Building consumer understanding by utilizing a Bayesian hierarchical structure within the Behavioral Perspective Model

Introduction

The Behavioral Perspective Model (BPM; Foxall, 1990/2004, 2010) has been used extensively to understand and predict consumer behavior e.g., Foxall, 2016a,b, 2017; Foxall & James, 2003; Foxall & Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). This study aims to build on this framework both theoretically and empirically. From a theoretical perspective, the study introduces a mixed effects hierarchical structure to the model which better resembles both the consumer purchase pattern and also the underlying structure of the data. This is the first instance of this structure being introduced to the BPM framework. The results are compared against a non-hierarchical framework and show the hierarchical nature better reflects the underlying consumer behavior theoretically and diagnostically.

A second theoretical advancement is the introduction of a Bayesian inference to estimate the parameters of the BPM. Hence, building on the demonstrated advantages of a hierarchical framework, two Bayesian hierarchical structures are evaluated and compered, relating to vague prior and informed prior models, with the informed priors calibrated from frequentist estimates. This shows the interpretation of the posterior distribution of the parameters can vary when different prior distributions are used and highlights the importance of prior information whilst utilizing a Bayesian approach. The text will argue the advantages of using both Bayesian and frequentist tools gives the researcher a larger analysis tool kit and agrees with Little (2006) that the 21st century should be about pragmatism in utilizing a broad range of methods for furthering consumer behavior. This demonstrates the flexibility of the BPM whilst constructing consumer behavior models and suggests further research to build a multi-hierarchical framework to encompass cross-category behavioral models within a similar hierarchical structure.

The Behavioral Perspective Model

Many studies have stemmed from the behavioral psychology and consumer behavior, with one of the earliest being that of the development of the Behavioral Perspective Model (BPM) (Wells, 2014). The Behavioral Perspective Model (BPM) has been used as a theoretical and methodological behavioral framework to explain consumer choice (e.g. Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). It "…provides an account of consumer choice founded on behavior analysis; that is, it explains purchase and consumption responses in terms of the contingent relationships among the behavior in question, its stimulus antecedents and its reinforcing and punishing consequences" (Foxall, 1999 p. 572).

The development of the Behavioral Perspective Model initiated a move to a more radical behavioral view (Foxall, 1987) and the resulting consumer behavior analysis program is now the most developed program of radical behaviorism principles to consumer behavior (Wells, 2014).

It is routed in the intercept of behavioral economics, economic psychology and marketing science and uses behavioral theory to interpret consumer behavior (Foxall, 2001).
The model (Figure 1), an extension of the Skinnerian three-term contingency states 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 similar kind (Foxall, 2005).

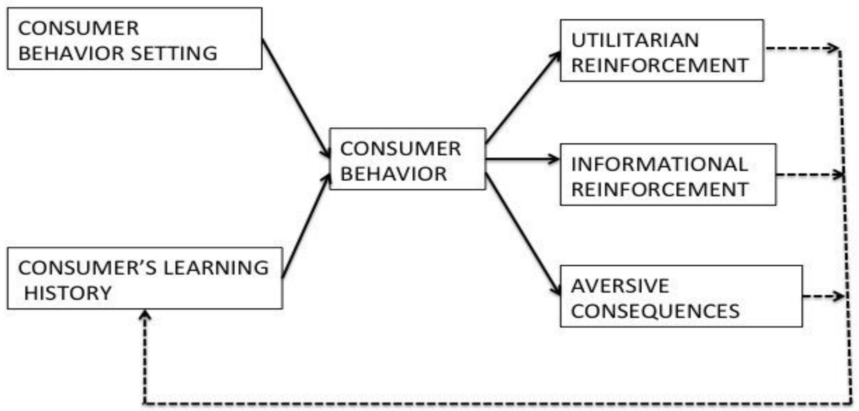

Figure 1: The Behavioral Perspective Model. *Source:* Foxall, G. R. (2010).

Utilitarian reinforcement is mediated by the product where its attributes and characteristics influence the rate of consumption of the product itself. Utilitarian reinforcements are usually functional where low utilitarian reinforcement usually constitute the basic product. Increased utilitarian reinforcements usually deliver a functional benefit above this base level. Informational reinforcement is mediated by more social aspects of the brands. Consumers may choose brands with similar utilitarian reinforcement but are deemed to have a higher social value. Foxall *et al.* (2004) show while some consumers maximize only the utilitarian or informational reinforcement, most consumers purchase a combination of both.

A further element of the BPM is an aversive consequence which may result from the behavior (e.g. monetary compensation) hence BPM studies often include elements of behavioral economics within its framework such as price elasticity (Foxall *et al.*, 2011; Oliveira-Castro *et al.* 2006).

The left hand side of the BPM contains the "Consumer Behavior Setting" and the "Learning History", which represent the antecedents of purchase behavior (see Foxall *et al.*, 2013).

One aspect explored within the BPM framework is the study by Oliveira-Castro *et al.* (2006) which looks at individual consumer elasticity rather than aggregated elasticity. Many consumer studies are built on aggregated data which can be an issue since the theory is built upon individual behavior (Kagel *et al.*, 1995). To that end Oliveira-Castro *et al.*, (2006) built individual models for 80 households using data from FMCG product categories comparing individual and household demand and found the general assumption of the similar trends across inter-consumer and intra-consumer could not be made.

While this is interesting, it is challenging for the market researcher to build many hundreds or even thousands of models to appertain to individual consumer levels. Also, this granularity can lead to coefficients with unreasonable sign and/or magnitude (Montgomery and Rossi, 1999). Also, many researchers are comfortable with calculating the consumer behavior to estimate sales of a product and hence the aggregated coefficients of models answer their needs (Oliveira-Castro *et al.*, 2006).

A middle ground may be the consideration of the structure of the data itself. Buyers form part of a household and therefore there is a hierarchical structure to the data, where purchases are made within household, as depicted in figure 2.

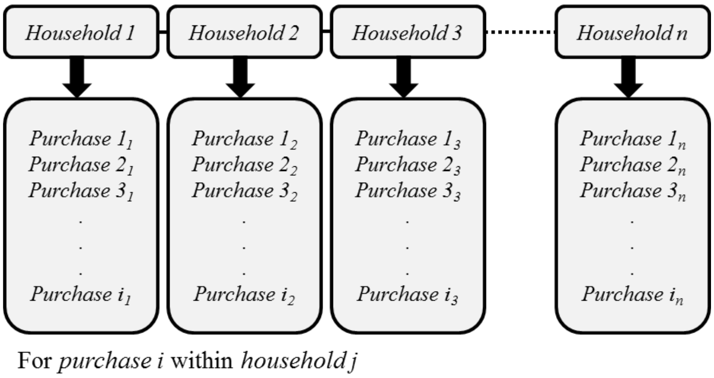

Figure 2: Hierarchical structure of the data.

Households make multiple purchases throughout the 52 week period from the same category and their purchases may form a certain habit. The implication is while the assumption of independence of purchases *across* household is realistic, the assumption of independence of purchases *within* household may not be. This is somewhat justified by conclusions from the Oliveira-Castro *et al.* (2006) study. Not accounting for the hierarchical structure may result in underestimated regression standard errors, leading to erroneously determination of a statistically significant causality between the independent and dependent variables (Browne and Rasbash, 2004).

This hierarchical structure has not been tested within the BPM framework and hence this would address a gap in the literature, i.e. whether a hierarchical model structure can provide increased model diagnostic benefits consumer understanding within the BPM framework.

Bayesian Inference

A statistical paradigm, which has become to be known as Bayesian statistics was first published post-humus in 1763 in a work by the Reverend John Bayes entitled “An essay towards solving a problem in the doctrine of chances”. The essay introduced the notion that the probability of an event could be an update of the current view, given the observation of new data (Koop et *al.*, 2007). Known as the Bayes theorem, it “is to the theory of probability what Pythagoras’ theorem is to Geometry” (Jeffreys, 1931, p. 7).

Bayesian inference differs from the Bayes’ theorem itself by omitting the denominator of the right hand side since this is just a normalizing constant (Jeffreys, 1931), hence can be constructed as such.

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

The left hand part of the equation is known as the *posterior* probability. The terms on the right hand are known as the *likelihood* and *prior* respectively. The *prior* is the initial belief of a parameter or event before any (new) data is considered (This is discussed in more depth next). The *likelihood* is the addition of new data to be evaluated. The *posterior* probability is the blend of both, resulting in an updated view of knowledge based on a combination of the current belief (*prior*) updated by the additional data (*likelihood*). (Koop et *al.*, 2007).

Prior Distribution

One fundamental difference between the frequentist and Bayesian paradigms is the explicit inclusion of this prior knowledge within the calculation of the posterior distribution. This poses an issue for the frequentist researcher, since this prior information would seem to act as a bias to the experiment. It means different results may be obtained from the same data if differing prior distributions are utilized (Little, 2006). In contrast, the Bayesian researcher views the prior distribution as an important element to the posterior calculation, since it mimics the learning process of the human mind by observing new data and comparing to what (s)he already knows (O' Hagan, 1998; Bernado, 1999; Bernado and Smith, 2000). Hence, differing results to the same study is an issue for the quality of the researchers' knowledge rather than the methods employed to inform the inference (Dunson, 2001). Researchers are not passive observers and experiments are designed to fit analytic models whether be it within a frequentist or Bayesian framework, hence the inclusion of the prior is an extension of this build (Efron, 2005). Furthermore, Leamer (1992) argues that, in practice, the frequentist researcher must have some prior incline as to the nature of parameters and would reject any absurd model outputs.

Rossi and Allenby (2003) say Bayesian methods offer an advantage, since assumptions are explicit and model assumptions in themselves are a form of prior information usually implicit under frequentist based models.

Defining the Prior Distribution

The prior distribution can be considered vague or informative. Vague prior distributions assign large variances to the estimate of the prior and hence these priors are weak in that they have little influence over the posterior estimates compared to the likelihood (Lunn *et al.*, 2000; Gelman, 2007), often resulting in estimates similar to that of maximum likelihood techniques (Dunson, 2001). Vague priors have been used for some time within Bayesian models (Lunn *et al.*, 2000) and some authors prefer them as the point estimates better the match those of a frequentist approach (Samaniego and Reneau, 1994).

Informative priors have a stronger influence on the posterior distribution. These informative priors can be obtained through deduction from previous studies, expert opinion or may be as simple as controlling for absurd results. (Hansen *et al.*, 2004; O’Hagan, 1994; Dunson, 2001). The Bayesian argues the inclusion and embracement of any prior information around an experiment to create realistic prior information is better than “relying on ignorance” (O’Hagan 1998 p.21; Aspinall, 2010).

The Bayesian considers the inverse of the square root of the variance as the precision, hence the stronger the precision, the more influence the prior commands on the posterior distribution. The term precision is standard within the Bayesian literature, replacing the variance of the distribution, though both are a mathematical derivation of the other (e.g. Lunn *et al.*, 2000).

Interpretation of the Posterior Distribution

(Dunson, 2001) claims the primary advantage of the Bayesian approach is the interpretation of the posterior distribution of the parameter estimates themselves. The frequentist views a parameter of a model as unknown but fixed and the inference obtained is the probability of observing the data given the estimated parameter value i.e. $P(X_i \mid \theta)$ (Abelard, 2012). Recall the Bayesian theorem which states

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

This means the Bayes' theorem calculates the probability of the parameter, given the data, i.e. $P(\theta | X_i)$ (Abelard, 2012). Hence the Bayesian interpretation of the posterior estimate as the direct probability of the event occurring is more intuitive, allowing management more transparent means of embracing the uncertainty of a parameter (O'Hagan, 1994; Dunson, 2001). Little (2006, p.218) agrees saying management would rather have "fixed probability intervals for unknown quantities" (the Bayesian posterior) than "random intervals for fixed quantities" (the frequentist outputs).

Pragmatism

Little (2006) says there are three groups of statisticians: frequentists, Bayesians and pragmatists. He claims pragmatists pick and choose from both the frequentist and Bayesian paradigms to suit their analysis needs. Efron (2005) claims models are imperfect in themselves and hence due diligence is required in checking them, hence they should not be constrained by a paradigm. Efron (2005, p.1) labels the "19th Century as generally Bayesian, the 20th Century as generally frequentist, and suggested that statistics in the 21st Century will require a combination of Bayesian and frequentist ideas". Little (2006) suggests this is currently the case with a dominance of the pragmatist. He says this is useful as the use of two paradigms increases the number of tools available for analysts to utilize.

