Consumer's revealed preferences for yogurt purchase in Catalonia: A Generalized Multinomial Logit Approach

Wajdi Hellali[1], Zein Kallas[2], José María Gil[2]

[1]Université Laval, Québec, Canada
[2]Centre for Agro-food Economy and Development (CREDA)
Esteve Terradas, 8, 08860 Castelldefels, Spain.

wajdi.hellali.1@ulaval.ca, Zein.kallas@upc.edu, chema.gil@upc.edu

Paper prepared for presentation at the 148th seminar of the EAAE, ''Does Europe need a Food Policy?" Brussels, Belgium, date as in:
30 November – 1 December, 2015

Abstract

The revealed-preferences of yogurt consumption in Catalonia was analyzed in order to identify the attributes affecting the purchase behavior and to suggest reliable marketing strategies for the yogurt industry. The Discrete Choice Modelling was applied using the Generalized Multinomial Logit Model (G-MNL) calibrated on 52 weeks of yogurt purchases by 987 households that belongs to the home-scan database of ©Kantar World panel during 2012. Eight different yogurt categories were created within the four main supermarket chains in Catalonia (Mercadona, Dia, Carrefour and Lidl) using the following attributes: brand, added ingredients, fat content and the price which was differentiated by two monthly dynamic price indicators. Results showed that the price was the major driving factor. Consumers were more sensitive to the price per product unit than for the price per kilogram showing that yogurt industry could take advantage of this tendency by producing smaller yogurt units with lower weight and cheaper price per unit rather than introducing products with greater volume and cheaper price per kilogram. There was also a clear tendency to consume more private label, especially within the consumers belonging to the lower social class, large families, immigrant populations and individuals aged less than 65 years old. Finally higher levels of randomness related to scale heterogeneity were more significant in the sale points where consumers face a broader variety of products.

Keywords: Yogurt preferences, Revealed choice experiment, home-scanner data, Generalized Multinomial Logit Model, Scale Heterogeneity, private label.

1. Introduction

Understanding consumers and what drives their purchasing behavior is a basic tool for developing a successful marketing strategy and efficient food policies. The proliferation of new information resources and technologies has stimulated the development of new and different marketing research methodologies. Marketing researchers handle a variety of approaches to study the consumers that may relies on the Stated (SP) or the Revealed (RP) preferences analysis. Each approach shows a different data origin and collection technique and presents advantages and drawbacks. The former involves questionnaires or hypothetical treatment by which respondents are asked to make choices or state their preferences in an experimental or questionnaire framework, while the latter is based on consumers' observed and actual behavior revealing implicitly their preferences among a number of available choices that are made in real decision contexts (Ryan and Farrar, 2000).

An important development within marketing research appeared when the companies store started to use data on the purchase behavior of their customers in databases providing managers great opportunities for improving sales results (Verhoef and Hoekstra, 1999; Winer, 2001) and for guiding marketers to establish more effective strategies (Peter and Olson, 2010). This type of data is normally collected by the mean of devices that scan the bar codes at checkout lines of retail stores obtaining electronic records of transactions as part of the operation of their businesses (Robert and Matthew, 2003). The data provide detailed information about quantities, characteristics and values of goods sold as well as their prices (OECD, 2004). In some cases, where the sales are associated with the discount grocery cards, consumers' profiles and their socio demographic are also identified. During the past two decades, food marketing research has accelerated considerably, and the use of the scanner data is becoming widespread, it has provided better opportunities for economic research and marketing decision making to deeply investigate the real food consumption behavior (Nayga, 1992).

A large literature of preference studies is based on SP by dint of its low cost of data collection and their ability to analyze future not existing products while RP analysis permit the work on actually made real choices. The strengths of RP methods are the weaknesses of the SP techniques and vice versa (McFadden, 1994; Whitehead *et al*., 2008). Notwithstanding their extensive use in the literature, considerable challenges subsist concerning the Stated-preferences methods employing self-reported results as to whether respondents are providing responses that are consistent with their true preferences apart (Ossa *et al*., 2007). Thus, individuals' SP may not match with their actual preferences, they may diverge because of systematic bias in SP responses (Bonsall, 1983). The experimental environment itself is considered as a central element in the consumer behavior studies, since the researcher must create an atmosphere similar to real purchasing environment in which participants consider choice alternatives seriously to be able to extract better-quality data (Ossa *et al*., 2007). Whereas the revealed preferences techniques avoid such effects through keeping track of actual behavior of their customers by collecting revealed preference data when behaving in a proper purchasing environment. It took some time for agricultural economists and economic researchers before they have started using scanner data in investigation. Indeed, only since 1979 that these data have been used for application in economic research (Jourdan, 1981). The use of this type of data is becoming more prevalent, it has provided huge opportunities for economic research and marketing decision making in one hand, and a managerial tool for supermarkets in the other (Nayga, 1992).

While utilization of scanner data for food preference investigation has become widespread in the United States, few food preference studies have been conducted using scanner data in Spain such as; (Martinez *et al.* ,2005; Labeaga *et al.* ,2006; Albisu, 2007; Resano, 2008). In 2009, Labeaga *et al*. used a multinomial Logit model including the loyalty variable of Guandani and Little (1983) on an ACNielsen Spanish household scanner panel data to analyze how different forms of heterogeneity help to explain consumers' decisions on non-fine laundry detergents. Even though scanner data revolution in the food marketing and consumption investigation is gaining relevance, there in only few studies in Spain addressing consumer behavior using supermarket or households scanner data. Therefore, there is a need for updated knowledge about consumer preferences using revealed preferences in Spain. According to the scare literature on the yogurt industry, it is revealed to be a worthy area to investigate.

