, 1–5.

Ehrenberg, A. S. C. (1972). *Repeat-buying: theory and applications*. New York: North-Holland.

Ehrenberg, A. S. C. and Goodhardt, G. (2000). New brands: near-instant loyalty. *Journal of Marketing Management* 16, 607–617.

Ehrenberg, A. S. C. and Goodhardt, G. (2002). Double Jeopardy revisited again. *Marketing Learnings* 7, The Southbank University, 1–2.

Ehrenberg, A. S. C., Goodhardt, G and Barwise, P. T. (1990). Double Jeopardy revisited. *Journal of Marketing* 54, 82–91.

Fader, P. S. and Schmittle, D. C. (1993). Excess behavioral loyalty for high-share brands: deviations from the Dirichlet model for repeat purchasing. *Journal of Marketing Research* 30, 478–493.

Fader, P. S. and Hardie, B. G. S. (1996). Modeling consumer choice among SKUs. *Journal of Marketing Research* 33 (November), 442–452.

Field, A., Miles, J. and Field, Z. (2012). *Discovering statistics using R*. London: Sage.

Fleury, M. (2013). *Yahoo to buy Tumblr for $1.1bn*. Available at http://www.bbc.co.uk/news/business-22591026.

Foxall, G. R. (1986a). Consumer theory: some contributions of a behavioral analysis of choice. *Management Bibliographic and Reviews* 12, 27–51.

Foxall, G. R. (1986b). The role of radical behaviorism in the explanation of consumer choice. *Advances in Consumer Research* 13, 187–191.

Foxall, G. R. (1987). Radical behaviorism and consumer research. *International Journal of Research in Marketing* 4, 111–129.

Foxall, G. R. (1990/2004). *Consumer psychology in behavioral perspective*. London and New York: Routledge. (Reprinted 2004 by Beard Books, Frederick, MD.)

Foxall, G. R. (2001). Foundations of consumer behavior analysis. *Marketing Theory* 1, 165–199.

Foxall, G. R. (2004). *Context and cognition: interpreting complex behavior*. Reno, NV: Context Press.

Foxall, G. R. (2005). *Understanding consumer choice*. New York: Palgrave Macmillan.

Foxall, G. R. (2007). Explaining consumer choice: coming to terms with intentionality. *Behavioural Processes* 75, 129–145.

Foxall, G. R. (2010). *Interpreting consumer choice: the behavioral perspective model*. New York: Routledge.

Foxall, G. R., Yan, J., Oliveira-Castro, J. M. and Wells, V. K. (2013). Brand-related and situational influences on demand elasticity. *Journal of Business Research*, 66, 73–81.

Foxall, G. R. and James, V. K. (2003). The behavioral ecology of brand choice: how and what do consumers maximize? *Psychology and Marketing* 20, 811–836.

Foxall, G. R., Oliveira-Castro, J. M. and Schrezenmaier, T. C. (2004). The behavioral economics of consumer brand choice: patterns of reinforcement and utility maximization. *Behavioural Processes* 66, 235–260.

Foxall, G. R., Oliveira-Castro, J. M., James, V. K., Yani-de-Soriano, M. and Sigurdsson, V. (2006). Consumer behavior analysis and social marketing: the case of environmental conservation. *Behavior and Social Issues* 15, 101–124.

Foxall, G. R. and Schrezenmaier, T. C. (2003). The behavioral economics of brand choice: establishing a methodology. *Journal of Economic Psychology* 25, 675–695.

Foxall, G. R., Yan, J., Oliveira-Castro, J. M. and Wells, V. K. (2013). Brand-related and situational influences on demand elasticity. *Journal of Business Research* 66, 73–81.

Goodhardt, G. J., Ehrenberg, A. S. C. and Chatfield, C. (1984). The Dirichlet: a comprehensive model of buying behavior. *Journal of the Royal Statistical Society A* 147, 621–655.

Hansen, M. H., Perry, L. T. and Reese, C. S. (2004). A Bayesian operationalization of the resource-based view. *Strategic Management Journal* 26, 1279–1295.