The concept of pragmatism suggests both paradigms are used in order to help the analysis process and this can be construed in two ways. Little (2006, 2011) claims the Bayesian paradigm better lends itself to assert model inference; however there is a lack of Bayesian tools to assess the model diagnostically. Hence, it is a natural compromise for the model development and assessment to incorporate frequentist tests. A further blend of combining both paradigms comes with calibrated Bayes which combines the Bayesian and frequentist approach to model construction and evaluation (Efron, 2005; Little, 2006). The approach involves deriving estimates of the prior distribution of a Bayesian model by using frequentist methods. Box (1980) states a similar argument that sampling theory be used for the exploration and criticism of the model. Rubin (1984) in Little (2006 p.7) agrees stating "The applied statistician should be Bayesian in principle and calibrated to the real world in practice". Where Bayesian models benefit from a thorough model specification encompassing the likelihood and prior dimensions (Little, 2011), the rigorous procedures for model evaluation as seen within the frequentist environment is less apparent (Rubin, 1984).

Hansen *et al.* (2004) make good use of relevant frequentist diagnostics when evaluating the relevance of the Bayesian model and parameter estimates.

The Bayesian methodology has been favored in this short text, though the wider philosophical view of this study is very much in line with Efron (2005) and Little (2006) view of a pragmatic approach to building solutions to statistical problems. Gelman (2010, p. 162) also wisely notes

"…the key to a good statistical method is not its underlying philosophy or mathematical reasoning, but rather what information the method allows us to use. Good methods make use of more information".

Two approaches are suggested, utilizing both vague and informative priors. The nature of these will be discussed later. The BPM could benefit from exploiting the flexibility of a Bayesian approach, both in terms of the prior distributions and how the individual estimates are interpreted.

Data Discussion

The study now continues with a data description and category analysis of the GB biscuits FMCG category before the research questions are defined. The data relate to a panel sample of 1,594 households and 75,563 purchases from the biscuits (sweet and savory) category. The data account for the period of week ending 17 July 2004 to 9 July 2005 and are assembled at SKU level, whereby each descriptor contains a string relating to the brand and the number of items within the pack.

Some records appear to have an extremely low price per SKU (as low as 1pence per item). The lowest value biscuits range are classed as supermarket own label or value brands. There is a minimum price of circa 20p per pack. Hence a minimum price of 20p is used as a minimum acceptable price for a packet of biscuits. Any transactions at the SKU level which place a biscuit pack lower than 20p per item are excluded from the study. In the same manner, there are transactions with a very low price per 100g. 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 floor for all transactions. This leaves a base sample of 61,087 records to analyze (80.8% of the original biscuits category panel data).
As well as the SKU name, there is a product description field. Figure 3 shows the distribution of data within this.

There are 19 many categories of biscuit which can be confusing to understand and arguably not how the consumer may classify the products. Also, some categories have a low number of transactions which may lead to mathematical sample size issues when analyzing. In order to overcome these, categories are grouped together to form logical macro categories. Chang (2007, p. 107) suggests a 5 band classification yielding the following groups which is deemed a sensible approach and is undertaken in this study. The descriptions shown in figure 4.

Subcategory	Definition
Chocolate countlines	Individually wrapped chocolate-covered cookie 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 cookies.
Plain sweet cookies	Plain sweet cookies 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 cookies and ginger.
Chocolate coated cookies	Plain sweet cookies coated partially, topped or completely with chocolate
Filled cookies	Sweet cookies which can either be filled, topped or sandwiched between plain cookies
Non-sweet cookies	Plain savory cookies, savory crackers and bread-like savory cookies. Often flavored or topped with salt, cheese or other savory products.

Figure 3: Description of the Redefined structure of the biscuit data.

This grouping of the categories result in the distribution of items shown in figure 4. These grouping are now more identifiable to the consumer and have sample sizes which allow statistical analysis to be undertaken.

	Count	% Count	Volume	% Volume
Countlines	17293	28.3%	5089771	28.3%
Chocolate Coat	9645	15.8%	3414240	19.0%
Plain Sweet	14153	23.2%	4696549	26.1%
Filled	5291	8.7%	1538860	8.6%
Non Sweet	14705	24.1%	3231637	18.0%

Figure 4: Redefines biscuit data.

Within the biscuits category, the nature of the packets of biscuits relies on the type of biscuit they contain. (E.g. whether they relate to individual based biscuits for example “Kit Kat” or many smaller biscuits such as digestives). A manual process is conducted to allocate the items per pack to each SKU. Biscuit packs which contain many standard biscuits (such as digestive) relate to the number of packets within the SKU, in this case 1. Where biscuits are individually wrapped single serve portions rather, then the number of individual biscuits is deemed to be the number in pack. For example a single packet of six Kit-Kat biscuits will be coded as a pack size of “6”.

There are also other larger formats such as drums, bags, barrels etc. which do not contain the actual number of items but all imply larger packs. These are grouped together and called “pack”. Hence the biscuit pack sizes are grouped as per figure 5 based on the distribution of transactions within group and also logical combinations of the pack sizes.

	Count	% Count	Volume	% Volume
1	33743	55.2%	9364867	52.1%
2-5	3922	6.4%	1195547	6.6%
6-7	6880	11.3%	1665044	9.3%
8-11	7349	12.0%	2056553	11.4%
12+	6771	11.1%	2597200	14.4%
pack	2422	4.0%	1101835	6.1%

Figure 5: Redefined pack type biscuit data

This resulting pack distribution is both logical and also appropriate for statistical analysis. Most of the category is constructed of the single pack size though larger packs (i.e. the 12 and the “pack”) sizes have a higher volume per transaction as may be expected from larger formats.

Category Overview Analysis

In order to better understand the category dynamics, an analysis is undertaken which will focus on the three sets of variables, namely behavioral psychology variables relating to the BPM and the marketing variables relating to seasonality and also the effect of a supermarket own brand and the behavioral economic variable (price per 100g).

BPM Variables

SKU count vs. Informational Reinforcement

Figure 6 shows the average informational scores per defined SKU. The x-axis represents the mean informational reinforcement score for the SKU while the y-axis represents the number of SKUs associated with the informational score. SKUs with a larger number of incidences tend to have a higher informational score. The outlier relates to the SKUs which are classified as supermarket own brand. These dominate the category in terms of number of SKUs, hence warrant a separate investigation as to how they relate to a consumer's behavioral psychology of purchase.

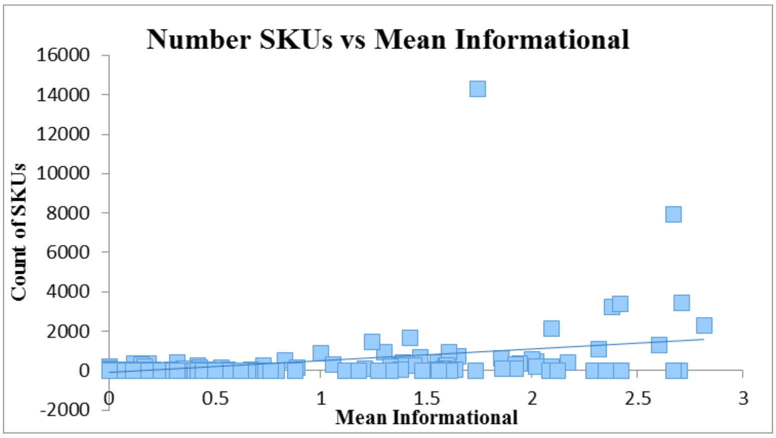

Figure 6: SKU count vs informational reinforcement score.

Size of Brand vs Informational Reinforcement

Figure 7 shows the relationship between SKU size (number of units sold) and informational score. There is a group of larger brands (indicated within the ellipse) which seems to score a higher informational score which may indicate that larger informational reinforcement brands tend to have a higher number of selling SKUs (or that larger SKUs attract a higher informational reinforcement score). Supermarket own brand score is the lowest amongst the larger brands with a mean score more reflective of the smaller SKUs. This indicates a possible difference in informational reinforcement when it comes to supermarket own brands

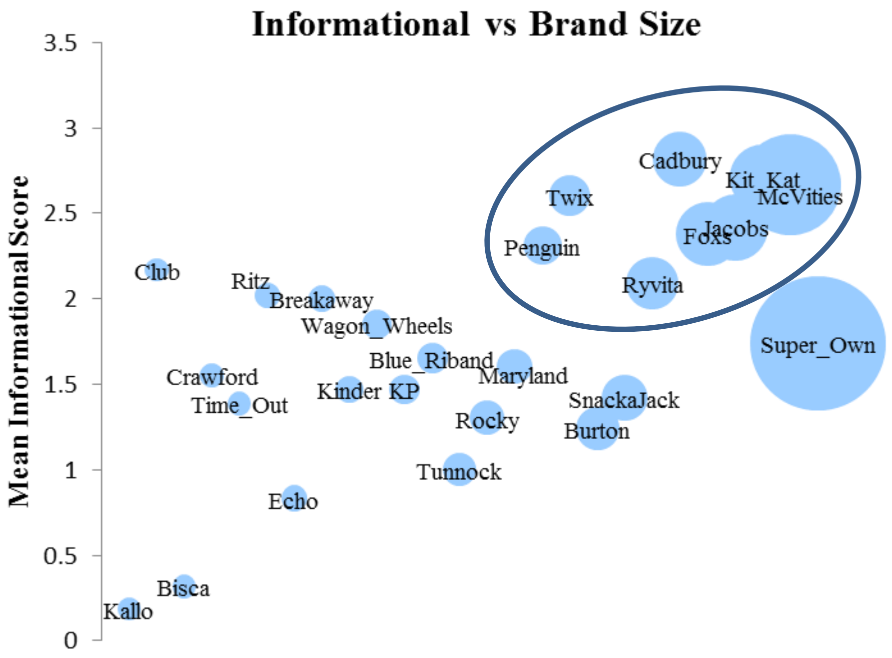

Figure 7: Informational reinforcement vs brand size.

Biscuit Type and Pack Size vs. Informational Reinforcement

Figure 8 shows the scatterplot of informational reinforcement score versus biscuit type and also informational score versus pack size. The number of units sold is reflective of the bubble size and ordered smallest to largest. There is no apparent relationship between the informational score and either the category size or the biscuit type or the pack size given the biscuit types and pack sizes are all of similar informational reinforcement levels. This implies these characteristics may be less relevant to the behavioral psychology purchases of this category.

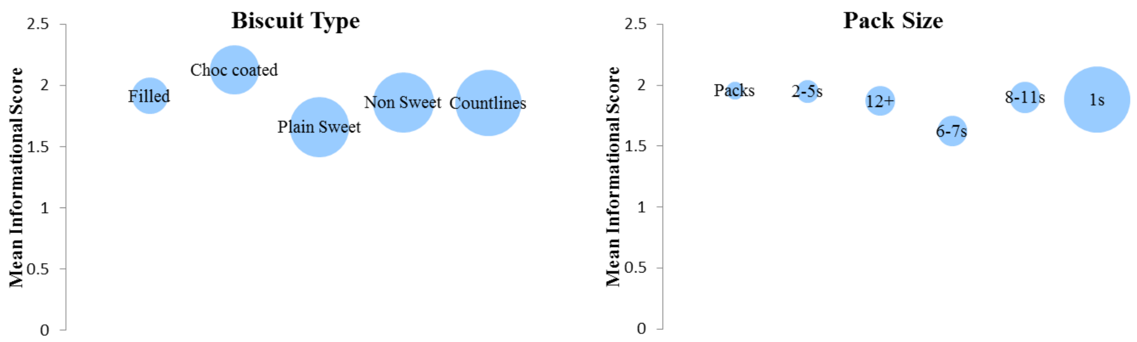

Figure 8: Informational reinforcement vs biscuit type and pack size.

Utilitarian Reinforcement

Utilitarian reinforcement variables are dichotomous, representing the lower and higher utilitarian reinforcement levels of the products.

