A Satisficing Perspective of MNL Choice Model

Shasha Han* Jussi Keppo† Melvyn Sim‡

July 22 2015

Abstract

The multinomial logit model (MNL) model has traditionally been derived from the theory of random utility maximizing choice with an assumption of i.i.d Gumbel noise. While random utility maximizing choice theory is effective in handling static choice sets, it is not applicable to dynamic choice sets, i.e., the actual set is a subset rather than the complete set. We propose a MNL model to estimate and predict individuals decision making with dynamic choice sets from the satisficing perspective. We first put forward a formula to model satisficing choice based on two factors: acceptance set and exposure intensity. Then, we utilise the logs transformation of acceptance set to get the Satisficing MNL model while maintaining the maximum likelihood optimization problem tractable. Based on real market data, we compare the validity of Satisficing MNL (S-MNL) and Utility MNL (U-MNL) in estimating in-sample and predicting out-of-sample. We find that in the non-brand specific case, the S-MNL model has better fitness and prediction than the U-MNL model. Also, the S-MNL model maintains the superior prediction ability in the brand specific case. Although the fitness of the two models become similar in the latter case, the U-MNL model is mistakenly estimated because the effect of exposure in the S-MNL model is combined into the attribute preference in the U-MNL model. Our study is among the first attempts to focus on decision making under dynamic choice sets and quantitatively estimate the satisficing choice. Furthermore, our study strengthens the importance of both attribute and exposure and enables firms to make effective marketing decisions based on these two factors.

*NUS Business School. Email: shashahan@u.nus.edu

†NUS Business School, NUS Risk Management Institute. Email: keppo@nus.edu.sg

‡NUS Business School, Singapore-MIT Alliance, NUS Risk Management Institute, National University of Singapore. Email: dscsimm@nus.edu.sg.

1 Introduction

In both economics and psychology, it has been noted that limited attention or limited cognition plays a significant role in people's choice (e.g. Broadbent 1958, Simon 1959, Kahneman 1973). Simon (1959) pointed out that under simple situations where choice comparisons are easily made, a decision maker behaves like a utility maximizer; but as the choice situation becomes more complicated, one becomes much less consistent. From utility maximising perspective, this inconsistence choices under complex situations are resulted from bounded rationality, which constraints decision makers maximising behaviour (Gul et al. 2014). However, as St*ü*ttgen et al. (2012) suggested, constrained utility maximizing model failed to explain the search behaviour pattern which appeared in their visual conjoint experiment. As a result, St*ü*ttgen et al. (2012) proposed a satisficing choice model for conjoint analysis based on Simon's (1955, 1959) satisficing theory.

In contrast to utility maximizing theory, which assumes a static choice set, satisficing rule allows a dynamic choice set (Simon 1955). Individuals evaluate choice alternatives sequentially and make choice decisions before searching all the alternatives in the choice set. Thus, if there is more than one satisfactory object, a decision maker chooses the one which first satisfies him/her. Obviously, the sequence of choice assessment affects the choice outcome. Further, due to this sequential evaluation, it is difficult to empirically investigate the satisficing choice. Recently, St*ü*ttgen et al. (2012) designed a conjoint experiment to examine individual's search patterns using eye tracking technique. However, this technique is not applicable to real market data.

We address the dynamic choice set from satisficing perspecitve by viewing decision maker's sequential evaluation as *stimulus-directed* and *goal-driven behaviour*. Specifically, we study the discrete choice behavior of a decision maker who can be either intentionally or unintentionally guided by the exposure of alternatives, as in Zajonc (1968), Janiszewski (1993) and Zajonc (2001). From the satisficing perspective, the exposure of an alternative increases the chances that it is seen at first sight, and in turn, increases its choice probability if it is acceptable. The default bias (Kahneman 1991), where a default one is usually opted for even when choosing an alternative is effortless is an illustration of the exposure effort.

In this paper, we propose a satisficing choice model based on satisificing theory. We first introduce *acceptance set* and *exposure intensity* and then propose the satisficing multi nominal logit (S-MNL) model in particular. The multinomial logit model (MNL) model has traditionally been derived from the theory of random utility maximizing choice with an assumption of i.i.d Gumbel noise by Marley (see Luce and Suppes 1965). We show the connection between the S-MNL model and the utility mulit nominal logit (U-MNL) model. Finally, we derive the corresponding tractable robust maximum log-likelihood problems, which are estimated in an optical scanner panel dataset on purchases of saltine crackers. Our estimation results show that the S-MNL model fits and predicts data significantly better than the U-MNL model in the non-brand specific case. Although the two models result in the similar fitness in the brand specific case, the U-MNL model is mistakenly estimated because exposure effect in the S-MNL model is *distorted* into attribute coefficient estimates. Nevertheless, the S-MNL still has significantly superior prediction than the U-MNL in the brand specific case.

The practical implications of S-MNL model are highly encouraging. After splitting the exposure effect and product attributes, it could help a firm to detect the true causes of its market share. The market failure of a product may be accounted by lack of the marketing related exposure activities or unpopular product attributes. Knowing the determinant causes of market failure, a firm is able to make wiser operational and marketing strategies.

The reminder of this paper is organized as follows. We first briefly review the relevant literatures in Section 2. In Section 3, we give the foundation of choice rule from satisficing perspective and then introduces the S-MNL model and the corresponding maximum likelihood problem. In Section 4, we estimate and compare the validity of the S-MNL and U-MNL models based on our dataset. In section 5, we extend the basic model to the case of piecewise approximation of attribute's value. Finally, we conclude with a general discussion.

2 Literature Review

We first briefly introduce the literatures on the satisficing choice. Then we review the traditional utility approach to multinominal logit (MNL) choice model. Furthermore, given the incomplete choice set of satisficing choice, which is related to consideration set and rational inattention, we review the relevant work on constrained utility optimization. Finally, because we use the ideal of *exposure*, we review evidences of the exposure effect in psychology field.

2.1 Satisficing Choice

Simon (1955, 1959) introduced the idea of satisficing decision makers "whose problem solving is based on search activity to meet certain aspiration levels". The merit of satisficing lies in a decision maker's awareness of the limited cognitive and computational abilities and his/her activities to simplify the choice process.

One important feature of satisficing choice is the non-compensatory property, i.e., an alternative is acceptable if and only if all the attributes are acceptable. Other non-compensatory rules include conjunctive and disjunctive rules (Coombs 1951) and elimination by aspects (Tversky 1972). Due to the non-compensatory feature, a decision maker's choice process is tremendously simplified. Individuals only need to compare the attributes value and their aspiration levels rather than comparing among all alternatives. In this paper, we develop acceptance set from the aspiration level so as to including discrete attribute value. We propose that individuals only need to identify whether the alternative he/she is exposed to is in his/her acceptance set.

2.2 Multinomial Logit Model

The multinominal logit model (MNL) model has been derived from the utility maximising perspective with i.i.d. Gumbel noise by Marley (see Luce and Suppes 1965), which bridges the Luce theory and random utility maximising choice. The MNL model from the utility maximising perspective inherits the shortcoming of Luce's choice theory: lack of consideration for alternative substitutability. Fishbein and Ajzen (1972) addressed this drawback by decomposing an alternative into a set of attributes (See also Lancaster 1966). However, the classical example of red bus and blue bus, which is initially described by Debreu (1960), shows that the Fishbein and Ajzen (1972)'s solution still yields implausible conclusions when there are strong similarities between alternatives. Thus, researchers further study constrained utility maximising choice.

2.3 Constrained Utility Maximising Choice

Although the constrained utility choice also recognizes the bounded rationality of the economical man, they work in the opposite direction, i.e., the more complicated choice rules. Here, we focus only on two aspects, consideration set and rational inattention, which are closely relevant to the satisficing choice.

I. Consideration Set

An implicit assumption in the utility maximising is that a decision maker considers all feasible alternatives. The marketing and psychology literatures, however, provide well-established evidence that consumers do not consider all brands in a given market before making a purchase. For example, Goeree (2008) showed that high markups in the PC industry can be explained by simply allowing an incomplete choice set in the computer market.

Consideration set uses a two-stage method to construct the incomplete choice set. In the first stage, the consideration set is selected from the complete choice set; in the second stage, the utility maximizer acts on the chosen set. Related research has long focused on formalising this consideration set. (See Roberts and Lattin 1991). For example, Masatlioglu et al. (2012) used this approach to qualitatively derive the ordinal preference when a choice reversal is observed. Recently, Manzini and Mariotti (2014) used a primitive but unobservable attention parameter to produce a consideration set and further inferred the parameter from observed choice behavior. However, their choice rule relies on the assumption that the complete choice set is knowable in the first stage and the decision maker deliberately eliminates some alternatives in the second stage. Thus, the consideration set approach does not apply to the situation where decision makers are unaware of some alternatives. In addition, previous models only focus on pure ordinary preference. However, a simple ordinal preference is not enough in practice. Our satisficing choice model illustrates an quantitative method to study categorical choice by estimating the exposure intensity from marketing display data.

