Erasmus University of Rotterdam

Erasmus School of Economics

Bachelor Thesis Econometrics and Operations Research

Sampling Based Approaches for Estimating Heterogeneous Discrete Choice Models

Author

Christina van Spaendonck

Student number

421913

Thesis Supervisor

Kathrin Gruber

Second Assessor

A. Castelein

Abstract

This paper aims to investigate the effect of marketing activities on the probability that an individual purchases a specific brand from a set of substitutable alternatives and the individual-specific heterogeneity of intrinsic preferences. To circumvent the Independence of Irrelevant Alternatives, the discrete choice model that is employed is the probit model. This paper estimates the parameters replicating the work of Chintagunta and Honore (1996) by using the Method of Simulated Moments. The results of this method show a trade-off between running-time and accuracy, making it an inefficient method. The research is extended by estimating the parameters using the Gibbs sampling approach, which proves more efficient. Incorporation of heterogeneity is done using either a brand-loyalty variable, which provides biased results due to the incorporation of purchase-history, or a random-effects model, which assumes that the heterogeneity is distributed normally over the population. This research shows that the method that provides the least biased results is the Method of Simulated Moments, with incorporation of heterogeneity by means of a random-effects model.

July 8, 2018

Contents

1 Introduction

In the current market, companies compete with each other to lure consumers to buy their products. In markets where the products are substitutable, companies engage largely in brand-management to achieve this aim. Such brand management includes brand positioning strategies, pricing strategies and marketing and promotion activities. Increased investment in these types of activities can improve brand-performance and hence allow for companies to stand out as a brand. For companies to know which activities reap the most reward, they need to know how consumers individually and collectively respond to these marketing activities.

In this study, I will investigate this unknown effect as analyzed and evaluated by Chintagunta and Honore (1996). A discrete choice model is employed to model the purchase behavior of individuals, specifically focusing on the multinomial logit and probit models. When there are more than four alternatives, estimation of the logit model can be done using parametric methods such as the Maximum Likelihood method, whereas the probit model requires estimation methods which can be done non-parametrically. With these estimation methods at the focus of this study, I pose the following research question:

To what extent can a discrete choice model be accurately and efficiently estimated using sampling based approaches while incorporating heterogeneity?

To answer the research question, this paper replicates and then extends the work of Chintagunta and Honore (1996). To do this, the discrete choice models are estimated using Maximum Likelihood and the Method of Simulated Moments, as done in their paper. Then, I extend their research by evaluating the Gibbs Sampling approach as a potential solution that tackles some of the problems of the other methods. The Gibbs Sampling approach can be compared to the Method of Simulated Moments as they are both Monte Carlo sampling methods. Additionally, the incorporation of heterogeneity to allow for different preferences across individuals further extends the research of Chintagunta and Honore (1996).

2 Problem Description and Literature Review

To answer the research question, a heterogeneous discrete choice model is used. There are several problems that come into play regarding the decision of which model and variable specification to use. In this section I address the problems and provide the solutions that have been found in existing literature.

2.1 Independence of Irrelevant Alternatives

To tackle the estimation of the effect of the marketing variables on choice probabilities, many researchers employ a multinomial logit model. Though this model allows for the choice between alternatives to be analyzed, it depends strongly on the Independence of Irrelevant Alternatives (IIA) assumption. This assumption states that an individual's preference of one brand over another should not be affected by the presence of other, irrelevant brands. When the products from different brands are hard to differentiate, which is the case for substitutable goods, Currim (1982) suggests that this may not be the case. If the IIA does not hold, this may be problematic in estimating the effect of marketing variables using the logit model. Chintagunta and Honore (1996) state that this assumption causes the prediction of the joint probability of selection of the similar brands to be too high. Specifically, the IIA hinders the accurate estimation of the effect due to the overestimation of the probability. This assumption underlies not only the logit model, but all models that do not incorporate correlation between the alternative brands, such as the independent probit. To circumvent the IIA, Hausman and Wise (1978) suggest to use the multinomial probit model which allows for general similarities across the brands and therefore acknowledges the dependence between the alternatives. This dependence is incorporated in that the covariance structure of a probit model allows for correlation across brand alternatives.

2.2 Estimation of the Probit Model

Though the multinomial probit model circumvents the problem of the IIA, it comes with complications that are not characterized by the logit model. The problem of the probit model is that the estimation of its parameters are not as straightforward as in the logit model. For the evaluation of the discrete choice models, a distribution of the error term needs to be assumed. The probit model is dependent on the normal distribution, which is further explained in section 3, where the probit model is specified. The normal distribution is more complex to use than the logistic distribution for the logit model. This results in computational difficulties when estimating the parameters, as the integral is taken with respect to the number of alternatives. McFadden (1989) states that the direct numerical integration to estimate the parameters of a probit model is still possible when there are less or equal to 4 alternative brands, suggesting that the use of maximum likelihood is then still feasible.

The focus that I would like to place in this study is the use of sampling based approaches to estimate the parameters of a probit model. In sampling approaches, the probabilities are not calculated using the integral, but rather estimated using sampling methods. This solves the problem of the complicated integration. One such sampling method is the Method of Simulated

Moments, as explored by Chintagunta and Honore (1996). McFadden (1989) and Gourieroux and Monfort (1990) exemplified this method, and use it to estimate the coefficients of a probit model. A downside of simulation methods is that they can be time-consuming and if not enough replications are made then this may cause the results to be inaccurate.

McCulloch and Rossi (1994) specify a different method to use for the estimation, the Gibbs sampler. This is also recommended by Chintagunta and Honore (1996) as it is a more recent method. The Gibbs sampler aims to provide an approximation of the exact likelihood by means of sampling from conditional distributions. This is a less time-consuming method of sampling, though it still uses Monte-Carlo chains, as in Method of Simulated Moments. However, the Gibbs Sampler works more efficiently. If sampling from the conditional distributions is done sufficiently, McCulloch and Rossi (1994) state that the draws converge to stable estimates.

