UiO : Department of Mathematics
University of Oslo

Discrete choice methods in market research for seafood. A statistical analysis

Lu Nguyen Nhung

Master's Thesis, Autumn 2021

This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Statistics*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Abstract

This master thesis reviews and discusses the literature on discrete choice methods. Discrete choice modelling is an effective tool for predicting choice behavior, analyzing consumer preferences, and determining consumption patterns and market trends. Nevertheless, increasing diversity in markets, customer preferences, and available products require more customized models that have high predictive power in different choice situations and behaviors. This also applies to the study of food choice. This thesis assesses the available methods from the point of view of statistics. It discusses the usefulness, strengths, and weaknesses of some of the most widely used discrete choice models, with a special focus on seafood preferences and food choice. We use survey data on the French seafood market for empirical assessments of the models, in order to determine the best model for predicting choice probabilities. By doing so, this thesis aims to improve the modelling of seafood preferences in particular, and of food choice in general.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Riccardo de Bin, who has taught me the importance of consistency in statistical modelling and in working with formulas and mathematical functions. I am grateful for his excellent guidance, detailed feedback, and kind support.

I am also grateful to Nguyen Tien Thong for sharing his seafood data with me. His research has inspired me to carry out this study. Before I started working on this thesis, Xianwen Chen introduced me to the study of discrete choice methods for fish and seafood. I would like to thank him for this, as well as his later help in explaining how survey and experimental data are collected for the study of food choice. I am also much indebted to the faculty and staff at the Department of Mathematics, University of Oslo. In particular, I would like to thank Annika Rigenholt for her kind support in administrative matters, making it much easier for me to complete the master program.

I would like to thank my husband, Aike Peter Rots, for his lifetime friendship and support. Finishing a difficult study program in the midst of a global pandemic has been challenging, and without his emotional support and his care for our two young children, I would not have finished this thesis. Finally, I would like to thank my family in Vietnam and in the Netherlands for encouraging me to follow my dream.

Oslo, September 2021
Lu Nguyen Nhung

Contents

List of Figures

CHAPTER 1

Introduction

In this thesis, we review the theory of discrete choice methods and the use of choice models in survey and experiment. In particular, we present and discuss three standard models, including multinomial logit, nested logit and mixed logit, together with recent developed discrete choice models such as latent class, generalized multinomial logit and logit-mixed-logit. In addition to describing the models, we point out the strengths and weaknesses of each model for analyzing choice behavior in general, and for predicting choice probabilities of different product alternatives and attributes in particular. The objective of the master thesis is to find out the best model that has powerful predictive power and at the same time avoid measurement errors in food choice analysis.

In summarizing the existing literature on choice studies, McFadden et. al. (2005) point out the sources of response errors and biases that have accumulated in survey and experiment research. They conclude that much of the literature on measurement errors assumes them to be "classical", i.e. the error is assumed to be independent of the latent true variable. However, empirical studies in the field suggest that survey response errors, in many cases, are correlated with unmeasured variables and do not confirm to the classical measurement assumption. For example, they can be correlated with unobservable factors such as the cognitive ability of participants in a survey. For identifying sources of correlation in discrete choice methods, Hensher, Rose & Greene (2005) and Train (2009) study different choice models and use statistical data for empirical assessment of the models. They found that choices can be

- correlated over alternatives;
- correlated over attributes of alternatives. For example, seafood alternatives may share the same attributes, such as product origin. In such cases, the attributes do not vary across alternatives;
- correlated over different choice situations responsed by the same individual;
- correlated over time periods.

In addition to correlated choices, preference heterogeneity (or taste variation) among consumers also poses another challenge to choice modelling. Discrete choice models that allow for individual heterogeneity by making coefficients of the model random, as in the mixed logit model by McFadden and Train (2000), can be good strategies for the statistical analysis of survey response behavior.

In order to test whether respondents show greater variability of preferences, researchers can implement within subject setting using repeated choices in stated preference experiments (i.e. choices made by one individual in several choice situations) that vary in difficulty and design. Stated preference experiments are studies in which respondents are presented with one or more choice experiments. In these studies, repeated measurement can provide additional variation and instruments, which are then used in analytical models (McFadden et. al. 2005). The measurement techniques are often performed by creating choices made by one individual in several choice situations and over several time periods. According to many researchers in the field, the approach that allows for coefficient heterogeneity is very flexible and can capture a variety of survey response effects. For this reason, Train (2009) suggests that stated reference studies can test the impact of new product attributes on consumers' choices and can fabricate new situations, for example in this analysis, for sustainable food choices (Chen et. al. 2015, Nguyen et. al. 2015).

We review discrete choice models in this thesis from a statistical point of view. For empirical assessment of the models, we use a seafood dataset that exhibits several challenges in choice modelling, as well as in survey and choice experiments. First of all, we begin with clarifying the settings for choice probabilities in chapter 2 in order to understand discrete choice methods clearly. In chapter 3, we introduce, describe, and discuss three standard choice models: multinomial logit, nested logit, and mixed logit. In addition, we present recent development of discrete choice methods such as latent class, generalized multinomial logit, and logit-mixed-logit models. These are important choice models in the field, the estimation of which will be described briefly in chapter 4. The dataset and empirical analysis in chapter 5 are obtained from an online survey of seafood consumption in France under a stated choice experiment. In this experiment, potential seafood consumers were asked for their preferences of different seafood alternatives and attributes. We also discuss the empirical analysis of discrete choice models and suggest some measures to improve the modelling of food choices in this chapter. We summarize our work in chapter 6.

CHAPTER 2

Modelling consumer choice behavior

2.1 The settings for modelling consumer choice

Commercial products have numerous characteristics and attributes. These include, but are not limited to, prices, brands, labels, origins, quality and quantity. The consumer's choice of a product among similar ones is assumed to be dependent on the product's characteristics or attributes (Lancaster 1966). These product attributes act as explanatory variables in discrete choice models (McFadden 1974), in which choice probabilities are calculated under a variety of behavioral specification. Based on choice contexts, product trends and consumer behaviors, researchers make assumptions about probabilities of a set of available outcomes, called the choice set. The consumer is assumed to choose one outcome among the alternatives within the choice set. Discrete choice models explain consumers' choices based on alternative products and their attributes. In particular, by using this method, researchers can separate the random effects of different product attributes on consumers' choice and predict consumer response in a new hypothetical situation, for example, a price change.

Researchers build discrete choice models using utility theory and derive choice probabilities from utility-maximizing behavior (Manski 2001). By observing how a consumer n $(n = 1, .., N)$ gives values to different alternatives j $(j = 1, ..., J)$ when this person faces a choice situation C_t, a researcher will define the consumer's utility function $u_t(.)$ for these alternatives. The choice situation C_t is also known as a choice set and t specifies a specific set out of a total of T sets. As choices are observed repeatly over time, t also specifies a specific time period. When a choice dataset has several time periods, we call it a panel dataset. In the given choice situation, the consumer obtains a certain utility level only if he/she chooses an alternative. If the consumer decides to choose alternative $i \in C_t$, the utility brought by choosing i is always higher than any other alternative $j \in C_t$, or $U_t(i) \geq U_t(j)$ $(i, j = 1, ..., J)$. Subscript j is used generally to address any alternative in the set of J alternatives. In our analysis, j is used to address alternatives i and any other alternatives in the choice set. When an alternative is chosen, we denote it i. The choice probability for the chosen alternative i is (Train 2009)

$$P_i = Prob[t : U_t(i) \geq U_t(j), j \neq i] \quad (2.1)$$

Equation (1) is known as a random utility model (RUM) in which the probability of a randomly chosen outcome i in a choice set C_t will specify a function for the observed part of utility V_{nit} for person n.

There are factors that collectively determine the consumer choice for alternative i in the choice set. In order to understand the behavioral process of this individual, researchers observe some of the factors that may lead to the person's choice. The observed factors are believed to have underlying relationship with the outcome y_{nj}, which we will specify later. There are also other factors in utility that can not be observed by the researcher, $V_{nj} \neq U_{nj}$. Consumer utility is then decomposed into two parts, the observed utility V_{nj} and unobserved error terms ϵ_{nj}

$$U_{nj} = V_{nj} + \epsilon_{nj} \quad (2.2)$$

In this introduction, we omit the subscript t for the sake of simplicity and describe the utilities U_{nj} for different alternatives j in one choice set C_t. The observed utility V_{nj} does not represent the consumer utility as a whole, ($V_{ni} \neq U_{ni}$). The error terms are, thus, defined as unknown factors affecting utility. They are the difference between true utility U_{nj} and the part of utility that the researcher captures in V_{nj}. Because researchers can calculate V_{nj} based on their measurement of some observable variables, it is called the observed utility.

In the particular choice set C_t, the researcher observes some product attributes labelled z_j and some characteristics of person n labeled s_n, such as income and education levels, which are also known as alternative specific variables and individual specific variables, respectively (Hensher, Rose & Greene 2005). Then the observed utility in general is

$$V_{nj} = V(z_j, s_n).$$

For simplicity, z_j and s_n are combined in a vector of observed variables x_{nj}. The observed utility V_{nj} in RUM models is assumed to be a linear function as

$$V_{nj} = \beta' x_{nj} + k_j, \quad (2.3)$$

where k_j is a constant that is specific to alternative j. Vector β is a vector of estimated coefficients reflecting the relationships between observed variables and outcome y_{nj}. The alternative-specific constant for an alternative captures the average effect on utility of all factors that are not included in the model. Thus they served a similar function to an intercept in a regression model, which also captures the average effect of all unincluded factors. We will further explain the logic of having the constant in the specified utility function as well as several modelling techniques related to it in section 2.3.

It is worth noting the difference between U_{nj} and y_{nj}. The outcome y_{nj} takes in the value of 1 or 0 if consumer n chooses or does not choose a specific alternative j in the choice set. For example, $y_{ni} = 1$ indicates the success of having alternative i as the chosen alternative while U_{ni} indicates the utility (or satisfaction) obtained by the consumer n for choosing alternative i. According to Train (2009), consumer's choice is not deterministic and can not be predicted exactly if researchers rely only on observed variables. He defines a function $y_{nj} = h(x_{nj}, \epsilon_{nj})$ to represent the consumer behavioral process based on the probability of any particular outcome as

$$P(y_{nj}|x_{nj}) = Prob(\epsilon_{ni} \text{ s.t. } h(x_{nj}, \epsilon_{nj}) = y_{nj}),$$

where the unobserved terms ϵ_{nj} are considered random with density $f(\epsilon)$. The probability that the consumer chooses a particular outcome from the set of all possible outcome is simply the probability that the behavioral process results in that outcome with such unobserved factors.

For the chosen alternative i, an indicator function $I[h(x_{ni}, \epsilon_{ni}) = y_{ni}]$ is defined, which takes the value of 1 when the statement in bracket is true and 0 otherwise. In particular $I[.] = 1$ if the value of ϵ_{ni}, combined with x_{ni} induces the agent to choose outcome y_{ni} with the chosen alternative i and $I[.] = 0$ if the value of ϵ_{ni}, combined with x_{ni} induces the agent to choose another outcome. Then the probability that the consumer choose outcome y_{ni} is simply the expected value of this indicator function, where the expectation is over all possible values of the unobserved factors

$$P(y_{ni}|x_{ni}) = \pi(I[h(x_{ni}, \epsilon_{ni}) = y_{ni}] = 1). \tag{2.4}$$

The distributions of error terms and the values of unobserved factors, in sum, are random and unknown. Nevertheless, through choice modelling, they can be approximated by the researchers depending on the specification of the choice situation and the calculation of observed utility V_{nj}. In most case, ϵ_{nj} are treated as random and assumed to be independently and identically distributed (i.i.d.). The i.i.d. assumption is part of the limitation in describing choice behavior, because choices in surveys and choice experiements are often made in different situations and time periods by the same decision maker, i.e. a panel data setting. In those cases, choice decisions may be identical but not independent. Thus, the error terms may be correlated, and may not be independently distributed over time and/or over alternatives. We will discuss this problem further in chapter 3 and chapter 5. In the next section, we will examine closely probabilities for choice.

2.2 Choice probabilities

The random Utility Models introduced in the previous section have two components: a systematic component denoted V_j and an unobserved component denoted ϵ_j. The second component is an error term that captures the impact of all unobserved variables on the utility of choosing a specific alternative j in a choice set. Therefore, we should note here that the choice is made by a random decision maker/consumer. From the researcher's point of view, the determinants of the decision maker's utility are partly unobserved. Choice models based on choice probabilities are built to fabricate the impacts of the error terms as well as of the determinants on consumer utility, as we have seen in the previous part. In this section, we will further describe choice probability for a chosen alternative i, which are derived from error terms in different utility functions for all alternatives in the choice set.

With a setting for modelling choice as in the previous section, beginning with equation 2.2 and 2.3 while ignoring the constant for simplicity, Croissant (2010) further clarifies utility levels of the consumer n for each alternative $j \in J$ as follows,

$$U_1 = V_1 + \epsilon_1 = \beta_1' x_1 + \epsilon_1$$
$$\vdots$$
$$U_J = V_J + \epsilon_J = \beta_J' x_J + \epsilon_J$$

Alternative i will be chosen if $\forall j \neq i$, $U_t(i) \geq U_t(j)$. We have the following $J-1$ conditions:

$$U_i - U_1 = (V_i - V_1) + (\epsilon_i - \epsilon_1) > 0$$
$$\vdots$$
$$U_i - U_{i-1} = (V_i - V_{i-1}) + (\epsilon_i - \epsilon_{i-1}) > 0$$
$$U_i - U_{i+1} = (V_i - V_{i+1}) + (\epsilon_i - \epsilon_{i+1}) > 0$$
$$\vdots$$
$$U_i - U_J = (V_i - V_J) + (\epsilon_i - \epsilon_J) > 0.$$

The general expression of the probability of choosing alternative i is then

$$(P_i|\epsilon_i) = P(U_i > U_1, ..., U_J)$$

$$(P_i|\epsilon_i) = F_{-i}(\epsilon_1 < V_i - V_1 + \epsilon_i, ..., \epsilon_J < V_i - V_J + \epsilon_i), \tag{2.5}$$

where F is the joint distribution of all error terms and F_{-i} is the multivariate distribution of $J-1$ error terms (i.e. all except ϵ_i). The unconditional probability for alternative i is then

$$P_i = \int (P_i|\epsilon_i) f(\epsilon_i) d\epsilon_i \tag{2.6}$$

$$P_i = \int F_{-i}(\epsilon_1 < V_i - V_1 + \epsilon_i, ..., \epsilon_J < V_i - V_J + \epsilon_i) f(\epsilon_i) d\epsilon_i, \tag{2.7}$$

where $f(\epsilon_i)$ is the marginal density function for the error term ϵ_i related to the chosen alternative i.

McFadden (1974) points out a challenge facing choice modellers in practice. It is particularly difficult to define a joint distribution F that allows the computation of probability P_i, such as the one in equation 2.4. He suggests an alternative approach for which researchers can specify formulas for the selection probabilities and then examine whether these formulas could be obtained via equation 2.4 from some distribution of utility maximizing consumers. Using this approach, researchers in the field have specified different choice models depending on varried choice contexts, utility maximizing behaviors and corresponding distribution for the density $f(\epsilon_i)$. For example, McFadden (1978) uses logit formula and Gumbel distribution to specify a multinomial logit model and analyze residential choice behavior.

Using function h as the generalization of utility function specifying the choice behavior, Train (2009) rewrites a general form for choice probability in which y_{ni} is the observed outcome for choosing alternative i over j as

$$P_i = \int I[h(x_{ni}, \epsilon_{ni}) = y_{ni}] f(\epsilon) d\epsilon, \tag{2.8}$$

where $f(\epsilon)$ is the joint density function for the vector of random error terms $\epsilon = (\epsilon_1, ..., \epsilon_J)$. Calculating choice probability is the first step to build models in order to explain choice behaviors. For certain specifications of density $f(\epsilon)$ for the error terms/ unobserved utility ϵ, the integral in equation 2.8 takes a closed form, which means that the choice probability can be calculated exactly. An example is when the unobserved portion of utility ϵ is assumed to be logit or nest logit. In some other cases, such as probit and mixed logit, the integral does not have a closed form and choice probability has to be calculated numerically.

2.3 Assumptions in discrete choice models

The distribution of the error terms $\epsilon = (\epsilon_1, ..., \epsilon_J)$ in discrete choice models, according to McFadden (1974), will rely on two hypotheses: one is about Gumbel distribution and the other one is about independent and identical distribution for the error terms.