II. Rational Inattention

For an utility maximizer, non-optimal choice behavior occurs if the choice process is costly due to limited attention (Carplin and Dean 2013). Shugan (1980) first incorporated the cost of thinking into choice models. Shugan showed that the cost of information processing is proportional to the number of brands the consumer evaluates and the difficulty of making comparisons. Sims (1998, 2003) adopted Shannon information theory (1948) to measure the amount of information a decision maker processes. Weibull et al. (2007) found that a state with higher payoffs could result in lower expected utility, since higher payoff are often accompanied by states which are difficult for decision makers to infer. They further showed that when the signals are i.i.d extreme-value distributed cantered on the true payoffs of the actions, the best policy then follows the MNL model.

Matejka and McKay (2015) extended Sims' work and achieved the same conclusion as Weibull's (2007). In Matejka and McKay's model, a decision maker is not irrational because of limited attention, but more strategic since he/she can control his/her attention. Hence, the decision maker needs to allocate his/her attention besides choosing among alternatives. However, a rational inattention optimizer is not enough to explain what we found in our dataset: the distance measure from marketing display frequency to exposure intensity is sufficiently small. The dataset recorded the households' purchase history for multiple brands of crackers over a period of two years. Because the choice set was simple and the time period was long enough, households should clearly know which brand would lead to the highest payoff if they were utility maximizers. In that case, the marketing display frequency would not significantly impact people's attention strategy, which is contrary to our findings. Our findings can be explained from satisificing perspective as individuals are more likely to choose products that they have more exposure to, provided the products are acceptable.

Also, it is difficult to rigorously define and measure the cost in rational inattention because a decision maker may intentionally or unintentionally filter out some alternatives to prevent her cognitive capacity from being overloaded (Broadbent 1958). It might be reasonable to construct a thinking cost for the intentional attention allocation, yet the same thinking cost for the unintentional filtering needs to be justified.

Moreover, utility maximising under constraints violates the bounded rationality by proposing a complicated choice process. As pointed out by Gigerenzer and Todd (1999, p.11) "The paradoxical approach of optimisation under constraints [i.e., the optimisation including a search or other sort of cost] is to model the 'limited' search by assuming that the mind has essentially unlimited time and knowledge with which to evaluate the costs and benefits to further information search." Compared to the utility maximising, the choice from the satisficing perspective is much simpler.

2.4 Mere Exposure Effect

Zajonc (1968) first proposed the mere exposure effect which states that mere repeated exposure of a stimulus enhances preference for the stimuli. Following Zajonc, evidences on the mere exposure has been demonstrated across various contexts, for a wide range of stimulus. Birch and Marlin (1982) and Pliner (1982) proved that preference is an increasing function of exposure frequency in a food investigation. Baker (1999) showed this phenomenon in an experimental setting, when competitors were not inferior in terms of attribute performance. Elliott and Dolan (1998) further confirmed this effect using PET (positron emission tomography) in their neuroanatomical studies. Though Zajonc (2001) claimed that "preference" can be acquired by simple repeated exposure to stimuli, he did not move further to explain how the "preference" is formalised by this absurdly simple mean. One potential reason is that a degree of cognitive mediation, where the mere exposure failed to increase "preference", was observed in the experiments (see Birnbaum and Mellers 1979). In the same vein, Lazarus (1982) critiqued Zajonc's "preference" lack of thoughts which is a necessary condition for emotion. However, it is easy to resolve this conflict by viewing Zajonc's "preference" as "satisficing". When an alternative is considered to be acceptable by a decision maker, the mere exposure directly enhances its choice share. However, if the alternative is not acceptable, the mere exposure does not help. This explains why Baker (1999) failed to observe the mere exposure effect when the competitors are inferior in terms of attribute performance.

3 Model

In this section, we first introduce the foundation of choice rule from the satisficing perspective. Then we study the S-MNL model and the corresponding maximum likelihood problem.

3.1 Basic Model

Let J be the choice set, which represents a collection of the products and K be an attribute set, which represents a collection of the physical attributes/characteristics associated with the products in the choice set. As in Gul et al. (2014), we use physical attributes/characteristics to differentiate from subjective features which will be introduced in Section 5. Each attribute $i \in K$ is represented by a value in real chosen from a finite set $\mathcal{A}_i \subseteq \Re$. A product $j \in J$ is characterized by a vector of attributes $\boldsymbol{x}_j \in \mathcal{A}$, $\mathcal{A} = \{\boldsymbol{x} \in \Re^{|K|} \mid x_i \in \mathcal{A}_i, i \in K\}$.

Under the satisficing choice model, we assume that an individual's decision to select a product $j \in J$, depends on whether the attributes associated with product j are acceptable to the individual, and the level of exposure the individual has to the product. In this model of choice, we postulate that an individual does not necessarily choose the product with the highest perceived utility, but would accept any product with attributes that are acceptable to the individual, provided that he/she has been exposed to the product.

Let Ω denote the sample space, which represents a finite set of individuals who are characterized by their attributes acceptance profiles. Specifically, whether a product $j \in J$ would be acceptable to an individual $\omega \in \Omega$ depends on whether $\boldsymbol{x}_j$ is an element of his/her attribute acceptance set, which we denote by $\mathcal{S}_\omega \subseteq \mathcal{A}$. We denote π_ω, $\omega \in \Omega$ as the probability of an individual chosen from the population with attribute acceptance set, $\mathcal{S}_\omega$. Given $\bar{\Omega} \subseteq \Omega$, we use the notation $\mathbb{P}(\bar{\Omega})$ to denote the probability of choosing an individual in $\bar{\Omega}$, i.e.,

$$\mathbb{P}(\bar{\Omega}) = \sum_{\omega \in \bar{\Omega}} \pi_\omega.$$

In particular, we denote

$$\mu_j := \mathbb{P}(\{\omega \in \Omega \mid \boldsymbol{x}_j \in \mathcal{S}_\omega\})$$

as the probability of a randomly chosen individual who would *accept* product $j \in J$. Suppose the acceptance of attributes is independent, then for a given attribute vector in $\boldsymbol{x} \in \mathcal{A}$, we have

$$\mathbb{P}(\{\omega \in \Omega \mid \boldsymbol{x} \in \mathcal{S}_\omega\}) = \prod_{i \in K} \nu_i(x_i)$$

where

$$\nu_i(x) := \mathbb{P}(\{\omega \in \Omega \mid \exists\ \boldsymbol{y} \in \mathcal{S}_\omega\ :\ x = y_i\}).$$

Hence, if independence is assumed, we have

$$\mu_j = \prod_{i \in K} \nu_i([\boldsymbol{x}_j]_i).$$

Note μ_j is utilised to quantify the probability of product $j \in J$ being acceptable by the sample space Ω. Supposet a product $\jmath \in J$ is characterised by two attributes, price and colour, with $[\boldsymbol{x}_\jmath]_p = 5$ and colour $[\boldsymbol{x}_\jmath]_c = red$, where p refers to price and c refers to colour. For connvience, we illustrate two individuals' acceptance set:

$$S_1 = \left\{ \left(\begin{array}{c} x_p \\ x_c \end{array} \right) \middle| \begin{array}{l} x_p \in \{0, 1, 2, 3, 4\} \\ x_c \in \{red, blue, white, green\} \end{array} \right\}, \quad S_2 = \left\{ \left(\begin{array}{c} x_p \\ x_c \end{array} \right) \middle| \begin{array}{l} x_p \in \{4.5, 5, 5.5, 6, 6.5\} \\ x_c \in \{red, black\} \end{array} \right\}.$$

According to the definition, the probability of a random chosen individual would accept the product $\jmath$ is:

$$\mu_\jmath = \mathbb{P}\left(\left\{ \omega \in \Omega \,\middle|\, \left(\begin{array}{c} 5 \\ red \end{array} \right) \in \mathcal{S}_\omega \right\} \right).$$