2.3 Individual-Specific Heterogeneity in Intrinsic Preferences

When considering the effect of marketing variables, not all individuals react the same. A feature of a brand in the newspaper for one individual may affect them more than when another individual is faced with the same feature. Similarly, when two individuals are confronted with a special display of a brand in the store, one may be more inclined to buy the product while the other remains indifferent between the alternatives. This reflects the intrinsic preferences of an individual, which can be interpreted as heterogeneity across the population. The concept of brand loyalty suggests that individuals may be inclined to buy the same products, regardless of the marketing of alternative brands. This difference in intrinsic brand preferences across individuals and households must be taken into account when estimating the effect on the probability of an individual buying one brand, as otherwise it can over- or under-estimated when considering different individuals.

The heterogeneity due to the differences in individuals is not implicitly considered in a normal multinomial probit model. This can be considered by a individual-specific variable, such as brand-loyalty, suggested by Guadagni and Little (1983). This variable takes into account the purchase history of an individual, which is updated every time a consumer buys a specific brand. The criticism that this variable has received is that the effect of the marketing variable on the purchase history is not filtered out when updating the brand-loyalty variable. What this suggests is that the promotion and price variables of the previous purchases affected those brand choices as well, and thus using the purchase history in the brand-loyalty variable will still contain the marketing mix effects. As a result, the intrinsic preferences of an individual towards one brand may be overstated by use of this variable.

Hausman and Wise (1978) suggest to use a random coefficients specification to take into account the heterogeneity, advocated by Heckman and Singer (1984). Such a specification assumes that the preferences are distributed across individuals following a specific distribution, such as the normal distribution. However, if the distribution is incorrectly specified then this yields biased results. Chintagunta et al. (1991) and Kamakura and Russell (1989) suggests the estimation and use of support points to estimate the distribution of the intrinsic preference.

The starting point is the general framework of a discrete choice model as introduced by McFadden (1989), which is specified to circumvent the IIA. Subsequently, the heterogeneous intrinsic preferences are incorporated.

3 Methodology

In section 3.1, the discrete choice model is explained. This general framework is then used to establish a multinomial logit, multinomial probit and independent probit model as different specifications of the discrete choice model. The multinomial logit model is estimated using Maximum Likelihood, and both the independent and multinomial probit model are estimated using the Method of Simulated Moments and the Gibbs Sampler, for which the methodology is presented in section 3.2 and 3.3, respectively. The incorporation of heterogeneity is explained in section 3.4.

3.1 Discrete Choice Model

For the general discrete choice model there are n individuals, K brands and T purchase occasions. Then, the utility U_{ikt} for individual i, where $i = 1, ..., n$, of buying brand k, where $k = 1, ..., K$ at time t, where $t = 1, ..., T$, is modeled as

$$U_{ikt} = X_{ikt}\beta_i + \alpha_{ik} + \epsilon_{ikt}. \tag{1}$$

X_{ikt} is a set of explanatory variables and β_i is a set of parameters. ϵ_{ikt} are the error terms. As can be seen, the explanatory variables are household and brand dependent, whereas the parameters only vary across households. α_{ik}'s are the intrinsic brand preferences of the households.

Chintagunta and Honore (1996) specify the identification of the model such that the interpretation of the intrinsic preference for one brand is presented as relative to a base brand. This identification provides the following final model specification:

$$\tilde{U}_{ikt} = (X_{ikt} - X_{iKt})\beta_i + (\alpha_{ik} - \alpha_{iK}) + (\epsilon_{ikt} - \epsilon_{iKt}), \tag{2}$$

which leads to:

$$\tilde{U}_{ikt} = \tilde{X}_{ikt}\beta_i + \tilde{\alpha}_{ik} + \tilde{\epsilon}_{ikt}, \tag{3a}$$

$$\tilde{U}_{iKt} = 0, \tag{3b}$$

where $\tilde{U}_{ikt}$ is the utility of brand k with respect to the base brand, K.

For simplicity, there are a few more vectors defined by Chintagunta and Honore (1996). First, μ_i is a vector of intrinsic preferences ($\tilde{\alpha}_i$) and slope coefficients($\tilde{\beta}_i$), $\mu_i = \{\tilde{\alpha}_i, \tilde{\beta}_i\}$. Additionally, the vector θ contains all parameters that must be estimated in the model. θ_0 then is the vector of true values of θ.

This set up then applies to the standard multinomial logit by assuming that the error terms, ϵ_{ikt}, of the latent variable follow a logistic distribution. For the probit model, the error terms are assumed to be normally distributed. This distribution in the error terms of the probit model allows for a correlation between the alternatives. The independent probit model is a special case of the multinomial probit model, as it assumes the same normal distribution of the error terms. However, the independent probit assumes that there is no correlation between the alternatives, which suggests that the off-diagonal values of the covariance are zero. This leads to the identity matrix being the covariance matrix.

3.2 Method of Simulated Moments

The use of the Method of Simulated Moments (MSM) is justified when there are more than four alternatives (McFadden, 1989). When this is the case, the integral that must be solved to find the estimates of the coefficients of the probit model is too complex, and thus methods such as Maximum Likelihood (ML) cannot be used. The MSM does not require the numerical integration, but rather uses simulation to provide an exact solution of an approximation of the likelihood.