Firstly, in order to have discrete choice models that are consistent with economic theory of utility maximization and at the same time have strong mathematical foundation, McFadden (1978) developed a process to generate Generalized extreme value (GEV) models for choice analysis. Let $Z_j \equiv e^{V_{nj}}$, and consider a function G that depends on Z_j for any alternative j in the set of J available alternatives. For the chosen alternative i, we have $Z_i \equiv e^{V_{ni}}$. Let G_i to be the derivative of G with respect to Z_i, such that:

$$G_i = \partial G / \partial Z_i,$$

$$G = G(Z_1, ..., Z_J).$$

If G satisfies four following conditions

1. For all $Z_j > 0 (\forall j), G \geq 0$,
2. G is homogenous of degree one, meaning: $G(\alpha Z_j) = \alpha G(Z_j)$ for any positive α,
3. $G \to \infty$ as $Z_j \to \infty$ for any j,

4. the cross partial derivatives of G change sign in a particular way, i.e. $G_i \geq 0$ for all i, $G_{il} = \partial G_i / \partial Z_l \leq 0$, and $G_{ilk} = \partial G_i l / \partial Z_k \geq 0$ for any distinct alternatives i, l, k and so on for higher-order cross-partials.

then we have the choice probability for the chosen alternative i as

$$P_i = \frac{Z_i G_i}{G} = \frac{Z_i}{G} \frac{\partial G}{\partial Z_i}.$$

Following this generating process, any function G that satisfies the four conditions can be used to derive a choice probability, which helps to constitute a GEV model. In the GEV model, each error term ϵ_j follows a *Gumbel distribution* with a location parameter μ and a scale parameter θ as

$$f(z) = \frac{1}{\theta} e^{-\frac{(z-\mu)}{\theta}} e^{-e^{-\frac{(z-\mu)}{\theta}}},$$

$$P(z < t) = F(t) = \int_{-\infty}^{t} \frac{1}{\theta} e^{-\frac{(z-\mu)}{\theta}} e^{-e^{-\frac{(z-\mu)}{\theta}}} dz = e^{-e^{-\frac{(t-\mu)}{\theta}}}. \tag{2.9}$$

The first two moments of the distribution are $E(z) = \mu + \theta\gamma$, where γ is the Euler-Mascheroni constant (0.577) and $V(z) = \frac{\pi^2}{6}\theta^2$ (McFadden 1974).

We have mentioned briefly the assumption of i.i.d. distribution for the error terms in 2.1. We will explain here the complexity of error terms in explaining unobserved utility. Train (2009) suggests two properties of choice probability that can affect the specification and estimation of any discrete choice model. These properties are

1. Only differences in utility matter, not the absolute utility level,
2. The scale of utility is arbitrary.

The first property implies that we can only estimate and identify coefficients capturing differences across alternatives. In other words, the overall scale of utility will not be identified when we account for the change in a decision maker's choice over alternatives. Recalling from equation 2.3 that k_j is a constant that is specific to alternative j which captures the average effect of all unincluded factors and has similar function to an intercept in a regression model. When alternative specific constants are included, the unobserved portion of utility ϵ_{nj} has zero mean by construction. We can then, without loss of generality, suppose that $\mu_j = 0 \; \forall j$. Indeed, when we do not have the constants k_j in the observed utility V_{nj} for the opposite case, the unobserved error terms ϵ_{nj} may have a nonzero mean. Suppose that from equation 2.2, we have a utility function

$$U^0_{nj} = V^0_{nj} + \epsilon^0_{nj} = \beta'_1 x_{nj} + \epsilon^0_{nj}, E(\epsilon^0_{nj}) \neq 0.$$

Adding the contants, as a result, may force the error to have zero mean

$$U^1_{nj} = V^1_{nj} + \epsilon^1_{nj} = \beta'_2 x_{nj} + k_j + \epsilon^1_{nj}, E(\epsilon^1_{nj}) = 0.$$

Thus, in discrete choice models, we have alternative specific constants acting like intercepts and at the same time capturing differences across alternatives.

Then the property *only differences in utility matter* can be translated to only differences in the alternative specific constants are relevant, not their absolute levels (Train 2009). This property often causes confusion in studying econometrics for choice modelling. And in discrete choice models, when researchers mention the technique of adding a constant or alternative specific constants, it means adding the intercepts.

The second property, *the scale of utility is arbitrary*, requires us to have a treatment for *scale heterogeneity*, i.e. variance of utility varries over different choice situations. Because utility has no unit or scale, the estimated coefficients are uniterpretable as numbers. This problem happens due to unidentified units for utility in discrete choice models and generally in economic theory. In order to solve the problem and account for scale differences between utilities gained from different choices, we have to rely on several techniques of normalizing the scale of utilities. The normalization process is, thus, considered as a way to account for scale heterogeneity as we specify, interpret, and compare different choice models (Greene and Hensher 2010) as well as different consumer groups (Swait and Louviere 1993) in chapter 5.

Vass et. al. (2017) clarify that the use of scale parameter λ in econometrics is similar to one of *dispersion parameter* in statistics. In particular, λ is explained to be inversely proportional to the variance of the error term σ^2. As scale parameter decreases, variance increases and the errors become more dispersed. As error variance increases, the random part of utility (ϵ_{nj}) becomes larger relative to the systematic part (V_{nj}), and choices become more random. Completely random choices occur when there is an equal probability of selecting any alternative.

According to Train (2009), a standard way of normalizing the scale of utility is to normalize the variance of the error terms in different choice models. When utility is multiplied by λ, the variance of each ϵ_{nj} changes by λ^2, or $Var(\lambda\epsilon_{nj}) = \lambda^2 Var(\epsilon_{nj})$. Therefore normalizing the variance of the error terms is equivalent to normalizing the scale of utility. In general, researchers can normalize the error variance to some number based on each model's assumptions about the error terms. For example, in the case of multinomial logit, the error terms are assumed to be i.i.d. and researchers can normalize the error variance of different multinomial logit models (with different explanatory variables) to 1, as we will explain in detail in the next paragraph. As all the errors in different multinomial logit choice models have the same variance by assumption, normalizing the variance of any error term in a particular choice model also sets the variance of other multinomial logit models. As we will discuss multinomial logit, nested logit, and mixed logit in chapter 3, we separate the assumptions about the error terms into three cases: i.i.d. errors, heteroskedastic errors, and correlated errors. Depending on assumptions for the error terms of each choice model following each of these cases, we have different approaches to normalize the variance of the error terms.

The first approach is used for models with error terms that are assumed to be independent and identically distributed with univariate distribution. We have

$$P(U_j - U_1) = F_1((V_j - V_1 + \epsilon_j)$$

$$P(U_j - U_2) = F_2((V_j - V_2 + \epsilon_j)$$

$$\vdots$$

$$P(U_j - U_J) = F_J((V_j - V_J + \epsilon_j),$$

where F_j is the cumulative density of ϵ_j. From the conditional choice probability in equation 2.5, we realize that $J-1$ error terms may be identified, and therefore, $J-1$ scale parameters θ in the Gumbel distribution in equation 2.9 have to be identified, one natural choice of normalization is to impose one of the θ_j to be equal to 1. For example, we have a utility function for individual n in a general case, when the variance has not been normalized as

$$U^0_{nj} = \beta' x_{nj} + \epsilon^0_{nj}, E(\epsilon^0_{nj}) = 0, Var(\epsilon^0_{nj}) = \sigma^2.$$

The researchers can normalize the scale of utility by setting the error variance to 1. We then have an equivalent model as

$$U^1_{nj} = \frac{\beta'}{\sigma} x_{nj} + \epsilon^1_{nj}, E(\epsilon^1_{nj}) = 0, Var(\epsilon^1_{nj}) = 1, \quad (2.10)$$

where the coefficients becomes β/σ due to the fact that the original coefficients β are divided by the standard deviation of the unobserved portion of utility σ. The new coefficient β/σ reflect the effect of the observed variables relative to the standard deviation of the unobserved factors.

In the second case of heteroskedastic errors, the variance of the error terms are different for different segments of population. We have the second approach for normalization. We can set the overall scale of utility by normalizing the variance or one segment, and then estimate the variance for each segment relative to the segment we have just normalized the variance. We can use the example of seafood choice again for this case. Suppose decision makers are divided into different groups of similar sociodemographic characteristics. Considering two groups, one group/segment above 45 years old (O) and another group under 45 (U), the model in its original form is

$$U_{nj} = \beta' x_{nj} + \epsilon^O_{nj} \text{ for above 45 years old,}$$

$$U_{nj} = \beta' x_{nj} + \epsilon^U_{nj} \text{ for under 45 years old.}$$

The variances for two groups are not the same. Suppose we have $k = Var(\epsilon^O_{nj})/Var(\epsilon^U_{nj})$, which is a rate between the two variances for the two groups of people. We can divide the utility for decision makers under 45 years old by $\sqrt{k}$ and do not have any impact on their choices (given the fact that absolute level of utility does not matter). The model is rewritten as

$$U_{nj} = \beta' x_{nj} + \epsilon_{nj} \text{ for above 45 years old}$$

$$U_{nj} = (\frac{\beta}{\sqrt{k}})' x_{nj} + \epsilon_{nj} \text{ for under 45 years old} \quad (2.11)$$

where variance of ϵ_{nj} is the same for all observations in two groups. The coefficient k has a function like a scale parameter in this case, which will be estimated along with β. The estimated value of $\hat{k}$ tells us the relative variance of unobserved factors in the group under 45 years old to compare with the variance of the other group.

In the third case where choice models have correlated errors, the normalization process is more complicated. Since in such case, normalizing the variance of the error for one alternative is not sufficient to set the scale of utility differences. We have an example for choices made over four alternatives, in which utility is expressed as

$$U_{nj} = V_{nj} + \epsilon_{nj}, j = 1, ..., 4.$$

The vector of error is

$$\epsilon_{nj} = \prec \epsilon_{n1}, ..., \epsilon_{n4} \succ$$

and is assumed to be normally distributed with zero mean and a covariance matrix

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ . & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ . & . & \sigma_{33} & \sigma_{34} \\ . & . & . & \sigma_{44} \end{bmatrix}. \tag{2.12}$$

Since only differences in utility matter, this model is equivalent to the model

$$\tilde{U}_{nj} = \tilde{V}_{nj} + \tilde{\epsilon}_{nj}, j = 2, 3, 4 \tag{2.13}$$

where

$$\tilde{U}_{nj} = U_{nj} - U_{n1}$$
$$\tilde{V}_{nj} = V_{nj} - V_{n1},$$

and the vector of error difference becomes

$$\epsilon_{nj} = \prec (\epsilon_{n2} - \epsilon_{n1}), (\epsilon_{n3} - \epsilon_{n1}), (\epsilon_{n4} - \epsilon_{n1}) \succ = \prec \tilde{\epsilon}_{n2}, \tilde{\epsilon}_{n3}, \tilde{\epsilon}_{n4} \succ$$

The variance of each error difference, thus, depends on the variance and covariance of the error that is normalized, which is in this case ϵ_{n1}. For instance, the variance of the difference between the first and second errors is

$$Var(\tilde{\epsilon}_{n21}) = Var(\epsilon_{n2} - \epsilon_{n1}) = Var(\epsilon_{n1}) + Var(\epsilon_{n2}) - Cov(\epsilon_{n2}, \epsilon_{n1}) = \sigma_{11} + \sigma_{22} - 2\sigma_{12}.$$

The covariance of $\tilde{\epsilon}_{n21}$ and $\tilde{\epsilon}_{n31}$ is calculated as

$$Cov(\tilde{\epsilon}_{n21}, \tilde{\epsilon}_{n31}) = E(\epsilon_{n2} - \epsilon_{n1})(\epsilon_{n3} - \epsilon_{n1}) = E(\epsilon_{n2}\epsilon_{n3} - \epsilon_{n2}\epsilon_{n1} - \epsilon_{n3}\epsilon_{n1} + \epsilon_{n1}\epsilon_{n1}) = \sigma_{23} - \sigma_{21} - \sigma_{31} + \sigma_{11}.$$

The covariance matrix for the vector of error differences, following the matrix in equation 2.12, becomes

$$\Omega = \begin{bmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14} \\ . & \sigma_{11} + \sigma_{33} - 2\sigma_{13} & \sigma_{11} + \sigma_{34} - \sigma_{13} - \sigma_{14} \\ . & . & \sigma_{11} + \sigma_{44} - 2\sigma_{14} \end{bmatrix}.$$

To set the scale of utility, normalization is done by setting the variance of one of the error differences to a number. Suppose we normalize the variance of $\tilde{\epsilon}_{n21}$ to 1, the covariance matrix for the error differences now has the form

$$\Omega = \begin{bmatrix} 1 & (\sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13})/m & (\sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14})/m \\ . & (\sigma_{11} + \sigma_{33} - 2\sigma_{13})/m & (\sigma_{11} + \sigma_{34} - \sigma_{13} - \sigma_{14})/m \\ . & . & (\sigma_{11} + \sigma_{44} - 2\sigma_{14})/m \end{bmatrix}, \quad (2.14)$$

where $m = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$. Utility is divided by $\sqrt{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}$ in order to achieve this scaling.

CHAPTER 3

Discrete choice models

3.1 Multinomial Logit

Logit choice probabilities

Suppose that the consumer n faces J alternatives and has a utility function U_{ni} by choosing alternative i. A multinomial logit choice model is obtained by assuming that each error term is independently and identically distributed (i.i.d.) with a Gumbel distribution (see section 2.3). The Gumbel distribution of each error term is also known as *Extreme Value type I* and has density $f(\epsilon_i) = e^{-\epsilon_i} e^{-e^{-\epsilon_i}}$ and cumulative distribution $F(\epsilon_i) = e^{-e^{-\epsilon_i}}$.

In order to derive choice probabilities for logit models, Croissant (2010) first simplifies the unconditional probabilities in equation 2.5 as a product of joint distribution F of all error terms, except for ϵ_i as

$$(P_i|\epsilon_i) = F_{-i}(\epsilon_1 < V_i - V_1 + \epsilon_i, ..., \epsilon_J < V_i - V_J + \epsilon_i) = \prod_{i \neq j} F(V_i - V_j + \epsilon_i) \quad (3.1)$$

The unconditional probabilities in equation 2.5 becomes

$$P_i = \int \prod_{i \neq j} F(V_i - V_j + \epsilon_i) f(\epsilon_i) d\epsilon_i, \quad (3.2)$$

where f is the marginal density with a Gumbel distribution. We then have the conditional probabilities of choosing alternative i over J alternatives as

$$(P_i|\epsilon_i) = Prob(\epsilon_j < V_i - V_j + \epsilon_i) = e^{-e^{-(V_i - V_j + \epsilon_i)}}. \quad (3.3)$$

The conditional probability of choosing i is then simply the product of probabilities in equation 3.1 for all the alternatives except for i

$$(P_i|\epsilon_i) = \prod_{i \neq j} e^{-e^{-(V_i - V_j + \epsilon_i)}}, \quad (3.4)$$

where the unconditional probability is the expected value of expression in equation 3.2 with respect to ϵ_i,

$$P_i = \int_{-\infty}^{+\infty} \prod_{i \neq j} e^{-e^{-(V_i - V_j + \epsilon_i)}} e^{-\epsilon_i} e^{-e^{-\epsilon_i}} d\epsilon_i = \int_{-\infty}^{+\infty} [\prod_{i \neq j} e^{-e^{-(V_i - V_j + \epsilon_i)}} e^{-e^{-\epsilon_i}}] e^{-\epsilon_i} d\epsilon_i.$$

The later has a better form, which is an intergral over probabilities of unobserved factors in utilities for all J alternatives in the choice set

$$P_i = \int_{-\infty}^{+\infty} \prod_{j=1}^{J} e^{-e^{-(V_i - V_j + \epsilon_i)}} e^{-\epsilon_i} d\epsilon_i. \tag{3.5}$$

Replacing $t = e^{-\epsilon_i}$ or $dt = -e^{-\epsilon_i} d\epsilon_i$ and following some algebra manipulation, the unconditional probability becomes

$$P_i = \int_{0}^{+\infty} e^{-t\Sigma_j e^{-(V_i - V_j)}} dt,$$

which has the closed form

$$P_i = - \frac{e^{-t\Sigma_j e^{-(V_i - V_j)}}}{\Sigma_j e^{-(V_i - V_j)}} \Bigg|_0^{+\infty} = \frac{1}{\Sigma_j e^{-(V_i - V_j)}}.$$

This is the logit choice probability for the chosen alternatives i over J alternatives

$$P_i = \frac{1}{\Sigma_j e^{-(V_i - V_j)}} = \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} = \frac{e^{\beta' x_i}}{\sum_{j=1}^{J} e^{\beta' x_j}}, \tag{3.6}$$

where the observed utility is specified to be linear: $V_i = \beta' x_i$. The vector β contains the estimated coefficients of the explanatory variables, reflecting the relationships between product attributes and the chosen outcome y_i. In the equation, x_i is a vector of explanatory variables or independent variables. These variables are often observable product attributes and consumer characteristics. As x_i's are specific to alternative $i \in J$, they are called alternative specific attributes (z_i). And as x_i's are specific to individual characteristics $n \in N$, they are called individual specific attributes (s_n). When individual specific attributes vary between N decision makers, we have a multinomial logit choice probability for the individual n as

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}}. \tag{3.7}$$

In assessing consumer's utilities for different alternatives, the researcher also has to take into account the property of *Independence from irrelevant alternatives* (IIA). For any two alternatives i and j, the ratio of the logit probabilities is

$$\frac{P_{ni}}{P_{nj}} = \frac{e^{V_{ni}}}{e^{V_{nj}}} = e^{V_{ni} - V_{nj}}. \tag{3.8}$$

This ratio must not depend on any alternatives other than i and j. It means that the relative probability of choosing i over j is the same no matter what other available alternatives as well as their attributes are. Since the probability ratio is independent from other alternatives than i and j, it is said to be independent from irrelevant alternatives and is known as the IIA condition in or property of choice modelling theory (Train 2009).