Brand Distribution by Utilitarian Reinforcement

The Venn diagram on the left in figure 9 shows the distribution of the number of defined brands split by the two utilitarian reinforcement levels. For this category, most brands (59.1%) are located within the lower utilitarian reinforcement level. Some brands can be located within either the lower or higher reinforcement level, depending on the individual SKU within the brand. For example, Adams Malted Milk biscuits are coded as the lower utilitarian reinforcement (level 1) while Adams Malted Milk with Chocolate is coded as the higher utilitarian reinforcement (level 2). This overlap accounts for 21.8% of the defined brands in the category. Brands which are solely defined as the higher utilitarian reinforcement level 2 account for 19.0% of the defined category.

The Venn diagram on the right in figure 9 shows the volume accounted for by the brands discussed above. The intersection of the two utilitarian reinforcement levels accounts for the largest share of volume (66.4%), hence 21.8% of brands are accounting for 66.4% of the volume of the defined category. This may suggest larger brands are able to offer SKU variants which appeal to both utilitarian reinforcement levels. The smaller volume size brands are located entirely within the lower utilitarian reinforcement group, where 59.1% of these brands account for only 10.1% of volume.

Bottom right of figure 9 shows the split of the supermarket own brand across the two utilitarian reinforcement levels. There is a relatively even split between the lower and higher utilitarian reinforcement levels which suggests the category volume is driven by the type of biscuit rather than just the branded nature of the biscuit. This implies the own supermarket own label products are offering a diversity of utilitarian reinforcement attributes within this category and the study should investigate this.

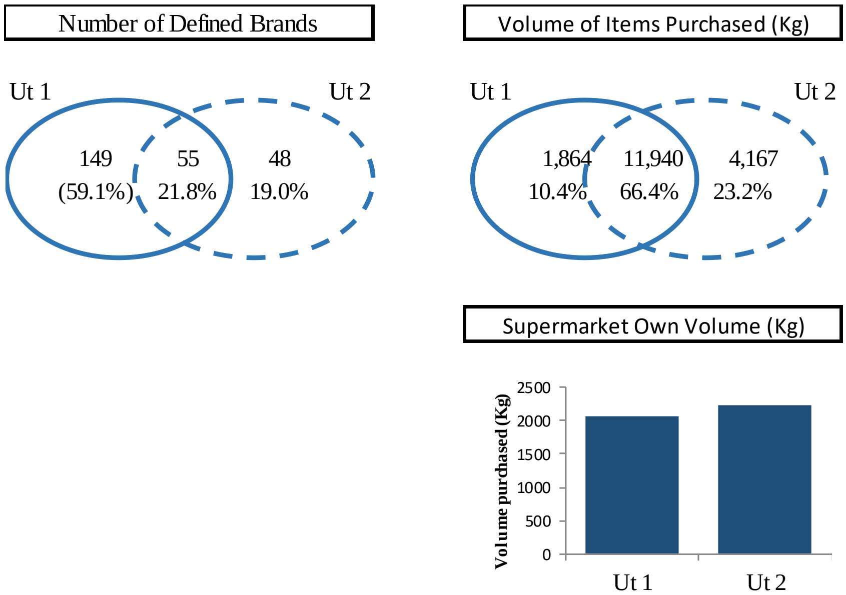

Figure 9: Utilitarian reinforcement distribution of the data.

Seasonal Analysis

Figure 10 shows the time series chart of volume of the biscuit category split by the utilitarian reinforcement variable. As volume increases during the build up to the Christmas period, it is the lower utilitarian group which increases market share. This seems to suggest consumers are purchasing more utilitarian products for the increased consumption period around the Christmas holidays. The category volume significantly falls during the Christmas week. This, at least in part is due to a shorter commercial week, though may also be due to a change in consumers' purchase behavior. The volume level returns to near average levels the following week.

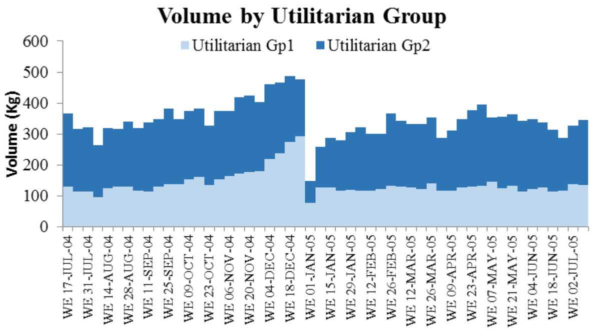

Figure 10: Time series by utilitarian reinforcement level.

Economic Behavioral Variables

Average Price

Past research on this dataset says there is a price elasticity of demand within the category (e.g. Foxall et *al*., 2013). This can be obtained using the following equation.

$$LN(Volume) = a + b\ LN(Price)$$

The coefficients obtained are compatible with economic theory and consistent over time (Oliveira-Castro *et al.*, 2006).

Recent studies have indicated the biscuit category data has a negative elasticity of demand and this value is close to –0.5 (Chang, 2007; Oliveira-Castro *et al.*, 2006). These coefficients being less than unitary value demonstrates the category is inelastic, which is consistent with food products in general (Driel *et al.*, 1997).

This analysis relates to the average price movement per 100g hence is comparable across pack format and physical size of the item. The natural logarithm is the change in price, which also is used as the elasticity of demand noted in the equation (e.g. Oliveira-Castro *et al.*, 2006).

The lack of promotional calendar information implies the price elasticity will be an average price elasticity, which will be the result of possible regular (long term) price changes, promotional (short term) price discounts and also changes in category mix.

Summary and Research Question Development.

The literature review and the category analysis has shown three areas of consumer behavior could be leading to varying demand of the biscuit category in terms of individual purchase. These are behavioral economics, the BPM psychological variables and the impact of the marketing based variables on the consumer psychological purchase behavior.

In terms of behavioral economics, the average price of the product seems to have a negative effect on demand in terms of an economic behavior and this is show across other studies (e.g. Oliviero-Castro *et al.*, 2006; Broadbent, 1980; Gabor, 1988; Nagle, 1987; Roberts, 1980; Telser, 1962; Chang, 2007; Foxall et *al*., 2013). Despite this featuring extensively in the literature, the study needs to confirm this is still the case when a more complex model structure is applied. Hence, this leads to the first research question.

RQ1: The average price of the products within the category influences consumer economic behavior.

The BPM has proven to be a good predictor of consumer behavior (e.g. Foxall and James, 2003; Foxall and Schrezenmaier, 2003; Foxall *et al.*, 2004; Foxall *et al.*, 2006; Oliveira-Castro *et al.*, 2005; Romero *et al.*, 2006). This study seeks to augment the predictability of the BPM by the inclusion of marketing variables, namely the branding of the supermarket own brand. This research aims to build on previous studies by exploring the nature of the psychological impact of products being formally branded as supermarket own brands. Does this impact the utilitarian and/or informational reinforcement within the BPM theoretical framework? Hence RQ2 is constructed as follows

RQ2: Is the nature of the supermarket own brand impacting consumer behavior through differing behavior at a consumer psychological level, either at a utilitarian and/or informational reinforcement level?

In a similar fashion to RQ2, the seasonal pattern of the Christmas week has a negative effect on total category volume. However, it is not clear whether this difference is down to individual consumer behavioral change during the period or whether it is due to a general drop in category purchase through less consumption across category and less shopping days during the period. Hence this research aims to test this by seeking to understand whether consumer psychology changes within the Christmas period with regards to the BPM variables. A further interaction term is built to address this, this time an interaction between the Christmas week and both the utilitarian and informational reinforcement elements of the BPM. Hence RQ3 is constructed.

RQ3: Is the seasonal Christmas week impacting consumer behavior within the BPM through different levels of utilitarian and/or informational reinforcement during the Christmas seasonal week?

The fourth research area focuses on how the structural development of the model itself within the BPM framework. This paper will bring two developments which will result in a further two research questions. Firstly, the study will model the BMP within a hierarchical framework and will compare the model performance and also the interpretation of the variables from a hierarchical and non-hierarchical framework.

RQ4: Will the BPM structure will benefit from a hierarchical model structure? What differences in interpretation would be included versus a non-hierarchical framework?

Finally, the BPM model will be run using Bayesian estimation using the MCMC algorithm. Two modes, built with vague prior distributions and also informative prior distributions will be compared in terms of predictive ability and also how variables are interpreted.

RQ5: How will Bayesian inference utilizing informative and vague priors impact the predictive nature of the model and the interpretation of the parameters?

Methods

Focal/Non Focal Parameters

This study focuses on the behavioral parameters as discussed covering the economic, psychology and marketing parameters. The BPM variables account for the differences in informational and utilitarian reinforcement of each if the biscuit SKUs and this is taken account in terms of how the utilitarian codes are allocated.

The characteristics of the product are included in the model. These will account for any intercept level differences between biscuit types and pack types and whether these attract a consistent difference in average sales of the category. This isolates the focal variables to account for any causality between behavior change and sales.

Therefore focal variables will relate to the price, BPM variables and their interception with supermarket own brand and interaction with the Christmas week.

Non-focal variables will refer to the biscuit type and the pack type.

Fixed and Random Effects

A hierarchical model may consist of fixed effects, random effects or both (known as a mixed effects model). A fixed effect is generally associated with the assumption that the range if possible attributes of a study relate to the rage of possible attributes within the wider population. A random effect, however, is used if inferences are to be generalized from a recognized sample to a wider population (Field, 2012). Within this study, the sample of 1,594 households is a representative sample of the GB population hence a random effect is assigned to these. The simplest form of a random term is that of a random intercept where the intercepts vary across the contextual group, in this case, the household (Field, 2012).

Consider the focal and non-focal variables of this study. While the results of the focal variables may result in working hypotheses for other product categories, the specific results of these focal variables are relevant to the biscuit category and are not intended to represent a generalization to other product categories available within the GB market place. Furthermore, the non-focal variables are specific to the categories they represent and cannot be generalized to all GB FMCG categories. Hence the focal and non-focal variables will be represented by a fixed effects parameters.

A regression model is constructed to simultaneously test the model structure incorporating the variables discussed. Three models are proposed which will allow the aspects of the research questions to be addressed. Each model will test the economic, behavioral psychology and marketing variables as addressed in RQ1, RQ2 and RQ3. The behavioral psychology BPM variables play an active role in RQ2 and RQ3. For these variables, the model is based on an offset approach, whereby the lower utilitarian reinforcement is regarded as the base coefficient and the higher utilitarian reinforcement variable is built as an offset to this. This allows the statistical evaluation of whether the two utilitarian reinforcement levels are statistically different or not. The three models, however will differ in the following ways.

Model 1 will be of a non-hierarchical structure with vague priors. Model 2 will be a hierarchical model with vague priors. This means that model 1 and 2 are identical apart from the hierarchical structure of model 2 allowing a comparison as to how the hierarchical structure impacts the model diagnostics and coefficient inference and interpretation. This hierarchical structure is based on purchases within household which can be identified in the data through the *panel id* variable, since it is unique to a household.

Model 3 will also be of a hierarchical nature, however the prior distributions will be constructed as informative rather than vague. This will allow the comparison of how a differing prior distribution may affect model and coefficient interpretation and evaluation.

Defining the Models' prior distribution

Defining the Prior Distributions

The nature of the Bayesian model requires the definition of a prior distribution. As discussed, the prior distribution is independent of the data and subject to the researcher's disposition.

Defining Vague Priors

The use of a vague prior has been used extensively to represent knowledge around a parameter (Lunn *et al.*, 2000). The study will utilize this vague prior information around each parameter in the form of the normal distribution (Lunn *et al.* 2000). The distribution will require a mean and a precision. The precision is the inverse of the standard deviation of the distribution, hence a small precision means a large variance for the distribution. The same prior distribution will apply to each parameter of the model and will have a mean of zero and a small precision at 0.001. This will imply the likelihood will have a strong influence, compared to that of the prior, in terms of the inference of the posterior estimates of the parameters. These will be used for Model 1 (non-hierarchical) and Model 2 (hierarchical vague).