To obtain the parameter estimates the moment conditions must be defined, as is done in the general method of moments. In the simple method of moments methodology, the aim is to obtain parameters for which the sample moments equal the true moment of the distribution of $D_i|X_i$:

$$E[D_i|X_i] = E_{\theta_0}[D_i|X_i], \tag{4}$$

where D_i is a vector of observed responses, such that D_{ikt} is 1 if individual i chooses alternative k at time t and 0 otherwise. Consequently, the moment condition that the method of moments uses is given by:

$$E[(D_i - P_\theta[D_i|X_i])W_i] = 0, \tag{5}$$

where $P_\theta[D_i|X_i]$ is the probability that individual i chooses alternative k at time t, dependent on the parameters, θ. This moment equation as specified by Chintagunta and Honore (1996) includes instruments, W_i, rather than only the explanatory variables, X_i. This is done due to the number of parameters to be estimated. Thus, solving the optimization problem is done using the following equation:

$$\theta_{mm} = \arg\min_\theta (D - P(\theta))' W' W (D - P(\theta)), \tag{6}$$

where $P(\theta) = P_\theta[D|X]$ for simplicity.

The MSM replaces the actual computation of probability or expectation, $P(\theta)$, with a simulator. This is done such that the expected value does not have to be calculated and thus there is no need to compute the complicated integral. The smooth unbiased estimators are random variables conditional on W_i, as exemplified by McFadden (1989). This suggests that it is calculated such that $\tilde{U}_{ik} = X_{ik}\tilde{\beta} + \epsilon_{ij}$ is computed using a fixed set of parameters, $\tilde{\beta}$. Then,

$$P(\theta) = \frac{1}{R}\sum_{r=1}^{R}(\tilde{U}_{i1} - \tilde{U}_{ik}) \tag{7}$$

is calculated. This provides enough information to calculate the moment condition, and thus to provide information on the the fit of the fixed $\tilde{\beta}$. This is then iteratively done until the $\tilde{\beta}$ gives the smallest moment condition.

3.3 Gibbs Sampling Approach

The Gibbs Sampling approach is a Bayesian method to sampling. In this paper I will not expand on the methodology to a great extent, but rather give an intuition of the method.

The specification of a prior, $p(\beta, \Sigma)$, is needed in order to do the Bayesian analysis of the multinomial probit model, as stated by McCulloch and Rossi (1994). The posterior density function is then obtained by:

$$p(\beta, \Sigma | y_1, ...y_N, X) \propto p(\beta, \Sigma) p(y_1, ...y_N, X | \beta, \Sigma), \tag{8}$$

where $p(y_1, ...y_N, X|\beta, \Sigma)$ is the model likelihood, $l(\beta, \Sigma)$. For the multinomial probit model, this is the product of the independent multinomial distributions. The direct evaluation of the likelihood of the normal distribution is costly and thus the Gibbs sampler is employed.

McCulloch and Rossi (1994) state that the Gibbs sampler does not sample from the joint distribution but rather from the conditional distributions. In practice, the Gibbs sampler does iterative sampling of one parameter from the conditional distribution of the parameters of the previous iteration. This process of repeated draws from the conditional distribution is done recursively, and eventually should converge in distribution to the joint distribution. In theory, this suggests that the method does not require the joint distribution, but relies fully on the conditional distributions in order to represent the full distribution.

3.4 Incorporating Unobserved Heterogeneity

The two methods to incorporate heterogeneity in the discrete choice models are explained in the following subsections.

3.4.1 Brand Loyalty

To account for unobserved heterogeneity in the models, the intrinsic brand preferences must be incorporated into the model. Guadagni and Little (1983) specify one method to do this, by the use of a $brandloyalty$ variable to represent these the household-specific differences. This variable is constructed using the purchase history of a household and updating this loyalty of a brand per purchase. It is calculated as follows:

$$Loy_{ikt} = \gamma Loy_{ik(t-1)} + (1-\gamma) D_{ik(t-1)}, \tag{9}$$

where Loy_{ikt} is the loyalty of individual i of brand k at purchase t, and $D_{ikt} = 1$ if individual i bought brand k at purchase t and 0 otherwise. γ is the carry-over constant. The use of brand loyalty accounts for heterogeneity in the logit model by incorporating purchase history.

3.4.2 Random Effects Model

Papers by Chintagunta et al. (1991) and Chintagunta and Honore (1996) state that the intrinsic preferences of individuals follow a distribution over the population. This suggests that if this distribution is known, the parameters can be estimated and the heterogeneity can be taken into account, such that the model becomes:

$$U_{ij} = \bar{U}(X_{ij}, a_i) + \epsilon(X_{ij}, a_i) = \bar{\beta} ij X_{ij} + \beta_{ij} X_{ij} + \gamma_{ij}. \tag{10}$$

As suggested by Hausman and Wise (1978), β are the taste parameters, and γ represents the truly random term.

There are multiple ways of defining the heterogeneity distribution. Chintagunta and Honore (1996) suggest the use of support points to estimate the underlying distribution. Doing this would allow for an accurate representation of the heterogeneity of the population. It is incorporated such that there are a finite amount of support points which each have a probability mass, a mean and standard deviation. This discrete distribution is estimated for the sake of interpretation, but is not used for the estimation of the model due to complexity in programming.

For this research, it is assumed that the heterogeneity is normally distributed over the population, and it is incorporated accordingly into the logit model using the Maximum Likelihood method of estimation. Additionally, I incorporate heterogeneity by means of the random-effects probit model using the Method of Simulated Moments to estimate the parameters.

4 Data

To implement the methodology, scanner data is needed on purchase behavior of households buying alternative brands that produce substitutable goods. For each purchase occasion, which is when an individual purchases one of the alternative brands, information that is brand specific but also household specific is optimal to have. Additional to this, it is preferable that the data is clean.