Normalization in multinomial logit models

Following the normalization process for the case of i.i.d. errors in equation 2.10 in section 2.3, researchers can choose any number to normalize error variances. Nevertheless, the error variances in multinomial logit model are traditionally normalized to $\pi^2/6$. The error variance in equation 2.10, thus, is $\pi^2/6$ rather than 1. The normalized scale of utility in multinomial logit model becomes

$$U^1_{nj} = \frac{\beta'}{\sqrt{\frac{\pi^2}{6}}} x_{nj} + \epsilon^1_{nj}, E(\epsilon^1_{nj}) = 0, Var(\epsilon^1_{nj}) = \pi^2/6. \tag{3.9}$$

The coefficients still reflect the variance of the unobserved portion of utility but these coefficients are scaled by a factor of $\sqrt{\frac{\pi^2}{6}}$.

Panel data setting in multinomial logit models

Suppose that we have a case in which a consumer has to make decisions over T choice situations and/or T time periods and J alternatives, known as a panel data setting. The utility obtained from alternative j in period t would be

$$U_{njt} = V_{njt} + \epsilon_{njt} \tag{3.10}$$

where ϵ_{njt} is the error terms. In this case, the chosen alternatives may change over choice situations, and so do the attributes z_{jt}. One may expect the unobserved error terms to be not independent over time as well as over alternatives, as in the case of the seafood data we will analyze in this thesis.

In some cases, the unobserved factors that affect the decision makers can be independent over repreated choices and multinomial logit models can be used to examine such panel data. The error terms ϵ_{njt} is then assumed to be i.i.d. extreme value, i.e. independent and identically distributed over n, j, and t. Using the same proof in section 3.1, the choice probabilities become

$$P_{nit} = \frac{e^{V_{nit}}}{\sum_{j=1}^{J} e^{V_{njt}}}, \tag{3.11}$$

where each choice situation and/or time period t by each decision maker n becomes a separate observation. We can proceed in two ways to analyze the data. In the first way, if utility for each period t is specified to depend only on variables for that period, then $V_{njt} = \beta' x_{njt}$ where x_{njt} are explanatory variables related to alternative j and individual n in period t. In this case, the multinominal logit model for panel data is not very different from cross-sectional data. In the second way, utility in each period can be specified to depend on observed variables from other periods. For example, utility in period t can be specified to be dependent on observed variables from the previous period $t-1$. A lagged independent variable $x_{nj(t-1)}$ (i.e. attributes of the chosen alternative in the previous period) can be used to define the observed utility in period t. For example, we can use a lagged price of the previously chosen alternative ($Price_{nj(t-1)}$) to be the lagged independent variable. The observed utility function in this case becomes

$$V_{njt} = V_{nit} = Price_{nj(t-1)} + \beta' x_{nit}, \tag{3.12}$$

where the lagged price acts like an intercept in our model. The previously chosen alternative can be any alternative in J alternatives while x_{nit} are attributes of the chosen alternative i in this period. Or we can use a lagged dependent variable $y_{ni(t-1)}$. The observed utility becomes

$$V_{njt} = V_{nit} = \alpha y_{nj(t-1)} + \beta' x_{nit}, \tag{3.13}$$

where $y_{nj(t-1)} = 1$ if individual n also chooses i in the previous period $t-1$ and 0 otherwise. If $\alpha > 0$, the utility of choosing alternative i in the current period will be higher if alternative i is also chosen in the previous period. If $\alpha < 0$, the utility in the current period will be higher if the individual switched to another alternative, or the previous choice must be different from alternative i.

The assumption of independent errors over choice situations or over time periods is, nevertheless, unreasonable. We will discuss the cases of panel data further in 3.3. In the next sections of this part, we will point out several important analytical aspects of multinomial logit as a starting point for the discussion of other choice models.

Interpreting discrete choice models' results

3.4.1 Marginal effects

While the coefficients in a linear model can be directly interpreted as marginal effects of the explanatory variables on the outcome y, coefficients in discrete choice models need to be transformed for analysis. To obtain meaningful results as marginal effects in multinomial logit model, for example, the coefficients for alternative specific attributes (z_j), individual specific attributes (s_n) and the price ($Price$) need to be transformed separately. In particular, the marginal effect of an individual specific attribute is

$$\frac{\partial P_{ni}}{\partial s_n} = P_{ni}(\beta_i - \sum_{j=1}^{J} P_{nj}\beta_j)$$

The sign of the marginal effect is given by $(\beta_i - \sum_{j=1}^{J} P_{nj}\beta_j))$, which is positive if the coefficient for the alternative i is greater than the weighted average of the coefficients for all the alternatives (Croissant 2010).

For an alternative specific attribute defining two different alternatives i and j, the marginal effect is

$$\frac{\partial P_{ni}}{\partial z_{ni}} = \gamma P_{ni}(1 - P_{ni})$$
$$\frac{\partial P_{ni}}{\partial z_{nj}} = -\gamma P_{ni} P_{nj}$$

Thus, the sign of the coefficients for alternative specific variables are directly intepretable. In this case, the marginal effect is obtained by multiply the coefficient by the product of two probabilities (which is $-\gamma 0.5 \times 0.5 = -1/4\gamma$).

3.4.2 Marginal rate of substitution

The term *marginal rate of substitution* is often used in economic models in order to interpret the marginal effect of price on the chosen outcome. Suppose we have $\hat{\beta}_{nAttribute}$ and $\hat{\beta}_{nPrice}$ as the estimated coefficients of an alternative specific attribute (*Attribute*) and price (*Price*). The marginal rates of substitution between product attributes are given as

$$WTP_n = -\frac{\hat{\beta}_{nAttribute}}{\hat{\beta}_{nPrice}} \tag{3.14}$$

where WTP is the willingness to pay level of person n. Ratios of coefficients such as the one in equation 3.14 usually provide economically meaningful information. For example, in the case of two seafood types with sustainable ecolabel and without ecolabel, the ratio between coefficients of ecolabel and seafood price enacts the rate of substitution for seafood with and without ecolabel. This ratio, in other words, reflects the WTP of person n for a desired seafood attribute (in this example, the ecolabel). Consumer willingness to pay is one way to assess different choices by saying that who wants the product the most will pay the highest price for that product. In orther words, utility brought by the chosen alternative with that specific attribute compared to utilities brought by other alternatives is the highest.

3.4.3 Matching choice probabilities

One way to assess the predictive power of a choice model is to see how closely the estimated average probabilities by the model (*average probabilities*) match with the shares of chosen alternatives in the experiment. The shares of chosen alternatives are the probabilities for chosen alternatives in practice, which are known as *frequencies of alternatives* or *the shares of consumers choosing each alternative*. When the model predicts the average probabilities that are close to the frequencies of alternatives in the experiment, we can say the predictive power of the model is high. And if the chosen participants are randomly selected in the experiment (which differs from the case of self-selected participant problem), we can make the inference that these predicted probabilities represent the market shares for product alternatives. When a new product alternative or its attributes are presented, we can use the good model to predict/estimate the market share for this new alternative or product attributes. This is often the reason why a choice experiment is carried out in the first place.

3.2 Nested logit models

The nested logit model is one of the well-known members of the GEV family for choice analysis, especially in choice studies of transports and vehicles. Nested logit models are developed to relax the assumption that unobserved factors ϵ_j are mutually independent (or i.i.d.). Each error term ϵ_j has an extreme-value type I distribution. A nested logit model is appropriate for a choice context when the set of alternatives J faced by a consumer can be divided into k non-overlapping subsets (denoted $B_1, B_2, ..., B_K$), called nests, so that the following properties hold

1. For any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes or existence of all other alternatives. That is the IIA property holds within each nest;
2. For any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests, i.e. the IIA property does not hold in general for alternatives in different nests.

These two properties imply that the unobserved utility ϵ_j's can be correlated within nests. For any two alternatives i and j in nest B_k, ϵ_i is correlated with ϵ_j. But for any two alternatives in different nests, the unobserved utility is still uncorrelated, $Cov(\epsilon_i, \epsilon_m) = 0$ for any alternative $m \in$ nest B_l, and alternative $i \in$ nest B_k, $k \neq l$ $(i\&m = 1, ..., J; k\&l = 1, ..., K)$. This kind of setting is called a two-level nested logit model. The first level is for choices over nests and the second level is for choices over alternatives within a nest.

In a nested logit model, vector of unobserved utility $\epsilon_j = (\epsilon_1, ..., \epsilon_J)$ is assumed to have cumulative distribution as

$$F(\epsilon_j) = exp(-\sum_{k=1}^{K}(\sum\nolimits_{j \in B_k} e^{-\epsilon_j/\lambda_k})^{\lambda_k}), \tag{3.15}$$

where λ_k is a measure of the degree of independence in unobserved utility among the alternatives in nest k. The unobserved factors of utility are less correlated and more independent if λ_k has a higher value. The quantity $1 - \lambda_k$ is a measure of correlation, as an increasing λ_k indicates less correlation. In some literature, λ_k is known as a log sum coefficient.

In order to derive a nested logit probability for alternative i among a set of j alternatives in nest B_k, we assume a logit structure for this probability, as the one in equation 3.6, instead of using the cumulative distribution of the unobserved utility $F(\epsilon_j)$. As in the case of i.i.d. ϵ_j, we have a logit probability for choosing alternative i in nest B_k as

$$P_{ki} = \frac{e^{V_{ki}}}{\sum\limits_{k=1}^{K}\sum\limits_{j=1}^{J} e^{V_{kj}}} = \frac{e^{\beta' x_{ki}}}{\sum\limits_{k=1}^{K}\sum\limits_{j=1}^{J} e^{\beta' x_{kj}}}. \tag{3.16}$$

However, it is possible that unobserved utility ϵ_j are not i.i.d. In nested logit models, this part of utility can be correlated within a nest. We then need to derive another form of probability for nested logit, which is an intersection probability. The intersection probability of alternative i in nest B_k is

$$P(A \cap B) = P(A|B)P(B)$$

Replacing A by alternative i and B by nest B_k, we have the probability for choosing $i \in B_k$

$$P(i \cap B_k) = P_{(i|B_k)} P_{(B_k)}$$

Written in this form, the nested logit probability can be decomposed into two logit probabilities: $P_{(B_k)}$ - a marginal probability of choosing an alternative in nest B_k and $P_{(i|B_k)}$ - a conditional probability of choosing alternative i given that an alternative in nest B_k is chosen (McFadden 1978). From equation 3.16 we have

$$P_{(i|B_k)} = \frac{e^{V_{ki}}}{\sum_{j=1}^{J} e^{V_{kj}}} \tag{3.17}$$

$$P_{(B_k)} = \frac{\sum_{i=1}^{J} e^{V_{ki}}}{\sum_{l=1}^{K} \sum_{m=1}^{M} e^{V_{lm}}}. \tag{3.18}$$

Suppose that we can further decompose the observed utility V_{kj} into two parts; one part is W_k - a constant utility for all alternatives in nest k, and another part is Y_{kj} - varried utility over alternatives within nest k, such that

$$V_{kj} = W_k + Y_{kj} = \gamma' w_k + \beta' x_{kj} \tag{3.19}$$

where w_k is a matrix of independent variables associated with nest k; γ, β are coefficients corresponding with observed nests' characteristics and the alternatives' attributes. The constant utility W_k differs over nests, but not over alternatives within each nest. For any W_k , Y_{kj} is defined as $V_{kj} - W_k$, which is the log of the denominator of the conditional logit probability.

In order to obtain the joint probability for the conditional and marginal probabilities in 3.17 and 3.18, Ben-Akiva (1973) suggests a quantity I_{nk} called inclusion value. This quantity is defined as a link between the two probabilities so that $\lambda_k I_k$ is the expected utility a decision maker receives from the choice among J alternatives in nest B_k. In particular, we have

$$I_k = ln \sum_{j=1}^{J} e^{V_{kj}}, \tag{3.20}$$

then a formula for the conditional probability is derived as

$$P_{(i|B_k)} = \frac{e^{V_{ki}}}{\sum_{j=1}^{J} e^{V_{kj}}} = \frac{e^{\beta' x_{ki}}}{e^{I_k}} = \frac{e^{Y_{ki}}}{e^{I_k}}. \tag{3.21}$$

Writen in this form, the conditional probability is slightly different from the formula suggested by McFadden (1978), which uses a scaling factor λ as

$$P_{(i|B_k)} = \frac{e^{Y_{ki}/\lambda_k}}{\sum_{j=1}^{J} e^{V_{kj}}} = \frac{e^{Y_{ki}/\lambda_k}}{e^{I_k}}. \tag{3.22}$$

Recalling that λ_k reflects the correlation among unobserved factors in nest k (correlation $= 1 - \lambda_k$), the scaling factor can differ over nests. Models using λ are also called "normalized nested logit" and are consistent with RUM setup, i.e. by scaling the coefficient of utility within each nest, the model allows utility to be compared across nests. In other words, without the scaling, utilities can only be compared for alternatives within the same nest. Having the option of choosing alternative m in another nest B_l, the marginal probability of nest B_k is then

$$P_{(B_k)} = \frac{\sum_{i=1}^{J} e^{V_{ki}}}{\sum_{l=1}^{K} \sum_{m=1}^{J} e^{V_{lm}}} = \frac{e^{\gamma' w_k} [\sum_{i=1}^{J} e^{\beta' x_{ki}}]}{\sum_{l=1}^{K} e^{\gamma' w_l} [\sum_{m=1}^{J} e^{\beta' x_{lm}}]} = \frac{e^{\gamma' w_k + I_k}}{\sum_{l=1}^{K} e^{\gamma' w_l + I_l}}. \tag{3.23}$$

Following the logic of scaling, instead of using the general observed utility V_{ki}, Train (2009) use the varried utility Y_{kj} and the degree of independence λ_k in unobserved utility among alternatives within a nest to define the inclusion value as

$$I_k = ln \sum_{j=1}^{J} e^{Y_{kj}/\lambda_k}. \tag{3.24}$$

Using the decomposition of known utility V_{kj}, the conditional and marginal probabilities then take the forms

$$P_{(i|B_k)} = \frac{e^{Y_{ki}/\lambda_k}}{\sum_{j=1}^{J} e^{Y_{kj}/\lambda_k}},$$

$$P_{(B_k)} = \frac{e^{W_k + \lambda_k I_k}}{\sum_{l=1}^{K} e^{W_l + \lambda_l I_l}}.$$

The choice probability for nested logit models is

$$P_i = \frac{e^{Y_i/\lambda_k}}{\sum_{j \in B_k} e^{Y_j/\lambda_k}} \frac{e^{W_k + \lambda_k I_k}}{\sum_{l=1}^{K} e^{W_l + \lambda_l I_l}} \tag{3.25}$$

After some steps of mathematical manipulation (see Train 2009, p. 86 for more detailed calculations), we come to the well known form for nested logit probability

$$P_i = \frac{e^{V_i/\lambda_k} (\sum_{j \in B_k} e^{V_j/\lambda_k})^{\lambda_k - 1}}{\sum_{l=1}^{K} (\sum_{j \in B_l} e^{V_j/\lambda_l})^{\lambda_l}}. \tag{3.26}$$

One advantage of the nested logit model is that choice probabilities take a closed form and can be estimated without relying on numerical approximations. In other words, standard maximum likelihood techniques can be used for estimation (which will be discussed in section 4.1). Until recently, only a small portion of nested logit models has ever been implemented to consumer choice problems. This is also the case for seafood choice studies. In their systematic literature review of purchasing behavior of fish and seafood products, Carlucci et. al. (2015) show no relevant study that uses nested logit. Nevertheless, Train (2009) believes that researchers may find more potentially powerful models to the consumer choice analysis in this model class. Hensher and Greene (2002), for example, have developed a three-level nested logit model. In this thesis, we attempt to analyze consumer seafood choice with nested logit models in section 5.4.