Defining Informative Priors

In order to compare the impact of the informative prior, a second hierarchical model is constructed with informed prior distribution for the focal parameters (Model 3 – hierarchical informative). This means the prior information of the model will have a degree of influence on the posterior estimate of the parameters. As discussed earlier, prior estimates may be taken from previous research or derived from the notion of a calibrated prior, whereby information taken from a past frequentist analysis is used to produce the prior distribution parameters. Given the structure of this model has not been considered previously, no prior knowledge exists about the parameter estimates, hence the method of calibrated priors is used to inform the estimates of the prior distribution for the focal parameters.

These prior distribution estimates are calibrated by running a linear model for each parameter in turn. The mean of the prior is set equal to the mean of the frequentist linear model. Similarly the precision of the prior distribution is calculated from the inverse of the standard error of the frequentist estimate. This is a similar approach by that seen in Rossi and Allenby (1993).

One issue, relevant to this data set, is that for large data sets, the influence of the likelihood relative to the prior becomes very strong Dunson (2001). However since the calibrated prior is estimated from the same large data set, the standard errors of the estimate are relatively small (due to large *n*), implying a relatively large precision which, arguably, goes to balance the influence of the large *n*. Figure 11 shows the point estimate, standard error and the calculated precision for the parameters in question.

Price	Beta	-0.72
	Std Error	0.00
	Precission	67061

Informational	Beta	0.14	**Supermarket Own x Informational Higher Ut**	Beta	0.09
	Std Error	0.00		Std Error	0.01
	Precission	91750		Precission	16016
Informational Higher Utilitarian	Beta	0.11	**Christmas**	Beta	0.05
	Std Error	0.00		Std Error	0.03
	Precission	192052		Precission	1175
Supermarket Own x Informational	Beta	0.04	**Chrsitmas Higher Utilitarian**	Beta	0.14
	Std Error	0.00		Std Error	0.04
	Precission	88794		Precission	511

Figure 11: Informative prior distributions of Model 3.

The above table is translated into the prior distributions for the independent focal variables of the model, shown below.

Price ~ N(-0.717, 67061)
Informational ~ N(0.137, 91750)
Informational Utilitarian Gp2 ~ N(0.106, 192052)
Supermarket Own x Informational ~ N(0.041, 88794)
Supermarket Own x Informational Ut 2 ~ N(0.088, 16016)
Christmas ~ N(0.05, 1175)
Chrsitmas Utilitarian Gp2 ~ N(0.141, 511)

Prior Distribution of the Model Variance Term

In both the non-hierarchical and hierarchical structure, the variance coefficient requires a prior distribution. The variance is non-negative therefore a Gamma distribution is commonly used as a prior distribution for both the variance term for the model τ (i.e. the variance across household) and also the hierarchical variance term τ_v (i.e. the variance between household) (Spiegelhalter *et al.*, 2002).

Defining the Models

Model 1 is therefore a non-hierarchical model with vague priors. This can be defined as indicated below in figure 12.

$$
\begin{aligned}
LN(Volume_j) = \beta_0 & \\
& + \beta_1 LN(\Pr ice_j) \\
& + \beta_2 * Informational_j * Utilitarian_j \\
& + \beta_3 Informational_j * Utilitarian_{jGroup2} \\
& + \beta_4 Supermarket * Informational_j \\
& + \beta_5 Supermarket * Informational_j * Utilitarian_{jGroup2} \\
& + \beta_6 Christmas * Informational_j \\
& + \beta_7 Christmas * Informational_j * Utilitarian_{jGroup2} \\
& + \sum_{i=1}^{a-1} \beta_{i+6} Flavour\_type_i \\
& + \sum_{i=1}^{b-1} \beta_{i+6+(a-1)} Pack\_type_i \\
& + \varepsilon_j
\end{aligned}
$$

$$
\begin{aligned}
&\beta_1 \sim N(0, 0.001) \\
&\beta_2 \sim N(0, 0.001) \\
&\beta_3 \sim N(0, 0.001) \\
&\beta_4 \sim N(0, 0.001) \\
&\beta_5 \sim N(0, 0.001) \\
&\beta_6 \sim N(0, 0.001) \\
&\beta_7 \sim N(0, 0.001) \\
&\beta_i \sim N(0, 0.001) \quad \text{for i} = 8,\ldots,8+(\text{a}-1)+(\text{b}-1)
\end{aligned}
$$

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

Figure 12: Structure of Model 1 (non-hierarchical).

Model 2 is also non –hierarchical though will utilize the informative prior distributions as indicated earlier. Therefore the model structure is as indicated below in figure 13.

$$
\begin{aligned}
LN(Volume_j) = \beta_0 \\
&+ \beta_1 LN(\mathrm{Pr}ice_j) \\
&+ \beta_2 * Informational_j * Utilitarian_j \\
&+ \beta_3 Informational_j * Utilitarian_{jGroup\,2} \\
&+ \beta_4 Supermarket * Informational_j \\
&+ \beta_5 Supermarket * Informational_j * Utilitarian_{jGroup\,2} \\
&+ \beta_6 Christmas * Informational_j \\
&+ \beta_7 Christmas * Informational_j * Utilitarian_{jGroup\,2} \\
&+ \sum_{i=1}^{a-1} \beta_{i+6} Flavour\_type_i \\
&+ \sum_{i=1}^{b-1} \beta_{i+6+(a-1)} Pack\_type_i \\
&+ \varepsilon_j
\end{aligned}
$$

$$\beta_1 \sim N(-0.717, 67061)$$
$$\beta_2 \sim N(0.137, 91750)$$
$$\beta_3 \sim N(0.106, 192052)$$
$$\beta_4 \sim N(0.041, 88794)$$
$$\beta_5 \sim N(0.088, 16016)$$
$$\beta_6 \sim N(0.05, 1175)$$
$$\beta_7 \sim N(0.141, 511)$$
$$\beta_i \sim N(0, 0.001) \quad \text{for i} = 8,\ldots,8+(\text{a}-1)+(\text{b}-1)$$

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

Figure 13: Structure of Model 2 (hierarchical vague).

Model three differs from model 2 as it employs a mixed effects model with random effect intercepts and fixed effects model parameters. The functional form, in figure 14, is as below.

$$
\begin{aligned}
LN(Volume_j) = \beta_0 & \\
& + \beta_1 LN(\Pr ice_j) \\
& + \beta_2 * Informational_j * Utilitarian_j \\
& + \beta_3 Informational_j * Utilitarian_{jGroup\,2} \\
& + \beta_4 Supermarket * Informational_j \\
& + \beta_5 Supermarket * Informational_j * Utilitarian_{jGroup\,2} \\
& + \beta_6 Christmas * Informational_j \\
& + \beta_7 Christmas * Informational_j * Utilitarian_{jGroup\,2} \\
& + \sum_{i=1}^{a-1} \beta_{i+6} Flavour\_type_i \\
& + \sum_{i=1}^{b-1} \beta_{i+6+(a-1)} Pack\_type_i \\
& + \varepsilon_j
\end{aligned}
$$

$$
\begin{aligned}
&\beta_0 = U_0 + \upsilon_0 \\
&\upsilon_0 \sim N(0, \sigma^2) \\
&where\ \varepsilon_j \sim N(0, \sigma^2) \qquad j = 1,2,\ldots,n
\end{aligned}
$$

$$
\begin{aligned}
&\beta_1 \sim N(0, 0.001) \\
&\beta_2 \sim N(0, 0.001) \\
&\beta_3 \sim N(0, 0.001) \\
&\beta_4 \sim N(0, 0.001) \\
&\beta_5 \sim N(0, 0.001) \\
&\beta_6 \sim N(0, 0.001) \\
&\beta_7 \sim N(0, 0.001) \\
&\beta_i \sim N(0, 0.001) \quad \text{for i} = 8,\ldots,8 + (\text{a - 1}) + (\text{b - 1}) \\
&\upsilon_0[k] \sim N(0, 0.001) \quad \text{for } k = 1,\ 2,\ldots,h \quad \text{where } h = \#\,\text{households}
\end{aligned}
$$

Figure 14: Structure of Model 3 (hierarchical informative).

Running the Bayesian Model

Modeling the Data

The parameter inference is calculated using the Markov Chain Monte Carlo (MCMC) Metropolis-Hastings method with the Gibbs sampler since it has been shown to converge at a geometric rate (Tierney, 1994). Consider the posterior distribution with θ_k elements $(\theta_1,\ldots,\theta_k)$. The Gibbs sampler works by drawing from conditional distributions of the posterior by cycling through each parameter, one at a time whilst maintaining the other parameters constant in the following fashion, shown in figure 15.

$$
\begin{aligned}
& p(\theta_{r,1} \mid \theta_{r-1,2}, \theta_{r-1,3}, \ldots, \theta_{r-1,k}, y) \\
& p(\theta_{r,2} \mid \theta_{r,1}, \theta_{r-1,3}, \ldots, \theta_{r-1,k}, y) \\
& \ldots \\
& p(\theta_{r,k} \mid \theta_{r,1}, \theta_{r,2}, \ldots, \theta_{r,k-1}, y)
\end{aligned}
$$

Figure 15: Gibbs sampler.

This continues until the joint posterior distribution converges. Inference can then be derived for each of the parameters $(\theta_1,\ldots,\theta_k)$ by calculating the estimate for the parameter from the iterations of the converged MCMC values.

The modeling process is conducted through the Rjags package, within the R software system. The Rjags package calls on the JAGS (Just another Gibbs sampler) software package (see Plummer (2013) for details). Furthermore, the CODA package within R is used to calculate this Bayesian inference of the parameters from the output of the Rjags package (see Plummer et *al.*, 2006).

Convergence Criteria

There is no mechanism whereby the Gibbs sampler “knows” it has converged and instead the *chain* of evolving MCMC estimates of each parameter value is plotted where a converged chain resembles a “hairy caterpillar” which is a random noise around a stationary value of the estimate, allowing a visual means of parameter convergence. Gelman and Rubin (1992) also offer a convergence diagnostic where a value of less than 1.1 is sufficient to indicate parameter convergence.

Two independent MCMC chains are run to estimate the coefficients and will converge to the same estimate of the parameter given sufficient number of draws. Visually comparing the convergence of more than one chain also offers further reassurance the estimate has converged (Gelman and Rubin, 1992; Rossi and Allenby, 2003).

A starting value for the MCMC algorithm is required and this is taken from a random draw of the posterior of the distribution, which is the default option within the Rjags package (Plummer, 2013).

Estimate of the parameter

Each parameter is estimated by taking an average of the draws within and then across both *converged* chains. To ensure convergence is met before parameter estimation, a “burn in” sample of draws is ignored and hence the inference is estimated only from the post burn-in converged draws of the chains. The burn in is set at 4,000 iterations with a further 2,000 iterations used as the basis of the parameter estimate. There is no rule as to the number of burn in draws and hence it is important to monitor the convergence criteria for all parameter estimates before estimation.

Interpreting the Parameter Inference Statistics

For this study, a combination of Bayesian and frequentist measures are calculated to understand the inference of the parameters given the discussion within the literature review and the views of Efron (2005). Figure 16 gives an illustration of the structure of the parameter inference and how these statistics can be interpreted.

	(1)	(2)	(3)	(4)	(5)	(6)
	Beta (SE posterior)	Bayes CI		t	sig	
Constant	4.489 (0.0096)	4.47, 4.507	^	467.6	0.000	**
Log Price	-0.701 (0.004)	-0.709, -0.693	^	-175.3	0.000	*
etc.	…					

Figure 16: Example of the model inference output.