Thus the foremost aim of this study is to analyze the consumption behavior of yogurt in Catalonia (Spain) in order to provide the policy makers with reliable patterns in marketing strategy research and to suggest recommendations to improve profitability in the yogurt industry. To fulfill the abovementioned main objective we planned different secondary objectives, among which, to first, explore the determinant factors influencing the purchase of private labeled yogurts. Second, to investigate the both sources of unobserved heterogeneity among consumers in each point of purchase and finally to calculate the willingness to pay of the major attributes that define consumers preference for yogurt in Catalonia (Spain).

To achieve this objective, we carried out a real choice experiment analysis using homescan data in order to better understand yogurt consumers and the actual tendencies driving their purchasing behavior. Considering that utility of a product is compound of separable utilities for their characteristics or attributes, we applied the recently developed Generalized Multinomial Logit (G-MNL) model. The attributes included were, the label, the presence of added ingredients, the fat content and the price. In a second stage we performed a random effects probit model to determine factors influencing the purchase of private label yogurts. Our investigation relies on a households' purchase scanner data that belong to ©Kantar World Panel which involves a representative sample of Catalan households that scan and transmit their store-bought food and beverage purchases on a daily basis through the use of bar codes and during the period extending between January and December 2012.

From one hand, at the empirical level, this study is the first paper that analyses consumers' preferences toward yogurt using Homescan data in Spain. On the other hand, at the methodological level, this paper contributes to the literature of the Discrete Choice Modelling (DCM) using the recently developed Generalized Multinomial Logit Model (GMNL) of Fiebig et al. (2010). In this context to our knowledge, this research is the first application, in the literature of yogurt preferences studies that analyses the impact of attributes' preferences on real yogurt purchase and how the scale and the preference heterogeneity vary among supermarkets in Spain.

2. Materials and methods

2.1. Database

The data used includes a stratified representative sample of the Catalan households that scan and transmit their store-bought food and beverage purchases on a daily basis through the use of bar codes and during the period extending between January and December 2012. A database involving only the purchases of yogurt has been isolated from the original database representing the households that realized at least one purchase of a yogurt product.

During the preparation of the data set, we dropped out from the dataset the liquid or drinkable yogurts, the dual-compartment format, special yogurts for children under 2 years old. Such filters were applied to retain only data concerning the most common and ordinary yogurt consumption. The panel provides extensive information about every single act of purchase with detailed socio-demographic variables of each household, the product acquired and its attributes and the place of purchase.

2.2. Empirical Application

Before starting the construction of the choice scenarios, we restricted the research to the four most important supermarkets in terms of market share, specifically, Mercadona, Dia, Carrefour and Lidl. These latter, together represent over 59% of the total revenue from the purchase of yogurt by the individuals in the sample during the year 2012. The final sample size was 987 households.

Several strides have been made before the database was able to fit the Real Choice Experiment. First concerning the product identification, a distinct product was defined on the basis of its own characteristics; label (private or producer brands), sub−brand (specific name of the producer or retailer), added ingredients, fat content (regular, semi-skimmed and skimmed), biological information (bifidus or normal), and packaging (plastic cup or crystal)

Table 1: References and Observations by Supermarket

	Mercadona	Carrefour	Dia	Lidl	Total
Number of References	97	128	80	43	348
Number of Observations	11.050	3.157	7.684	2.160	24.051

Since the variables representative of added ingredients, packaging, biological information and sub-brands present each an important number of levels, this led to a large number of alternatives (i.e. yogurt product) for each purchase occasion obtaining too many references by retailer as can be seen in Table 1. In order to remedy this situation we decided to perform a series of aggregations that were basically based upon the different levels of; packaging, sub-brands, added ingredients and biological information. All the attributes were coded with effect coding as discrete variables except for the price. The base levels were; semi-skimmed and skimmed, with added ingredients and private label respectively. Regarding the price, since the yogurt units have different weights and forms, we decided to construct two price indices for every yogurt category; the price per unit and the price per kilogram, where the price per unit was calculated by dividing the reference price by the number of units in each package and by the weight (kg) for the price per kilogram. In addition, due to the aggregation of different product reference in each category, we calculated the monthly-fluctuating average prices per yogurt alternative for each purchase occasion.

Finally, the choice set construction lies in transforming every purchasing record in the database into a purchasing situation that belong to the different alternatives identified as shown in Table 2. To identify the choice sets, we first determined the number of alternatives to be included. Due to the aggregation, we finally considered 8 alternatives in order to ensure the availability with all the recorded purchases in all the points of purchase analyzed. The DCE calibrated on the homescan data, would be as the consumer was confronted to a situation when he/she is in a point of purchase aiming to buy yogurt, faced to eight different yogurt categories and has to choose one in every purchasing occasion.