3.3 Mixed logit models

Error component approach

The i.i.d. asumption in discrete choice modelling is restrictive because it does not allow unobserved error terms of differerent alternatives to be correlated. In order to take this correlation into account, we can decompose ϵ_{nj} into two parts with a joint density $f(\epsilon_{nj}) = f(\epsilon_{nj1}, \epsilon_{nj2})$. The error terms, thus, are conveniently partitioning into one part that is heteroskedastic and correlated over alternatives and another part that is i.i.d. over alternatives as well as individuals. The utility function in equation 2.2 is rewritten as

$$U_{nj} = V_{nj} + \epsilon_{nj1} + \epsilon_{nj2} = \beta' x_{nj} + \epsilon_{nj1} + \epsilon_{nj2}, \tag{3.27}$$

where ϵ_{nj1} is a random term with zero mean whose distribution over individuals and alternatives depends on underlying parameters and observed data relating to alternative j and individual n; ϵ_{nj2} is a random term with zero mean that is i.i.d. over alternatives and does not depend on underlying parameters or data (Hensher and Greene 2003).

The joint density can be expressed as a product of a marginal and a conditional density $f(\epsilon_{nj1}, \epsilon_{nj2}) = f(\epsilon_{nj2} \mid \epsilon_{nj1}) f(\epsilon_{nj1})$ (Train 2009). And the choice probabiltiy in equation 2.1 can then be rewritten as

$$P_j = \int_{\epsilon_{nj1}} [\int_{\epsilon_{nj2}} I(h(x, \epsilon_{nj1}, \epsilon_{nj2}) = y) f(\epsilon_{nj2} \mid \epsilon_{nj1}) d\epsilon_{nj2}] \underbrace{f(\epsilon_{nj1}).d\epsilon_{nj1}} \tag{3.28}$$

The integral in bracket has a closed form, which helps to calculate exactly the integral over ϵ_{nj2} given ϵ_{nj1}. This is an advantage because we can use the derived choice probability in multinomial logit model in 3.1, which has a closed form. Based on an assumption of i.i.d. extreme value for ϵ_{nj2}, we can derive a logit form for the conditional probability of the chosen alternative i for a given value of ϵ_{nj1} following equation 3.7. Thus, mixed logit models assumes an i.i.d. extreme value type 1 distribution for ϵ_{nj2} and a general distribution for ϵ_{nj1}. It means that ϵ_{nj1} can take on a number of distributional forms such as normal, lognormal or other forms depending on researchers' assumption for their choice models. Denote the density of ϵ_{nj1} by $f(\epsilon_{nj1} \mid \theta)$ where θ are the fixed parameters of the assumed distribution of ϵ_{nj1}. We have the conditional probability for choosing alternative i as

$$(P_{ni} \mid \epsilon_{nj1}) = \frac{e^{(\beta_n' x_{ni} + \epsilon_{ni1})}}{\sum_{j=1}^{J} e^{(\beta_n' x_{nj} + \epsilon_{nj1})}}. \tag{3.29}$$

Moreover, learning from equation 2.6, we can derive the unconditional probability as

$$P_{ni} = \int (P_{ni} \mid \epsilon_{nj1}) f(\epsilon_{nj1} \mid \theta) d\epsilon_{nj1} = \int \frac{e^{(\beta_n' x_{ni} + \epsilon_{ni1})}}{\sum_{j=1}^{J} e^{(\beta_n' x_{nj} + \epsilon_{nj1})}} f(\epsilon_{nj1} \mid \theta) d\epsilon_{nj1}. \tag{3.30}$$

Thus, the unconditional probability would be a multinormial logit probability integrated over all value of ϵ_{nj1} weighted by the density of ϵ_{nj1} as

$$P_{ni} = \int L_{ni}(\epsilon_{nj1} \mid \theta) f(\epsilon_{nj1} \mid \theta) d\epsilon_{nj1}, \tag{3.31}$$

where L_{ni} denotes a logit form for the choice probability. Models in this form are called mixed logit because the choice probability P_{ni} is a mixture of logits with $f(\epsilon_{nj1} \mid \theta)$ as the mixing distribution. The mixing distribution has a set of parameters θ depending on the assumption about the general distribution of the random error term ϵ_{nj1}. Thus, the probability of the chosen alternative i will be dependent on the choice of θ by the researchers. The probabilities, nevertheless, satisfy the IIA property and different substitution patterns may be obtained by appropriate specification of f. This can be done in two ways. The first way is to treat the unobserved information as a single separate error component in the random component, known as *error component approach*. We have just addressed this approach, starting from equation 3.27. The second way is to specify each element of β_n in equation 3.29 associated with attributes of an alternative as having both a mean and a standard deviation, known as *random coefficient approach*. The random coefficient specification is widely used in recent applications of mixed logit models, and will be described in the next part.

Random coefficient approach

Suppose that utility of person n is derived from his/her behavior of choosing alternative i among J alternatives as

$$U_{ni} = \beta_n' x_{ni} + \epsilon_{ni}, \tag{3.32}$$

where x_{ni} are observed variables that are related to both attributes of alternative i and characteristics of person n, β_n is a vector of coefficients of these variables representing the person's tastes, and ϵ_{ni} is a random term that is i.i.d. extreme value. The coefficients β_n vary over decision makers in the population with density $f(\beta_n \mid \theta)$, which is a function of fixed parameters θ that present the mean and covariance of the β_n's in the population. To compare with multinomial logit models, the specification is similar except that β_n varies over decision makers rather than being fixed. This is why we use "random coefficients" rather than "coefficients" in general. In mixed logit models with random coefficient approach, the researcher observes the x_{ni}'s but not β_n or the ϵ_{ni}'s. As such, these two components are treated as stochastic influences. For each value of β_n, the conditional choice probability for the chosen alternative i is logit based on the i.i.d. assumption of extreme value for ϵ_{ni}. Thus, we specify another form for equation 3.29 as

$$P_{ni}(\beta = \beta_n) = L_{ni}(\beta_n \mid \theta) = \frac{e^{\beta_n' x_{ni}}}{\sum_{j=1}^{J} e^{\beta_n' x_{nj}}}. \tag{3.33}$$

The unconditional choice probability would be the logit L_{ni} integrated over all values of β_n and weighted by the density $f(\beta_n \mid \theta)$. This leads us to a well-known mixed logit choice probability (with some modification which we believe to be necessary)

$$P_{ni} = \int L_{ni}(\beta_n \mid \theta) f(\beta_n \mid \theta) d\beta_n = \int \frac{e^{\beta_n' x_{ni}}}{\sum_{j=1}^{J} e^{\beta_n' x_{nj}}} f(\beta_n \mid \theta) d\beta_n. \tag{3.34}$$

Since parameters θ have unknown distribution, selecting the right specification for the density $f(\beta_n \mid \theta)$ is empirically challenging. Hensher and Greene (2003) list several empirical issues that they have investigated in their paper about mixed logit models, which include

1. selecting the random coefficients,
2. selecting the distribution of the coefficients,
3. selecting the number of points on the distributions,
4. preference heterogeneity around the mean of a random coefficient,
5. correlated choice situations for observations drawn from the same individual,
6. correlation between coefficients,
7. willingness to pay challenges.

For the first four empirical issues, researchers can generally specify a distribution for the random coefficients and estimates the parameters of that distribution depending on choice contexts and situations. For most applications, the density $f(\beta_n \mid \theta)$ is specified to be continuous (Train 2009). For instance, a normal distribution $\phi(\beta_n \mid b, \Omega)$ with mean b and covariance Ω is assumed to be the density and the choice probability for normal distributed mixed logit models becomes

$$P_{ni} = \int \frac{e^{\beta_n' x_{ni}}}{\sum_{j=1}^{J} e^{\beta_n' x_{nj}}} \phi(\beta_n \mid b, \Omega) d\beta_n. \tag{3.35}$$

In other contexts and situations, the researcher may choose gamma, lognormal or uniform to specify the mixing distribution of $f(\beta_n \mid \theta)$, instead of using a normal distribution.

If the mixing distribution $f(\beta_n \mid \theta)$ is specified to be discrete, in which the random coefficients β_n take a finite set q of distinct values labelled $b_1, ..., b_Q$ with probability ρ_{nq} that $\beta_n = b_q$, mixed logit becomes a latent class model. This model relaxes the requirement that specific assumptions about the density $f(\beta_n \mid \theta)$ have to be specified. In latent class model, choice probabilities is derived from the logit probabilities in equation 3.33 as

$$P_{ni} = L_{ni}(\beta_n) = \sum_{q=1}^{Q} \rho_{nq}(\frac{e^{b'_q x_{ni}}}{\sum_{j=1}^{J} e^{b'_q x_{nj}}}). \tag{3.36}$$

Comparing the two approaches for mixed logit models

According to Train (2009), the two approaches, random coefficients and error components, are equivalent. Under the random coefficient specification, β_n can be decomposed into their mean μ and deviation v_n (or deviation from the mean μ) so that utility in equation 3.32 is specified

$$U_{ni} = \mu' x_{ni} + v'_n x_{ni} + \epsilon_{ni}, \tag{3.37}$$

with x_{ni} as observed variables. Under the error component specification, utility in equation 3.27 may be specified as

$$U_{ni} = \mu' x_{ni} + v'_n z_{ni} + \epsilon_{ni}, \tag{3.38}$$

where x_{ni} and z_{ni} are vectors of observed variables relating to alternative i. In this case, vector v_n comprises random terms with zero mean while vector μ has fixed coefficients. The error terms ϵ_{ni} are assumed to be i.i.d. extreme value type 1. We can infer from this specification that $v'_n z_{ni} + \epsilon_{ni} \equiv \epsilon_{nj1} + \epsilon_{nj2}$. In this way, models have fixed coefficients for variables x_{ni} and random coefficients with zero means for variable z_{ni}. Thus, correlation over alternatives (if any) in the error component approach depends on the specification of observed variables z_{ni}.

Panel data in mixed logit models and the heterogeneity problem

In mixed logit models, preference heterogeneity is often an issue in a panel data setting (N persons with T choice situations and/or time periods). The heterogeneity problem due to correlated random effects can happen in two scenarios:

- correlation over alternatives or correlation over attributes of alternatives;
- correlation over choice situations and/or time periods made by the same individual.

Both of these are likely the case for our seafood data. In other words, there is possible correlation among choices made in different situations, as a decision maker is given two or five choice sets. Following Hensher and Greene (2003), to address correlated choice situations and correlated random coefficients in the previous section (empirical issue no. 5 and 6), we chose the random coefficient approach and specify a structure for the (likely) correlated random coefficients β_{nt} for person n in choice situation t as

$$\beta_{nt} = \mu + Lv_{nt}, \tag{3.39}$$

where the underlying parameters are: $\theta = (\mu, v_{nt})$. In this specification, μ is the mean of β_{nt}, v_{nt} is a set of correlated random components with variance on the diagonals of a covariance matrix Ω, and L is a lower triangular matrix so that $\text{Var}[\beta_{nt}] = L\Omega L'$.

Furthermore, we specify the utility obtained from the chosen alternative i for this person as

$$U_{nit} = \mu' x_{nit} + v'_{nt} x_{nit} + \epsilon_{nit} \tag{3.40}$$

This utility function is not different from the one in equation 3.37 except that the subscript t is now added to describe choice situations.

Correlation over random coefficients of alternative specific variables can happen because v_{nt} (as being the deviation from the mean μ) may be the same for all alternatives in the first scenario. We can think of such a situation when different alternatives have similar attributes or when alternative specific variables are invariant. The covariance matrix for unobserved utility in this cenario is $\text{Cov}[v'_{nt}x_{nit} + \epsilon_{nit}, v'_{n(t+1)}x_{ni(t+1)} + \epsilon_{ni(t+1)}] = \sigma^2(v_{nt})$ where $\sigma^2(v_{nt})$ is the variance of v_{nt}.

Correlation over random coefficients of individual specific variables may appear for each alternative in the shape of unchanged socio-demographic variables. These variables are also known as individual specific variables. In this case, covariance matrix is more complex, since we have to specify the random coefficient as β_{nt} and assume it to be i.i.d. over choice situations for each individual. This is often a unreasonable assumption. Even though the random coefficients β_{nt} may be identical across T choice situations/time periods as preferences vary, they are not independent with each other, because the preferences are given by the same individual n in this case. This leads to the conclusion by Hensher and Greene (2003) that a trade-off between the two types of correlation exists in mixed logit models. In particular, by using empirical evidences from analyzing different datasets, they show that mixed logit can accommodate correlation over alternatives but not over choice situations by just using the i.i.d. assumption. And according to these two researchers, correlation over choice situations (but not over alternatives) may be resolved by using alternative specific constants, i.e. using intercepts in our models.

In the presence of correlation due to choices made in different choice situations/choice sets, as in the case of our seafood data (which will be described in chapter 4), the utility that decision maker n obtains from alternative i in the choice set t is

$$U_{nit} = \beta'_{nt} x_{nit} + \epsilon_{nit}$$

where β_{nt} is a vector of random coefficients. One would expect ϵ_{nit} to be correlated over choice sets as well as over alternatives. Denote the vector of errors for all chosen alternatives in T choice sets as

$$\epsilon_{nit} = \prec \epsilon_{n11}, ..., \epsilon_{nJ1}, \epsilon_{n12}, ..., \epsilon_{nJ2}, ..., \epsilon_{n1T}, ..., \epsilon_{nJT} \succ$$

where the density of ϵ_{nit} belongs to a distribution of assumption with zero mean. Instead of having a normal estimation procedure for the random coefficients and their variances, Train (2009) and later, Bliemer and Rose (2013) specify a method to derive a covariance matrix Ω for the correlated random coefficients based on *Cholesky decomposition.* This covariance matrix has dimension $JT \times JT$. Compared to the logit probability for one time period, the integral in equation 3.34 is expanded to be over JT dimension rather than J.

The distribution of random coefficients $f(\beta_{nt})$ may have a normal distribution with mean b_{nt} and covariance Ω. From equation 3.39, we obtain a normal multivariate β_{nt} with mean b_{nt} and variance σ^2_{nt} for the random coefficients of explanatory variables x_{ni} as

$$\beta_{nt} = b_{nt} + \sigma_{nt}\eta_{nt} \tag{3.41}$$

where η_{nt} is a draw from a standard normal distribution $N(0, 1)$.

In order to analyze the survey data of seafood and address correlated random coefficients β_k with K elements, we use β_k as a vector of random coefficients corresponding to attributes of seafood alternatives (or alternative specific variables). Random coefficients β_k are assumed to be normal distributed with mean b_k and covariance matrix Ω: $\beta_k \sim N(b_k, \Omega)$. In order to test whether these random coefficients are correlated, we derive the covariance matrix and look at the non-diagonal elements on this matrix. If these elements are non-zero, we can conclude that the random coefficients are correlated. For calculating the elements of the covariance matrix, a Cholesky factor of Ω, which is defined as a lower-triangular matrix L such that $LL' = \Omega$, is derived. The matrix L is often called the *generalized square root* of Ω (Train 2009). According to Bliemer and Rose (2013), a draw of β_k from $N(b, \Omega)$ is obtained as

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_K \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ ... \\ b_K \end{bmatrix} + \begin{bmatrix} s_{11} & 0 & ... & 0 \\ s_{21} & s_{22} & ... & 0 \\ ... & ... & ... & 0 \\ s_{K1} & s_{K2} & ... & s_{KK} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ ... \\ \eta_K \end{bmatrix}. \tag{3.42}$$

We modify the steps used to derive a K-dimensional vector of error terms (in a probit model) by Train (2009) to obtain the K-dimensional vector of random coefficients β_n as follows:

1. Take K draws from a standard normal and label the vector of these draws

$$\eta = (\eta_1, ..., \eta_K)'.$$

2. By using equation 3.42, each single random coefficient can be calculated as

$$\beta_k = b_k + \sum_{c=1}^{K} s_{kc}\eta_k,$$

where β_k is normally distributed (because the sum of normals is normal). The estimated parameters $\hat{\theta}_k$ of the random coefficients β_k have a mean b_k as

$$E(\beta_k) = b_k + LE(\eta_k) = b_k,$$

and the variance

$$Var(\beta_k) = E(L\eta_k(L\eta_k)') = LE(\eta_k\eta_k')L' = LVar(\eta_k)L' = LIL' = LL' = \Omega,$$

which is also the covariance matrix Ω we want to derive. The diagonal elements of this matrix specify variances of random coefficients

$$Var(\beta_k) = s_{11}^2 + s_{22}^2 + ... + s_{kk}^2 = \sigma_1^2 + \sigma_2^2 + ... + \sigma_k^2.$$

The square roots of these diagonal elements specify the standard deviations of the random coefficients. Futhermore, the non-diagonal elements of the covariance matrix will specify the covariances between random coefficients. For example, the covariance between β_1 and β_2 would be

$$Cov(\beta_{n1}, \beta_{n2}) = s_{11}s_{21} = \sigma_{11}\sigma_{21}.$$

As mentioned above, if the covariance matrix Ω has non zero covariances (or the non-diagonal elements are not zero), the random coefficients are correlated. Thus β_k no longer just depends on only η_k but also $\eta_1, \eta_2, ..., \eta_K$. The K random coefficients will depend on $K + 1$ distributional parameters $\hat{\theta}_K = (b_k, s_{K1}, ..., s_{KK})$. And estimation of the vector of random coefficients produces not only the parameter estimates, $\hat{\theta}_k$, but also yields an asymptotic covariance matrix Ω.