Each of the metrics of figure 16are interpreted as follows

(1) Point estimate of the parameter (and it's standard error) calculated from the posterior distribution of the MCMC
(2) The 95% Bayesian posterior confidence interval of the parameter
(3) The symbol ^ denotes the interval does not straddle zero (and hence means the parameter has at least a 95% probability it is contributing to the model fit).
(4) The frequentist t-statistic denotes the ratio of the parameter estimate and its standard error
(5) The frequentist statistical two-tailed significance level associated with the computed t-statistic.
(6) Indication of the statistical significance with * denoting significance at 10% level and ** denoting significance at the 5% level (two tailed). Lack of stars indicate the level of statistical significance is >10%

Assessing the Model Criteria

The Deviance Information Criteria (DIC) is a generalization of AIC, useful when the posterior distribution has been generated from an MCMC estimation. It is the combination of the "*goodness of fit*" adjusted for model "*complexity*" (Spiegelhalter *et al.* 2002). The "*goodness of fit*" element is defined as

$$D(\theta) = -2 \log L(data \mid \theta)$$

The "*complexity*" is defined as the posterior mean deviance plus the deviance of each of the means of each parameter, forming a penalty imposed for a more complex model, i.e.

$$\begin{aligned} p_D &= E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) \\ &= \overline{\mathrm{D}} - D(\overline{\theta}) \end{aligned}$$

The DIC is then constructed in similar means to the AIC as in

$$\begin{aligned} DIC &= D(\overline{\theta}) + 2p_D \\ &= \overline{\mathrm{D}} + p_D \end{aligned}$$

The smaller the DIC the better the models supports the underlying data.

MAPE

The Mean Average Percentage Error is a statistic diagnostic statistic which expresses the average percent difference between the actual and modeled values of a series. The statistic is calculated as thus, where *A* indicates actual values and *M* indicates modeled values.

$$MAPE(\%) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - M_i}{A_i} \right| \times 100$$

Variance Partition Coefficient

Given the hierarchical nature of the model, the variance will be partitioned into two parts, namely the variance between household and the variance between purchaser (Browne and Rasbash, 2004). Let the variance between household be defined as σ_u^2 and the variance between purchasers defined as σ_e^2 then the variance parturition coefficient, which can be expressed as a percentage is defined as

$$VPC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Interpreting the Models

The study now proceeds with a detailed discussion of the results of the three models.

Model Diagnostics

From a Bayesian perspective the Deviance Information Criteria (DIC) can be seen from figure18. The penalty for the hierarchical models (vague and informative) is higher than the non-hierarchical model (1,323 for the hierarchical vague, 1,318 hierarchical informative and 18 for the non-hierarchical). The mean deviance for each respectively is 69,379, 69,988 and 81,152 resulting in DIC calculations of 70,702, 71,306 and 81,170 respectively. Therefore the increased penalty incurred by the hierarchical model structure is outweighed by the increase in the predictive nature of the model. This suggests the hierarchical models would better predict a replicated data set of the same structure (Spiegelhalter *et al.* 2002). The difference between the hierarchical models (>5) suggests the vague model is better representing the data than the informative model (Spiegelhalter *et al.* 2002).

From the frequentist perspective, figure 18 shows the R-squared (adj) figures of 55.863% (hierarchical vague), 55.398% (hierarchical informative) and 45.291% (non-hierarchical) suggesting the hierarchical models are explaining a higher proportion of the variance, having accounted for the additional complexity of the models. The Mean Average Percentage Error (MAPE) values in respective order are 6.55%, 5.98% and 5.93% showing similar average absolute deviance for the models, though the hierarchical vague model has a larger MAPE. The total model variance for the hierarchical models is lower than that of the non-hierarchical models (0.182, 0.184 and 0.221 respectively) suggesting the hierarchical structure is representing more of the variability of the data within the model structure. The coefficients of the hierarchical variance term have high t-values when considering their ratio with their standard errors offering sufficient evidence to reject the null hypotheses these values are equal to 0. Additionally, the hierarchical variance partition coefficients are

$\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} = 17.582\%$ for the hierarchical vague model and 17.413% for the hierarchical informative model.

Despite all three models seeming adequate representations of the underlying data, these statistics suggest the functional form of the hierarchical models is benefitting the model fit above and beyond that of the non-hierarchical form.

Model Coefficients (Biscuits)

The convergence of the parameters needs to be assessed. Figure 28 in appendix A shows the convergence "hairy caterpillar" type charts and their nature indicates convergence has been achieved. Furthermore the Gelman statistics in figure 17 also indicate convergence of the parameters, with confidence intervals <1.1.

	Hierarchical Vague		Hierarchical Informative		Non Hierarchical	
	Point Estimate	Upper CI	Point Estimate	Upper CI	Point Estimate	Upper CI
Constant	1	1.01	1	1	1	1
Log Price	1	1	1	1	1	1
Informational x Utilitarian Gp1	1	1	1	1	1	1
Informational x Utilitarian Gp2	1	1.01	1	1	1	1
SuperOwn x Informational	1	1	1	1	1	1
SuperOwn x Informational GP2	1	1	1	1	1	1
Chrsitmas	1	1	1	1	1	1
Chrsitmas Ut Gp2	1	1	1	1	1	1
Chocolate Coated	1	1	1	1	1	1
Plain Sweet	1	1.01	1	1	1	1
Filled	1	1.01	1	1	1	1
Non Sweet	1	1.01	1	1	1	1
Size 2-5	1	1	1	1	1	1
Size 6-7	1	1	1	1	1	1
Size 8-11	1	1	1	1	1	1
Size 12+	1	1	1	1	1	1

Figure 17: Gelman convergence measures

	Non Hierarchical				Hierarchical Vague				Hierarchical Informative			
	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig	Beta (SE posterior)	Bayes CI	t	sig
Constant	4.489 (0.0096)	4.47, 4.507 ^	467.6	0.000 **	4.541 (0.0105)	4.52, 4.561 ^	432.5	0.000 **	4.390 (0.0094)	4.372, 4.409 ^	467.0	0.000 **
Log Price	-0.701 (0.004)	-0.709, -0.693 ^	-175.3	0.000 **	-0.702 (0.004)	-0.71, -0.695 ^	-175.6	0.000 **	-0.705 (0.0027)	-0.71, -0.7 ^	-261.1	0.000 **
Informational x Utilitarian Gp1	0.027 (0.0035)	0.02, 0.034 ^	7.7	0.000 **	0.033 (0.0034)	0.026, 0.039 ^	9.6	0.000 **	0.055 (0.002)	0.051, 0.059 ^	27.4	0.000 **
Informational x Utilitarian Gp2	0.074 (0.0042)	0.065, 0.082 ^	17.5	0.000 **	0.054 (0.004)	0.046, 0.062 ^	13.5	0.000 **	0.102 (0.0007)	0.101, 0.104 ^	146.1	0.000 **
SuperOwn x Informational	0.008 (0.0036)	0.001, 0.015 ^	2.3	0.030 *	-0.001 (0.0035)	-0.008, 0.005	-0.4	0.368	0.010 (0.0023)	0.005, 0.014 ^	4.1	0.000 **
SuperOwn x Informational GP2	-0.093 (0.005)	-0.102, -0.083 ^	-18.5	0.000 **	-0.081 (0.0049)	-0.09, -0.071 ^	-16.6	0.000 **	-0.061 (0.0037)	-0.068, -0.054 ^	-16.5	0.000 **
Chrsitmas	0.058 (0.0292)	0.001, 0.117 ^	2.0	0.055	0.043 (0.0266)	-0.009, 0.094	1.6	0.111	0.027 (0.0186)	-0.009, 0.064	1.5	0.136
Chrsitmas Ut Gp2	0.008 (0.0439)	-0.08, 0.091	0.2	0.393	-0.015 (0.0405)	-0.094, 0.067	-0.4	0.374	0.039 (0.0276)	-0.014, 0.092	1.4	0.145
Chocolate Coated	0.152 (0.0069)	0.139, 0.166 ^	22.1	0.000 **	0.143 (0.0066)	0.13, 0.156 ^	21.6	0.000 **	0.123 (0.0067)	0.11, 0.136 ^	18.4	0.000 **
Plain Sweet	0.160 (0.0093)	0.142, 0.178 ^	17.2	0.000 **	0.123 (0.009)	0.105, 0.14 ^	13.6	0.000 **	0.212 (0.007)	0.198, 0.225 ^	30.3	0.000 **
Filled	-0.011 (0.0085)	-0.028, 0.005	-1.3	0.162	-0.027 (0.0084)	-0.043, -0.01 ^	-3.3	0.002 **	-0.030 (0.0084)	-0.046, -0.013 ^	-3.6	0.001 **
Non Sweet	0.039 (0.0104)	0.019, 0.059 ^	3.7	0.000 **	-0.017 (0.01)	-0.036, 0.003	-1.7	0.099	0.086 (0.0071)	0.072, 0.099 ^	12.1	0.000 **
Countlines	base				base				base			**
Size 2-5	0.207 (0.0083)	0.19, 0.223 ^	24.9	0.000 **	0.200 (0.0079)	0.184, 0.215 ^	25.3	0.000 **	0.204 (0.008)	0.188, 0.22 ^	25.6	0.000 **
Size 6-7	0.086 (0.0072)	0.072, 0.1 ^	12.0	0.000 **	0.101 (0.0067)	0.089, 0.115 ^	15.1	0.000 **	0.124 (0.0069)	0.11, 0.137 ^	17.9	0.000 **
Size 8-11	0.195 (0.0078)	0.179, 0.21 ^	24.9	0.000 **	0.190 (0.0077)	0.175, 0.205 ^	24.7	0.000 **	0.199 (0.0076)	0.184, 0.214 ^	26.2	0.000 **
Size 12+	0.360 (0.0071)	0.347, 0.374 ^	50.7	0.000 **	0.332 (0.0068)	0.318, 0.345 ^	48.8	0.000 **	0.333 (0.0068)	0.32, 0.347 ^	49.0	0.000 **
Size packs	0.590 (0.01)	0.571, 0.61 ^	59.0	0.000 **	0.583 (0.0093)	0.564, 0.6 ^	62.6	0.000 **	0.585 (0.0094)	0.566, 0.603 ^	62.2	0.000 **
Size 1s	base				base				base			
R-Squared (adj)	45.291%				55.863%				55.398%			
Mean Deviance	81,152				69,379				69,988			
Penalty	18.2				1323.0				1318.0			
DIC	81170				70702				71306			
MAPE	5.93%				6.55%				5.98%			
Variance (between purchases)	0.221				0.182				0.184			
Variance (between housholds)					0.039				0.039			
between household t-stat (sig)					23.135(0)				23.458(0)			
Variance Partition Coeficient					17.582%				17.413%			

* significant 5%
** significant 1%
^ 95% Bayesian estimates do not include zero

Figure 18: Model diagnostics and parameter inference of the three models.

The posterior estimates of the parameters are shown in figure 18 and the focal variables are visualized graphically in figure 19 below, demonstrating the differences between the hierarchical and non-hierarchical estimates.

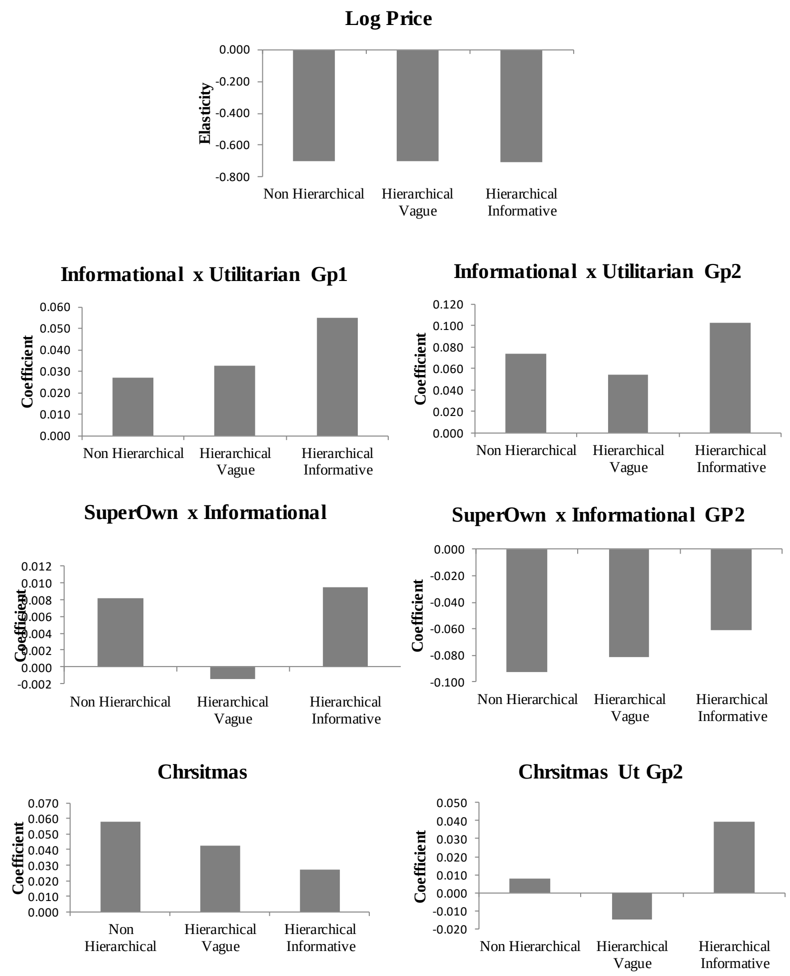

Figure 19: Graphical parameter estimates of the three models.