Table 2: The choice sets

The eight alternatives within each purchase occasion	**Price /kg**	**Price /unit**
Private-label, regular in fat, without added ingredients	1,65	0,17
Private-label, regular in fat, with added ingredients	1,90	0,36
private-label, semi or fully skimmed, without added ingredients	1,12	0,14
Private-label, semi or fully skimmed, with added ingredients	1,78	0,23
Producer brand, regular in fat, without added ingredients	3,40	0,42
Producer brand, regular in fat, with added ingredients	3,95	0,49
Producer brand, semi or fully skimmed, without added ingredients	2,61	0,32
Producer brand, semi or fully skimmed, with added ingredients	3,66	0,46

2.3. Theoretical foundation of the Discrete Choice Experiment

2.3.1. Utility theory

As it is well known, the DCE rely on Lancaster's Theory of Value which proposes that utility of a product is decomposed into separable utilities for their attributes and the Random Utility Theory (RUT) which assumes that subjects choose among alternatives according to a utility function with an observable component and a random error term:

$$U_{jn} = V_{jn}(X_j, S_n) + \varepsilon_{jn} \quad (1)$$

where U_{jn} is the utility of alternative j to subject n, $V_{jn}$is the systematic component of the utility, X_j is the vector of attributes of alternative j, S_n is the vector of socio-economic characteristics of the subject n and ε_{jn} is the random term.

McFadden (1974) developed the multinomial logit (MNL) model which is the base model for to predict the subjects' preferences for attributes (k). it estimates the "probability of choice" that an individual n chooses the alternative i rather than the alternative j within choice sets T). According to this model the utility is given by:

$$U_{njt} = \beta x_{njt} + \varepsilon_{njt} / \sigma_n \qquad n = 1, \ldots, N \quad j = 1, \ldots, J \quad t = 1, \ldots, T \quad (2)$$

Where, β is a vector of mean of the observed attributes (x_{njt}) and ε_{njt} is the "idiosyncratic" error term or the residual heterogeneity for the unobserved latent attributes (Fiebig et al., 2010). This term follows a Type 1 extreme value distribution with scale parameter σ_n which is normalized to one for identification. The probability that an individual n will choose alternative j among other alternative of an array of choice set T is formulated as follows:

$$(P_j \left| X_{nt} \right.) = \frac{\exp(\beta x_{njt})}{\sum_{j=1}^{J} \exp(\beta x_{njt})} \qquad \forall j \in T \qquad (3)$$

However, this model imposes homogeneity in preferences as it only allows for the average attributes' utilities which is often unrealistic as consumers' preferences are, by nature, heterogeneous. Therefore, the mixed or heterogeneous logit models (MIXL) has been introduced.

The Mixed, Heterogeneous Logit models (MIXL) or Random Parameter Logit model (RPL) allows for unobserved heterogeneity by introducing random coefficients on attributes (Ben-Akiva et al., 1997). However, it still assumed that the scale parameter is one for identification. The utility in this case is specified by a vector of n specific deviations (η_n) from the mean value of attributes β and follows a continuous distribution defined by the researcher. In most applications the multivariate normal distribution is used.

Recently Louviere & Mayer (2007), Louviere *et al*. (2008) argued that much of the preference heterogeneity captured by random parameters can be better captured by the scale term. Besides, the normal distributions of the random attributed do not appear to be very close to it. Thus, this model turns to be likely a poor approximation to data if scale heterogeneity is not accounted for (Fiebig et al., 2010). In this context, the analysis of the scale heterogeneity is important, where consumers may have varying levels of attention paid to the purchasing situation they are facing, as well as the level of certainty in their choice (Train & Weeks, 2005). Thus, it would be expected that the scale of the error term could be greater for some consumers than for others and accounting for tend to be highly relevant.

Feibig *et al*. (2010) developed the Generalized Multinomial Logit model (GMNL) that mainly account for the scale heterogeneity. Within this approach, the utility to person n from choosing alternative j on choice set t is given by:

$$U_{njt} = [\sigma_n \beta + \gamma \eta_n + (1-\gamma)\sigma_n \eta_n] X_{njt} + \varepsilon_{njt} \qquad (4)$$

where γ is a mixing parameter between 0 and 1 that determines the level of interaction between the scale heterogeneity coefficient σ_n and the parameter heterogeneity coefficient η_n.

Finally, the GMNL model is specified by default to consider the η_n as uncorrelated. However, the repeated choice situations containing the same attributes and levels may have unobserved effects that are correlated among alternatives in a given choice situation and the correlation would be estimated (Hensher *et al*., 2005).

In our case study, we used a GMNL model with correlated random parameters as it showed the best goodness of fit compared to other specification in terms of Pseudo-R^2, AIC and improvement in the Likelihood functions. We used the GMXLOGIT procedure in NLOGIT 5.

Once parameters are estimated, the implicit price (willingness to pay) of the attributes are determined as follows:

$$IP_{\text{Product\_attribute}} = -\left(\frac{\beta_{\text{Product\_attribute}}}{\beta_{\text{monetary\_attribute}}}\right) \qquad (5)$$

The implicit price to move from the base level of each attribute to the analyzed level is multiplied by 2.

3. Results and discussion

3.1 Impact of attributes' preferences on yogurt purchase

Focusing on how the concerned attributes present within the offered yogurt categories, the G-MNL model results affords, in the table below, detailed information allowing deeply analyze the households' preference of yogurt in each point of purchase. Marginal utilities of the attributes, the attributes heterogeneity terms and the scale parameters resulting from the G-MNL models with correlated random parameters were provided separately for every point of purchase and with both price per unit and price per kilo. As can be seen, at a 99% confidence level, we can reject the null hypothesis that all coefficients are jointly equal to zero with a Log-Likelihood ratio test highly significant for all the supermarkets. The goodness of fit is assessed through the McFadden's pseudo-R^2 (0.30 and 0.43) considered as highly accepted values.