The data that is used in this paper is scanner information on four Saltine Cracker brands that are sold in the US. The names of these brands are Nabisco, Sunshine, Keebler and a Private Label and they provide a substitutable product. The variables that are included in the discrete choice models are price, display and feature. Price of each alternative of Nabisco, Keebler, Sunshine and the private label are reported for each purchase occasion. Price is a continuous variable. The variables display and feature are dummy variables. For display, $display_{ijt}$ presents whether there was a display for household i for brand j at purchase occasion t. For $feature_{ijt}$ it has the same set-up, representing whether the brand was featured in a newspaper at time of purchase. Table 1 presents the summary table of the variables for each brand.

	Share	Price	Display	Feat
Sunshine	0.073	0.957	0.129	0.038
Keebler	0.069	1.126	0.106	0.043
Nabisco	0.544	1.079	0.340	0.087
Private	0.314	0.681	0.099	0.047

Table 1: Summary of Saltine Crackers Data, where the share of the amount of time that the brand was purchased is presented in the first column. The average price, amount of time that the brand was on display or had a feature are presented respectively.

This table clearly shows that Nabisco is the most commonly bought brand, with the private label following. Sunshine and Keebler appear to have a less large market share, being bought only on about 7% of the occasions. The highest priced brand is Keebler, followed by Nabisco. The lowest priced brand is the private label. Nabisco was both displayed and featured most frequently throughout time, suggesting a prevalent marketing strategy. Contrasting to this is the private label, with a low average price and fewer marketing activities. This suggests that the private label is a cheap variant, such as a house-brand.

5 Results

The estimation of the multinomial logit, multinomial probit and independent probit models are chosen with the following reasoning. The multinomial logit and independent probit model are compared to show the effect of the method of estimation of the probit model, and each method's ability to accurately estimate the parameters of the model. Both of these models suffer from the Independence of Irrelevant Alternatives (IIA) assumption. The multinomial probit model can be compared with the independent probit model to show the influence that the correlations between the brands have on the estimates.

The results are given in the following order. The results of the replication are presented first. The multinomial logit model, independent probit model and the multinomial probit model are estimated using both the Maximum Likelihood (ML) approach and the Method of Simulated Moments (MSM). Consequently, the estimates of these results are analyzed and compared. Hereafter, the results of the Gibbs Estimation Approach are given, alongside a comparison of its performance compared to the MSM. Last, the incorporation of heterogeneity by means of random-effects is included, presented by the mixed logit model and the random-coefficient probit model.

These ML results are produced using the package mlogit by Croissant (2018). For the MSM, there are no available packages or code to use to estimate the estimates. Thus, the largest bulk of time for this study was spent on exploring this method, finding an algorithm and writing the code to run the Method of Simulated Moments. The R code for the MSM that is used in this paper is presented in Appendix A. To estimate the results for the Gibbs Sampling, the package bayesm by Rossi (2017) is used. For the mixed logit model, the gmnl package by Sarrias and Daziano (2017) is used. The random-coefficient probit model required an adaptation of the MSM code. This adaptation is presented in Appendix B.

The results for the elasticities models are calculated using the code presented in Appendix C. Due to the complexity of the multinomial probit model, the elasticities are calculated by scaling the estimates of the probit models to fit the logit model. Though this is not entirely correct, it allows for the interpretation of the estimates, and thus it is useful.

Table 2: Parameters Estimates

Parameters estimates:	Multinomial Logit (ML)	Multinomial Probit (ML)	Independent Probit (MSM)	Multinomial Probit (MSM)	Multinomial Probit (MSM: 50 Replications)	Multinomial Probit (Gibbs)
Keebler:(intercept)	0.618	0.451	0.62	0.619	-0.624	-2.072
	(0.127)	(0.098)				(0.541)
Nabisco:(intercept)	1.137	0.616	1.309	1.137	2.446	-3.004
	(0.106)	(0.079)				(0.486)
Private:(intercept)	-0.434	-0.143	-0.434	-0.43	0.647	0.447
	(0.119)	(0.084)				(0.067)
Price	-3.777	-1.823	-3.77	-3.778	-3.36	0.094
	(0.299)	(0.211)				(0.086)
Display	0.299	0.158	0.423	0.303	0.042	-0.069
	(0.091)	(0.045)				(0.033)
Feature	0.737	0.385	0.737	0.741	0.489	-0.132
	(0.134)	(0.073)				(0.048)
Loyalty	3.534	1.973	3.516	3.533		-1.126
	(0.082)	(0.176)				(0.151)

Parameters estimates and their standard errors (presented in the parenthesis below the estimates) for the models that account for heterogeneity through the brand loyalty variable. The standard errors for the MSM are not calculated in the code, thus also not presented here.

5.1 Maximum Likelihood

McFadden (1989) states that the direct numerical integration to estimate the parameters of a probit model is still possible when there are less or equal to 4 alternative brands. For the data of the Saltine Crackers, there are only four brands. This suggests that it is possible to estimate the probit model using Maximum Likelihood. Thus, I report the results of the probit model using Maximum Likelihood (ML) in this section for comparison to the simulation methods. These results are presented in the first two columns of Table 2.

The results in Table 2 show that the signs for each coefficient are the same in both the multinomial logit and probit model. This is quite a crucial observation, as it confirms that the marketing variables have a positive influence on the utility of each brand. If the signs of the marketing variables were negative, it could suggest a problem in the model or a bias in the estimates. This statement is under the presumption that marketing activities have either no effect or a positive effect on the probability that a consumer buys a product. Though it can be seen that the estimates for the multinomial logit are approximately double the size of those of the probit model, the estimates cannot be interpreted directly with respect to the probabilities. Thus, the marginal effects are analyzed in the following section.

5.1.1 Choice-Price Elasticities

The marginal effects of price are mentioned in the first and third columns of Table 5, where they show each marginal effect of price of one brand on the probability of each alternative being chosen. These marginal effects I refer to as choice-price elasticities. The results for the estimates using ML in the choice-price elasticities table shows that an increase in price of each brand results in a decreased probability of an individual purchasing that brand. This is known as the demand elasticity of a brand. Intuitively, this is straight forward as individuals are less likely to buy brands with increased prices, which is a law in economics.