Under the independence assumption,

$$\mu_\jmath = \mathbb{P}\left(\left\{ \omega \in \Omega \,\middle|\, \left(\begin{array}{c} 5 \\ x_c \end{array} \right) \in \mathcal{S}_\omega, \forall\ x_c \in \mathcal{A}_c \right\} \right) \times \mathbb{P}\left(\left\{ \omega \in \Omega \,\middle|\, \left(\begin{array}{c} x_p \\ red \end{array} \right) \in \mathcal{S}_\omega, \forall\ x_p \in \mathcal{A}_p \right\} \right).$$

Let $e_j \in [0, 1]$ denote the *exposure intensity*, which represents the level of exposure a randomly chosen individual has to product $j \in J$, where $\sum_{j \in J} e_j \leq 1$. Suppose $|N|$ draws are made from Ω, where N is an index set. Let $r_{nj} \in \{0, 1\}$ indicates whether there is a marketing display activity for product $j \in J$ in draw n. We denote d_j as the normalised marketing display frequency of product $j \in J$ in $|N|$ draws, i.e., $d_j = \frac{\sum_{n \in N} r_{nj}}{\sum_{j \in J} \sum_{n \in N} r_{nj}}$. We postulate that the distance between exposure intensity and normalised marketing display frequency is small. Specifically, $\mathrm{dist}(e_j, d_j) \leq \delta$, where δ represents a small number. In particular, we use Kullback Leibler divergence measure, represented as

$$\mathrm{dist}(\boldsymbol{e}, \boldsymbol{d}) = \sum_{j \in J} d_j \ln \frac{d_j}{e_j} \leq \delta.$$

The distance between the exposure intensity and the normalised marketing display frequency could be due to an individual's intentional control of his/her attention (Sims 1998, Sims 2003) and/or factors other than marketing display activities which also impact the exposure.

A high level of exposure could increase the choice share of a product if it is acceptable. Suppose two products are equally acceptable, then their relative choice share is simply determined by their exposure intensities. Obviously, a product that has high exposure intensity and is more likely to be acceptable would occupy a larger market pie. Hence, we have

$$\rho_j \propto e_j \mu_j.$$

where ρ_j denotes the probability of product $j \in J$ would be chosen by a random chosen individual from Ω.

Let N_j be a collection of index with associated individuals who chose product j in real draws. The choice share of product j is then given by $f_j = \frac{|N_j|}{|N|} \in [0, 1]$. Motivated by this, we normalize ρ_j and get theorem 1.

Theorem 1 ***Choice Probability under the Satisficing Choice Model****: The choice probability of product $j \in J$ under the satisficing choice model is*

$$\rho_j = \frac{e_j \mu_j}{\sum_{l \in J} e_l \mu_l}. \tag{1}$$

Remark 1 Properties The satisficing choice model mainly contributes in two fields: the choice process and the choice set. Under the satisficing choice model, the choice is viewed as a process of a hybrid of *stimulus-directed* and *goal-driven* behaviour of individuals. An individual's choice is not only determined by his/her own acceptance set (goal-driven), but also impacted by the level of exposure he/she has to products (stimulus-directed).

Moreover, the actual choice set is dynamic. Different from conventional random utility maximising models, which usually assume an individual would compare among all alternatives in a static choice set, the satisficing choice model assumes he/she choses the acceptable one which he/she has been exposed to. Hence, the actual choice set is only a subset of the complete choice set.

3.2 Satisficing MNL

The literature on the random utility maximising (RUM) have derived the MNL model by assuming an i.i.d. Gumbel error term. In this section, we derive the MNL model under the satisficing choice, namely, Satisficing MNL (S-MNL).

Suppose the acceptance of attributes is independent, take logs[1] for $\mu_j, j \in J$. Following theorem 1, the choice probability for S-MNL is given by the following theorem.

[1] The motivation to take the logs is to get a tractable robust counterpart for the maximum likelihood problem.

Theorem 2 ***Choice Probability under the S-MNL****: The choice probability of product $j \in J$ under the S-MNL model is*

$$\rho_j = \frac{e_j \exp(\sum_{i \in K} \ln \nu_i([\boldsymbol{x}_j]_i))}{\sum_{l \in J} e_l \ \exp(\sum_{i \in K} \ln \nu_i([\boldsymbol{x}_l]_i))}. \tag{2}$$

From theorem 2, the satisficing choice model provides a new perspective to elicit the MNL model, which starts from an extended Luce's axiom, other than the conventional parametric RUM model with i.i.d. Gumbel error term.

Lemma 3 ***Choice Probability under the S-MNL (Linearization)****: If $\ln \nu_i(x) = \beta_i x$, the choice probability of product $j \in J$ under the S-MNL model is*

$$\rho_j = \frac{e_j \exp(\sum_{i \in K} [\boldsymbol{x}_j]_i \beta_i)}{\sum_{l \in J} e_l \exp(\sum_{i \in K} [\boldsymbol{x}_l]_i \beta_i)}. \tag{3}$$

In particular, if $\nu_i(x) = 0$, we assume $\ln \nu_i(x) = -\infty$. The linear form of $\ln \nu_i(x)$ bridges the S-MNL and the utility maximising MNL (U-MNL) in a fixed alternative specific case. Under alternative specific U-MNL (U-MNL *alt*), utility of product $j \in J$ is given by

$$u_j = \gamma_j + \boldsymbol{x}_j^{'} \boldsymbol{\beta} + \epsilon_j.$$

where γ_j denotes a fixed alternative specific value, and error term ϵ_j follows i.i.d Gumbel distribution.

Connection between the S-MNL and the U-MNL ***alt*** : If $\delta = \infty$, $\log(e_j) = \gamma_j, e_j > 0, \forall j \in J$, then the S-MNL model and the U-MNL *alt* model share the similar form of MNL.

$$\rho_j = \frac{\exp(\gamma_j + \sum_{i \in K} [\boldsymbol{x}_j]_i \beta_i)}{\sum_{l \in J} \exp(\gamma_l + \sum_{i \in K} [\boldsymbol{x}_l]_i \beta_i)}. \tag{4}$$

Remark 3 On Zero Choice Probability The S-MNL model is capable of explaining zero choice probability, while the U-MNL *alt* can not. Note the transformation of $\log(e_j) = \gamma_j$ requires $e_j > 0, \forall j \in J$. However, if $\exists \ e_\jmath = 0, \jmath \in J$, according to theorem 1, $\rho_\jmath = 0$. As evident in equation (4), there is no way to explain zero choice probability under the U-MNL *alt*. Nevertheless, under the S-MNL model, the associated maximum likelihood optimisation problem will not be impacted if product $\jmath$ is excluded in the dynamic choice set.

Remark 4 On Parameter Estimation and Interpretation If $\delta < \infty$, The S-MNL model and the U-MNL *alt* are distinct in parameter identification. In U-MNL *alt*, only $|J| - 1$ of γ_j are identifiable while $|J|$ of e_j can be estimated in S-MNL, due to the condition of Kullback Leibler divergence measure. Since only difference matters under the U-MNL *alt*, γ_j in fact denotes the difference between product j's alternative specific value and that of a restricted one, which is usually 0.

The second difference between the U-MNL model and the S-MNL model is the attribute coefficient β. The U-MNL model interpret β as individuals' attributes preference. However, in our empirical study, it will be shown that β could be a mixture of the exposure effect and attribute coefficients under the U-MNL model. In this respect, β under the U-MNL model could be a "disguised preference". Unlike the U-MNL model, the S-MNL model is capable of separating the attribute coefficients from the exposure effect.

3.3 Maximum Likelihood Problem

Next, we give the concrete form of maximum likelihood problem under the S-MNL model. We then show the corresponding counterpart of maximum likelihood optimisation problem. For convenience, we also illustrate the counterpart of maximum likelihood optimisation problem under the U-MNL model. For simplicity, we use the same parameter notation α, β for both models.

Based on lemma 3, the maximum likelihood optimisation problem under the S-MNL model is given by

Proposition 1 ***Maximum Likelihood Optimisation under the S-MNL***

$$\begin{aligned}
\max_{e_j, \beta_j, j \in J} \quad & \sum_{n \in N} \sum_{j \in J} y_{nj} \ln \frac{e_j \exp(\boldsymbol{x}'_{nj} \boldsymbol{\beta})}{\sum_{l \in J} e_l \exp(\boldsymbol{x}'_{nl} \boldsymbol{\beta})}. \\
\text{s.t.} \quad & \sum_{j \in J} d_j \ln \frac{d_j}{e_j} \leq \delta, \\
& \sum_{j \in J} e_j \leq 1, \\
& e_j \geq 0, \qquad j \in J.
\end{aligned} \tag{5}$$

where $\boldsymbol{x}_{nj}$ represents the product j's attributes value in draw n.