We will see the method mentioned above more clearly in chapter 5, section 5.4 for the seafood data. In this particular dataset, we have three random coefficients that are correlated. Therefore, we need a three-dimensional vector of correlated random coefficients and a covariance matrix for efficient estimation of mixed logit model as well as for correct interpretation of WTP (which is, in fact, the empirical issue no. 7 we have mentioned in the previous section).

Thus, for a general case with K-dimension random coefficients, the Cholesky factorization expresses K correlated terms as arising from K dependent components, with each component loading differently onto each random coefficient. For any pattern of covariance, there is some set of loadings M from dependent components that reproduces that covariance. From the specification in equation 3.39, we can infer that μ is now a k^{th} dimension vector of mean coefficients estimated from the relations between choices and observed attributes as well as consumer characteristics. Vector v_{nt} is a m^{th} dimension vector of correlated random variables that are dependently distributed with standard normal density and Ω is a $K \times M$ matrix of factor loadings (McFadden and Train 2000).

In sum, any behavioral specification whose derived choice probabilities take the particular form in equation 3.34 is called a mixed logit model. Each derivation depends on a variety of behavioral specifications, and the derived choice probability provides a particular interpretation of consumer behavior in a market, as we will see in the next chapter. One advantage of mixed logit models is that they allow preferences to vary across individuals and pairs of alternatives to be correlated through specifying a random error term. For estimation, the choice probabilities in mixed logit models cannot be calculated exactly but have to be approximated through numerical approximation, which will be discussed in the next chapter.

3.4 Latent Class models

Latent class models (LCM) is a special case of mixed logit, with mixing distribution to be specified as discrete. In these models, individuals are sorted implicitly into a set of Q classes with a probability ρ_q that is unknown to researchers. Suppose that individual n with a set of observable charateristics called s_n has a probability ρ_{nq} to be sorted to class $q (q = 1, ..., Q)$. Even though it is unknown, the probability can take a semiparametric multinomial logit formulation suggested by Greene and Hensher (2003) as

$$\rho_{nq} = \frac{e^{s_n \theta_q}}{\sum\limits_{q=1}^{Q} e^{s_n \theta_q}}. \tag{3.43}$$

In the formulation, θ_q is the latent class parameter vector, $\sum\limits_{q=1}^{Q} \rho_{nq} = 1$ and $\rho_{nq} > 0$. The parameter of the first class are normalized to zero: $\theta_1 = 0$ for the identification of the model, as described in 2.3 (Sarrias and Daziano 2017).

Considering y_{ni} as a specific choice made by individual n by choosing alternative i among a set of available alternatives $j = 1, ..., J$. This specific choice is observed in several choice situations/time periods called $T (t = 1, ..., T)$. The probability of individual n in class q choosing alternative i is

$$P_{ni|q} = Prob(y_{ni} = i \mid class = q) = \frac{e^{b'_q x_{ni}}}{\sum\limits_{j=1}^{J} e^{b'_q x_{nj}}}. \tag{3.44}$$

Since there are T choice situations/time periods, which gives T observations for each individual, the conditional contribution of each individual to the likelihood of joint probability has the form

$$P_{t|q} = \prod_{t=1}^{T} P_{ni|q} = \prod_{t=1}^{T} \frac{e^{b'_q x_{ni}}}{\sum\limits_{j=1}^{J} e^{b'_q x_{nj}}}. \tag{3.45}$$

Using the logit probabilities in equation 3.36, in which q is seen as possible values for β, we make some modification and specify a new form for choice probability in latent class models

$$P_{ni} = \sum_{q=1}^{Q} \rho_{nq} (\frac{e^{b'_q x_{ni}}}{\sum_{j=1}^{J} e^{b'_q x_{nj}}}) \prod_{t=1}^{T} \frac{e^{b'_q x_{ni}}}{\sum_{j=1}^{J} e^{b'_q x_{nj}}}. \tag{3.46}$$

Seen in this light, LCM is a logit model for discrete choice among J alternatives by individual n observed in T choice situations with probability specified in equation 3.46. Data is considered to be panel data in the sense that the individual is observed in several choice situations or events. It is, nevertheless, not clear whether there is dependency between T and Q, in the sense that sorting probability is correlated to time period/choice situation. Further research on this type of model needs to address the preference heterogeneity due to invariant decision makers over T situations. Nevertheless, due to the fact that participants are sorted into different classes and groups with unknown probabilities, and socio-demographic variables vary in different groups, the latent class parameters θ_q are likely independent of each other, unlike the case of random coefficients β'_{nt} discussed in the previous section.

3.5 Other discrete choice models

In recent years, researchers have used new techniques to solve several problems in modelling choice and improve the choice models mentioned above. Fiebig and his colleagues (2010), for example, raise the problem of heterogenous preferences that are not well-captured in mixed logit models. The density of the random coefficients $f(\beta)$ in equation 3.34 is often assumed to have a multivariate normal distribution, such as the one in equation 3.35. This misspecifies choice contexts and different tastes, which can be well observed nowadays by the diversified product attributes and consumer characteristics. They suggest a "scale heterogeneity" σ for the error term in the utility function (specified in section 2.3) and develop a generalized multinomial logit model (G-MNL) to solve the problem. The scale heterogeneity σ can explain the different scales across individuals and choice contexts. It is a function of consumer characteristics s_{nt} and choice events t

$$\sigma_{nt} = exp(\bar{\sigma} + \theta s_{nt} + \tau \epsilon_0),$$

where ϵ_0 is an error term for the scale heterogeneity and $\overline{\sigma}$ is the average scale heterogeneity over n observations. θ is the estimated coefficients of consumer characteristics s_{nt}. Researchers can let the vector s_{nt} contain demographics, or some factors that differentiate consumers from each other.

Another way to solve the problem of heterokedasticity is allowing the random coefficients β in mixed logit to have flexible mixing distribution. Train (2016) suggests a new model called Logit-Mixed-Logit (LML). Researchers can specify the shape of distribution as polynomials, splines, steps or other functions to the probability of each coefficient. The probabilities of β are positive, sum to one with a discrete mixing distribution over a finite support set S

$$Prob(\beta = \beta_r) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}},$$

where $z(\beta_r)$ is a vector-valued function of any coefficient $\beta_r \in S$, that captures the shape of the probabiltiy mass function. Vector α is a vector of coefficients for β in this new model. Caputo et. al. (2018) explore the LML model in their analysis of food choices and conclude that this model captures very well food quality attributes, which often have asymatric and multi-modal distribution features.

CHAPTER 4

The estimation of discrete choice models

4.1 Maximum likelihood estimation

The estimation of choice models relies on the calculation of choice probabilities, which are specific for each choice model as we have explained in chapter 3. Depending on whether or not the choice probabilities are in closed form, researchers will use maximum likelihood estimation (MLE) or numerical approximation to estimate the models' coefficients or/and parameters.

As choice probabilities have closed forms, such as those in logit models, a log-likelihood function is derived from probability $P_n(\beta)$ of the observed outcome for individual n:

$$LL(\beta) = \sum_{i=1}^{N} lnP_n(\beta), \tag{4.1}$$

where N is the sample size and β is a vector of coefficients. Finding the value of β that maximizes the log-likelihood function can be done by using the usual methods of numerical maximization such as using gradient descent or Newton Raphson algorithms. In logit choice models, the log-likelihood function derived from logit choice probability is concave (McFadden 1974) and it is often simple to find a global maximum for the estimation of the coefficient β. The Newton-Raphson algorithm, in this case for example, is guaranteed to provide an increase in the likelihood function at each iteration. The estimator of β that maximizes the logit function is consistent and efficient.

Nested logit models also have closed forms for choice probabilities and are relatively easy to estimate (Hensher and Greene 2002). According to Train (2009), there are two ways to estimate β coefficients in these models. The first way is to use maximum likelihood and follow the procedure described in the previous paragraph for logit models. Nevetheless, the log-likelihood function derived from nested logit probability is not always globally concave and multiple peaks may exists. As a consequence, gradient based optimization methods may give local optimal solution. We can then try different algorithms and starting values in order to find the global maximum. This step takes time and researchers need to have certain skills to find a good value to begin with. In addition, this solution is difficult when a large number of coefficients need to be estimated, as of the case in dynamic behavioral models. In order to find the starting value systematically, Liu and Mahmasani (2000) suggest a procedure that incorporates genetic algorithms (GAs) and nonlinear programming (NLP) techniques, to search for a global optimum. The GAs are used to search for good starting points systematically and globally through the possible solution region while maintaining a positive definite variance-covariance matrix. At the same time the nonlinear programming algorithm is used to fine-tune the solutions obtained from the GAs procedure.

The second way to estimate β coefficients for nested logit models is to apply a sequential method. As shown in 3.2, the nested logit probability is decomposed into two logit probabilities, with which the coeffcients can be estimated separately in two different models. Recalling that P_{nB_k} is marginal probability of choosing an alternative in nest B_k and $P_{ni|B_k}$ is conditional probability of choosing alternative i given that an alternative in nest B_k is chosen, we have

$$P_{ni} = P_{ni|B_k} P_{nB_k} = P_{ti} = \frac{e^{Y_{ti}/\lambda_k}}{\sum_{j \in B_k} e^{Y_{tj}/\lambda_k}} \frac{e^{W_{ti}+\lambda_k I_{tk}}}{\sum\limits_{l=1}^{K} e^{W_{tl}+\lambda_l I_{tl}}}. \tag{4.2}$$

The model with the conditional logit probability for choice of alternatives within a nest is estimated first and is called lower model. The model for choice of nests is estimated afterwards and is therefore called upper model. For more details of this sequential estimation method, please see Train (2009, p. 81-86). Sequential estimation, as in this way, may be easier in the sense that the log-likelihood function is a sum of two logit log-likehoods and is maximized in two stages. In particular, choice probabilities in each stage can be estimated with a standard multinomial logit model. The log-likelihood function in the lower model has an advantage of being definitely concave. Nevertheless, the estimated values of coefficients β are consistent but may not be as efficient as the ones in conventional maximum likelihood methods for choice probability given in equation 3.26. Approximation using Hessian matrix for the log-likelihood function of the lower model may be imprecise and thus the estimated standard errors in the first stage is not correct. As a consequence, the estimated coefficients β in the upper models are biased downward. In addition, several overlapping coefficients may exist in two stages. Different values for the same coefficients can be estimated in the two stages because the estimation procedures are carried out in two separated models. And researchers who used this sequential method made no effort to constrain these overlapping coefficients to be equal. Last but not least, different ways of specifying nests can produce very different results for estimation. Regardless of the limitation, estimated coefficients from the sequential estimation method can be used to improve the maximum likelihood estimation mentioned above, specially for nested logit models. For example, we can use them as starting values for the MLE of the joint probabilities in equation 4.2, and the estimation procedure afterwards is similar to the first way for nested logit estimation mentioned above.

4.2 Simulation and maximum likelihood estimation

In mixed logit models, under the random component specification, the random coefficients β_n are assumed to vary from one individual to another. If conventional maximum likelihood methods can be used to estimate β_n, we need to compute the average of the unconditional probabilities in equation 3.34 for all the value of β_n as

$$P_{ni} = E(P_{ni} \mid \beta_n) = \int\limits_{\beta_1}\int\limits_{\beta_2} ... \int\limits_{\beta_N} (P_{ni} \mid \beta_n) f(\beta \mid \theta)\, d\beta_1\, d\beta_2 ...\, d\beta_N. \tag{4.3}$$

This is a N-dimensional integral which cannot be easily estimated without resorting to numerical approximation. In this approximating process, for a given value of the parameters θ, a number of R Monte Carlo draws for β_n are taken from its distribution $f(\beta \mid \theta)$ $(r = 1, ..., R)$. The parameters θ are unknown and rely on researchers' assumption about the distribution of random coefficients. Based on the value of β_n^r in each draw, we use the logit formula in equation 3.34 to calculate $L_{ni}(\beta_n^r \mid \theta)$. This is a logit probability evaluated at parameters θ as alternative i is chosen in a choice situation, given the explanatory variables x_{ni}. The process is repeated for R draws, and the mean of the resulting L_{ni}'s is taken as the approximated choice probability for the chosen alternative i as

$$P_{ni} = (1/R) \sum_{r=1}^{R} L_{ni}(\beta_n^r \mid \theta), \tag{4.4}$$

where β_n^r is the value of the rth draw. The resulting approximated probabilities in 4.4 is a positive, unbiased estimator of the mixed logit choice probability (McFadden and Train 2000)

For empirical issue no. 4 listed in Hensher and Greene (2003) and specified in section 3.3, in order to overcome the difficulty of finding the location of each individual's preferences on the distribution, one can retrieve estimates of individual-specific preferences by deriving the individual's conditional distribution (within sample) based on their choices (i.e. prior knowledge) and *Bayesian Theorem*. This suggestion is helpful, but the researchers mispecified the Bayes Rule. With the help of Train (2009), we modify the function of conditional distribution for individual n as

$$H_{ni}(\theta \mid \beta_n) = \frac{L_{ni}(\beta_n \mid \theta) g(\theta)}{P_{ni}(\beta_n)}, \tag{4.5}$$

where $L_{ni}(\beta_n \mid \theta)$ is now the logit likelihood of an individual's choice if they had this specific β_n, $g(\theta)$ is the distribution of the population of β_n with θ as the set of parameters, and $P_{ni}(\beta_n)$ is the choice probability function defined in the open form specified in equation 3.34 as

$$P_{ni}(\beta_n) = \int_{\beta_n} L_{ni}(\beta_n \mid \theta) g(\beta_n \mid \theta) d\beta_n. \tag{4.6}$$

This form shows how one can estimate the person's specific choice probabilities as a function of the underlying parameters in the random coefficients' distribution.

The estimation of mixed logit model can take into account the time dimension of a panel setting (Train 2009). Considering a sequence of chosen alternatives, one for each time period, $i = i_1, ..., i_T$, we will compute one probability for each individual, which is then included in the log-likelihood function. In particular, for a given vector of random coefficients β_{nt}, the conditional probability that individual n chooses alternative i in the t^{th} observation has a logit form

$$(P_{ni} \mid \beta_{nt}) = L_{ni}(\beta_{nt}) = \frac{e^{\beta'_{nt} x_{nit}}}{\sum_{j=1}^{J} e^{\beta'_{nt} x_{njt}}}. \tag{4.7}$$

Define a dependent variable y_{nit} that has a value of 1 when alternative i is chosen and 0 otherwise. The probability for individual n in the t^{th} observation is then

$$P_{ni} = \prod_{i=1}^{J} (P_{ni} \mid \beta_{nt})^{y_{nit}}.$$

The joint probability for the total observations in T time periods of individual n is:

$$P_{nit} = \prod_{t=1}^{T} \prod_{i=1}^{J} (P_{ni} \mid \beta_{nt})^{y_{nit}}.$$

Nevertheless, the probabilities for the mixed logit P_{nit} are the integrals with no closed form, which degree of integrations is the number of of random coefficients. The computation is done using the following steps in practice:

1. Make an initial hypothesis about the distribution of the random coefficients β_{nt},
2. Draw R numbers of this distribution,
3. For each draw β^r_{nt}, compute the conditional probability

$$(P_{ni} \mid \beta_{nt}) = \frac{e^{\beta'^{r}_{nt} x_{nit}}}{\sum_{j=1}^{J} e^{\beta'^{r}_{nt} x_{njt}}},$$

4. Compute the average of the conditional probabilities in step 3

$$P_{nit} = 1/R \sum_{r=1}^{R} (P_{ni} \mid \beta_{nt}).$$

The random coefficients β_{nt} can be serial correlated over different choice situations ($T > 1$) given to the same decision maker n. In such case Train (2009, p.147) specifies a utility function as

$$U_{njt} = \beta_{njt} x_{njt} + \epsilon_{njt}.$$

The random coefficients for that person in choice set C_t are specified as

$$\beta_{nt} = \mu + \tilde{\beta}_{nt},$$

where μ is fixed and $\tilde{\beta_{nt}}$ has a function similar to a variance

$$\tilde{\beta}_{nt} = \rho\tilde{\beta}_{nt-1} + v_{nt},$$

where v_{nt} is i.i.d. over n and t and ρ represents the serial correlation over choice situation/time period (if any). Simulations of the probability for the sequence of choices i_t in the presence of serial correlation proceed as follows:

1. Draw v_{n1}^r for the initial period with choice i_1, and calculate the logit fomula for this period using β_{n1}^r as

$$\beta_{n1}^r = \mu + v_{n1}^r,$$

2. Draw v_{n2}^r for the second period, calculate the logit formula using β_{n2}^r as

$$\beta_{n2}^r = \mu + \rho v_{n1}^r + v_{n2}^r$$

   with the serial correlation p,

3. Continue for all T time periods and choice situations,
4. Take the product of the T logits,
5. Repeat steps 1-4 for numerous sequences of draws,
6. Average the results.