The marginal effects of the other prices can be interpreted as cross-price elasticity and this is an interesting finding, as it allows for interpretation as to how a price of one brand can effect the purchase probability of another brand. This concept is especially useful for brands to understand the competing brands that they are most price-competitive with. Additionally, this gives insight into how the brands interact.

Analyzing these choice-price elasticities for the models estimated using ML, it can be seen that the change in price of Sunshine specifically affects the probability that individuals purchase Nabisco. This suggests that they are seen as very similar alternatives for households, as a slight price in one can encourage individuals to switch to the other. Additionally, the private label appears to increase most in purchase probability if there is a Nabisco price increase.

5.2 Method of Simulated Moments

The results obtained by estimating the parameters of independent and multinomial probit model using Method of Simulated Moments (MSM) are presented in the third and fourth columns of Table 2 respectively, where the amount of random draws is set to 10.

The results in this table show that the estimates obtained using MSM are different in size from the results obtained by Chintagunta and Honore (1996). It should be noted that the sign of each estimate is the same as presented in their paper. This difference in size of estimates may be due to the way that the results were obtained back in 1996, which may have been less accurate or done differently. Additionally, the method of Simulated Moments never produces the same estimates, thus this may just be due to the simulation.

To compare for credibility of the MSM, I compare the results to those found by the ML method. Comparing these two estimates, there is quite a large difference. The signs of each estimate is the same. However, the estimates produced by ML are approximately double the size of the estimates that are the result of the MSM. This may suggest that the MSM is not a very credible method. The results are relatively similar to the multinomial logit model. This may be due to the fact that the initial estimates for the parameters provided to the optimization program in R are those of the multinomial logit. As the optimization method searches for points close to this starting value, it can explain why these results are so similar.

Comparing the independent probit to the multinomial probit model again shows extremely similar results. Some results only differ by one hundredth. This may be due to the small amount of random draws taken, which would suggest that the correlation between the variables was not taken into account enough through the simulation procedure and thus the difference is not noticeable. The effect of IIA can therefore not be extracted from these results.

To evaluate the method, it must be noted how slow the program runs. To obtain estimates with a replication set at 10, the running time was more than 7 hours. Any increase in R resulted in even longer waiting times. As 10 is not a large amount of randomly drawn error vectors, the results may not be reliable. This trade-off between time and accuracy is a large drawback of the use of MSM, and a critical note must be placed on it.

An additional limitation mentioned by Chintagunta and Honore (1996) is that the usage of the exponential distribution for smoothing can cause bias in the estimates. This is not the case for the estimates presented here.

5.2.1 Sensitivity Analysis: 50 Random Draws

In an attempt to produce more accurate results, the MSM with $R = 50$ results are also reported. The loyalty variable is removed as otherwise the method takes too long. The program required more than 12 hours to produce output. The results are presented in the fifth column of Table 2.

These results appear different than those with $R = 10$ suggesting that an increased number of random draws may change the parameters significantly (though this cannot be said as a fact as the standard errors are not reported in this method). Such differences can be seen in, for example, the brand-specific intercept of Nabisco, that is 1.137 with $R = 10$ and increases to 2.446 when $R = 50$. Such a change in coefficients due to a change in R questions the reliability and robustness of the MSM as a method for estimating the probit model. With an R too low, the estimates appear biased and inaccurate, however increasing the amount of random draws will produce a running time that is even longer. Due to time constraints the results could not be done with a larger amount of random draws, though it is not infeasible.

5.2.2 Choice-Price Elasticities

The choice-price elasticities for the probit models estimated using MSM are presented in the second and fourth columns of Table 5. Comparing the results obtained using MSM for the multinomial probit model with those obtained with ML show a similarity. The estimates appear to not differ largely from each other, providing the same conclusions about cross-price elasticities as mentioned previously. The cross-price elasticities of Nabisco and the private label remain largest, while all brands are negatively impacted by their own price increases.

When comparing the independent probit model and the multinomial logit model, one can make inferences about the effect of the IIA on estimates of the model. It can be seen that the price-choice elasticities are much larger in the multinomial logit model than in those of the independent probit model, such as the price effects of sunshine on its own probability are -0.219 and -0.089, respectively. This effect may be due to the violation of the IIA.

5.3 Gibbs Sampling Approach

The Gibbs sampling method is a way to produce estimates for the probit model in a much faster way, thus allowing for more accurate estimates. The results using this approach are presented in the last column of Table 2.

The estimates returned by the Gibbs sampler are significantly different from those obtained by both the ML estimation and the MSM estimates. This difference is specifically noticeable when considering the price coefficient. Those of MSM and ML both provided estimates ranging between -1.7 and -4, which is contrary to the estimate for price of the Gibbs sampler, which is around 0.094. A positive influence of price on the utility of a brand appears inconsistent with the results obtained in this study.

The results provided for the choice-price elasticities, as shown in the fifth column of Table 5, are equally as illogical. This provides positive price-elasticities for some brands, and negative cross-price elasticities throughout the analysis.

This difference in estimates and cross-price elasticities can be explained by a number of reasons. As the simulation draws from conditional distributions, there is a chance that these may be wrongly defined. Additionally, if not enough draws are performed, the results do not converge to a stable estimate, resulting in biased and inaccurate estimates. For the results presented, the amount of draws is set to 1000. Results for $R = 10000$ provide estimates of coefficients larger than 10. It may be explained by the idea that the chain did not converge within 10000 draws. This behavior can be avoided by applying a burn-in phase of 1000 draws. These results, however, appeared similar to the previous ones. My knowledge of Bayesian econometrics is, however, limited and thus I cannot further explain this.