The objective function follows from the above analysis and lemma 3. The first constraint specifies the Kullback Leibler distance measure between the exposure intensity and the normalised marketing display frequency. The second and third constraint say the exposure intensities should be nonnegative and the sums should be no more than 1.

After some algebra, we transform problem (5) into an equivalent optimization problem (see Appendix A for the detail):

Proposition 2 ***Counterpart of Maximum Likelihood Optimisation under the S-MNL*** *The counterpart of maximum likelihood optimisation problem under the S-MNL model is given by*

$$\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})'\boldsymbol{\beta} - y_{nj} z_{nj}) \le t_{nj}, \qquad j, l \in J, n \in N, \\
& \sum_{j \in J} t_{nj} \le 1, \qquad n \in N, \\
& \sum_{n \in N} \sum_{j \in J} y_{nj} z_{nj} \le a, \\
& \sum_{j \in J} d_j (\ln d_j - \alpha_j) \le \delta, \\
& \exp(\alpha_j) \le w_j, \qquad j \in J, \\
& \sum_{j \in J} w_j \le 1.
\end{aligned} \tag{6}$$

where a, z_{nj}, t_{nj}, w_j are auxiliary variables introduced in our optimisation method.

The counterpart of maximum likelihood optimisation under the U-MNL model is given by the following corollary.

Corollary 1 ***Counterpart of Maximum Likelihood Optimisation under U-MNL***

$$\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \exp((\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})'\boldsymbol{\beta} - y_{nj} z_{nj}) \le t_{nj}, \qquad j, l \in J, n \in N, \\
& \sum_{j \in J} t_{nj} \le 1, \qquad n \in N, \\
& \sum_{n \in N} \sum_{j \in J} y_{nj} z_{nj} \le a.
\end{aligned} \tag{7}$$

4 Empirical Result

In this section, we analyze a subset of cracker dataset by the S-MNL and U-MNL models respectively and compare the performance of each model in terms of in-sample and out-of-sample.

Table 1: Basic descriptive statistics: Cracker (136 panelists, 13 observations each, 1768 observations)

Variables	Sunshine	Keebler	Nabisco	Private Label
Average Price($oz)	90.652	108.761	104.170	66.035
Display [a]	0.174	0.101	0.282	0.073
Brand shares %	8.8	7.9	56.0	27.3

a. Fraction of purchase occasions that a brand was on display

4.1 Data

The dataset is accessed from the R package (mlogit)[2]. This is an optical scanner panel dataset on purchases of saltine crackers in the Rome (Georgia) market, collected by Information Resources Incorporated. Jain et al. (1994) and Paap et al. (2000) have analyzed this dataset.

As in Jain et al. (1994), we consider a subset dataset to keep the same observations for each household. The dataset contained information on all purchases of crackers of 136 households [1768 observations, 13 observations for each household] over a period of two years, including brand choice, actual price of the purchased brand, shelf price of other brands, and whether there was a display at the time of purchase.

Table 1 shows a summary statistics for the dataset we used. There were three major national brands in the database, Sunshine, Keebler and Nabisco, with market shares of 8.8%, 7.9%, and 56.0% respectively. The local brands were collected under 'Private Label', which had a market share of 27.3%. 'Display' refers to the fraction of purchase occasions that a brand was on display. The 'Average Price' denotes the mean of the price of brand over the 1768 observations.

The dataset is classified into four small subsets, which are indicated by observation indexes as (1, 2, 3, 4), (4, 5, 6, 7), (7, 8, 9, 10) and (10, 11, 12, 13). Partial overlap between the consecutive subsets is allowed, as the last observation is used as out-of-sample. Furthermore, we apply the method of *leave 1-out non exhaustive cross-validation* in each subset. Specifically, the subset is further split into two parts, with the first three observations for each household as in-sample data, and the rest one as out-of-sample data. For example, the first subset contains the data with indexes 1,2 and 3 as-in-sample (408 observations), index 4 as out-of-sample (136 observations). This particular cross-validation is chosen due to forecast comparison, as we require that the acceptance set is the same in out-of-sample as that in in-sample. Also, since the exposure effect has a relatively moderate-term effect, it can be assumed

[2] We acknowledge Yves Croissant for offering the dataset online.

stable if the out-of-sample is immediately afterwards the in-sample.

4.2 Estimation

We first estimate the S-MNL and U-MNL models in the case with non-brand specific parameters (where attribute coefficients are identical across brands) and then in the case with the brand specific parameters (where price coefficients are different for different brands). The maximum likelihood problems are solved by ROME package created by Goh and Sim (2011), which is available online (http://www.robustopt.com).

4.2.1 The Non-brand Specific Case

We estimate the S-MNL model and the U-MNL model according to proposition 2 and corollary 1 respectively. In the S-MNL model, because the Kullback Leibler divergence bound is a small number, we estimate the two case with $\delta = 0.1, 0.2$ and 0.3. We also estimate the case when $\delta = 0$, where $e_j = d_j$. In addition, two maximum likelihood problems under the U-MNL model are estimated: U-MNL *pure*, where only price is explanatory variable, and U-MNL *disp*, where price and display are explanatory variables.

We use $\hat{\beta}$ to denote the estimates for coefficients of explanatory variables, where subscript p refers to price and d refers to display and the supscript s and u denote the estimates associated with the S-MNL model and the U-MNL model respectively. Further, four exposure estimates $\hat{e}_i^s$ are estimated under the S-MNL model, where $i = Sun., Kee., Nab., Pri.$ corresponds to brand Sunshine, Keebler, Nabisco and Private respectively.

Table 2 gives a summary of the estimation results under the S-MNL model and the U-MNL model. First, we found that the S-MNL model performs better than the U-MNL model in all three scenarios. The Negative log-likelihood under the S-MNL model is smaller than that under the U-MNL model, and significantly smaller when $\delta > 0$. Although the U-MNL *disp* is more valid than the U-MNL *pure*, the U-MNL*disp* is inferior to the S-MNL model. It suggests that the market display information can improve in-sample performance, but given the same market display information, the S-MNL model fits the data significantly better than the U-MNL model.

Second, the S-MNL and the U-MNL models give distinct β estimates. On average, a lower price estimate is given under the former than that under the latter, because the satisficing perspective views product alternatives from two factors, attribute and exposure. Intuitively, allowing the uncertainty only in attribute dimension instead of two dimensions is restricted to give a mixed effect from the two dimensions. In this respect, the satisficing perspective helps to reveal the *true* coefficient for attribute through variation in exposure.

More importantly, the use of two dimensions has practical implications. It enables firms to identify the cause of its market share changes and make relevant strategies. For example, if a product has a small market share, the reason may be that some of its attributes are less likely to be acceptable or the exposure related marketing activities are rare. The S-MNL model provides both a theoretical and practical method to distinguish these two possibilities. Moreover, with comparison of the exposure estimates in table 2 and the brand share of data statistics (see Table 7), it is clear that the exposure plays a significant role in determining the market share for crackers.

Table 2: Estimates of the S-MNL and the U-MNL in the Non-brand Specific: Cracker (136 panelists, 408 observations in-sample)

In-sample	Parameter	Variable	S-MNL					U-MNL
		S-MNL(U-MNL)	$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	*pure*	*disp*
1,2,3								
	Price Coefficient	$\hat{\beta}^s_p(\hat{\beta}^u_p)$	-0.028	-0.035	-0.043	-0.046	-0.007	-0.007
	Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	—	0.813
	Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.257	0.108	0.071	0.059	—	—
		$\hat{e}^s_{Kee.}$	0.154	0.128	0.122	0.113	—	—
		$\hat{e}^s_{Nab.}$	0.477	0.652	0.722	0.692	—	—
		$\hat{e}^s_{Pri.}$	0.112	0.112	0.085	0.073	—	—
	Neg Loglkhood		490.212	438.018	433.086	432.660	562.464	549.030
4,5,6								
	Price Coefficient	$\hat{\beta}^s_p(\hat{\beta}^u_p)$	-0.025	-0.026	-0.031	-0.036	-0.011	-0.010
	Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	—	0.734
	Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.367	0.183	0.126	0.092	—	—
		$\hat{e}^s_{Kee.}$	0.126	0.115	0.120	0.117	—	—
		$\hat{e}^s_{Nab.}$	0.400	0.633	0.739	0.686	—	—
		$\hat{e}^s_{Pri.}$	0.107	0.126	0.108	0.081	—	—

	Neg Loglkhood		533.723	447.435	431.076	423.749	556.239	542.812
7,8,9								
	Price Coefficient	$\hat{\beta}^s_p(\hat{\beta}^u_p)$	-0.021	-0.019	-0.020	-0.023	-0.002	-0.002
	Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	—	0.556
	Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.309	0.159	0.111	0.080	—	—
		$\hat{e}^s_{Kee.}$	0.213	0.153	0.130	0.114	—	—
		$\hat{e}^s_{Nab.}$	0.390	0.560	0.628	0.683	—	—
		$\hat{e}^s_{Pri.}$	0.088	0.128	0.131	0.124	—	—
	Neg Loglkhood		542.661	450.837	434.453	428.475	565.377	558.247
10,11,12								
	Price Coefficient	$\hat{\beta}^s_p(\hat{\beta}^u_p)$	-0.016	-0.024	-0.032	-0.034	-0.005	-0.005
	Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	—	0.687
	Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.221	0.091	0.061	0.054	—	—
		$\hat{e}^s_{Kee.}$	0.143	0.111	0.109	0.100	—	—
		$\hat{e}^s_{Nab.}$	0.482	0.670	0.742	0.690	—	—
		$\hat{e}^s_{Pri.}$	0.154	0.128	0.088	0.077	—	—
	Neg Loglkhood		465.782	419.671	416.031	415.829	563.437	551.906

[1] Neg Loglkhood refers to Negative Log-likelihood.