We will show in the next chapter how the dependence and correlation among random coefficients are resolved in the case of repeated choices for the same decision maker. In this chapter, in order to describe choice modelling of seafood preference, real data are analyzed with the use of different discrete choice models.

CHAPTER 5

Choice modelling in a market research for seafood

5.1 Description of data

The data was collected by a research group at the University of Southern Denmark in August 2011 in a stated choice preference study. The study was carried out in a form of the online survey of French consumers. Details of how the survey was carried out can be found in Nguyen et. al. (2015). In the survey, the consumers were asked for choices of different seafood products in a hypothetical choice experiment. There were 960 participants who were divided into 16 blocks by the researchers. There were 60 participants in each block. Each respondent answered six choice sets, in which the first set is a practice set. This practice set is kept unchanged over respondents. The other choice sets have different number of fish and shellfish alternatives, which is six as the minimum and thirteen as the maximum number. Every choice set has a *None* option alternative.

We chose a data subset which represents choices for 10 seafood alternatives, including the *None* option alternative (which is also known as *no option* alternative) because the choice sets in the original dataset do not have equal number of alternatives. The data subset shows the choices of 840 individuals, out of a total of 960 participants in the survey. Information on dependent variable seafood choice is 1 if an alternative is chosen and 0 otherwise.

The dataset used for this analysis, thus, has 16800 rows, listing 1680 choice sets. Each choice set has 10 rows with 10 seafood alternatives. Information on seafood preferences is given by 840 participants. Each participant answers either 1, 2, or 5 choice sets. The fish alternatives assigned for choice sets are chosen among 12 fish types, including 8 finfish and 4 shellfish. The 8 finfish types are salmon, cod, sole, seabream, saithe, pangasius, monkfish and tuna. The 4 shellfish types are oyster, mussel, langoustine and crab. The choice models are built on nine variables, which are listed in the next part.

5.2 Description of variables

We have information on participants's choices based on one dependent variable and eight explanatory variables. These explanatory variables are divided into two types: four are alternative specific variables and four are individual specific variables.

1. 'Choice' is the dependent variable in our models. This variable shows us the chosen alternative in a choice set. It is assigned with value 1 if an alternative is chosen and 0 otherwise.

2. 'Seafood Origin' is an alternative specific variable. The variable gives information about whether a seafood is produced domestically in France (assigned as '1') or is imported (assigned as '2').

3. 'Seafood Price' is an alternative specific variable. It shows the prices of different seafood alternatives. 'Seafood Price" is collected based on both market positioning and retail system. On the one hand, the researchers used market price in euros at the time of the survey in order to illustrate the market position and pricing strategy of different seafood producers (figure 3). On the other hand, they also used three price levels in order to adjust the given price in accordance with retail and distribution system (figure 2, chart d.). For our modelling, we use market price in euros for the 'Seafood Price' variable.

4. 'Product Form' is an alternative specific variable specifying package forms for different seafood alternatives. This variable is coded as '1' for fillet, loin, live, or raw seafood and as '2' for steak, tail or whole fish and chilled or cooked (with shell on) seafood. In our analysis, we consider this variable as having two forms: 'Handy Form and Fresh' category and 'Unhandy and Cooked' category.

5. 'Production Method' is also an alternative specific variable. It is differentiated from farmed (assigned as '1') or wild catch seafood (assigned as '2'),

6. 'Children' is an individual specific variable. It shows the number of children that each participant has in his/her household. No children in the household is the minimum number and six is the maximum number of children.

7. 'Education level' is an individual specific variable. The participants give information about their highest education level that they received in the French education system. These levels are: elementary, secondary, and high school, 1-2 years at bachelor level, 3-4 years at bachelor level, or 5 years and above.

8. 'Income' is an individual specific variable. It shows the level of income each participant has. The income levels are divided into 13 ranges with less than 1000€/month to more than 5000€/month.

9. 'Age' is also an individual specific variable. It is divided into 5 different categories, depending on the age of participants.

Tables 5.1 and 5.2 summarize the alternative specific as well as individual specific variables. It is worth keeping in mind the coding for alternative specific variables in the survey. Researchers used values 1 and 2 to describe categorical variables such as seafood origin, seafood price, forms of product and methods of production. Thus, three out of four alternative specific variables are dummy variables. The opt-out choice naturally does not have information on seafood products as well as participants and was coded as "NA" in the survey. It is the nooption alternative in each choice set, meaning that not choosing any seafood type is also an option and is listed as a choice alternative in every choice set. As a way to overcome the problem of missing values of this alternative (in statistical software), we turn the value of the opt-out choice to 0 in this analysis. We describe these variables further in figure 2, in which value 0 accounts for 10% of the statistics of variables 'Seafood Origin', 'Seafood Price', 'Product Form', and 'Production Method'.

Table 5.1: Design for Alternative Specific Variables

Variables	**Type**	**Value 1**	**Value 2**
Seafood Origin	Categorical	French	Imported
Production Method	Categorical	Farmed	Wild catch
Product Form	Categorical	Fillet, lion, live, raw	Steak, whole, chilled, cooked
Seafood Price	Continuous	(min)	(max)

Table 5.2: Design for Individual Specific Variables (demographic characteristics of participants)

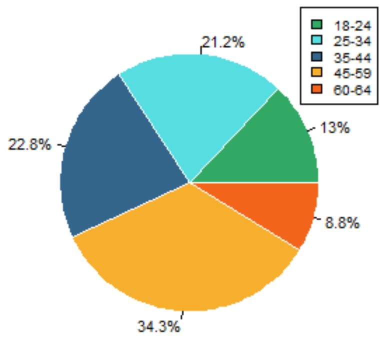
Range	**Age**	**Education**	**Number of children**	**Income**
0	-	-	No Children	Not disclosed
1	18-24	Elementary	1	1-12 levels
2	25-34	Secondary	2	12000-36000 Euros a year
3	35-44	High School	3	
4	45-59	Lower college	4	
5	60-64	College/ University	6 children	
6	-	Post-graduate	-	

5.3 Description of choice sets

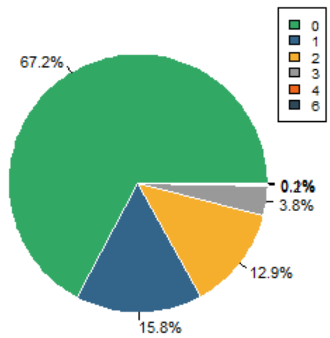
Participants are given 1, 2, or 5 choice sets. Each choice set C_t has 9 seafood alternatives and an opt-out choice, $j = 1, ..., 10$. Thus, each participant has to choose one alternative among 10 available seafood types in each choice set. The total set T varies between 1, 2 or 5 randomly depending on the number of sets that researchers gave to each participant. The 9 seafood alternatives varies in different sets, and are chosen from a list of 12 seafood types. The 9 seafood types in each set are characterised by alternative specific variables, also known as products' attributes. These attributes are labeled as z_j in 2.1 in the theory part. The data subset has a total of 1680 choice sets, which are given to 840 participants, $n = 1, ..., 840$.

Figure 5.1: Individual Specific Variables (demographic characteristics of participants)

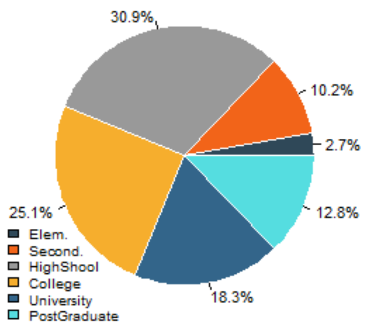
(a) Participants' Age



(b) Number of children in the household



(c) Participants' Education Level



(d) Participants' Income Range

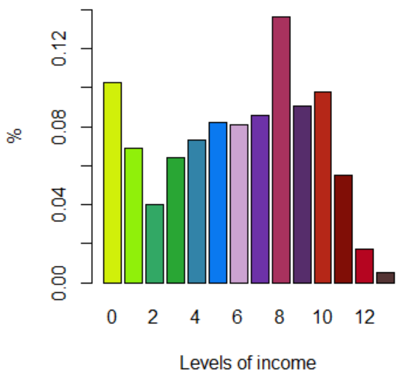
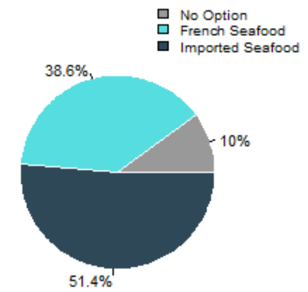
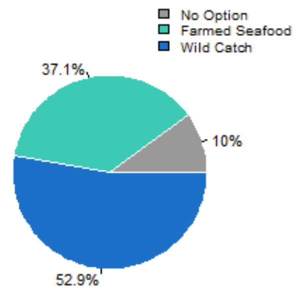


Figure 5.2: Alternative Specific Variables (seafood attributes)

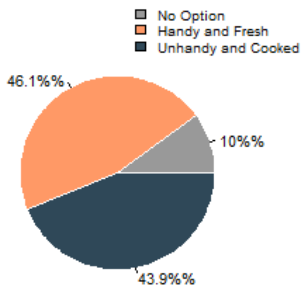
(a) Seafood Origin



(b) Production Method



(c) Product Form



(d) Seafood Price

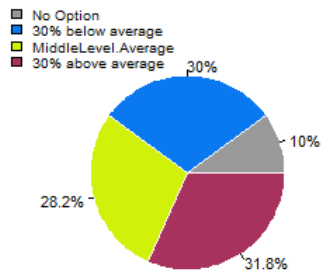
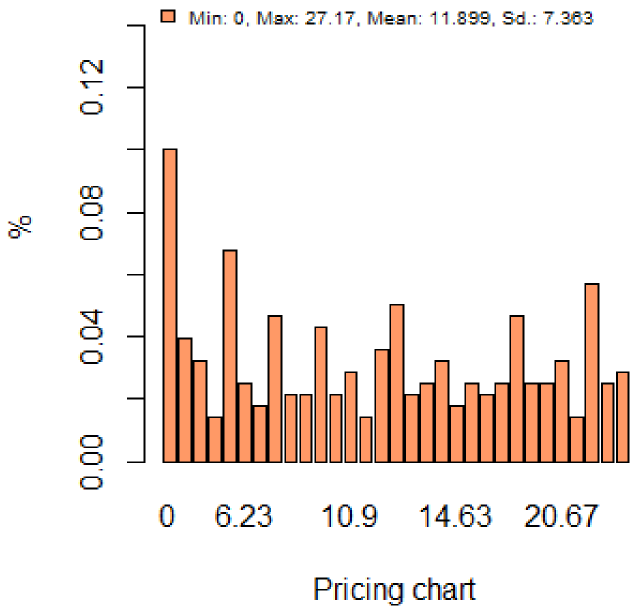


Figure 5.3: Descriptive statistics of Seafood Price



The choice sets in the survey satisfy three condition for the choice sets listed in Train (2009). The three conditions are

1. The alternatives must be mutually exclusive. Choosing one alternative in one choice set necessarily implies not choosing any of the other alternatives,
2. The choice set must be exhaustive in that all possible alternatives are included, even the option of choosing none of the alternatives,
3. The number of alternatives is finite. Nevertheless, this condition is restrictive because there would be unlimited amount of alternatives. In an experiment, researchers often give consumers a constraint on budget or shopping basket to limit the number of alternatives.

5.4 Description of discrete choice models and estimation results

Multinomial logit

In multinomial logit, we assume the error terms ϵ_{nj} to be independently and identically distributed. We use forward selection to choose a model with the best Akaike information criterion (AIC). The best multinomial logit model has the form

$U_{nj} = \beta_0 + \beta_1$ Seafood Price $+ \beta_2$ Production Method $+ \beta_3$ Product Form $+$ β_4 Seafood Origin $+ \gamma_1$ Age $+ \epsilon_{nj}$

where β_k is coefficients of alternative specific variable k, γ_k is coefficients of individual specific variables k. In this model, α_1 is an alternative specific constant used as a baseline. We have $\beta_0 = \sum \alpha_j$. The estimation results for the model are shown on table 5.3.

The results first show α_{j-1} alternative specific constants. The first constants for the first seafood alternative in a choice set is α_1, acting as a baseline for normalizing the constant. For example, the first person chooses an alternative in the first choice set, with 'Salmon' as the first fish alternative. The constant for alternative 'Salmon' is set to zero in this case and the constants for nine other alternatives in the person's choice set are interpreted as utility being relative to 'Salmon', with values -1.175 for 'Seabream', -1.678 for 'Saithe' and so on. In general, these are the constants specific to alternatives 2,...,10 and known as the differences in utilities between the first and the other nine alternatives (as we have discussed in chapter 2).

Secondly, we examine the results for four alternative specific variables. All four variables are statistically significant (at 0.001 level) and their four coefficients are significantly different from zero. The t-statistics for all four coefficients in absolute value are greater than 1.96, which is the critical level for 0.05 confidence level.

The coefficient for seafood price has an expected negative sign -0.107. As the food cost for an alternative increases (and the cost of other alternatives remain the same), the probability of that seafood alternative being chosen falls. It means that a higher price will discourage the purchase of more expensive seafood alternatives. Furthermore, this negative impact of seafood price on choice is highly significant (at 0.001 level).

The ratios of coefficients for seafood price and other alternative specific variables, as discussed in 3.1, shows marginal rates of substitution between seafood attributes. In particular, the marginal rate of substitution between farmed and wild catch seafood is

$$WTP = -\frac{0.465}{(-0.107)} = 4.35$$

which is the coefficient ratio of 'Production Method' and 'Seafood Price'. This marginal rate of substitution implies that consumers prefer wild catched seafood and natural products from the sea to those harvested from aquaculture. With similar interpretation, this ratio is -8.79 for 'Seafood Origin' and 'Seafood Price', reflecting a marginal rate of substitution between domestic and imported seafood. It is negative in this case because the French seafood is specified as 1 and imported seafood as 2, meaning a switch from a French product to an imported product will bring negative utility. In other words, the marginal rate of substitution lets us know consumer preference towards French products. In fact, this result is consistent with the conclusion of Nguyen et. al. (2015) that

Table 5.3: Estimation results for multinomial logit and nested logit models

	Dependent variable: Choice	
	Multinomial logit	Nested logit
(Intercept):2	-1.175^{***} (0.251)	-1.115^{***} (0.285)
(Intercept):3	-1.678^{***} (0.277)	-1.589^{***} (0.349)
(Intercept):4	-2.085^{***} (0.297)	-1.971^{***} (0.390)
(Intercept):5	-1.428^{***} (0.299)	-1.353^{***} (0.342)
(Intercept):6	-2.354^{***} (0.350)	-2.223^{***} (0.440)
(Intercept):7	-2.110^{***} (0.270)	-2.026^{***} (0.346)
(Intercept):8	-1.509^{***} (0.276)	-1.441^{***} (0.295)
(Intercept):9	-2.597^{***} (0.355)	-2.475^{***} (0.462)
(Intercept):10	-3.305^{***} (0.322)	-3.152^{***} (0.475)
Seafood Price	-0.107^{***} (0.006)	-0.101^{***} (0.015)
Production Method	0.465^{***} (0.061)	0.442^{***} (0.084)
Product Form	-0.180^{***} (0.057)	-0.171^{***} (0.058)
Seafood Origin	-0.879^{***} (0.058)	-0.833^{***} (0.122)
Age:2	0.294^{***} (0.078)	0.278^{***} (0.084)
Age:3	0.366^{***} (0.084)	0.347^{***} (0.095)
Age:4	0.354^{***} (0.090)	0.335^{***} (0.096)
Age:5	0.131 (0.094)	0.124 (0.093)
Age:6	0.291^{***} (0.106)	0.274^{***} (0.104)
Age:7	0.236^{***} (0.082)	0.220^{**} (0.090)
Age:8	0.152^{*} (0.086)	0.139 (0.086)
Age:9	0.168 (0.110)	0.155 (0.113)
Age:10	0.053 (0.088)	0.047 (0.086)
Log sum λ_k		0.943^{***} (0.140)
Observations	1,680	1,680
R^2	0.089	0.089
Log Likelihood	$-3,438.186$	$-3,438.106$
LR Test	671.983^{***} (df = 22)	672.144^{***} (df = 23)
Note:		*p<0.1; **p<0.05; ***p<0.01

French consumers value domestic seafood highly.