5.4 Incorporating Unobserved Heterogeneity: Random Effects

As stated previously, the methodology for using the brand-loyalty variable is flawed due to the use of purchase history. The random-effects methodology does not depend on this but rather depends on a specified distribution of heterogeneity over the population. In this section the discrete support points are analyzed and the random-effects logit and probit model results are presented.

5.4.1 Heterogeneity Distribution

The implementation of using a discrete heterogeneity distribution for the random-effects model is too complex for the scope of this research. However, the support points for a discrete heterogeneity distribution are interesting to observe and can provide a valuable interpretation. The distribution is presented in Table 3.

	Probability	Keebler:(intercept)	Nabisco:(intercept)	Private:(intercept)	Price
Point 1	0.675	0.549	3.075	-2.062	-7.147
		(0.853)	(2.621)	(2.810)	(0.456)
Point 2	0.325	0.098	1.470	1.749	-0.389
		(3.709)	(3.472)	(3.053)	(0.774)

Table 3: Heterogeneity Distribution with Two Support Points. This shows that for point 1, which is reached with a probability of 0.675, the estimates for the price and brand-specific intercepts are presented horizontally next to it, with their standard deviations reported in the parenthesis below. The results for the second support point are below.

The interpretation of these results is done as in Kamakura and Russell (1989), who presented this as a method for market segmentation. The table shows that there is a 67.5% chance of an individual to be at point 1, where the price sensitivity is much higher than at point 2. Due to the model specification, it is difficult to make any conclusions about the intercepts, especially considering the large standard errors.

The log-likelihood of using 2 discrete support points is -1567, which is better than using 1 support point. Increasing the amount of points did not yield much better results, with 3 points having a likelihood of -1553.7 and 4 points being -1555. Thus, using 2 points appears appropriate, confirming results found by Chintagunta and Honore (1996).

	Keebler:(intercept)	Nabisco:(intercept)	Private:(intercept)	Price	Display	Feature
			Mixed Logit Model			
Estimate	0.506	3.471	-0.871	-4.771	0.353	0.719
	(0.155)	(0.135)	(0.146)	(0.367)	(0.095)	(0.152)
Standard Dev.	1.668	2.893	3.376	5.530		
	(0.134)	(0.120)	(0.141)	(0.342)		
			Random Effects Probit Model			
Estimate	0.509	2.295	0.7651	-3.191	0.092	0.496

Table 4: Estimation Results Mixed Logit model and random-effects Probit Model. For the mixed logit, the mean values for the estimates of the price and the brand-specific coefficients are presented, with the standard deviations of these random variables presented below these. The display and feature variables were not considered random, and thus their standard deviations are not reported. For the random-coefficient probit model the no standard deviations nor standard errors were recorded.

5.4.2 Random Coefficients Logit Model (Mixed Logit)

The coefficients for the logit model that incorporates heterogeneity of intrinsic preferences by means of a random effects method are presented in the top part of Table 4. The variables considered to be random parameters are price and the brand-specific constants, as display and feature are not of interest. This idea is based on Chintagunta and Honore (1996).

The results for the mixed logit model are interesting to compare to those of the multinomial logit model. The logit model does not account for heterogeneity, and thus we can observe the effect of heterogeneity on the estimates of this mixed logit.

This table shows that increasing prices maintains having a negative effect on the utility of a brand. In the multinomial logit, this effect appears to be smaller than in the mixed logit model. For the other estimates, the results appear similar to those of the multinomial logit. The display and feature variables maintain a positive influence on the utility of a brand. The standard deviations of each of the random variables are also interesting to observe. These show that the private label has a much larger standard deviation for its intercept, which suggests that individuals have clear preferences that are distinct for this private brand. For the price, a standard deviation of 5.5 is very large, which may suggest that individuals could have a positive relation between price and utility.

Comparing the choice-price elasticities of the mixed logit model, as presented in the last column of Table 5, with those of the logit model, it can be seen that the elasticities are much larger. Significantly for the price-elasticity of the brands themselves, some elasticities are almost 5 times larger in the mixed logit than in the multinomial logit, as shown for the brand sunshine increasing from -0.219 to -1.019 respectively. A noticeable distinction as well is that in the mixed logit model, the cross-price elasticity of Nabisco on Keebler is the largest, rather than that of Nabisco on the private label. This change may be due to the incorporation of heterogeneity.

5.4.3 Random Coefficient Probit Model

The random coefficient probit model, as previously explained, uses the Method of Simulated Moments to estimate the probit model, while accounting for heterogeneity by assuming a normal distribution of preference across the population by means of random coefficients. To incorporate the random effects into a probit model, the code for the MSM for the probit model is adapted. The changes in the code are presented in the appendix, as mentioned previously.

The results of the random-effects probit model are presented in the bottom part of Table 4. These results can be compared to the probit model, allowing to extract the effect of heterogeneity. Additionally, these results can also be compared to those of the mixed logit model as they both use the random-effects methodology. The mixed logit model uses the same methodology for random-effects and thus it is interesting to understand what effect heterogeneity has on the coefficients and the utility.

The coefficients for the estimates of the random-effects probit model differ largely from those obtained in the multinomial probit model that excludes the brand-loyalty variable. This suggests that the effect of heterogeneity is not prominent. Comparing the results to the multinomial probit model that does include the brand-loyalty variable, the estimates do differ, however not consistently. This suggests that there is an effect of heterogeneity. Conclusions cannot be drawn with certainty as the MSM method has not shown consistent results, as previously mentioned. The coefficients of the random-effects probit model and mixed logit model appear similar, with some minor differences in the display and private intercept parameter estimates.