Table 1. The Generalized Multinomial model (G-MNL) results

	Mercadona		Día		Carrefour		Lidl	
	€/kg	*€/unit*	*€/kg*	*€/unit*	*€/kg*	*€/unit*	*€/kg*	*€/unit*
Random parameters in the utility function (β)								
Regular fat content	0.26***	0.53***	0.34***	0.15***	0.02	0.26***	-0.06	0.32***
Private label	0.41***	0.41***	0.64***	0.21***	0.31**	0.06	0.56***	0.69***
With added ingredient	0.32***	0.07**	-0.21***	-0.31***	-0.31***	-0.12***	-0.36***	-0.34***
Price	-1.82***	-6.91***	-0.70***	-4.38***	-0.33**	-4.81***	-0.82***	-2.92***
Covariances of Random Parameters								
Producer brand: Regular fat content	-0.35***	0.13***	-0.72***	-1.34***	1.58***	0.47***	-0.52***	0.36***
Without added ingredients: Regular fat content	0.57***	0.01	0.32***	0.29***	-0.28***	-0.30***	0.19***	-0.06*
Without added ingredients: Producer brand	0.09*	0.28***	-1.03***	-0.12	0.10	-0.10	-0.13	0.38***
Price : Regular fat content	-0.97***	0.33	-0.02	1.40*	0.47***	0.10	1.11	1.41***
Price : Producer brand	-0.32***	0.07	0.49***	-0.63	-0.17	-2.80***	-0.10	5.31
Price : Without added ingredients	-1.27***	-0.01	0.61***	-2.13***	-0.44***	0.37	0.40	-0.18
Variance parameter tau in G-MNL scale parameter								
TauScale (τ)	0.19***	0.31***	0.14***	0.07***	0.27***	0.35***	0.12***	0.14***
Weighting parameter gamma in G-MNL model								
Gamma in G-MNL (γ)	0.10	0.10***	0.09	0.11***	0.10	0.09	0.09	0.08
Standard deviations of parameter distributions								
Fat content	1.39***	1.33***	1.08***	1.12***	0.91***	0.84***	0.70***	0.73***
Label	0.59***	1.30***	1.83***	2.07***	2.04***	1.62***	0.75***	1.42***
Added ingredients	1.36***	0.89***	1.17***	1.10***	0.90***	0.90***	1.26***	1.15***
Price	1.27***	0.29	0.83***	3.33***	1.24***	2.55***	1.50***	3.92***
Model adjustment								
LL ratio test	18,004***	18,136***	13,346***	13,480***	4,017***	3,949***	3,615***	3,863***
McFadden Pseudo R^2	0.39	0.39	0.42	0.41	0.31	0.30	0.40	0.43
Observations	713	713	460	460	376	376	292	292

****, ** and * refer respectively to the rejection of the null hypothesis at 1%, 5% and 10% significance level*

To better understand the attributes preferences, the utilities of each attribute from the G-MNL estimation were obtained and organized in the table above. The positive/negative sign of the coefficient implies higher/lower levels of utility associated with these attributes' levels. In this context, the model estimates showed that the price attribute is statistically significant for all the supermarkets analyzed. As expected, results show negative signs on the statistically significant marginal utilities of the price which implies that an increase in the price attribute will decrease the utility of the yogurt presented to consumers. It is relevant mentioning also that the price is the most important attribute in both Mercadona and Carrefour supermarkets which means that the consumers buying yogurt from these points of purchase focus more on the price than any other attribute when choosing between the offered yogurt products.

In addition, when comparing both specifications of the price: per kilo and per unit price' marginal utilities, we notice that the latter shows remarkably higher magnitudes than the former in all the points of purchase with the highest difference in Mercadona. This trend implies that the consumers accord more importance to the price per unit than to the price per kilo when making the purchase of yogurt. Such tendency could be emphasized when deciding in marketing strategies in a way to make it profitable, by producing smaller yogurt units, with lower content and cheaper price rather than introducing products with greater volume and cheaper price per kilo.

Concerning, the fat content attribute preference, the G-MNL outcomes show positive signs and statistically significant marginal utilities in the different points of purchase for the regular fat content level. These results are in accordance with the univariate analysis where in all the points of purchase, the most bought yogurts are regular in fat.

Despite of the approval in the literature implying that indications of health claims on food packaging favorably influence the purchase intentions (Viana *et al*., 2014; Wagner *et al*., 2015), regular fat yogurt showed to be the most preferred in our case. These results are in accordance to what Kähkönen and Tuorila (1999) claimed in their research about consumers preferences stating that reduced-fat information decreased consumers' pleasantness of yogurt and chocolate.

Focusing on the label attribute preference, results display positive signs on all the statistically significant marginal utilities in the different models, which refer to the consumption of retailer-labeled yogurts. This implies that, households show a preference for the private labels rather than the manufacturer brands.

The point of purchase with the highest magnitude on the label attribute utility is Lidl and this could be explained by the fact that the own-branded yogurts purchased from Lidl by the households in the sample represent 92% of the total yogurts in terms of purchasing occasions against only 8% for the other brands in this point of purchase.