Thirdly, based on the estimated coefficients, the first participant in the dataset obtains a level of utility of -3.795 for choosing 'Pangasius' alternative in his/her first choice set. This utility is negative because it is actually the utility difference between "Pangasius" and the first alternative "Salmon" in the choice set. Moreover, as mentioned in 2.3, we should interpret the effects of individual specific variable with caution, since it is related to a normalization process and the two properties *only differences in utility matter* and *the scale of utility is arbitrary*. In case the first seafood alternative in the choice set is chosen as the baseline, the coefficient for "Age 2" (0.294) is interpreted as the effect of participants' age on the indirect utility of the second alternative *relative* to the effect of age on the indirect utility of the first alternative (which is set at 0).

Last but not least, estimated average probabilities for the ten alternatives in the best multinomial logit model are shown on table 4. We can see that the average probabilities closely match the shares of customers choosing each alternative. As discussed in section 3.1.5, the matching probabilities show the predictive power of a choice model. In this case, most of the average probabilities match exactly the frequencies of chosen alternatives in the survey. Thus, the best multinomial logit model also has high predictive power.

The role of alternative specific constants was discussed in 3.1. For application and analysis, we fit a multinomial logit model without alternative specific constants in order to examine the effects of these constants more clearly. In particular, the new model has the form

$U_{nj} = \beta_1$ Seafood Price $+ \beta_2$ Production Method $+ \beta_3$ Product Form $+ \beta_4$ Seafood Origin $+ \epsilon_{nj}$

Estimated average probabilities of the new model are shown on table 5.4, in comparison with those estimated by the best multinomial logit model. We see that the average probabilities in the model without the intercepts do not match well with the share of customers choosing each alternatives.

Table 5.4: Comparison of matching probabilities in multinomial logit models

Fish Alternatives	Sample frequencies	Without alt. specific constants	Best Multinomial logit
1	0.1417	0.0694	0.1417
2	0.1518	0.0942	0.1518
3	0.1214	0.0871	0.1214
4	0.0917	0.0974	0.0917
5	0.0744	0.0882	0.0744
6	0.0565	0.0991	0.0565
7	0.1185	0.1298	0.1185
8	0.1071	0.0982	0.1071
9	0.0476	0.1154	0.0476
Opt-out choice	0.0893	0.1213	0.0893

It is worth noticing that seafood alternatives in the model are ones which fall into the positions of the alternative 1 to the alternative 10 ($j = 1, 2, ..., 10$) in each choice set C_t (given to each participant). Furthermore, seafood alternatives in different choice sets are not fixed in the ten positions due to the researchers' intention and initial purposes of designing the choice sets for the experiment. We do not know these purposes clearly. Nevertheless, we can observe the alternative positions for four (out of eight) finfish and four shellfish by having a brief look at the choice sets. Due to the variation of alternatives in different choice sets, estimating choice probabilities for consumer preference of seafood seems not to give as much comparison meaning as in other studies, such as a study of transport modes.

As discussed in 3.1, unobserved factors in utility ϵ_{nj} are not likely well-captured in multinomial logit due to taste variation and preference heterogeneity. We wish to examine other choice models as well in this analysis. We will specifically estimate a nested logit model in the next part.

Nested logit

We will now examine the data with nested logit models. One advantage of nested logit is that it relaxes the i.i.d. assumption by allowing some correlation among unobserved factors of utility in different nests. It means that the IIA condition can be alleviated for alternatives in different nests even though this condition has to be strictly fulfilled for alternatives in the same nest.

In order to estimate the nested logit models, we partition 10 alternatives in each choice set into 2 nests. One nest includes fish alternatives (known as *fish nest*) and one nest includes shellfish alternatives (known as *shellfish nest*). The opt-out choice is also included in the shellfish nest. We choose the two nests based on the suggestion by Nguyen et. al. (2015) about differences in tastes and consumer preferences of fish and shellfish. Using the same model selection technique, the best nested logit model with the best AIC has the form

$U_{nj} = \beta_0 + \beta_1$ Seafood Price $+ \beta_2$ Production Method $+ \beta_3$ Product Form $+$ β_4 Seafood Origin $+ \gamma_1$ Age $+ \epsilon_{nj}$

Table 5.3 shows estimation results for the best nested logit model, which has a similar structure to the best multinomial logit model. By comparing the two models, we realize that the multinomial logit tends to overestimate most of the coefficients, including those of alternative specific and individual specific variables.

The log sum coefficient λ_k is estimated to be 0.943. As discussed in 3.2, a high value for log sum coefficient implies that unobserved factors of utility in a nest are less correlated to each other. It also means that choice probabilities over alternatives are more independent within each nest. In this case, if there is any correlation or dependence in the fish nest or shellfish nest, it is estimated to be $1-0.943 = 0.057$. This is not high correlation or not very strong dependence, implying that the IIA condition is fulfilled within a nest in this model.

We can test if the within-nest correlations are statistically different from those implied by the multinomial logit model, because a log sum coefficient equal to 1 means a similar condition to multinomial logit. We have here $\lambda_k = 0.943$, which is near 1. We use a Likelihood Ratio Test in this case. The test results are shown on table 5.5. We fail to reject the null hypothesis, meaning nested logit is not very different from multinomial logit.

Table 5.5: Likelihood Ratio test for Nested Logit and Multinomial Logit models

	DF	Log- Likelihood	DF	Chi Square	Pr(>Chisq)
Multinomial logit	22	3438.1864			
Nested logit	23	3438.1060	1	0.160729	**0.688486**

Using table 5.6, we can compare average choice probabilities for the ten seafood alternatives in nested logit and multinomial logit models. The table shows that the estimated probabilities in nested logit do not differ significantly from the shares of customers choosing each alternative. Nevertheless, these average probabilities are not as precise as those estimated by the best multinomial logit model.

Table 5.6: Comparison of matching probabilities in different approaches

Fish Alternatives	Sample frequencies	Multinomial logit	Nested logit	Mixed logit
1	0.1417	0.1417	0.1418	0.1417
2	0.1518	0.1518	0.1516	0.1528
3	0.1214	0.1214	0.1214	0.1223
4	0.0917	0.0917	0.0917	0.0909
5	0.0744	0.0744	0.0743	0.0744
6	0.0565	0.0565	0.0566	0.0566
7	0.1185	0.1185	0.1182	0.1167
8	0.1071	0.1071	0.1073	0.1072
9	0.0476	0.0476	0.0477	0.0483
Opt-out choice	0.0893	0.0893	0.0893	0.0891

Mixed logit

Choices made in different choice sets/choice situations by the same individual are not independent. As mentioned above, some participants in the survey are given more than one choice set. In particular, some participants are given 2 and 5 choice sets ($T = 2$ or $T = 5$). Multinomial logit and nested logit have some limitations in accomodating the dependence of choice probabilities for different alternatives. Mixed logit overcomes these limitations by allowing preferences to vary among individuals and unobserved factors in utility among alternatives to be correlated over choice situations. The mixed logit model achieves this flexibility by not using a set of fixed coefficients for the entire population. It assumes that there is a distribution of coefficients throughout the population, in this case, a normal distribution. The best mixed logit model based on AIC has the form

$U_{nj} = \beta_0 + \beta_{n1}$ Seafood Price $+ \beta_{n2}$ Production Method $+ \beta_{n3}$ Seafood Origin $+ \beta_{n4}$ Product Form $+ \gamma_1$ Age $+ \epsilon_{nj}$

where the coefficients for the three alternative specific variables (Seafood Price, Production Method, and Seafood Origin) are assumed to be normally distributed with mean $b_k \neq 0$ and variance σ_k^2. All random coefficients are estimated using a panel specification.

$$\beta_{n1} \sim N(b_1, \sigma_1^2)$$
$$\beta_{n2} \sim N(b_2, \sigma_2^2)$$
$$\beta_{n3} \sim N(b_3, \sigma_3^2)$$

Most of the variables of the best mixed logit model are statistically significant at 0.001 level. The estimation results in table 5.7 show significant random coefficients for the three alternative specific variables. We run the mixed logit models with 100 *Halton draws* and take into account panel structure of the data (for more details of *Halton sequences* and the Halton draws, see Train, 2009, p. 221). The random effects of the three alternative specific variables are shown on tables 5.9. The variances for these coefficients are statistically significant at 0.001 level (which are 'sd.Fish Price', 'sd.Prod. Method', 'sd.Fish Origin' on table 5.7). It means that we can not ignore the random effects. The fixed effects on table 5.7, are ones for alternative specific intercepts, coefficients for individual specific variable 'Age' as well as coefficient for 'Product Form'. Most of these fixed effects are also statistically significant.

Table 5.7: Estimation results for Mixed Logit models

	Dependent variable: Choice	
	Mixed logit	Mixed logit with correlated random effects
(Intercept):2	-1.245^{***} (0.270)	-1.251^{***} (0.270)
(Intercept):3	-1.838^{***} (0.301)	-1.822^{***} (0.300)
(Intercept):4	-2.284^{***} (0.311)	-2.334^{***} (0.314)
(Intercept):5	-1.630^{***} (0.326)	-1.683^{***} (0.330)
(Intercept):6	-2.507^{***} (0.345)	-2.524^{***} (0.346)
(Intercept):7	-2.487^{***} (0.303)	-2.482^{***} (0.300)
(Intercept):8	-1.771^{***} (0.291)	-1.758^{***} (0.288)
(Intercept):9	-2.782^{***} (0.376)	-2.777^{***} (0.376)
(Intercept):10	-4.688^{***} (0.460)	-4.602^{***} (0.467)
Seafood Price	-0.131^{***} (0.008)	-0.129^{***} (0.008)
Production Method	0.532^{***} (0.070)	0.540^{***} (0.072)
Product Form	-0.221^{***} (0.061)	-0.214^{***} (0.061)
Seafood Origin	-1.044^{***} (0.076)	-1.033^{***} (0.076)
Age:2	0.314^{***} (0.084)	0.317^{***} (0.084)
Age:3	0.408^{***} (0.091)	0.407^{***} (0.091)
Age:4	0.389^{***} (0.094)	0.395^{***} (0.094)
Age:5	0.165 (0.103)	0.171 (0.104)
Age:6	0.312^{***} (0.105)	0.314^{***} (0.106)
Age:7	0.280^{***} (0.091)	0.284^{***} (0.090)
Age:8	0.178^{*} (0.092)	0.184^{**} (0.091)
Age:9	0.182 (0.117)	0.186 (0.117)
Age:10	0.033 (0.111)	0.039 (0.111)
sd.Seafood Price	0.091^{***} (0.012)	
sd.Production Method	0.633^{***} (0.173)	
sd.Seafood Origin	0.860^{***} (0.160)	
chol.Price:Price		0.087^{***} (0.012)
chol.Price:Method		0.309^{**} (0.120)
chol.Method:Method		0.513^{**} (0.214)
chol.Price:Origin		-0.109 (0.116)
chol.Method:Origin		-0.231 (0.212)
chol.Origin:Origin		0.841^{***} (0.165)
Observations	1,680	1,680
R^2	0.102	0.103
Log Likelihood	$-3,389.231$	$-3,385.761$
LR Test	769.893^{***} (df = 25)	776.834^{***} (df = 28)
Note:		*p<0.1; **p<0.05; ***p<0.01

In the best multinomial logit, all coefficients are assumed to be fixed, or variances $\sigma_k^2 = 0 \forall k$. We can test if the random coefficient variances of the mixed logit model are statistically different from those implied by the multinomial logit model. For example, we can use a likelihood ratio test with a null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0$. From the test results on table 5.8, we can reject the null hypothesis at 0.01% confidence level. There are significant random effects and we need the mixed logit model to estimate choice probabilties as well as consumer utilities.

Table 5.8: Likelihood Ratio test for Mixed Logit and Multinomial Logit models

	DF	Log- Likelihood	DF	Chi Square	Pr(>Chisq)
Multinomial logit model	22	-3438.1864			
Mixed logit model	25	-3389.2314	3	99.91	**<2.2e-16 *****

Table 5.9: Random effects – Mixed logit model

	1st Qu.	**Median**	**Mean**	**3rd Qu.**
$\beta_{\text{Seafood Price}}$	-0.1926278	-0.1309361	-0.1309361	-0.06924443
$\beta_{\text{Prod. Method}}$	0.1047443	0.5318100	0.5318100	0.95887576
$\beta_{\text{Seafood Origin}}$	-1.6236560	-1.0439228	-1.0439228	-0.46418970

In addition to the Likelihood Ratio Test mentioned above, we also use a Chi-Square test to examine the dependence in random effects (known as the Score test in the *m.logit* package) in the best mixed logit model. The null hypothesis is H_0: correlated random effects. Based on the statistic $\tilde{\chi_3}^2 : 5.014$ on table 5.10, we fail to reject the null hypothesis and there is indeed some level of dependence among random coefficients. In orther words, the random coefficients of the three alternative specific variables in this mixed logit model are correlated.

Table 5.10: Chi-Square test for correlation in the mixed logit model that has not accounted for correlated random effects

The Score test	*Ho: correlation = TRUE*		
	chisq = 5.014	df = 3	**p-value = 0.1708**

In order to account for the dependency of random coefficients, we fit a new mixed logit model. Estimation results for the new mixed logit model are also shown on table 5.7. For observation of whether random coefficients are correlated, we derive a correlation matrix (table 5.11). Elements of this matrix signify some degree of correlation between the two pairs: Price and Method, Origin and Method. As discussed in section 3.3, we can test whether the random coefficients are correlated by deriving a covariance matrix Ω, which is shown on table 5.12. The non-diagonal elements of this matrix is not zero. We can conclude that the three random coefficients in the new estimated mixed logit model are, indeed, correlated. These correlated random effects are shown on table 5.13. This conclusion is consistent with the Score test mentioned above.

The asymptotic covariance matrix Ω in the new mixed logit model is calculated based on equation 3.42 and the Cholesky decomposition method discussed in section 3.3. The three random coefficients of Seafood Price, Seafood Origin and Production Method are assumed to be normal distributed with mean and

Table 5.11: Correlation matrix of random coefficients in mixed logit

	$\beta_{\text{Seafood Price}}$	$\beta_{\text{Prod. Method}}$	$\beta_{\text{Seafood Origin}}$
$\beta_{\text{Seafood Price}}$	**1**	0.51593	-0.12374
$\beta_{\text{Prod. Method}}$	0.51593	**1**	-0.44102
$\beta_{\text{Seafood Origin}}$	-0.12374	-0.28923	**1**

Table 5.12: Covariance matrix of correlated random coefficients in mixed logit

	$\beta_{\text{Seafood Price}}$	$\beta_{\text{Prod. Method}}$	$\beta_{\text{Seafood Origin}}$
$\beta_{\text{Seafood Price}}$	0.0076	0.027	-0.0095
$\beta_{\text{Prod. Method}}$	0.027	0.3584	-0.1522
$\beta_{\text{Seafood Origin}}$	-0.0095	-0.1522	0.7729

Table 5.13: Coefficients of correlated random effects – Mixed logit model with dependently distributed random coefficients

	1st Qu.	**Median**	**Mean**	**3rd Qu.**
$\beta_{\text{Seafood Price}}$	-0.1877400	-0.1287962	-0.1287962	-0.06985242
$\beta_{\text{Prod. Method}}$	0.1361892	0.5399901	0.5399901	0.94379106
$\beta_{\text{Seafood Origin}}$	-1.6262973	-1.0333342	-1.0333342	-0.44037120

standard deviations shown on table 5.15. The variance σ_k^2 of the three correlated random coefficients are the elements on the diagonal of the matrix Ω shown on table 5.12. The covariances of the correlated random coefficients σ_{kc} are the non-diagonal elements of the same matrix. In addition to the covariance matrix Ω, we also derive standard deviations for the estimated variances and covariances of the random coefficients on table 5.14. Relating the results on the two tables 5.14 and 5.15 to those on table 5.7, we have made an improvement for the calculation of standard deviations for correlated random effects in the new mixed logit model that takes the correlation into account.