Next we use the the in-sample parameter estimates to predict the performance of each MNL model in out-of-sample. As mentioned in section 4.1, we assume that the exposure effect is the same in out-of-sample as that in in-sample.

Table 3 illustrates the performances of the S-MNL model and the U-MNL model in out-of-sample. It is obvious that both of the U-MNL *pure* and *disp* make worse predictions than the S-MNL model in terms of negative log-likelhiood, and significantly worse when $\delta > 0$. Also, for the S-MNL model, we find that the impact of Kullback Leibler divergence bound δ on the negative log-likelihood is diminished as δ increases. In terms of Cracker dataset, the prediction result is much better when $\delta = 0.2$ than when $\delta = 0.1$ and 0, however the improvement from $\delta = 0.2$ to $\delta = 0.3$ is small. It implies that $\delta = 0.2$ is an appropriate Kullback Leibler measure bound for the Cracker dataset.

Table 3: Predicts of S-MNL and U-MNL in the Non-brand Specific: Cracker (136 panelists, 136 observations out-of-sample)

Out-of-sample	Parameter	S-MNL				U-MNL	
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	*pure*	*disp*
4							
	Neg Loglkhood	156.245	144.158	135.710	135.306	184.791	177.787
7							
	Neg Loglkhood	185.552	156.911	152.957	151.939	191.036	190.647
10							
	Neg Loglkhood	176.056	143.050	136.119	132.828	188.189	184.440
13							
	Neg Loglkhood	156.885	143.323	142.293	142.546	184.791	176.751

4.2.2 The Brand Specific Case

We further estimate the S-MNL and the U-MNL models in the brand specific case. Table 4 presents the estimation results. First, the S-MNL model with a nonzero δ performs slightly better than the U-MNL *pure* and U-MNL *disp* in terms of negative log-likelihood, and the differences between them when $\delta = 0$ are negligible. Compared with the obvious advantage of the S-MNL model in the non-brand specific case, the reason for this similar fittness is that the effect of exposure in the S-MNL model is *distorted* into price preference estimates in U-MNL model. If this hypothesis is correct, the brand with relative high exposure intensity in the S-MNL model would follow a relatively high price coefficient in the U-MNL model, and vice versa. This is evident in the contrast of the four price coefficients under the S-MNL model and the U-MNL model. For example, in the in-sample (4,5,6) when $\delta = 0.2$, $\hat{e}^s_{Nab.}$ with the value 0.624 is relatively high , while $\hat{\beta}^u_{pNab.}$, which is -0.026 for the U-MNL *pure* and -0.022 for the U-MNL *disp*, are both much higher than $\hat{\beta}^s_{pNab.}$ with the value -0.041.

Table 4: Estimates of S-MNL and U-MNL in the Brand Specific Case: Cracker (136 panallists, 408 observations in-sample)

In-sample	Parameter	Variable	S-MNL			U-MNL	
		S-MNL(U-MNL)	$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	*pure*	*disp*
1,2,3							
	Price Coefficient	$\hat{\beta}^s_{pSun.}(\hat{\beta}^u_{pSun.})$	-0.059	-0.054	-0.053	-0.058	-0.058
		$\hat{\beta}^s_{pKee.}(\hat{\beta}^u_{pKee.})$	-0.045	-0.049	-0.053	-0.049	-0.049
		$\hat{\beta}^s_{pNab.}(\hat{\beta}^u_{pNab.})$	-0.038	-0.034	-0.033	-0.030	-0.030
		$\hat{\beta}^s_{pPri.}(\hat{\beta}^u_{pPri.})$	-0.046	-0.058	-0.061	-0.057	-0.057
	Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	0.027
	Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.257	0.178	0.149	—	—
		$\hat{e}^s_{Kee.}$	0.154	0.231	0.314	—	—
		$\hat{e}^s_{Nab.}$	0.477	0.342	0.274	—	—
		$\hat{e}^s_{Pri.}$	0.112	0.249	0.263	—	—
	Neg Loglkhood		431.706	430.154	430.529	430.749	430.831
4,5,6							
	Price Coefficient	$\hat{\beta}^s_{pSun.}(\hat{\beta}^u_{pSun.})$	-0.057	-0.056	-0.054	-0.051	-0.048
		$\hat{\beta}^s_{pKee.}(\hat{\beta}^u_{pKee.})$	-0.037	-0.044	-0.047	-0.043	-0.040
		$\hat{\beta}^s_{pNab.}(\hat{\beta}^u_{pNab.})$	-0.031	-0.038	-0.041	-0.026	-0.022
		$\hat{\beta}^s_{pPri.}(\hat{\beta}^u_{pPri.})$	-0.036	-0.029	-0.028	-0.048	-0.042
	Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	0.379
	Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.367	0.227	0.163	—	—
		$\hat{e}^s_{Kee.}$	0.126	0.161	0.180	—	—
		$\hat{e}^s_{Nab.}$	0.400	0.568	0.624	—	—
		$\hat{e}^s_{Pri.}$	0.107	0.044	0.033	—	—
	Neg Loglkhood		424.719	422.009	420.948	427.358	424.934
7,8,9							

Price Coefficient	$\hat{\beta}^s_{pSun.}(\hat{\beta}^u_{pSun.})$	-0.043	-0.038	-0.038	-0.037	-0.035
	$\hat{\beta}^s_{pKee.}(\hat{\beta}^u_{pKee.})$	-0.032	-0.039	-0.042	-0.030	-0.029
	$\hat{\beta}^s_{pNab.}(\hat{\beta}^u_{pNab.})$	-0.020	-0.018	-0.018	-0.013	-0.011
	$\hat{\beta}^s_{pPri.}(\hat{\beta}^u_{pPri.})$	-0.021	-0.024	-0.024	-0.033	-0.029
Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	0.321
Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.309	0.194	0.154	—	—
	$\hat{e}^s_{Kee.}$	0.213	0.412	0.508	—	—
	$\hat{e}^s_{Nab.}$	0.390	0.299	0.255	—	—
	$\hat{e}^s_{Pri.}$	0.088	0.095	0.084	—	—
Neg Loglkhood		426.936	426.095	425.711	427.104	425.472
10,11,12						
Price Coefficient	$\hat{\beta}^s_{pSun.}(\hat{\beta}^u_{pSun.})$	-0.041	-0.050	-0.055	-0.041	-0.041
	$\hat{\beta}^s_{pKee.}(\hat{\beta}^u_{pKee.})$	-0.030	-0.038	-0.043	-0.034	-0.034
	$\hat{\beta}^s_{pNab.}(\hat{\beta}^u_{pNab.})$	-0.023	-0.030	-0.033	-0.017	-0.017
	$\hat{\beta}^s_{pPri.}(\hat{\beta}^u_{pPri.})$	-0.032	-0.022	-0.018	-0.038	- 0.038
Display Coefficient	$(\hat{\beta}^u_d)$	—	—	—	—	0.052
Exposure / Intercept	$\hat{e}^s_{Sun.}$	0.221	0.266	0.286	—	—
	$\hat{e}^s_{Kee.}$	0.143	0.183	0.206	—	—
	$\hat{e}^s_{Nab.}$	0.482	0.511	0.489	—	—
	$\hat{e}^s_{Pri.}$	0.154	0.041	0.020	—	—
Neg Loglkhood		416.157	411.492	410.087	418.810	418.907

Table 5 compares the prediction performance of the S-MNL model and the U-MNL model in the brand specific case. It shows that the S-MNL model predicts significantly better than the U-MNL model, which is consistent with what we observe in the non-brand specific case. However, unlike the result in the non-brand specific case, the prediction differences between the S-MNL model with different δ are negligible. Again, this is due to the distortion of exposure effect into price coefficient under the S-MNL model in the brand specific case. Nevertheless, the inferior performance of the U-MNL model indicates that not all the choice shares could be attributed to product attribute preferences. Clearly, the exposure plays an important role in determining the choice for crackers.