Regarding the added ingredients attribute preference, the G-MNL results show highly statistically significant and negative signs for all the points of purchase except for Mercadona where the signs on the marginal utilities for the added ingredients' attribute are positive. This means that Mercadona's customers reveal a preference for the added ingredients in the yogurts they purchase, unlike for the other points of purchase, specifically, Día, Carrefour and Lidl where the households show a preference for the yogurt without added ingredients.

As commented, the scale parameter from the G-MNL describes the potential level of uncertainty with respect to the consumers' choices. We base our analysis on the hypothesis which consists that the more variety of products is offered; the more there is an increasing level of uncertainty in the decision making. In other words, if the consumer, when making the purchase of yogurt, finds himself in front of a large variety of products facing, the levels of choice randomness in the final decision would be higher. Interpreting the tau parameter (τ) that captures the scale heterogeneity, results showed a significant scale heterogeneity in the data in all the points of purchase. High values of τ parameter are shown for each of Mercadona and Carrefour against less important ones for Día and Lidl which could be explained by the broader variety of products that the consumers would face when trying to choose a yogurt type in Mercadona and Carrefour against the relatively fewer variety of products provided by the others. While comparing between the tau parameter for both price specification (€/kg and €/unit) we realized that there is a tendency to have higher tau scale values on price per unit than price per kilo. This could be explained by the fact that the consumer may be more uncertain when comparing unit prices of two different products than doing the same with price per Kg.

Regarding the unobserved preference heterogeneity, the estimated models showed statistically significant results. Thus, as mentioned by Lenk (2011), when the estimated standard deviation of parameters distribution are close to zero, then the unobserved heterogeneity is mostly due to heterogeneity in the scale parameter and not preference. In our results the values are far from zero and thus there is a mixture between both sources of heterogeneities. This can be verified analyzing the gamma mixing estimate. Results showed that in both model are relatively close to zero. This implies that both heterogeneity sources are not independent.

For the economic interpretation, the implicit prices of the attributes levels were calculated. Since these estimates are stochastic we used the Krinsky & Robb (1986) simulation to calculate their confidence intervals through 1,000 random repetitions. To avoid any misinterpretation, results must be considered as the willingness to pay (€/kg and €/unit) to shift preference from the base level of the attribute to the evaluated one.

Results showed positive and statistically significant values of willingness to pay for almost all the attributes in all points of purchase. For instance, in Mercadona consumers who made their purchase of yogurt were willing to pay 0.34 €/kg or 0.02 €/unit to make the purchase of yogurt with added content (added ingredients) and to pay an 0.28 €/kg or 0.16 €/unit for regular fat yogurt.

Since descriptive analysis shows that private-Label represents 48% of the yogurt purchases in Cataluña and since the results of the real DCE reveals that the label is one of the most important attributes entering in the purchase decision of Yogurt, we decided to investigate the factors influencing the purchase of private-label yogurts.

Table 4. The willingness to pay from the GMNL Model

Willing to pay for yogurt' attributes and levels								
	Mercadona		**Día**		**Carrefour**		**Lidl**	
	€/kg	*€/unit*	*€/kg*	*€/unit*	*€/kg*	*€/unit*	*€/kg*	*€/unit*
Regular fat content	0.28^{***} (0.16; 0.30)	0.16^{***} (0.14; 0.17)	0.96^{***} (0.76; 1.14)	0.06^{***} (0.04; 0.08)	0.14 (-0.18; 0.44)	0.10^{***} (0.08; 0.14)	-0.14 (-0.26; -0.02)	0.21^{***} (0.30; 1.06)
Private label	0.44^{***} (0.15; 0.85)	0.12^{***} (0.08; 0.14)	1.80^{***} (0.74; 2.88)	0.10^{***} (0.04; 0.14)	1.88 (-1.24; 3.53)	0.02 (-0.02; 0.06)	1.36^{***} (0.22; 4.14)	0.47^{***} (0.09; 2.06)
With added ingredient	0.34^{***} (0.28; 0.42)	0.02^{***} (0.02; 0.04)	-0.60^{**} (-0.96; -0.22)	-0.14^{***} (-0.18; -0.10)	-1.87 (-4.09; 0.44)	-0.06^{**} (-0.10; -0.01)	-0.87^{***} (-1.6; -0.22)	-0.23^{**} (-1.20; -0.08)

***, ** *and* * *refer respectively to the rejection of the null hypothesis at 1%, 5% and 10% significance level*

3.2 Factors influencing the purchase of private yogurt brand

A panel probit model has been carried out in order to identify the possible variables influencing the purchasing behavior of the private labeled yogurts.

$$Y_{it}^{*} = x_{1t}'\beta^{0} + x_{2t}'\beta^{0} + x_{3t}'\beta^{0} + \dots + x_{it}'\beta^{0} + \varepsilon_{i} \quad Y_{it} = 1(Y_{it}^{*} > 0) \quad (9)$$

The dependent variable is a dummy created from the panel data, which equals one if the household purchases private labeled yogurts and equals zero when the yogurt purchase recorded by the scanner concern a producer branded product. The independent variables included in the panel probit model comprise, a dummy from the legal status variable indicating if the individual is a Spanish or immigrant, a dummy from social class variable specifying if the household belongs to the higher and middle higher class, a dummy from the age range variable describes if the individual is older than 65 years old, and other variables from each, family size, yogurt budget, habitat and the EGM class. A continuous independent variable denotes the weight in value of yogurt in the shopping basket.