In order to illustrate the Cholesky decomposition method for a three-dimensional vector β_3 of correlated random coefficients, we derive a lower triangular matrix L as

$$L = \begin{bmatrix} s_{11} & 0 & 0 \\ s_{21} & s_{22} & 0 \\ s_{31} & s_{32} & s_{33} \end{bmatrix} \tag{5.1}$$

The covariance matrix on table 5.12 has the form

$$\Omega = LL' = \begin{bmatrix} s_{11}^2 & s_{11}s_{12} & s_{11}s_{13} \\ s_{11}s_{21} & s_{21}^2 + s_{22}^2 & s_{21}s_{13} + s_{22}s_{23} \\ s_{11}s_{31} & s_{21}s_{31} + s_{22}s_{32} & s_{31}^2 + s_{32}^2 + s_{33}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \tag{5.2}$$

Table 5.14: Standard deviations for the variances and covariances of the correlated random effects

	Estimation	**standard deviation**	**"z-value"**	**"Pr(>z)"**
var($\beta_{\text{Seafood Price}}$)-$\sigma_1^2$	0.0076	0.0021	3.6973	2e-04
var($\beta_{\text{Prod. Method}}$)-$\sigma_2^2$	0.3584	0.2154	1.6636	0.0962
var($\beta_{\text{Seafood Origin}}$)-$\sigma_3^2$	0.7729	0.2752	2.808	0.005
cov(Price:Method)-σ_{12}	0.027	0.0107	2.528	0.0115
cov(Price:Origin)-σ_{13}	-0.0095	0.0103	-0.9263	0.3543
cov(Method:Origin)-σ_{23}	-0.1522	0.122	-1.248	0.212

Table 5.15: Mean and variance of correlated random coefficients in mixed logit model that accounts for correlation

Random coefficients	**Mean - b_k**	**Standard deviation-σ_k**
$\beta_{\text{Seafood Price}}$	-0.129	0.087
$\beta_{\text{Prod. Method}}$	0.540	0.599
$\beta_{\text{Seafood Origin}}$	-1.033	0.879

Table 5.16: Comparison of matching probabilities in mixed logit models

Fish Alternatives	**Sample frequencies**	**Not account for correlation**	**Correlated random effect**
1	0.1417	0.1417	0.142
2	0.1518	0.1528	0.1529
3	0.1214	0.1223	0.1225
4	0.0917	0.0909	0.091
5	0.0744	0.0744	0.0748
6	0.0565	0.0566	0.0564
7	0.1185	0.1167	0.1167
8	0.1071	0.1072	0.1068
9	0.0476	0.0483	0.0482
Opt-out choice	0.0893	0.0891	0.0886

Table 5.17: Estimated probabilities for the first 10 choice sets – The best model

Salmon	**Fish_2**	**Fish_3**	**Pangasius**	**Monkfish**	**Tuna**	**Oyster**	**Mussels**	**Languostine/ Crab**	**Opt-out**
0.298	0.169	0.097	0.05	0.089	0.016	0.043	0.141	0.006	0.09
0.081	0.159	0.062	0.034	0.037	0.053	0.06	0.338	0.067	0.109
0.255	0.189	0.12	0.062	0.086	0.018	0.047	0.14	0.006	0.078
0.065	0.176	0.075	0.041	0.036	0.056	0.066	0.331	0.063	0.091
0.301	0.169	0.098	0.049	0.09	0.016	0.043	0.143	0.006	0.085
0.082	0.159	0.062	0.034	0.037	0.055	0.061	0.337	0.067	0.107
0.305	0.169	0.094	0.049	0.089	0.015	0.042	0.139	0.006	0.092
0.079	0.159	0.062	0.033	0.038	0.053	0.06	0.338	0.067	0.111
0.255	0.192	0.119	0.062	0.087	0.017	0.046	0.139	0.006	0.077
0.066	0.177	0.073	0.041	0.035	0.058	0.065	0.33	0.064	0.091

5.5 Implication for the choice modelling of seafood preference

After considering three criteria, AIC, matching probabilities, and tests for correlated random coefficients, we decide that the mixed logit model with correlated random coefficients is the best model so far. We have used this model to report choice probabilities derived from 1680 choice sets. The results for the first 10 choice sets are shown on table 5.17. In addition, we also compare the matching probabilities between two types of mixed logit models (correlated and uncorrelated random effects) on table 5.16.

Though the analysis of our best model, we come to the conclusion that correlation between random coefficients in mixed logit models is a serious problem to which researchers need to pay attention. In many studies instead, the modelling of consumer preferences often neglect the correlated random coefficients when researchers interpret marginal effects, willingness to pay, or marginal rates of substitution. This has an immediate consequence because these indicators for consumer preferences are not precisely described. In particular, the estimation of these indicators rely on correlated random coefficients, whose mean and standard deviations are not efficiently estimated. Taking the seafood study as an example, the calculation of *willingness to pay* relies on a correct estimation of two coefficients, which are the coefficient of price and the coefficient of one of the seafood attributes. If we did not take into account correlated random coefficients, and did not derive a covariance matrix for the calculation of standard deviations of these (alternative specific) random coefficients, willingness to pay would be misspecified.

The modelling of food choice using mixed logit models, in addition to the preference heterogeneity problem, needs to account for a possible correlation among random coefficients. And one of the proper ways to do this, as we have pointed out, is to test for the correlation. If there is, indeed, any dependence among random coefficients, we use Cholesky decomposition method to derive a covariance matrix. It is not adequate if we only assume a distribution for the random coefficients and report estimated parameters, i.e. mean and variances, for the assumed distribution. In sum, the correlated random coefficients need a covariance matrix so that we can caculate precisely their mean and standard deviations.

In addition to the conclusion about the best discrete choice model for describing the seafood data mentioned above, we would like to mention another finding in our empirical analysis of seafood choice. Estimation results of the model on table 5.7 shows significant fixed effect for individual specific variable 'Age', which is different from the seafood preference study of Nguyen et. al. (2015). Using the same data, researchers in this study estimated a latent class model and found significant effects for 'Income' and 'Education level'. Given the bias in survey questionnaire with regards to income (Moore 2000), which are difficult to collect and rarely disclosed precisely by survey participants, our finding suggests that participants' age can be obtained instead of income or other sensitive information. Nowadays, it is not a big challenge to have access to information on consumers' age. Based on the knowledge of utility differences and consumer preferences of different age groups, seafood producers can target the market for different customer groups. The companies can then improve their business based on effective pricing strategies for consumers in different age groups.

CHAPTER 6

Discussion and future research of seafood markets

In this thesis, we have looked at different discrete choice models and applied them in the choice modelling of seafood preferences, using real data from an online survey of seafood consumption in France. We have discussed ways of estimating choice models that can explain well consumer preference of seafood and how to overcome the models' limitation in dealing with preference heterogeneity as well as correlation among unobserved error terms. We have specifically examined the i.i.d. assumption and the IIA property and considered in which ways the models can predict correctly choice probabilities of different product alternatives. As such, the models are useful in unveiling the consumer preference and in explaining well choice behavior in market research for seafood.

Our main goal for this thesis, as stated in the introduction, is to describe and compare different discrete choice models. In order to find the best model that has powerful predictive power and at the same time avoid measurement errors in seafood choice analysis, we estimate three discrete choice models: multinomial logit, nested logit and mixed logit. In the seafood dataset, we have a choice setting in which a survey participant makes decisions over seafood alternatives in several choice situations. Thus, in addition to preference heterogeneity among individuals and taste variation among seafood alternatives as well as their attributes, we have to take into account correlation over choice situations. Modelling problems can arise due to the fact that choices are made by the same decision maker over many choice situations and therefore their probabilities are dependent upon each other.

We began the data analysis with estimating a standard multinomial logit model, assuming unobserved factors of utility are independently and identically distributed. This is a good model in terms of matching average choice probabilities. It shows that the estimated choice probabilities match well with shares of consumers choosing each seafood alternative. The model, however, does not allow random taste variation, as we have pointed out in the theory discussion. In addition, it cannot handle the dependence in choice probabilities due to similar individual specific variables. These demographic characteristics describe the same decision maker and thus are invariant over different choice sets.

With an attempt to address taste variation among seafood alternatives, we estimated a nested logit model. To the best of our knowledge, few studies in seafood choice have exploited the nested logit model. Therefore, we attempted to do so and hoped to solve part of the preference heterogeneity by directing the dependence between alternatives into dependence between nests. It means that choice probabilities can be dependent on attributes of alternatives in different nests but this dependence is resolved for alternatives within a nest. In sum, the unobserved factors of utility in different nests can be correlated, but not within a nest. With this logic in mind, we estimated a two–level nested logit model, which shows some advance over multinomial logit. The unobserved factor of utility is, indeed, uncorrelated within each nest in this model. However, a test for our nested logit model with two nests does not differ it much from multinomial logit.

For the purpose of handling dependence across alternatives and individuals, we estimated a mixed logit model, first with the uncorrelated random effects. The model confirms the existence of preference heterogeneity problem and we need models with random effects to explain choice behavior that varies among individuals in a population. Due to this existence, the mixed logit model was estimated with a specific hope to produce more accurate probability prediction for seafood alternatives when there are choice dependent effects. Nevertheless, it is difficult to ignore a possible correlation among unobserved factors of utility in our mixed logit model. By testing for the correlation, we found that the random effects are in fact correlated. Even after taking into account the preference heterogeneity problem, we still have the dependency among random effects.

By acknowledging the inevitable correlation, we estimate a new mixed logit model with correlated random effects. In this model, standard deviation of the random coefficients can not be estimated directly. The Cholesky decomposition and method of deriving an asymptotic covariance matrix were selected to calculate precisely the standard deviations of the random coefficients. In believing that the mixed logit model with correlated random effects is the best one we could find to analyze the seafood data, we calculate choice probabilities for different seafood alternatives. We judge the adequacy of our solution based on its ability to produce dynamic prediction, measured by matching average probabilities with the shares of consumers choosing each seafood alternative. The model works well for the seafood data compared to standard multinomial logit and nested logit. Furthermore, mixed logit model, given the correlated random effects, is considered as a solution to the problem of preference heterogeneity in a panel data setting. Thus, we believe that this model is the best discrete choice method in describing seafood preferences as well as choice behaviors in seafood survey and choice experiment. Last but not least, levels of willingness to pay in mixed logit models with a panel data setting should be interpreted with caution. Researchers should first adopt a routine of testing for correlation among random coefficients. If these random coefficients are correlated, researcher should explicitly include an asymptotic covariance matrix in estimation results of the mixed logit models. This technique will definitely help to estimate efficiently the mean and standard deviation of the correlated random effects, describe precisely levels of willingness to pay, and therefore, improve the modelling of food choices.

For further work, we wish to assess generalized multinomial logit and logit mixed logit models empirically. These are recent developed models for improving multinomial logit and mixed logit models. Moreover, as Hensher and Greene (2003) point out the complexity of individual specific random effects, it may be challenging but interesting to estimate these random effects for individual specific variables in food choice models. Estimation results of our best model show that participants' age is a significant variable indicating utility differences between seafood alternatives. Seafood producers can target their customers by this knowledge of utility differences and willingness to pay levels of different consumer groups in the population. Choice modelling should also takes into account preference heterogeneity that is dependent on consumer demographic characteristics, such as age. Other individual specific variables, for instance, income and education level, also signify random taste variation for different individual groups in many choice studies. Therefore, it will be interesting to see how the random effects of these individual specific variables, in addition to the random effects of alternative specific variables, can address preference heterogeneity in seafood choices as a whole.

Bibliography

Ben-Avika, M. (1973). *The structure of travel demand models*. PhD Thesis. MIT.

Bliemer, M. C. J., Rose, J. M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B*, 58, pp. 199-214.

Brownstone D., Small, K. A. (1989). Efficient Estimation of Nested Logit Models. *Journal of Business & Economic Statistics* 7(1), pp. 67-74.

Caputo, V., Scarpa, R., Nayga Jr., M. R., Ortega, D. L. (2018). Are preferences for food quality attributes really normally distributed? An analysis using flexible mixing distributions. *Journal of Choice Modelling* 28(2018), pp. 10-27.

Carlucci, D., Nocella, G., Devitiis, B., Viscecchia, R., Bimbo, F., & Nardone, G. (2015). Consumer purchasing behaviour towards fish and seafood products: Patterns and insights from a sample of international studies. *Appetite*, 84, pp. 212-227.

Chen, X., Alfnes, F., Rickertsen, K. (2015). Consumer Preferences, Ecolabels, and Effects of Negative Environmental Information. *AgBioForum*, 18(3), pp. 327-336.

Cosslett, S. (1983), Distribution-free Maximum Likelihood Estimation of the Binary Choice Model, *Econometrica* 51, pp. 765-782.

Cosslett, S. R. (1997). Nonparametric maximum likelihood methods. In Maddala, G. S. & Rao, C. R. (eds.). *Handbook of Statistics*, *vol.15*, pp. 385-404. Elsevier Science BV.

Croissant, Y. (2010). *Estimation of multinomial logit models in R: The mlogit Packages*. Université de la Réunion.

Croissant, Y. (2012). *mlogit: multinomial logit model*. R package version 0.2-3, URL http://CRAN.R-project.org/package=mlogit.

Fiebig, D. G., Keane, M. P., Louviere, J., Wasi, N. (2010). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science*, 29(3), pp. 393-421.

Greene, W. H., Hensher, D. A. (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research*, 37(8), pp. 681–698.

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

Hensher, D. A., Greene, W. H. (2002). Specification and estimation of the nested logit model: alternative normalisations. *Transportation Research Part B,* 36(2002), pp. 1-17.

Hensher, D. A., Greene, W. H. (2003). The Mixed Logit model: The state of practice. *Transportation* 30, pp. 133-176.

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

Horowitz, J. (1993). Semiparametric and Nonparametric Estimation of Quantal Response Models, in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics: Econometrics*, Vol. 11, North-Holland, Amsterdam.

Lancaster, K. J. (1966). A new approach to consumer theory. *The Journal of Political Economy*, pp. 132–157.

Liu, Y H., Mahmasani, H. S. (2000). Global maximum likelihood estimation procedure for multinomial probit (MNP) model parameters. *Transportation Research Part B*, 34, pp. 419-449

Louviere, J. J., Flynn, T. N., Carson, R. T. (2000). Discrete Choice Experiments Are Not Conjoint Analysis. *Journal of Choice Modelling*, 3(3), pp. 57-72.

Manski, C. (1975). Maximum Score Estimation of the Stochastic Utility Model of Choice. *Journal of Econometrics*, 3, pp. 205- 228.

Manski, C. F. (2001). Daniel McFadden and the Econometric Analysis of Discrete Choice. *The Scandinavian Journal of Economics*, 103(2), pp. 217-229.

McFadden, D. (1974).The measurement of urban travel demand. *Journal of Public Economics*, 3, pp. 303-328.

McFadden, D. (1978). Modelling the Choice of Residential Location. In Karlqvist, A., Lundqvist, L., Snickars, F. and Weibull, J. (eds.). *Spatial Interaction Theory and Planning Models*. North-Holland, Amsterdam, pp. 75-96.

McFadden, D., Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15, pp. 447-470.

McFadden, D. L., Bemmaor, A. C., Caro, F. G., Dominitz, J., Jun, B. H., Lewbel, A., Matzkin, R. L., Molinari, F., Schwarz, N., Willis, R. J. & Winter, J. K. (2005). Statistical Analysis of Choice Experiments and Surveys. *Marketing Letters*, 16, pp. 183-196.

Moore, J. C, Stinson, L. L., Welniak, E. J. (2000). Income Measurement Error in Surveys: A Review. Journal of Official Statistics, 16(4), pp. 331-361.

Nguyen, T. T., Haider, W., Solgaard, H. S., Ravn-Jonsen, L., Roth, E. (2015). Consumer willingness to pay for quality attributes of fresh seafood: A labeled latent class model. *Food Quality and Preference*, 41, pp. 225-236.

Sarrias, M., Daziano, R. A. (2018). Individual-specific point and interval conditional estimates of latent class logit parameters. *Journal of Choice Modelling*, 27, pp. 50-61.

Swait, J., Louviere, J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research*, 30 (3), pp. 305-314

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press, New York.

Train, K. E. (2016). Mixed logit with a flexible mixing distribution. *The Journal of Choice Modelling*, 19, pp. 40-53.

Vass et. al. (2017). Scale Heterogeneity in Healthcare Discrete Choice Experiments: A Primer. *Patient*, 11(2), pp. 167-173.

Whitehead, J. C., Pattanayak, S. K., Van Houtven, G. L., Gelso, B. R. (2008). Combining revealed and stated preference data to estimate the nonmarket value of ecological services: An assessment of the state of the science. *Journal of Economic Survey*, 22(5), pp. 872-908.