Moreover, the coefficient distortion effect raises the question of how much of the choice share differences should be explained by exposure differences or by acceptance differences in the brand attributes. Answers to it have a significant impact on commercial and operational strategies. For example, if consumers do not choose the brand due to its lower exposure intensity, the firm should invest more on marketing display activities. If it is because that consumers are less likely to accept some attributes, it may be more wise for the firm to improve the attributes. Our model offers a practical method to address this issue by setting a proper δ. Note adjusting δ would change the exposure distribution, hence the degree of coefficient distortion.

However, with comparison of the log-likelihood in table 3 and table 5, we find that the improvement on prediction accuracy in brand specific case is negeligble, let alone more parameters are estimated in the latter. It suggests that consumers do not show significantly different price acceptance for the four cracker brands.

Table 5: Predicts of S-MNL and U-MNL in Brand Specific Case: Cracker (136 panelists, 136 observations out-of-sample)

Out-of-Sample		S-MNL			U-MNL	
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	pure	disp
4						
	Neg Loglkhood	136.990	138.328	138.778	185.629	185.202
7						
	Neg Loglkhood	152.074	152.000	152.957	188.852	187.951
10						
	Neg Loglkhood	131.252	130.974	130.757	188.103	185.375
13						
	Neg Loglkhood	141.829	140.840	140.355	188.687	187.545

5 Extension: Piecewise Approximation Attributes Acceptance

Following the terminology in Gul et al. (2014), we consider subjective attributes. Subjective attributes mean that an individual's perception on attributes is not based on physical characteristics of a product. Brand specific case in section 4.2 is such an example, where an individual has different price acceptance set for different products. Here we study another case where an individual holds different acceptance sets for different values of the same attribute.

In deriving lemma 3, recall we assume the logarithm transformation of an attribute set identically linearly depend on the attribute value $[\boldsymbol{x}_j]_i \in \mathcal{A}_i, j \in J, i \in K$. Let $\mathcal{A}_{is} \subseteq \mathcal{A}_i$, where $\bigcup_{s\in S} \mathcal{A}_{is} = \mathcal{A}_i$ and $\mathcal{A}_{is} \bigcap \mathcal{A}_{is'} = \emptyset, \forall s, s' \in S, s \neq s'$. In the case where an individual's acceptance set depends on the subset $\mathcal{A}_{is}$ which the attribute value $[\boldsymbol{x}_j]_i$ belongs to, the S-MNL model is able to piecewise approximate the acceptance sets for attributes. For notation simplicity, we illustrate the piecewise approximation on one attribute $\imath \in K$.

Based on theorem 2, the piecewise approximation for attribute $\imath$ is given by

$$\ln \nu_\imath(x) = \sum_{s\in S} \beta_{\imath s} x I_{(x\in\mathcal{A}_{\imath s})}$$

where $\beta_{\imath s}$ denotes the coefficient for attribute $\imath$ associated with piece $s \in S$.

The maximum likelihood optimization problem in the case of piecewise approximation for attribute $\imath$ is given by

$$\begin{aligned}
\max_{e_j,\beta_j,j\in\mathcal{J}} \quad & \sum_{n\in N}\sum_{j\in\mathcal{J}} y_{nj} \ln \frac{e_j \exp(\sum_{s\in S} \boldsymbol{x}'_{nj}\boldsymbol{\beta}_s I_{([\boldsymbol{x}_{nj}]_\imath \in \mathcal{A}_{\imath s})})}{\sum_{l\in\mathcal{J}} e_l \exp(\sum_{s\in S} \boldsymbol{x}'_{nl}\boldsymbol{\beta}_s I_{([\boldsymbol{x}_{nl}]_\imath \in \mathcal{A}_{\imath s})})}. \\
\text{s.t.} \quad & \sum_{j\in\mathcal{J}} d_j \ln \frac{d_j}{e_j} \leq \delta, \\
& \sum_{j\in\mathcal{J}} e_j \leq 1, \\
& e_j \geq 0, \qquad j \in \mathcal{J}.
\end{aligned} \tag{8}$$

Note $\mathcal{J} = J \times S$ and $\boldsymbol{\beta}_s$ differentiates only on attribute $\imath$.

Following proposition 2, the corresponding counterpart optimization problem is given by corollary 2.

Corollary 2 ***Counterpart of Maximum Likelihood Optimization for the Piecewise S-MNL***

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \exp\left(\alpha_l - \alpha_j - y_{nj} z_{nj} + \sum_{s \in S} \left(\boldsymbol{x}'_{nl} \boldsymbol{\beta}_s I_{([\boldsymbol{x}_{nj}]_\imath \in \mathcal{A}_{\imath s})} - \boldsymbol{x}'_{nj} \boldsymbol{\beta}_s I_{([\boldsymbol{x}_{nj}]_\imath \in \mathcal{A}_{\imath s})} \right) \right) \le t_{nj}, \qquad j, l \in \mathcal{J}, n \in N, \\
& \sum_{j \in \mathcal{J}} t_{nj} \le 1, \quad n \in N, \\
& \sum_{n \in N} \sum_{j \in \mathcal{J}} y_{nj} z_{nj} \le a, \\
& \sum_{j \in \mathcal{J}} d_j (\ln d_j - \alpha_j) \le \delta, \\
& \exp(\alpha_j) \le w_j, \qquad j \in \mathcal{J}, \\
& \sum_{j \in \mathcal{J}} w_j \le 1.
\end{aligned} \tag{9}
$$

In estimation, a technical issue on piecewise approximation is how to replace the unoffered price in the dataset, as a brand has only one piece of price each time. We split the price attribute into two pieces: low price piece p_l (where $p < 80$) and high price piece p_h (where $p \ge 80$). We then use the average low prices to replace the unoffered low prices and average high prices to replace the unoffered high prices. The splitting strategy impacts the unoffered price and the display frequency. Under the above splitting strategy, we use the brand display frequency multiplied by the frequency of high price as the display frequency for high price, and the brand display frequency multiplied by the frequency of low price as the display frequency for low price.

Table 6 presents the estimation results for the two piecewise S-MNL model when $\delta = 0$, 0.2, 0.5, 1 respectively. Note that the price coefficient for the low price piece is much smaller than that for high price piece. However, it does not mean that consumers show a higher acceptability for the high price than that for the low price, since the associated prices are never equal. We suggest computing the estimated acceptability by $\rho_{jl} = \exp(\beta_{pj} p_{jl})$ and $\rho_{jh} = \exp(\beta_{pj} p_{jh})$ and comparing the acceptability directly. For example, we can use the average lower and higher prices for each brand (see Table 7), and compute their relative acceptability.

As in Table 2, the impact of Kullback Leibler divergence bound δ on the negative log-likelihood is diminished as δ increases. In the case of two piecewise approximation, the estimation result is much better when $\delta = 0.5$ than when $\delta = 0.2$ or 0, however the improvement from $\delta = 0.5$ to $\delta = 1$ is small. It implies that $\delta = 0.5$ is an appropriate Kullback Leibler measure bound for the Cracker dataset.

Generally speaking, the less one is confident on the display information, the higher the Kullback Leibler divergence bound one should set. How to better set this δ is an open question for future study.

In practice, a piecewise approximation helps a firm make price decisions roughly, such as whether to offer low price piece p_l or high price piece p_h.