The panel probit model output shows, as displayed in the Table 5, statistically significant coefficients (p-value < 0.05) on the independent variables. Positive signs appear on two variables, specifically, the "immigrant" and "large family" which mean that immigrants in Spain have more probability to purchase private labeled yogurts rather than the branded ones; besides, large-size households tend more to buy yogurts with private labels.

Table 5. Determinants factor for private label preference

Variables	Coefficient	P-value
Immigrant	0.63^{***}	0.006
Higher and middle higher class	-0.25^{*}	0.085
Age range 65+	-0.39^{***}	0.016
Large families	0.26^{**}	0.028
High yogurt budget	-1.27^{***}	0.000
Habitat 200,000+	-0.23^{**}	0.050
Yogurt weight in the shopping basket	-0.15^{***}	0.000
High EGM class	-0.34^{***}	0.007
Constant	0.96^{***}	0.000

Likelihood-ratio test of rho=0: chibar2(01) = 1.9e+04 Prob >= chibar2 = 0.000

***, ** and * refer respectively to the rejection of the null hypothesis at 1%, 5% and 10% significance level

Negative signs show up at the coefficients of the remaining explanatory variables, which are, the higher and higher middle class, high yogurt budget, habitat 200,000+, weight of the yogurt in the shopping basket and high EGM class. In other words, households belonging to the higher and higher middle class are less likely to consume private labeled yogurts. Furthermore, the older the homemaker is, the less probability it has to consume private labeled yogurts rather than other brands' yogurts which could be explained by their willing to care about their health and opting for a healthier diet, so, better quality of yogurt.

As expected, households with higher yogurt budget during a purchase occasion are more likely to buy other brands rather than private labels as they have the perception that these latter are cheaper. Moreover, results show that households living in highly populated areas are less likely to buy supermarkets' own branded yogurts.

Individuals, who have a greater yogurt weight in the shopping basket, have more probability to make the purchase of a white label. And finally, the household with higher EGM class, are less likely to consume private labeled yogurts, which may be related to the influence of the mass media and advertising on their purchasing decisions.

Once the results of the panel probit discussed we decided to profile the private-label yogurt' consumers via a cluster analysis in order to highlight possible consumers' profiles based on sample segmentations according to some purchasing data which we generated from the database and that we took into consideration in our analysis.

4. Conclusions

The foremost aim of this thesis is to investigate the consumption behavior of Catalan households while making the purchase of yogurt in four different supermarket chains located in the province of Catalonia, with the intention of proposing trustworthy patterns in marketing strategy research and suggesting recommendations to improve the profitability of the yogurt industry.

To achieve this objective, we carried out a real choice analysis using homescan data in order to better understand yogurt consumers and the actual tendencies driving their purchasing behavior. Considering that utility of a product is compound of separable utilities for their characteristics or attributes, we performed first recently developed Generalized Multinomial Logit (G-MNL) model calibrated on homescan data. The attributes included in the model were, the label, the presence of added ingredients and the fat content in one hand and the price specification in the other. Then we applied a random effects probit model to determine factors influencing the purchase of distributor brand yogurts.

During the choice design construction, we faced several issues concerning the number of alternatives and the price construction process, and hence the model didn't converge. Therefore, we rearranged the different yogurt products obtained in a way to have a single choice set with different purchase occasion so we could later compare between the four supermarkets with the highest importance in terms of markets share in the database which we will consider in our research. Finally, we aggregated the resulting products into eight final yogurts categories that fit in all the supermarkets. In the other hand, the yogurt prices are flexible and are likely to vary during the year, we calculated monthly different prices for every category in every point of purchase.

As in any investigation, we are aware that this research has some limitations, among-which we recognize, first the issue of aggregation decision where the existence of different products within the same category would have been investigated by defining a different strategy like the one followed in the study of (Keane et al., 2012) involving 100 different alternatives of pizza types, proposing to use a random subset of alternatives among the full choice. Another limitation related to the study of the observed heterogeneity would have been considered through the introduction of the socio-demographic variables into the G-MNL model due to the slowness of the iterations and the huge size of the data.

As a future line of the research, we suggest including a wider set of product categories within the dairy sector to be able to study the difference between the attributes perceptions that could appear between the different products, for instance the fat content for the milk and the health attributes for the cheese compared to the yogurt. More attributes and more alternatives would provide a better understanding of the sector, and more reliable outcomes. A second line would be to incorporate the observed heterogeneities in the study for a more complete research. Finally, we suggest to, carry out a hypothetical discrete choice experiment covering the same dimensions and objectives and comparing the both approaches in order to be able to evaluate any possible gap between the revealed and the stated preferences methods and propose thus possible corrections to improve the stated preferences approach.

References:

Albisu L. M. and Mtimet N., (2007), “Spanish wine consumer behaviour: A stated and revealed preferences analysis” Paper prepared for presentation at the I Mediterranean Conference of Agro-Food Social Scientists. 103rd EAAE Seminar ‘Adding Value to the Agro-Food Supply Chain in the Future Euromediterranean Space’. Barcelona, Spain, April 23rd - 25th, 2007.