Observing the choice-price elasticities, as presented in the second last column of Table 5, they are all of logical sign. The values of the elasticities are consistently smaller than the mixed logit elasticities. This is a difference that may be present due to the IIA, showing a promising improvement of the probit model. For the cross-price elasticities, the results again show the strong interaction between Nabisco and the private label, as previously found in the multinomial logit models. However, the own price elasticities are shown to be higher for Sunshine, Keebler and the private label, but lower for Nabisco. This shows that the heterogeneity in brand-preferences is removed, allowing for more accurate estimates.

Table 5: Choice-Price Elasticities

Change in Probability of Alternatives:	Multinomial Logit (ML)	Independent Probit (MSM)	Multinomial Probit (ML)	Multinomial Probit (MSM)	Multinomial Probit (Gibbs)	Heterogeneous Multinomial Probit (MSM)	Mixed Logit (SML)
Increase in Price of Sunshine							
Sunshine	-0.219	-0.089	-0.195	-0.114	0.037	-0.503	-1.019
Keebler	0.014	0.001	0.018	0.002	-0.001	0.064	0.130
Nabisco	0.149	0.076	0.123	0.092	0.000	0.151	0.815
Private	0.057	0.012	0.053	0.020	-0.036	0.289	0.075
Increase in Price of Keebler							
Sunshine	0.014	0.001	0.018	0.002	-0.001	0.064	0.130
Keebler	-0.207	-0.081	-0.232	-0.103	0.002	-0.508	-0.728
Nabisco	0.139	0.069	0.149	0.083	0.000	0.152	0.548
Private	0.054	0.011	0.064	0.018	-0.001	0.292	0.050
Increase in Price of Nabisco							
Sunshine	0.149	0.076	0.123	0.092	0.000	0.151	0.815
Keebler	0.139	0.069	0.149	0.083	0.000	0.152	0.548
Nabisco	-0.875	-0.816	-0.705	-1.002	0.000	-0.994	-1.679
Private	0.588	0.672	0.433	0.827	0.000	0.691	0.316
Increase in Price of Private label							
Sunshine	0.057	0.012	0.053	0.020	-0.036	0.289	0.075
Keebler	0.054	0.011	0.064	0.018	-0.001	0.292	0.050
Nabisco	0.588	0.672	0.433	0.827	0.000	0.691	0.316
Private	-0.699	-0.695	-0.551	-0.866	0.037	-1.272	-0.441

Choice Price Elasticities: presented for the different models that are studied in this paper. These elasticities show the change in the probability that an individual purchases each brand due to a change in price of each of the brands.

6 Discussion

The theory depicts a clear preference towards the probit model for the estimation of the effect of marketing variables on the probability of individuals buying specific brands. The Independence of Irrelevant Alternatives assumption is clearly broken due to the correlation between the variables. When analyzing the results found by the Method of Simulated Moments for the estimation of the coefficients, it can be seen that the estimates are quite similar. However, comparing their elasticities, it is shown that there is a strong violation of IIA assumption, as these differ largely. When considering the results from the simulation with 50 draws, the results appear more different, suggesting that the probit model is more fitting to analyze substitutable alternatives. This also suggests that running the model with more replications will further improve the coefficients, and allow for the accurate estimation of the coefficients of the marketing-mix variables. In turn, this will allow companies to make inferences on the effects of their marketing activities on the probability of an individual purchasing their brand.

For the Method of Simulated Moments, the discussion can be extensive. The method is slow, time-consuming and tedious to work with. Increasing the amount of replications increases the time that the program needs to run significantly. However, only with a higher amount of replications does the MSM appear more accurate and reliable. This trade-off for the MSM method provides a strong negative aspect to the method. For further research, the code that I wrote for the MSM program can be revised and improved. I suspect that one of the reasons for the extensive running-time is that the code is not written efficiently. Due to time-constraints, no improvements could be made during the duration of this study.

The Gibbs estimator, though more complex to understand, allows for the estimation of the parameters to be less time-consuming. The sampling from conditional distributions is a more efficient solution, which is the reason that the Gibbs sampler is a faster approach. However, from the estimation done in this paper, it appears that the coefficients are not very reliable and depend largely on the amount of random draws taken, and the amount of draws in the burn-out phase. Even after removal of the burn-in phase, the results did not improve. Due to my limited Bayesian econometrics knowledge, I could not improve this estimation further. However, this is an interesting method to delve deeper into.

Lastly, I evaluate the incorporation of the heterogeneity in the data. The use of the brand loyalty variable, as theoretically shown, depends largely on the purchase history which may already be affected by the marketing activities. Assuming that the heterogeneity is distributed over the population following a normal distribution, and incorporating this by means of a random-effects model is a solution that removes the purchase history dependency. However, if the heterogeneity does not follow the normal distribution over the population, then the results found in this study are biased due to the assumption of normality. A discrete support-point heterogeneity distribution would then be a better distribution to assume. However, in this study this is not done as it was too complex to incorporate. This suggestion by Chintagunta and Honore (1996) is intuitively interesting, and should be evaluated in further research.

To conclude, Chintagunta and Honore (1996) provide an interesting research, going into depth of the Method of Simulated Moments and the discrete support-point heterogeneity distribution. Their contribution is interesting, and worth noting. The comparison to the Gibbs sampler, and the incorporation of a normal-distributed heterogeneity is an addition to their study. Time should be further spent in perfecting the estimation methods of the probit model, as this provides interesting information to both companies and researchers alike.

References

Chintagunta, P. K. and Honore, B. E. (1996). Investigating the effects of marketing variables and unobserved heterogeneity in a multinomial probit model. *International Journal of Research in Marketing*, 13(1):1–15.

Chintagunta, P. K., Jain, D. C., and Vilcassim, N. J. (1991). Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research*, 28(4):417.

Croissant, Y. (2018). *mlogit: Multinomial Logit Models*. R package version 0.3-0.