Table 6: S-MNL : Cracker (136 panelists, 8 subjective brands, 408 observations in-sample)

In-sample	Parameter	Variable	$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1$
1,2,3						
	Price Coefficent	$\hat{\beta}_l$	-0.044	-0.061	-0.069	-0.076
		$\hat{\beta}_h$	-0.037	-0.042	-0.036	-0.030
	Exposure Intensity	$\hat{e}_{Sun.l}$	0.081	0.062	0.113	0.151
		$\hat{e}_{Sun.h}$	0.176	0.038	0.020	0.010
		$\hat{e}_{Kee.l}$	0	0	0	0
		$\hat{e}_{Kee.h}$	0.154	0.096	0.056	0.027
		$\hat{e}_{Nab.l}$	0.022	0.056	0.108	0.152
		$\hat{e}_{Nab.h}$	0.455	0.533	0.319	0.156
		$\hat{e}_{Pri.l}$	0.112	0.213	0.384	0.504
		$\hat{e}_{Pri.h}$	0.001	0.003	0.002	0.001
	Neg Loglkhood		604.387	542.713	539.553	538.483
4,5,6						
	Price Coefficent	$\hat{\beta}_l$	-0.069	-0.083	-0.096	-0.096
		$\hat{\beta}_h$	-0.051	-0.057	-0.055	-0.054
	Exposure Intensity	$\hat{e}_{Sun.l}$	0.117	0.070	0.109	0.080
		$\hat{e}_{Sun.h}$	0.250	0.068	0.024	0.015
		$\hat{e}_{Kee.l}$	0	0	0	0
		$\hat{e}_{Kee.h}$	0.126	0.113	0.073	0.047
		$\hat{e}_{Nab.l}$	0.010	0.018	0.041	0.030
		$\hat{e}_{Nab.h}$	0.390	0.558	0.421	0.271
		$\hat{e}_{Pri.l}$	0.102	0.159	0.317	0.233
		$\hat{e}_{Pri.h}$	0.006	0.021	0.016	0.011

	Neg Loglkhood		640.613	522.717	516.049	515.876
7,8,9						
	Price Coefficent	$\hat{\beta}_l$	-0.053	-0.065	-0.082	-0.093
		$\hat{\beta}_h$	-0.039	-0.043	-0.043	-0.039
	Exposure Intensity	$\hat{e}_{Sun.l}$	0.073	0.054	0.103	0.186
		$\hat{e}_{Sun.h}$	0.236	0.071	0.029	0.014
		$\hat{e}_{Kee.l}$	0	0	0	0
		$\hat{e}_{Kee.h}$	0.213	0.135	0.079	0.038
		$\hat{e}_{Nab.l}$	0.005	0.013	0.037	0.071
		$\hat{e}_{Nab.h}$	0.385	0.570	0.462	0.235
		$\hat{e}_{Pri.l}$	0.083	0.134	0.276	0.446
		$\hat{e}_{Pri.h}$	0.006	0.025	0.020	0.011
	Neg Loglkhood		647.869	520.066	505.026	502.408
10,11,12						
	Price Coefficent	$\hat{\beta}_l$	-0.044	-0.081	-0.100	-0.111
		$\hat{\beta}_h$	-0.031	-0.050	-0.050	-0.048
	Exposure Intensity	$\hat{e}_{Sun.l}$	0.029	0.057	0.159	0.261
		$\hat{e}_{Sun.h}$	0.192	0.038	0.019	0.010
		$\hat{e}_{Kee.l}$	0	0	0	0
		$\hat{e}_{Kee.h}$	0.144	0.100	0.059	0.030
		$\hat{e}_{Nab.l}$	0.007	0.023	0.061	0.094
		$\hat{e}_{Nab.h}$	0.475	0.605	0.384	0.196
		$\hat{e}_{Pri.l}$	0.137	0.154	0.303	0.402
		$\hat{e}_{Pri.h}$	0.018	0.025	0.016	0.008
	Neg Loglkhood		575.966	506.032	503.031	501.313

[1] $\hat{\beta}_l, \hat{\beta}_h$ denote the estimation of coefficient for low price and high price respectively.

[2] $\hat{e}_{ml}, \hat{e}_{mh}$ (where $m = Sun., Kee., Nab., Pri.$ corresponds to brand Sunshine, Keebler, Nabisco and Private respectively) denotes exposure estimates for high price and low price of four brand respectively.

6 Conclusion and Future Directions

In this paper, we introduce acceptance set and exposure intensity to model individuals' choices from the satisficing perspective. The choice of decision makers is too complicated to be described adequately by giving estimates for attribute parameters $\boldsymbol{\beta}$. The choice set does matter. The utility maximizing perspective is not applicable to the dynamic choice set. However, from our satisficing perspective, this dynamic choice set can be ivestigated from real market dataset by exposure intensity.

More importantly, we utilize the logs transformation of acceptance set to gain the Satisficing MNL model while maintaining the maximum likelihood optimization problem tractable. Specifically, we derive the maximum likelihood optimization problem from the minimum Kullback Leibler distance measure problem and apply a new optimization method to make estimation with a real market dataset. We show that the S-MNL model performs significantly better than U-MNL model both in terms of in-sample and out-of-sample in the non-brand specific case. Also, the S-MNL model has superior prediction than the U-MNL in the brand specific case. Although the fitness between the two models becomes small in the latter case, the U-MNL model is mistakenly estimated because the effect of exposure is *distorted* into attribute coefficient estimates.

Despite the superior performance of the S-MNL model, we do not claim that the S-MNL model fits and predicts any dataset better than U-MNL model, or any choice models from the satisficing perspective would outperform models from the utility maximising perspective. What we emphasise here is that the two different perspectives could provide different estimates, which will further impact decisions that rely on these estimates. A possible way is to compare the performance of each model based on the real dataset and then choose the more appropriate one. The superior performances of the S-MNL model in our dataset suggest that individuals' choices such as purchasing crackers are more likely to be based on the satisficing than the utility maximising theory. To further compare the satisficing and utility maximising choices, one future direction would be to study other choice models from these two perspectives, e.g. probit model.

Finally, the practical implication of the satisficing choice model is highly encouraging. By splitting the exposure and product attribute, our satisficing choice model enables a firm to identify the true causes of its market share. If a product is uncompetitive, the reason may be that the marketing related exposure activities are rare or the product attributes are unacceptable. Having a clear knowledge of this difference, a firm is able to make wiser marketing strategies. In addition, the satisficing choice opens a new door to piecewise approximate acceptance set. The piece-wise approximation provides a more vivid way to analyse individuals' choices, and enables firms to make more flexible marketing strategies.

Appendix

A Transformation S-MNL Optimization Problem to Convex Problem

The transformation procedure is illustrated as follows:

$$\begin{aligned} \max_{e_j,\beta_j,j\in J} \quad & \sum_{n\in N}\sum_{j\in J} y_{nj} \ln \frac{e_j \exp(\boldsymbol{x}'_{nj}\boldsymbol{\beta})}{\sum_{l\in J} e_l \exp(\boldsymbol{x}'_{nl}\boldsymbol{\beta})}. \\ \text{s.t.} \quad & \sum_{j\in J} d_j \ln \frac{d_j}{e_j} \leq \delta, \\ & \sum_{j\in J} e_j \leq 1, \\ & e_j \geq 0, \qquad j \in J. \end{aligned} \tag{10}$$

$$\begin{aligned} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad & \sum_{n\in N}\sum_{j\in J} y_{nj} \ln \frac{\exp(\alpha_j + \boldsymbol{x}'_{nj}\boldsymbol{\beta})}{\sum_{l\in J} \exp(\alpha_l + \boldsymbol{x}'_{nl}\boldsymbol{\beta})}. \\ \text{s.t.} \quad & \sum_{j\in J} d_j \ln d_j - \sum_{j\in J} d_j \alpha_j \leq \delta, \\ & \sum_{j\in J} \exp(\alpha_j) \leq 1. \end{aligned} \tag{11}$$

$$\begin{aligned} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad & \sum_{n\in N}\sum_{j\in J} y_{nj} \ln \frac{1}{\sum_{l\in J} \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})'\boldsymbol{\beta})}. \\ \text{s.t.} \quad & \sum_{j\in J} d_j \ln d_j - \sum_{j\in J} d_j \alpha_j \leq \delta, \\ & \sum_{j\in J} \exp(\alpha_j) \leq 1. \end{aligned} \tag{12}$$

$$\begin{aligned} \min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad & \sum_{n\in N}\sum_{j\in J} y_{nj} \ln \sum_{l\in J} \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})'\boldsymbol{\beta}). \\ \text{s.t.} \quad & \sum_{j\in J} d_j \ln d_j - \sum_{j\in J} d_j \alpha_j \leq \delta, \\ & \sum_{j\in J} \exp(\alpha_j) \leq 1. \end{aligned} \tag{13}$$

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \sum_{n \in N} \sum_{j \in J} y_{nj} \ln \sum_{l \in J} \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})' \boldsymbol{\beta}) \leq a, \\
& \sum_{j \in J} d_j \ln d_j - \sum_{j \in J} d_j \alpha_j \leq \delta, \\
& \sum_{j \in J} \exp(\alpha_j) \leq 1.
\end{aligned}
\tag{14}
$$

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \ln \sum_{l \in J} \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})' \boldsymbol{\beta}) \leq z_{nj}, \qquad \forall j \in J, n \in N, \\
& \sum_{n \in N} \sum_{j \in J} y_{nj} z_{nj} \leq a, \\
& \sum_{j \in J} d_j \ln d_j - \sum_{j \in J} d_j \alpha_j \leq \delta, \\
& \sum_{j \in J} \exp(\alpha_j) \leq 1.
\end{aligned}
\tag{15}
$$

Since z_{nj} is only constrained when $y_{nj} = 1$ and $y_{nj} \in \{0, 1\}$, we can replace z_{nj} by $y_{nj} z_{nj}$ in the first constraint.