Coefficient Discussion

A discussion of the coefficient of the three models is now offered.

Price Elasticity

Figure 20 shows the density plots and box plots of the hierarchical and non-hierarchical models. There is little difference between the elasticity measures of the models. As discussed, the Bayesian nature of the parameter estimate implies the posterior distribution is the probability distribution of the parameter itself and the density plots can be used to understand the shape of the posterior estimates. The point estimates for hierarchical vague, hierarchical informative and non-hierarchical models are -0.702, -0.705 and -0.701 respectively. The 95% Bayesian confidence interval (i.e. between the 2.5% and the 97.5% points on the posterior density plot) for the hierarchical vague model is (-0.71, -0.695) for the hierarchical informative (-0.710, -0.700) and for the non-hierarchical (-0.709, -0.693), none of which include the value zero, hence it can be stated with 95% probability, this parameter is non-zero and hence contributing to the model.

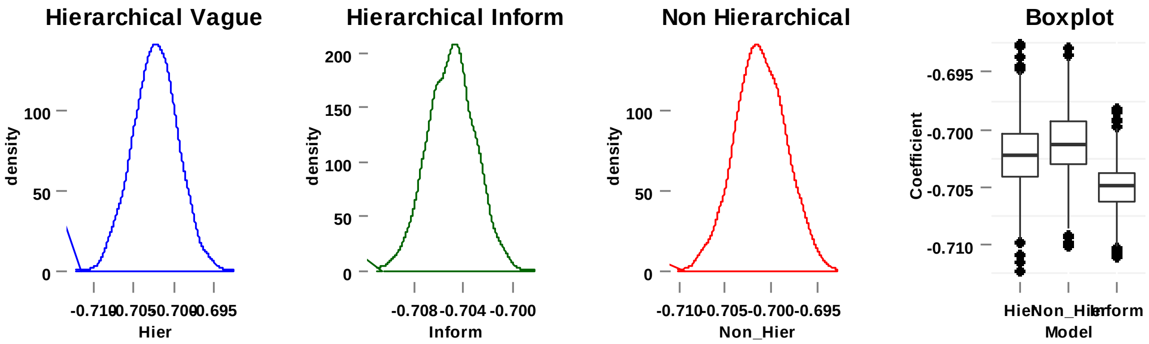

Figure 20: Density and Box plots of price elasticity.

From a frequentist perspective, the t-statistics of -175.6 for the hierarchical vague and -261.1 for the hierarchical informative and -175.3 for the non-hierarchical which are all statistically significant, all at $p<0.001$, which leads us to reject the null hypothesis the parameter is equal to zero, offering further evidence the parameter is significantly contributing to the model. This estimate is aligned with other studies (e.g. Oliveira-Castro *et al.*, 2006; Chang, 2007)[1].

Informational and Utilitarian 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.

Figure 21 shows the posterior distribution density plots and boxplot of the informational variable for the hierarchical and non-hierarchical models.

[1] For non-hierarchical models

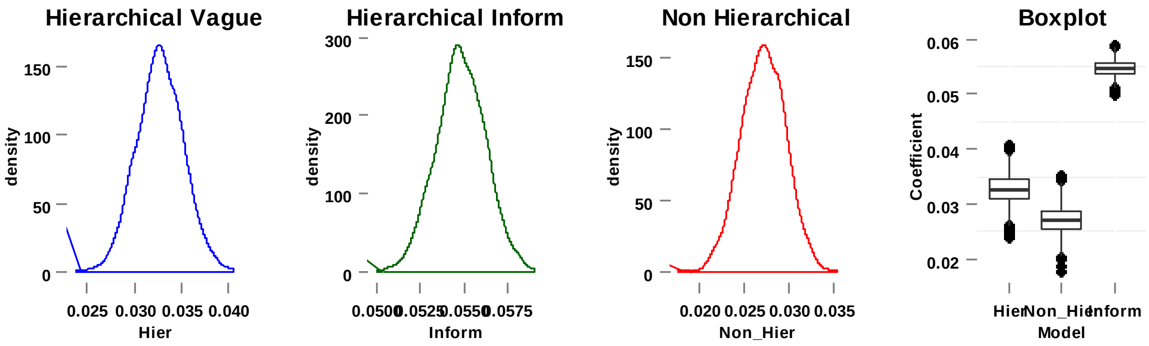

Figure 21: Density and Box plots of Informational reinforcement in the lower utilitarian reinforcement group.

The point estimates for the lower utilitarian groups are 0.033, 0.055 and 0.027 for the hierarchical vague, hierarchical informative and non-hierarchical models respectively. In each case, there is very little evidence to suggest this parameter is zero given the Bayesian confidence intervals of (0.026, 0.039) for the hierarchical vague model, (0.051, 0.059) for the hierarchical informative model and (0.020, 0.034) for the non-hierarchical model. None of the models' posterior confidence interval contains the value zero suggesting the parameters are significant in each case. There is some overlap in the posterior confidence intervals of the non-hierarchical model and the hierarchical vague model. This is due to agreement between the prior distribution of the hierarchical vague model and the likelihood based on the data. Also the frequentist t-statistics are 9.6, 27.4 and 7.7 respectively, all yielding $p<0.001$, hence strong evidence to suggest the parameter is non-zero in each case. Therefore 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 price.

Informational Reinforcement in the Higher Utilitarian Reinforcement Group

Figure 22 shows the hierarchical and non-hierarchical posterior distribution for the offset informational reinforcement variable for higher utilitarian reinforcement group as a density plot and as a box plot.

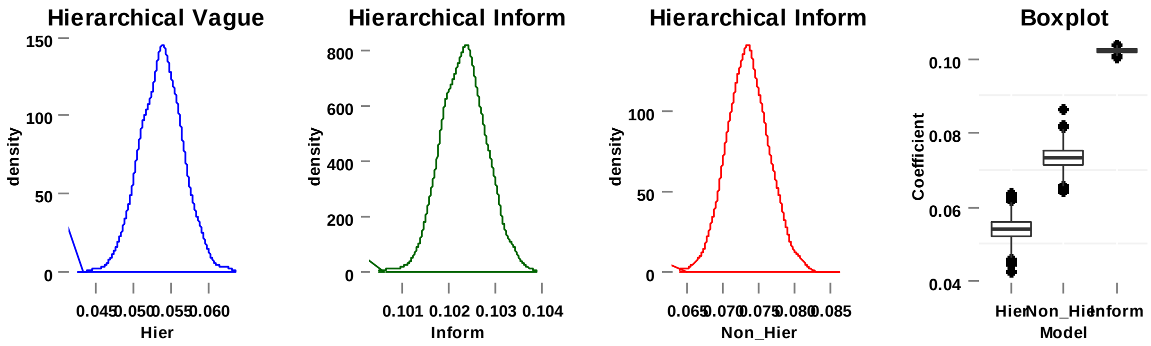

Figure 22: Density and Box plots of the offset for Informational reinforcement in the higher Utilitarian reinforcement group.

The offset value of the coefficient is 0.054 and 0.102 for the hierarchical models in turn and 0.074 for the non-hierarchical model. The Bayesian posterior confidence intervals are (0.046, 0.062), (0.101, 0.1014) and (0.065, 0.082) respectively. Given the intervals do not contain the value zero, there is a 95% probability the parameters are non-zero, hence benefitting model prediction. Also the frequentist t-statistics for each model are 13.5, 145.1 and 17.5 for the hierarchical vague, hierarchical informative and non-hierarchical models respectively, rejecting the null hypothesis of a zero value parameter. This suggests the informational benefit within the higher utilitarian group is contributing positively to the volume of the category above and beyond the informational benefit within the lower utilitarian group. Despite broad agreement between the models as to the positive nature of the coefficients, all models are suggesting a different magnitude of effect and given the lack of overlap in the posterior confidence intervals, this would imply these are statistically different. Hence the nature of model selected both in terms of structure and prior distribution selection has a differing outcome on the magnitude of the effect of the variable. This is in line with discussions around using the Bayesian paradigm and the importance of prior selection (Rossi and Allenby, 2003).

Combining the results of the two informational variables, it can be seen that, within the BPM structure, having taken account of the price variable, the informational and utilitarian variables are contributing positively to the volume of the biscuit category. The higher the informational values, the higher the volume and the higher utilitarian group is having a higher impact on volume per purchase. This is true for all three model structures.

Supermarket Own x Informational Reinforcement

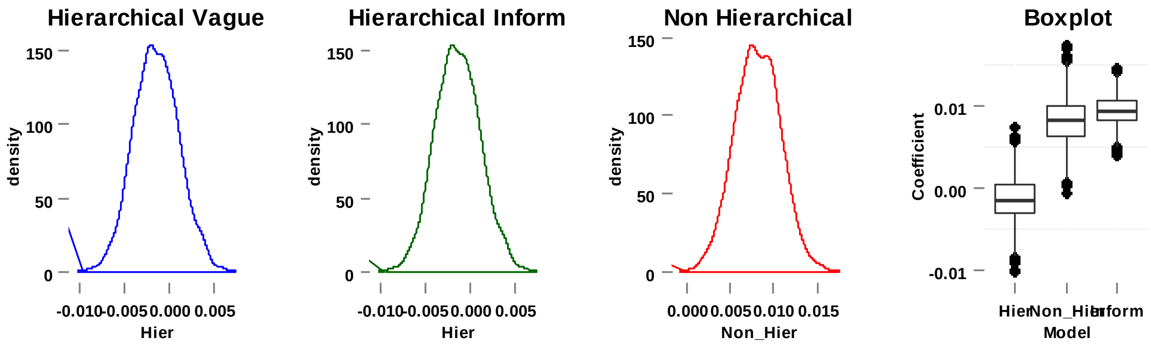

Figure 23: Density and Box plots of Informational reinforcement in the lower Utilitarian reinforcement group for supermarket own brand.

Figure 23 depicts the density and box plots for the hierarchical (vague and informative) and non-hierarchical models. From table 19 the coefficient for the hierarchical vague model is -0.001, 0.010 for the hierarchical informative and 0.008 for the non-hierarchical model. The 95% Bayesian confidence intervals for the three models in turn are (-0.008, 0.005), (0.005, 0.014) and (0.004, 0.015), with frequentist t-statistics of -0.4, 4.1 and 2.3 in each case respectively. This demonstrates the hierarchical vague model's parameter is not different from zero, given the Bayesian confidence interval straddles zero and the t-statistics is non-significant (p=0.368). However, the hierarchical informative model and the non-hierarchical model suggest the parameter is positive and statistically significant from both a Bayesian and frequentist standpoint. The informative nature of the hierarchical prior has influenced the result of the hierarchical informative model to have a positive estimate which differs to the hierarchical vague model estimate. This again demonstrates the importance of the prior distribution in model build.

Therefore differing conclusions as to the nature of the variable and how it may affect sales. The evidence suggests it will be a positive effect or no effect, depending on the model chosen to represent the data.

Supermarket Own x Informational Reinforcement in the higher Utilitarian Reinforcement Group

Figure 24 shows the density plots and box plots for the parameter estimates of this variable. The point estimates for the three models (in the usual order) are -0.081, -0.061 and -0.093.

The Bayesian posterior confidence intervals for the hierarchical vague and non-hierarchical models overlap, (-0.090, -0.071) and (-0.102, -0.083) suggesting these is agreement between the likelihood and the prior. The confidence interval of the hierarchical informative is higher at (-0.068, -0.054). All intervals do not straddle zero, also the t-statistics are all significant at p<0.001 (values re -16.6, -16.5 and -18.5 respectively). Hence these coefficients are statistically significant in the models. The models suggest the informational reinforcement variable associated with the supermarket own brands within the higher utilitarian reinforcement group are negatively contributing to the volume of the category, above and beyond the effect observed in the lower utilitarian reinforcement group.