Ben-Akiva, M., D. McFadden, M. Abe, U. Böckenholt, D. Bolduc, D. Gopinath, T. Morikawa et al. 1997. Modelling methods for discrete choice analysis. Marketing Letter, 8(3) 273–286.

Blattberg, R.C. and K.J. Wisniewski (1986), Analyzing Intra-Category, Inter-Brand Price competition; A Theory of Price Tier Competition, Working paper. University of Chicago January.

Bonsall, P. W. (1983): "Transfer Price Data - Its Use and Abuse". Proceedings of Seminar M, PTRC Summer annual Meeting, Brighton.

Capps Jr., O. (1989) I "Using Scanner Data to Estimate Retail Demand Functions for Meat Products"; American Journal of Agricultural Economics 71, 750- 60.

Capps Jr., O. and R.M. Nayga, Jr. (1990), "Effect of Length of Time on Measured Demand Elasticities: The Problem Revisited" , Canadian Journal of Agricultural Economics November; 499-512.

Capps Jr., O. and R.M. Nayga, Jr. (1991b), II Demand for Fresh Beef products in Supermarkets: A Trial with Scanner Data" Agribusiness: An International Journal May, 241-51.

Chen Zhen, Muth K.M. and Siegel P.H. (2007) “Homescan Data Description”, RTI

Eastwood, D.B., M. Gray and J. Brooker (1990) I Using Scan Data to Evaluate Advertising; A Suggested Methodology and Preliminary Results, Paper presented during the Southern Agricultural Economics Association meetings, Little Rock, Arkansas, February.

FAO 2010: Status of and Prospects for Smallholder Milk Production – A Global Perspective, by T. Hemme and J. Otte. Rome

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Marketing Science, 29(3), 393-421.

Gibbons Robert D. and Hedeker Donald (1994) “Application of Random-Effects Probit Regression Models” Journal of Consulting and Clinical Psychology1994, Vol. 62, No. 2, 285-296.

Greene, W. H., & Hensher, D. A. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. Transportation, 37(3), 413-428.

Greene, William (2002) “The behavior of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects” Econometrics Journal, Royal Economic Society, vol. 7(1), pages 98-119, 06.

Guadagni, P.M. and J.D.C. Little (1983) "A Logit Model of Brand Choice Calibrated on Scanner Data" , Marketing Science 2, 203-38.

Gurumurthy, K and John D.C. Little (1989) "A price Response Model Developed From Perceptual Theories", Sloan School Working Paper #3038-89, Marketing Center Working paper 89-5.

Harald J. van Heerde/Peter S. H. Leeflang/ Dick R. Wittink (2002), “How promotions work: Scan*pro-based evolutionary model building”, Schmalenbach Business Review, Vol. 54, July 2002, pp. 198 – 220

Hensher, D., Rose, J. & Greene, W. (2005) Applied choice analysis: A primer. Cambridge University Press: Cambridge.

IDRE, (2015). Institute for Digital Research and Education, http://www.ats.ucla.edu/stat/mult_pkg/faq/general/effect.htm

Jourdan D.K.. (1981) I Elasticity Estimates for Individual Retail beef Cuts Using Electronic Scanner Data, Unpublished M.S. Thesis, Texas A&M University.

Juan C. Gazquez-Abad and Manuel Sanchez-Perez (2009), "Factors Influencing Olive Oil Brand Choice in Spain: An Empirical Analysis Using Scanner Data", Agribusiness, Vol. 25 (1) 36–55 (2009)

Kai-Brit Bechtold (2013), " Consumers' Preferences and Willingness-To-Pay for Functional Dairy Products in Germany", Unpublished M.Sc. Thesis, Universität Zu Kiel

Kähkönen, P., & Tuorila, H. (1999). Consumer responses to reduced and regular fat content in different products: effects of gender, involvement and health concern. Food Quality and Preference, 10(2), 83-91.

Kallas, Z. & Gil, J.M. (2012). A dual response choice experiments (DRCE) design to assess rabbit meat preference in Catalonia A heteroscedastic extreme-value model. British Food Journal, 114 (10-11), 1394-1413.

Keane, M. 2006. The generalized logit model: Preliminary ideas on a research program. Presentation, Motorola-CenSoC Hong Kong Meeting, October 22, Motorola, Hung Hom, Kowloon, Hong Kong.

Keane M.P. and Wasi N. (2012) "Estimation of Discrete Choice Models with Many Alternatives Using Random Subsets for the Full Choice Set: With and Application to Demand for Frozen Pizza" Working Paper. University of Oxford.

Krinsky, I. & Robb, L. (1986). On approximating the statistical properties of elasticities, The Review of Economics and Statistics, 68(4), 715-719.

Kumar, Amresh G (2007), "Use of retail scanner data to gain strategic market insights: Two applications" Unpublished PhD Thesis, Faculty of the Graduate School of State University of New York.

Kumar, V. and Leone, R.P. (1988), "Measuring the effect of retail store promotions on brand and store substitution", Journal of Marketing Research, Vol. 25 No. 5, pp. 178-85

Labeaga, J M, Lado, N and Martos M. (2006), "Behavioural loyalty towards store brands", Journal of Retailing and Consumer Services (2007), Volume: 14, Issue: 5, Pages: 347-356.