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \ln \sum_{l \in J} \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})' \boldsymbol{\beta}) \leq y_{nj} z_{nj}, \qquad \forall j \in J, n \in N, \\
& \sum_{n \in N} \sum_{j \in J} y_{nj} z_{nj} \leq a, \\
& \sum_{j \in J} d_j \ln d_j - \sum_{j \in J} d_j \alpha_j \leq \delta, \\
& \sum_{j \in J} \exp(\alpha_j) \leq 1.
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \sum_{l \in J} \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})' \boldsymbol{\beta} - y_{nj} z_{nj}) \leq 1, \qquad \forall j \in J, n \in N, \\
& \sum_{n \in N} \sum_{j \in J} y_{nj} z_{nj} \leq a, \\
& \sum_{j \in J} d_j \ln d_j - \sum_{j \in J} d_j \alpha_j \leq \delta, \\
& \sum_{j \in J} \exp(\alpha_j) \leq 1.
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})'\boldsymbol{\beta} - y_{nj} z_{nj}) \le t_{nj} \qquad \forall j \in J, n \in N, \\
& \sum_{j=1}^{J} t_{nj} \le 1, \qquad n = 1, \ldots, |N|, \\
& \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} z_{nj} \le a, \\
& \sum_{j=1}^{J} d_j (\ln d_j - \alpha_j) \le \delta, \\
& \sum_{j=1}^{J} \exp(\alpha_j) \le 1.
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
\min \quad & a. \\
\text{s.t.} \quad & \exp(\alpha_l - \alpha_j + (\boldsymbol{x}_{nl} - \boldsymbol{x}_{nj})'\boldsymbol{\beta} - y_{nj} z_{nj}) \le t_{nj}, \qquad \forall j \in J, n \in N, \\
& \sum_{j \in J} t_{nj} \le 1, \qquad \forall n \in N, \\
& \sum_{n \in N} \sum_{j \in J} y_{nj} z_{nj} \le a, \\
& \sum_{j \in J} d_j (\ln d_j - \alpha_j) \le \delta, \\
& \exp(\alpha_j) \le w_j, \qquad \forall j \in J, \\
& \sum_{j \in J} w_j \le 1.
\end{aligned} \tag{19}
$$

B Sample Statistics

Table 7 illustrates the sample statistics for the dataset we use.

Table 7: Basic descriptive statistics: Cracker (136 panelists, 13 observations each, 1768 observations)

Sample	Sunshine	Keebler	Nabisco	Private Label
1,2,3				
Price($oz)	85.284	99.142	97.468	64.395
Brand shares	0.098	0.088	0.542	0.272
4,5,6				
Price($oz)	86.855	107.003	101.208	66.314
Brand shares	0.100	0.071	0.537	0.292
7,8,9				
Price($oz)	91.395	111.449	106.147	65.963
Brand shares	0.081	0.078	0.579	0.262
10,11,12				
Price($oz)	96.547	114.811	109.380	67.061
Brand shares	0.073	0.074	0.576	0.277

Acknowledgements

We acknowledge Yves Croissant for offering the dataset we used online. We would like to thank the reviewers of the paper for valuable insightful comments and editorial suggestions.

References

[1] Baker WE(1999). When can affective conditioning and mere exposure directly influence brand choice? *Journal of Advertising*, 28(4): 31-46.

[2] Birch LL, Marlin DW (1982). I don't like it; I never tried it: effects of exposure on two-year-old children's food preferences. *Appetite*, 3(4): 353-360.

[3] Birnbaum MH, Mellers BA(1979). Stimulus recognition may mediate exposure effects. *Journal of Personality and Social Psychology*, 37(3): 391.

[4] Broadbent DE(1958). *Perception and Communication*. New York: Pergamon Press.

[5] Carplin A, Dean M (2013). Rational inattention and state dependent stochastic choice. working paper.

[6] Coombs CH (1951). Mathematical models in psychological scaling. *J. Amer. Statist. Assoc*, 46(256):480-489.

[7] Debreu G (1960). Review of individual choice behavior by R. D. Luce. *American Economics Review*, 50: 186-8.

[8] Elliott R, Dolan RJ(1998). Neural response during preference and memory judgments for subliminally presented stimuli: a functional neuroimaging study. *The Journal of Neuroscience*, 18(12): 4697-4704.

[9] Fishbein M, Azjen I (1972). Attributes and opinions. *Annual Review of Psychology*, 23: 487-544.

[10] Gigerenzer G, Todd PM (1999). Fast and frugal heuristics: The adaptive toolbox. Gigerenzer G, Todd PM, ABC Research Group, eds. *Simple Heuristics That Make Us Smart*, (Oxford University Press, New York). 3-36.

[11] Goeree MS(2008). Limited Information and Advertising in the US Personal Computer Industry. *Econometrica* 76: 1017-1074.

[12] Goh J, Sim M(2011). Robust optimization made easy with ROME.*Oper. Res.*, 59(4): 973-985.

[13] Croissant Y(2013). Package 'mlogit'. http://www.r-project.org. Date/Publication 2013-12-27 13:28:20.

[14] Gul G, Natenzon P, and Pesendorfer W(2014). Random choice as behavioural optimization,*Econometrica*, 82(5),1873-1912.

[15] Jain DC, Naufel JV and Pradeep KC (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business and Economics Statistics*, 12(3): 317-328.

[16] Janiszewski C(1993). Preattentive mere exposure effects. *Journal of Consumer Research*, 376-392.

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

[18] Lazarus RS(1982). Thoughts on the relations between emotion and cognition. *American psychologist*,37(9):1019.

[19] Kahneman D(1973). Attention and Effort. *Prentice Hall*, New Jersey.

[20] Kahneman D, Knetsch JL, and Thaler RH (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives*, 5(1):193-206.

[21] Luce, RD and Suppes, P (1965). Preference, utility, and subjective probability. In Luce, R.D; Bush, R and Galanter, E, editors, *Handbook of Mathematical Psychology*(Wiley, New York.), 3: 249- 410.

[22] Manzini P, Mariotti M (2014). Stochastic choice and consideration sets. *Econometrica*. 82(3):1153-1176.

[23] Masatlioglu Y, Nakajima D, and Ozbay EY (2012). Revealed attention. *The American Economic Review*, 102(5):2183-2205.

[24] Matjka F, Alisdair M (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model.*American Economic Review*, 105(1): 272-98.

[25] Paap R Philip HF (2000). A dynamic multinomial probit model for brand choices with different short-run effects of marketing mix variables. *Journal of Applied Econometrics*, 15(6): 717-744.

[26] Pliner P (1982). The effects of mere exposure on liking for edible substances. *Appetite*, 3(3): 283-290.

[27] Roberts JH, Lattin JM (1991). Development and testing of a model of consideration set composition.*J. Marketing Res*, 28(4):429-440.

[28] Shannon CE (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379-423,623-656.

[29] Shugan SM (1980). The cost of thinking. *J. Consumer Res*, 7(2):99-111.

[30] Simon HA (1955). A behavioral model of rational choice. *Quart. J. Econom*, 69(1):99-118.

[31] Simon HA (1959). Theories of decision-making in economics and behavioral science. *The American Economic Revie*, 49(3): 253-283.

[32] Sims, CA (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49:317-456.

[33] Sims, CA (2003). Implications of rational inattention. *Journal of Monetary Economics*,50(3):656-690.

[34] Stütgen P, Boatwright P, Monroe RT (2012). A satisfying choice model.*Marketing. Sci*, 31(6):878-899.

[35] Tversky A (1972). Elimination by aspects: A theory of choice. *Psych. Rev*, 79(4):281-299.

[36] Weibull JW, Mattsson LG, and Voorneveld M (2007). Better may be worse: Some monotonicity results and paradoxes in discrete choice under uncertainty. *Theory and Decision*, 63:121-151.

[37] Zajonc RB(1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, 9(2)p2: 1-27.

[38] Zajonc RB(2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science*, 10(6): 224-228.