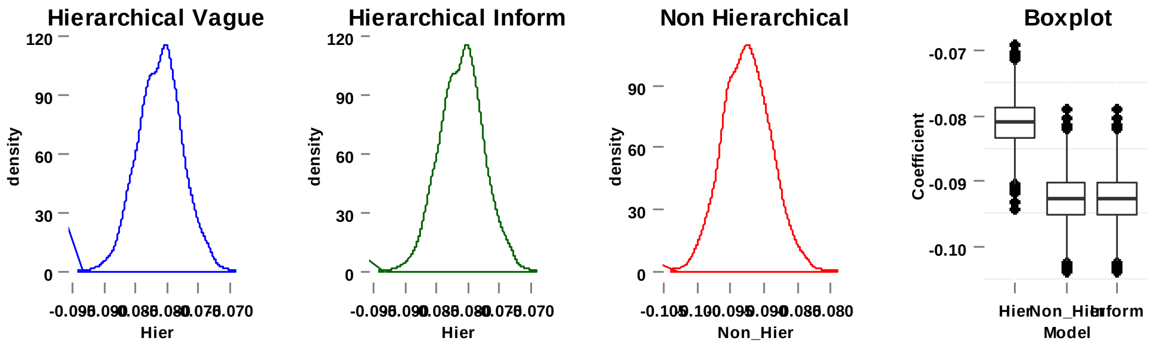

Figure 24: Density and Box plots of Informational reinforcement in the higher Utilitarian reinforcement group for supermarket own brand.

Christmas Week effect (Lower Utilitarian Reinforcement Group)

The earlier category analysis suggests the week containing the Christmas holiday is a noticeably lower volume than other weeks and the inclusion of the dummy variable to test this is discussed in the methodology chapter. Figure 25 shoes the usual charts of the inference

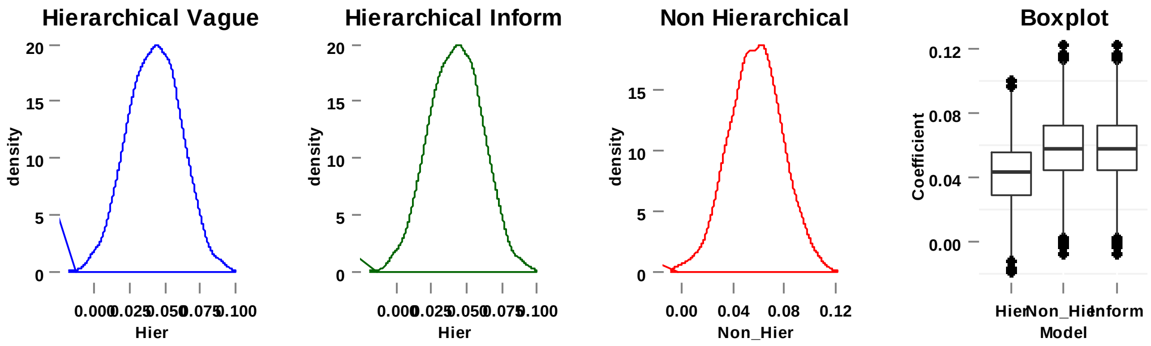

Figure 25: Density and Box plots of Informational reinforcement in the lower utilitarian reinforcement group during the Christmas week.

The models' estimates of the parameter are 0.043, 0.027 and 0.058 in turn. The Bayesian posterior confidence intervals are (-0.009, 0.094) for the hierarchical vague model, (-0.009, 0.064) for the hierarchical informative model and (0.001, 0.117) for the non-hierarchical model. The respective t-statistics are 1.6 (p=0.111), 1.5 (p=0.136) and 2.0 (p=0.055) for the three modes, suggesting the hierarchical structured models conclude no effect. The non-hierarchical model shows the Bayesian confidence interval does not straddle zero however the frequentist p-value at a strict 95% level is not significant. This does show some disagreement between the paradigms, strictly speaking, though given the proximity of the lower confidence interval to zero and also the marginal significance level (p=0.055). Therefore a collective viewpoint would be to accept this parameter is having a positive effect on volume purchases.

The variable relates to the volume purchased per transaction and hence despite a lower volume in the period, it would suggest this is due to lower number of transactions rather than lower volume per transaction. This implies the number of transactions (and hence volume) is much lower for this period, however, when transaction are made, the volume bought per transaction is higher. This may be reflective of the deals which are prevalent within the category immediately post-Christmas and consumers are possible making the most of these offers above and beyond what can be explained by the underlying price elasticity measure.

Christmas Week effect (Higher Utilitarian Reinforcement Group)

Figure 26 shows the density plots and box plots of the posterior distributions of the parameters of the three models.

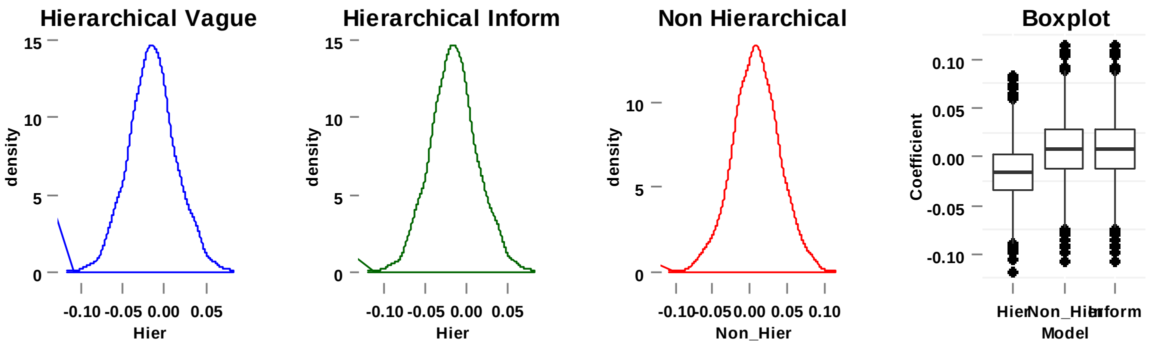

Figure 26: Density and Box plots of Informational reinforcement in the higher Utilitarian reinforcement group during the Christmas week.

The point estimates for the three models are -0.015, 0.039 and 0.008 respectively. The Bayesian posterior confidence intervals for the three models in turn are (-0.094, 0.067), (-0.014, 0.092) and (-0.080, 0.091) and the t-statistics are -0.4, 1.4 and 0.2 in turn, all non-significant at p >=0.145. Therefore there is no evidence from a Bayesian or frequentist perspective to suggest the Christmas week has an effect on volume sales within the higher utilitarian reinforcement group, above and beyond what can be accounted for by the effect within the lower Utilitarian reinforcement group.

Product Characteristic Variables

The product characteristics are dummy variables and the coefficient adjusts the intercept of the model for higher or lower volume levels. The characteristics are biscuit type and pack size. The base biscuit type is countlines and the other variants are offsets to this. The coefficients of the hierarchical models are almost identical. The non-hierarchical model differs with the sign of Non-Sweet being opposite to the hierarchical models. Though the coefficients are small they are statistically significant from a Bayesian and Frequentist perspective. Therefore the type of biscuits makes a difference to the volume bought per purchase.

The base category for the pack size is the single serve packs. The volume sold in other packs is all statistically significantly higher which makes logical sense given the volume per pack is higher in every case. Consistently across all three models the “pack” type has the higher coefficient which contains the larger weight purchases.

Conclusions and Future Considerations

RQ1

The first question was to test the economic behavior price variable and how this compares to past studies with differing functional form. For all three models, the price elasticity is similar across models and also similar to other studies involving the BPM with different functional forms (Oliveira-Castro *et al.*, 2006; Chang, 2007). The inclusion of a more complex model or Bayesian estimation has not changed the fundamental understanding of the price elasticity measure. This underlines the benefit of the BPM which allows economic behavior to be included alongside the psychology variables of the BPM without collinearity impacting the values of the price elasticity variable.

RQ2

This research area focused on the inclusion of a supermarket own brand interaction term with both the informational and utilitarian reinforcement elements of the BPM. The interaction was constructed using a base and an offset variable. The base represented the impact of the supermarket own brand effect on the informational reinforcement variable. The offset represented the impact of the informational reinforcement for the higher utilitarian reinforcement (i.e. an offset to the base variable).

Considering first the base variable, the supermarket own brand indicator interaction terms are statistically relevant for the non-hierarchical model and also the hierarchical vague model. For both these statistically relevant models, the effect is positive and of a similar magnitude which indicates that supermarket own brand is having a positive effect for brands within the lower utilitarian reinforcement group.

For the supermarket own brand within the higher utilitarian group, this is an offset to the lower utilitarian group. Here there is consistency across all three models in terms of direction, with each model indicating a negative coefficient for this variable. There is some difference in magnitude across the three models, however figure 19 demonstrates this magnitude is similar.

The combination implies that volume per purchase would be positively influenced by higher informational reinforcement within the lower utilitarian group, however negatively influenced within the higher utilitarian group.

This suggests consumers are actively seeking a supermarket own brand offering whilst shopping amongst the lower utilitarian brand and are being negatively influenced by the offering for higher utilitarian group brands.

RQ3

The second interaction variable focused on the seasonal Christmas week, having uncovered a significant drop in total volume for that period within the category analysis. As with the supermarket own brand, the Christmas variable was divided into a base (the interaction with the informational reinforcement in the lower utilitarian group) and the offset (the additional impact of the informational reinforcement within the higher utilitarian group for the Christmas week).

For the base measure, there is no statistical impact for either of the hierarchical variables. For the non-hierarchical variable there is also no evidence (at the 5%) from a frequentist perspective to reject the hypothesis the parameter is zero. However, the Bayesian 95% confidence intervals do not include the zero term and hence would suggest there is a statistically relevant and negative effect associated with this variable. Hence differing conclusions (albeit marginal) are arrived at whether the Bayesian or frequentist inference it to be believed.

Analyzing the influence of the offset variable for the Christmas period, all models suggest this impact is not statistically different to zero.

Hence the Christmas period does not have a significant effect on individual consumer behavior as far as the BPM variables are concerned. It is likely the decrease in volume can be attributed to less shoppers (buying in the same manner) and also the fewer shopping days within the period. When consumers do shop within the biscuit category, it would appear their behavior does not differ, psychologically from an average week.

RQ4:

RQ4 focused on the diagnostic difference a hierarchical structure would bring to the model and how interpretation may be affected. In order to address this, the study will compare the non-hierarchical model to the hierarchical model with vague priors, as this isolates the difference attributed by the hierarchical structure alone. In terms of the diagnostics of both models, the hierarchical functional form has a higher predictive power than the non-hierarchical from both a Bayesian and frequentist standpoint.

The focal coefficients are the same in direction for all but the supermarket own brand interaction with informational reinforcement in the lower utilitarian informational group. However the magnitude of the coefficients are varying. Hence a similar conclusion would be reached under both models, though the parameters of the hierarchical would give a better fit to the data. Hence the removal of the assumption of independence within household is providing a better understanding of the consumer behavior.

RQ5

The final research area was the consideration of the Bayesian informative and vague prior selection and how this impacts the model performance and interpretation. From a model diagnostic perspective, there is very little to separate the two models. The frequentist measures of R-squared (adj) , the variance measures and MAPE are all similar and from a Bayesian perspective, the DIC difference is less than 5 (Spiegelhalter, 2002).

The direction of the coefficients are the same apart from the supermarket own brand interaction with informational reinforcement in the lower utilitarian informational group (as seen with the hierarchical vs non-hierarchical comparison). The magnitude of the estimates are different due to the increased precision given to the prior distribution which has had a stronger influence on the data than that of the vague prior distributions. This is what would be expected of course and demonstrates both challenges and opportunities the informed Bayesian models represent to researchers.

Summary, future research and limitations

This study further highlights the flexibility of the BPM framework in incorporating more complex structures and also a differing statistical paradigm in estimating the model parameters. This adds to the literature within through bringing different model structures and Bayesian estimation techniques into the BPM theoretical framework.