Jobber, D. (2004). *Principles and practice of marketing* (4th ed.). London: McGraw-Hill.

Jones, G. and Morgan, N. J. (1994). *Adding value brands and marketing in food and drink*. London: Routledge.

Koop, G. (2006). *Bayesian econometrics*. Chichester: Wiley.

Koop, G., Poirier, D. J. and Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge: Cambridge University Press.

Kotler, P., Armstrong, G., Saunders, J. and Wong, V. (1999). *Principles of marketing*. 2nd European edition. Upper Saddle River, NJ: Prentice Hall.

Labeaga, J. M., Lado, N. and Mercedes, M. (2007). Testing the double jeopardy loyalty effect using discrete choice models. *Documento De Trabajo* 21, 1–22.

Leamer, E. E. (1992). Bayesian elicitation diagnostics. *Econometrica* 60, 919–942.

Lichtenstein, S. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology* 4, 557–578.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modeling framework: concepts, structure and extensibility. *Statistics and Computing* 10, 325–337.

McPhee, W. N. (1963). *Formal theories of mass behavior*. New York: The Free Press.

Merino, M. and Vargas, D. (2013). How consumers perceive globalization: a multilevel approach. *Journal of Business Research* 66, 431–438.
Miller, D. (1995). Consumption and commodities. *Annual Review of Anthropology* 24, 141–161.
O'Hagan, A. (1994). *Kendall's advanced theory of statistics, volume 2b: Bayesian inference*. Cambridge: Cambridge University Press.
O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society Series D* 47, 21–35.
Oliveira-Castro, J. M., Foxall, G. R. and Schrezenmaier, T. C. (2005). Patterns of consumer response to retail price differentials. *Service Industries Journal* 25, 309–335.
Oliveira-Castro, J. M., Foxall, G. R. and Schrezenmaier, T. C. (2006). Consumer brand choice: individual and group analysis of demand elasticity. *Journal of the Experimental Analysis of Behavior* 85, 147–166.
Oliveira-Castro , J. M., Foxall, G. R., James, V. K. B., Pohl, R. H. B. F., Dias, M. B. and Chang, S. W. (2008). Consumer-based brand equity and brand performance. *The Service Industries Journal* 28, 445–461.
Poirier, D. J. (2006). The growth of Bayesian methods in statistics and economics since 1970. *Bayesian Analysis* 1, 969–980.
Romero, S., Foxall, G. R., Schrezenmaier, T. C., Oliveira-Castro, J. M. and James, V. (2006). Deviations from matching in consumer choice. *European Journal of Bevavior Analysis* 7, 15–39.
Sharp, B. and Riebe, E. (2005). *Does Triple Jeopardy exist for retail chains?* Adelaide: Marketing Science Centre University of South Australia, School of Marketing.
Sharp, B., Riebe, E. and Dawes, J. (2002). A marketing economy of scale – big brands lose less of their customer base than small brands. *Marketing Bulletin* 13, 1–8.
Sharp, B., Wright, M., Dawes, J., Driesener, C., Meyer-Waarden, L., Stocchi, L. and Stern, P. (2012). It's a Dirichlet world. Modeling individuals' loyalties reveals how brands compete, grow, and decline. *Journal of Advertising Research*, 203–213.
Singh, V. P., Hansen, K. T. and Gupta, S. (2005). Modeling preferences for common attributes in multicategory brand choice. *Journal of Marketing Research* 42, 195–209.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* 64, 583–639.
Taylor, S. A., Celuch, K. and Goodwin, S. (2004). The importance of brand equity to customer loyalty. *Journal of Product and Brand Management* 13, 217–227.
Viscusi, W. K. (1985). Are individuals Bayesian decision makers? *The American Economic Review, Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association* 75, 381–385.
Webster, F. E. Jr. (1994). *Market-driven management*. New York: Wiley.
Wood, L. (2000). Brands and brand equity: definition and management. *Management Decision* 38, 662–668.
Wright, M. and Sharp, B. (2001). The effect of a new brand entrant on a market. *Journal of Empirical Generalizations in Marketing Science* 6, 15–29.