Labeaga J.M, Lado Nora and Martos Mercedes (2009), "Brand loyalty and heterogeneity in the Spanish market of non-fine laundry detergents", capítulo 10 en N. Kalai Seldan (ed.) Building Brand Loyalty. Issues and Perspectives, The Icfai University Press, 2009, págs. 99-116.

Lancaster, K. (1966) A new approach to consumer theory. Journal of Political Economy, 74, 132-57.

Lenk, P. (2011). A comparison of generalized multinomial logit (GMNL) and latent class approaches to studying consumer heterogeneity with some extensions of the GMNL model by Peter J. Lenk. Applied Stochastic Models in Business and Industry, 27(6), 580-583.

Louviere, J. J., D. Street, L. Burgess, N. Wasi, T. Islam, A. A. J. Marley. 2008. Modelling the choices of individuals decision makers by combining efficient choice experiment designs with extra preference information. Journal of Choice Model. 1(1) 128–163.

Louviere, J. J., R. J. Meyer. 2007. Formal choice models of informal choices: What choice modeling research can (and can't) learn from behavioral theory. N. K. Malhotra, ed. Review of Marketing Research. M. E. Sharpe, New York, 3–32.

Martínez-Ruiz, M.P. Mollá-Descals, A. Gómez-Borja, M.a. Rojo-Álvarez, J.L. (2006), "Evaluating temporary retail price discounts using semiparametric regression", Journal of Product & Brand Management, Volume 15 · Number 1 · 2006 · 73–80.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In P. Zarembka, (Ed) Frontiers in econometrics. New York: Academic Press.

McFadden, D. (1994) Contingent valuation and social choice, American Journal of Agricultural Economics, vol. 76, pp. 689-708.

Moriarty, M.M. (1985) ,"Retail Promotional Effects on Intra and Inter-brand Sales Performance", Journal of Retailing 6 p, 27-47.

Nayga, Jr. (1992). “Scanner data in supermarkets: Untapped data source for agricultural economists”, Review of Marketing and agricultural economics, Vol. 60, No. 2, August 1992, Part 1.

Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.

OECD, ILO, IMF, Eurostat, UNECE, World Bank, 2004, Consumer Price Index Manual: Theory and Practice, International Labour Office, Geneva.

OECD-FAO (2011), “OECD-FAO Agricultural Outlook 2011-2020”, chapter 9, Dairy.

Óscar González-Benito María Pilar Martínez-Ruiz Alejandro Mollá-Descals, (2008),"Latent segmentation using store-level scanner data", Journal of Product & Brand Management, Vol. 17 Iss 1 pp. 37 - 47

Ossa DF, Briggs A, et al., Recombinant erythropoietin for chemotherapy-related anaemia: economic value and health-related quality-of-life assessment using direct utility elicitation and discrete choice experiment methods. Pharmacoeconomics 2007;25:223-37.

Peter J. P. and Olson J.C. (2010) " Consumer Behavior & Marketing Strategy", Ninth Edition, p 6, 9-10, 234-235.

Rencher, A.C. (2002), Methods of Multivariate Analysis, Second edition, Wiley.

Resano H., Sanjuan A.I., 2008. “An hedonic approach applied to scanner data on cured ham purchases in Spain” Proc. XII^{th} Congress of the European Association of Agricultural Economists (EAAE). Gent, Belgium, August 26-29. 196 pp.

Robert C. Feenstra and Matthew (2003) "Introduction to: Scanner data and Price indexes", University of Chicago Press 2003, 0-226-23965-9.

Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ 2000;320:1530-3.

Shugan, S.M. (1987), "Estimating Brand positioning Maps Using Supermarket Scanning Data" ,Journal of Marketing Research 24, 1-18.

Stevens, J. P. (1999) “Intermediate statistics: A modern approach (2nd ed.)”. Post Hoc Tests in ANOVA. p1,4.

Thurstone, L. (1927) “A law of comparative judgement”. Psychological Review, 34, 273-286.

Torrisi, Francesco, Stefani, Gianluca, Seghieri, Chiara (2006). “Use of Scanner Data to Analyze the Table Wine Demand in the Italian Major Retailing Trade”, Agribusiness, Vol. 22, No. 3, 2006, p 391-403

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In Alberini, A. and Scarpa, R. (eds), Applications of Simulation Methods in Environmental and Resource Economics. Boston/Dordrecht/London: Kluwer Academic Publishers, 1–16.

Verhoef P.C. & Hoekstra J.C. (1999), The status quo of database marketing in the Dutch FMCG-Industry, Journal of Market Focused Management, 3, 313-332

Viana, M. M., Dos Santos Silva, V. L., & Trindade, M. A. (2014). Consumers' perception of beef burgers with different healthy attributes. LWT-Food Science and Technology, 59(2), 1227-1232.

Wagner, H. S., Howland, M., & Mann, T. (2015). Effects of subtle and explicit health messages on food choice. Health Psychology, 34(1), 79.

Whitehead, J.C., Pattanayek, S., Van Houtven, G. and Gelso, B. (2008) Combining revealed and stated preference data to estimate the non-market value of ecological services: an assessment of the state of the science. Journal of Economic Survey (forthcoming)

Wittink, D.R., M.J. Addona, W.J. Hawkes and J.C. Porter (l988), "The Estimation, Validation, and Use of promotional effects Based on Scanner Data", Working Paper, Cornell university.