The differing parameter posterior estimates of the three models underlines the importance of model definition when considering consumer psychology studies. These are due to differences in functional form and differences in the choice of prior distribution.

The hierarchical structure of the model would seem to be a logical way of assessing consumer behavior whereby the removal of the assumption of independence between household has a theoretical and inferential benefit. Currently the hierarchical nature is restricted to the intercept of panel id. It is conceivable the slope parameters of the focal variables may also benefit from a hierarchical structure as an extension of the model. This does however create a very complicated model. This leads on to a limitation of the Bayesian inference (at least currently) which is the increased computational time required to estimate the parameters. Due to the nature of the estimation where the model estimates the parameters posterior values from a large number of draws rather than from OLS or MLE techniques means that computation time can be extensive and the hierarchical structure compounds this. Hence a balance would need to be sought.
However the use of Bayesian inference demonstrates this form of inference is well suited to estimating problems of this nature and again can be used within the BPM framework. The use of the Bayesian estimation has a wider theoretical impact on the whole subject of consumer psychology as it gives means for a researcher to input past studies, “common sense” or elicitated views into the model process. This has already been suggested as both a good and dangerous practice, however the empowerment for a researcher to input widely recognized or known facts surely must be a benefit as it makes the interpretation of the posterior parameters more intuitive. Furthermore, the researcher has a means of controlling the strength of this prior information rather than accepting or rejecting models after observing results. A further benefit would be studies which are repeated on an ongoing basis whereby the evolution of the parameter values from previous studies could be traced,

Multiple Categories

This study focuses on the biscuit category and the hierarchical structure which lies within the purchase history for the 52 week period. It has been demonstrated this hierarchical structure benefits the consumer behavior understanding both from a theoretical and explanatory power perspective. The Venn diagram on figure 27 suggests consumers are likely to purchase across multiple categories during the year since the overlap of all categories contains the most number of transactions. Therefore the behavior of each category may not be independent as the consumers may be showing similar behavioral psychology, marketing and economic traits across category.

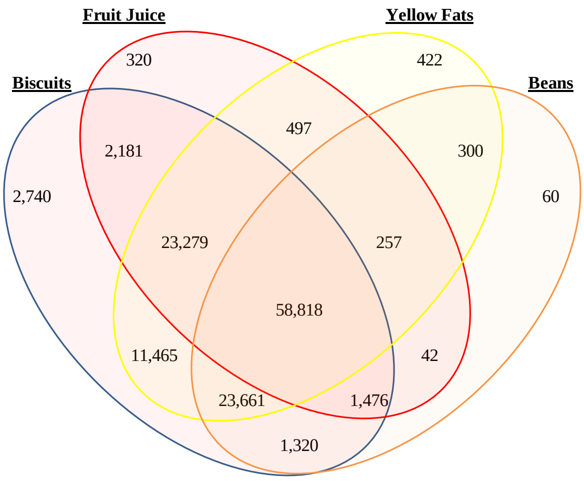

Figure 27: Four-way Venn diagram of the interaction of the four product categories by household.

The behavior within the categories may therefore be influenced by the type of household shopping and their frequency of shopping the category. This is an extension to the argument presented earlier whereby purchases within household may not be independent.
Hence a further build on this study would be to incorporate multiple categories into one hierarchical model where the household panel id would be the hierarchical term across households for cross-category purchase history.
There is a limit of course on how complex the models will become in terms of both resources to run the model and appetite to interpret such large data model outputs. It may be conceivable that sub-categories which are deemed to be more similar in behavior may benefit from such analysis.

References

Abelard (2012). Cause, chance and Bayesian statistics. Available at http://www.abelard.org/briefings/bayes.htm [Last accessed 03/01/2016].

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

Bernado, J. M. and Smith, A. F. M. 2000. *Bayesian theory*, Chichester: Wiley.

Bernardo, J. M. (1999). Bayesian methods in the sciences. *Special Issue of Rev. Acad. Cien Madrid 93(3)*, available at http://www.uv.es/~bernardo/academia.html.

Box, G. E. P. (1980), Sampling and Bayes Inference in Scientific Modeling and Robustness (with discussion). *Journal of the Royal Statistical Society, Series A*, 143, pp. 383-430.

Broadbent SR. (1980) *Price and advertising: volume and profit*. Admap. p. 532–40.

Browne, W. and Rasbash, J. (2004). New developments in modeling, multilevel modeling In: *Hardy, M and Bryman, A. 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.

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*, pp. 509-532.

Dunson, D. B. (2001). Commentary: practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology, 153*(12), 1222-1226.

Efron, B. (2005). Bayesians, frequentisits and scientists. *Journal of the American Statistical Association 100(469)*, pp. 1-5.

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

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

Foxall, G. R. (1999). The behavioral perspective Model consensibility and consensuality, *European Journal of Marketing 33(5/6)*, pp. 570- 596.

Foxall, G. R. (2001). Foundations of consumer behavior analysis. *Marketing Theory 1(2)*, pp. 165-199.

Foxall, G. R. (2005). *Understanding Consumer Choice.* London and New York: Palgrave Macmillan.

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

Foxall, G. R. and Schrezenmaier, T. C. (2003). The behavioral economics of brand choice: Establishing a methodology. *Journal of Economic Psychology 25(5)*, pp. 675-695.

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

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. *Behavioral Processes 66*, pp. 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(1)*, pp. 101-124.

Foxall, G. R. and Oliveira-Castro, J. M. and James, V. K. and Schrezenmaier, T. C. (2011) Consumer behavior analysis and the behavioral perspective model. *Management Online Review*

Gabor A. *Pricing*. London: Heinemann; 1988.

Gelman, A. (2007). Struggles with Survey Weighting and Regression Modeling. *Statistical Science, 22*(2), 153-164

Gelman, A. (2010). Bayesian statistics then and now. *Statistical Science*, *25(2)*, 162-165.

Gelman, A., & Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian statistics, 4*, 625-631.

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

Jeffreys, H. (1931) *Scientific inference*, Cambridge University press

Kagel, J. H., Battalio, R. C., and Green, L. (1995). *Economic choice theory: An experimental analysis of animal behavior*. Cambridge: Cambridge University Press.

Leamer, E. E. (1992). Bayesian elicitation diagnostics. *Econometrica 60(4*), pp. 919-942.

Little, R. J. (2006). Calibrated Bayes: a Bayes/frequentist roadmap. *The American Statistician, 60(3)*, 213-223.

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

Montgomery, A., & Rossi, P. (1999). Estimating Price Elasticities with Theory-Based Priors. *Journal of Marketing Research*, *36(4)*, 413-423

Nagle TT. (1987). *The strategy and tactics of pricing*. Englewood Cliffs, NJ: Prentice-Hall.

O'Hagan, A. (1994). *Kendall's advanced theory of statistics, volume 2b: Bayesian inference*. Cambridge: University press.

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society Series D 47(1)*, pp. 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(3)*, pp. 309-335.

Oliveira-Castro, J. M.,, Foxall, G. R. and Schrezenmaier, T. C. (2006). Consumer brand choice: Individual and group analyses of demand elasticity, *Journal of the Experimental Analysis of Behavior, 85*, pp. 147–166.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* 124, p. 125.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R news, 6(1)*, 7-11.

Roberts A. (1980) The decision between above- and below-the-line. *Admap*. pp. 588–92.

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 Behavior analysis 7(1)*, pp. 15-39.

Rossi, P. E., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science, 22(3)*, 304-328.

Rossi, P. E., & Allenby, G. M. (1993). A Bayesian approach to estimating household parameters. *Journal of Marketing Research*, pp. 171-182.

Rubin, D. B. (1984), Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics, 12*, pp. 1151-1172.

Samaniego, F. J., and Reneau, D. M. (1994), Towards a Reconciliation of the Bayesian and Frequentist Approaches to Point Estimation. *Journal of the American Statistical Association, 89*, pp. 947-957.

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(4)*, pp. 583–639.

Telser, Lester G. (1962). The demand for branded goods as estimated from consumer panel data. *The Review of Economics and Statistics*, pp. 300-324.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, *22(4)*, pp. 1701-1728

Wells, V. K. (2014). Behavioral psychology, marketing and consumer behavior: a literature review and future research agenda. *Journal of Marketing Management, 30(11–12)*, pp. 1119–1158.

Appendix A

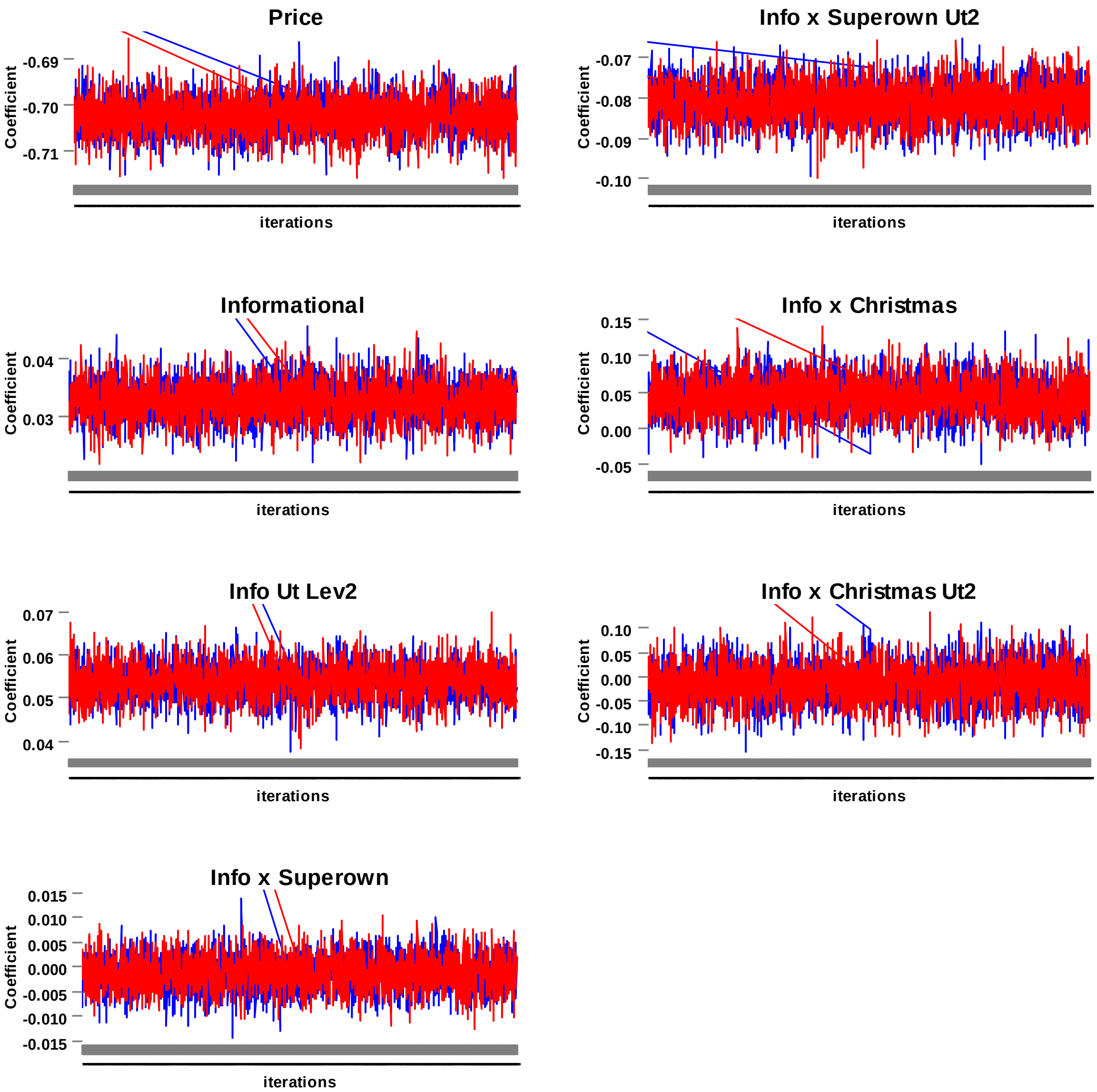

Figure 28: Four-way Venn diagram of the interaction of the four product categories by household.