Currim, I. S. (1982). Predictive testing of consumer choice models not subject to independence of irrelevant alternatives. *Journal of Marketing Research*, 19(2):208.

Gourieroux and Monfort (1990). Simulation based inference in models with heterogeneity. *Annales d'Économie et de Statistique*, (20/21):69.

Guadagni, P. M. and Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(3):203–238.

Hausman, J. A. and Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46(2):403.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271.

Kamakura, W. A. and Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26(4):379.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026.

Rossi, P. (2017). *bayesm: Bayesian Inference for Marketing/Micro-Econometrics*. R package version 3.1-0.1.

Sarrias, M. and Daziano, R. (2017). Multinomial logit models with continuous and discrete individual heterogeneity in R: The gmnl package. *Journal of Statistical Software*, 79(2):1–46.

A Appendix: MSM Code

```
#Generate R draws of e for each i (Independent Probit)
error_vector <- replicate(R, rnorm(K,0,1))
#Generate R draws of e for each i (Multinomial Probit)
# error_vector <- replicate(R, rnorm(K,0,1))
# covar_error_vector_7 <- matrix(c(1,0.5,0.5,0.5,0.5,0.5,0.5
#                                   ,0.5,1,0.5,0.5,0.5,0.5,0.5
#                                   ,0.5,0.5,1,0.5,0.5,0.5,0.5
#                                   ,0.5,0.5,0.5,1,0.5,0.5,0.5
#                                   ,0.5,0.5,0.5,0.5,1,0.5,0.5
#                                   ,0.5,0.5,0.5,0.5,0.5,1,0.5
#                                   ,0.5,0.5,0.5,0.5,0.5,0.5,1),
#                                nrow=7,ncol = 7)
# error_vector <- mvrnorm(6, mu = c(0,0,0,0,0,0,0) , Sigma = covar_error_vector_7)

#Function to calculate the moment condition values
msm.objective <- function(Y, modelmatrix, param, R, K, J, error_vector, identity_
    matrix){
  #Obtain eta
  probability <- NULL
  probability.derivative <- NULL
  #For loop starts here
  for (i in seq(from = 1, to=N, by=4))
  {
    eta.i <- data.matrix(modelmatrix[i:(i+J-1),]) %*% data.matrix(identity_matrix)
        %*% data.matrix(error_vector)
    #holding beta fixed, generate utility
    utility.i <- replicate(R, as.data.frame(data.matrix(modelmatrix[i:(i+J-1),])%*
        %data.matrix(param))) + as.data.frame(eta.i)
    utility.ih <- replicate(R, as.data.frame(data.matrix(modelmatrix[i:(i+J-1),])%
        *%data.matrix(param-h))) + as.data.frame(eta.i)
    prob.ih <- utility.ih
    prob.i <- utility.i
    #find relative frequency
    for (m in 1:(J-1)) {
      prob.i[m,] <- utility.i[J,] - utility.i[m,]
      prob.ih[m,] <- utility.ih[J,] - utility.ih[m,]
    }
    probability.i <- rowSums(prob.i)/R
    probability.ih <- rowSums(prob.ih)/R
    probability <- rbind(probability,probability.i)
    probability.derivative.i <- (probability.ih - probability.i)/2*h
```

```
    probability.derivative <- rbind(probability.derivative, probability.derivative.i)
  }
  moment <- sum(rowSums((Y-probability)*(probability.derivative/ probability)))
  return(moment)
}
#Running the optimization function
results <- optim(par = beta.0, fn = msm.objective, method = "BFGS",Y = Y, modelmatrix = modelmatrix, R=R, K=K,
                    J=J, error_vector=error_vector,identity_matrix = identity_matrix)
```

B Appendix: MSM Adapted Code

```
#Generating heterogeneity error term
heterogeneity_error <- mvrnorm(N/4, mu = c(0,0,0,0), Sigma = covar_heterogeneity_error)
heterogeneity_error <- as.vector(heterogeneity_error)
modelmatrix <- cbind(modelmatrix,heterogeneity_error)
#specify starting point
beta.0 <- append(summary(logitMSM)$coefficients,0)
param <- beta.0
```

C Appendix: Calculation of Elasticities

```
AverageValues <- with(crackersmdf, data.frame(price = tapply(price, index(logitCracker)$alt, mean),
                                        disp = tapply(disp, index(logitCracker)$alt, mean),
                                        feat = tapply(feat, index(logitCracker)$alt, mean),
                                        loyalty = tapply(loyalty, index(logitCracker)$alt, mean)))

#Coefficients for the logit model
coef <- coefficients(logitCracker)

#variables
c <- 6

#Coefficients for the probit model
```

```
coefprobit <- resultsHeterogeneityProbit$par[1:6]
scaling <- 4/sqrt(2*pi)
coef <- coefprobit * scaling

#Calculating the probabilities of each choice
exp.sunshine <- exp(sum(coef[c(4:c)]*AverageValues[1,]))
exp.keebler <- exp(coef[1] + sum(coef[c(4:c)]*AverageValues[2,]))
exp.nabisco <- exp(coef[2] + sum(coef[c(4:c)]*AverageValues[3,]))
exp.private <- exp(coef[3] + sum(coef[c(4:c)]*AverageValues[4,]))
exp.sum <- exp.sunshine + exp.keebler + exp.nabisco + exp.private
probability.sunshine <- exp.sunshine / exp.sum
probability.keebler <- exp.keebler / exp.sum
probability.nabisco <- exp.nabisco / exp.sum
probability.private <- exp.private / exp.sum
probs <- rbind(probability.sunshine,probability.keebler,probability.nabisco,
    probability.private)

#Calculating the elasticities
elasticities <- coef[4] * probs * (1-probs)
crossprice <- -coef[4]* data.matrix(probs)%*%t(data.matrix(probs))
diag(crossprice) <- elasticities
```