Analysis of the Nested Logit Model under the BLP Framework through Simulation and Empirical Applications

Master Thesis Presented to the

Department of Economics at the

Rheinische Friedrich-Wilhelms-Universität Bonn

In Partial Fulfillment of the Requirements for the Degree of

Master of Science (M.Sc.)

Economics

Supervisors: Prof. Dr. Christoph Breunig

Submitted in July 2024 by:
Raunak Mehrotra
Matriculation Number: 3391327

Contents

Abstract

Discrete choice models are essential in econometrics for analyzing decision-making where individuals choose from a set of alternatives. These models, including the Multinomial Logit (MNL) and Nested Logit (NL) models, use the Random Utility Maximization (RUM) framework to predict choices based on observable and unobservable factors. While the MNL model is popular due to its simplicity and ease of estimation, it is often limited by the Independence of Irrelevant Alternatives (IIA) property. The NL model overcomes this by allowing correlated error terms within groups of similar alternatives, providing a more realistic representation of consumer behavior. However, the application of these models is challenged by the limited availability of individual-level data, prompting reliance on more accessible market-level data. The Berry, Levinsohn, and Pakes (BLP) model addresses this by extending the RUM framework to differentiated product markets and incorporating random coefficients to account for consumer heterogeneity and price endogeneity. Despite its robustness, the BLP model's computational intensity has led to the integration of the NL model within its framework to streamline estimation while maintaining accuracy in capturing nested choice structures. This paper explores the theoretical foundations and practical applications of these discrete choice models, highlighting their advantages and limitations. Through discussing the BLP model's capabilities and the NL model's enhancements, the potential for these models to provide valuable insights into market behavior and consumer preferences is also demonstrated.

Keywords— Discrete Choice Models, Multinomial Logit, Nested Logit, Independence of Irrelevant Alternatives, BLP Model, Market-level Data, Competition Analysis

1 Introduction

In the real world, there are often circumstances wherein an individual has to choose from a set of alternatives. For instance, an employee decides between the mode of transportation (car, bus, or metro) to reach the office, or a customer selecting a product from a list of brands.

The decision-makers finalize their choice based on several behavioral and observable characteristics. Hence, such a model whereby the individual selects an outcome from a list of alternatives based on certain factors is known as the 'Discrete Choice Model'.

To fit within the model framework, the set of alternatives, also known as, the choice set, needs to satisfy three features[21]:

- The choice set must be 'mutually exclusive', that is, choosing an alternative necessarily implies not selecting other alternatives.

- The choice set must be 'exhaustive', that is, all the possible alternatives are included in the choice set.

- The choice set must be finite.

While the first and second criteria are not restrictive, the third criterion is actually restrictive and considered the defining characteristic of the discrete choice models. The conditions of 'mutually exclusive' and 'exhaustive' can be sufficed by the researcher by appropriately defining the choice set based on the research question and availability of the data. For instance, research on the choice of drinks consumers buy while shopping can ask the decision-maker to select from tea, coffee, soft drinks, and beer. Such a choice set is neither 'mutually exclusive' nor 'exhaustive'. Since the consumer can buy multiple types of drinks, the choice set violates the first feature. The second feature is violated as it's quite possible that the consumer doesn't buy any drinks at all. In such a scenario, the condition of 'mutually exclusive' can be reestablished by redefining the choice set such that all possible combinations are covered, like choosing between caffeinated drinks and non-caffeinated drinks and so on. The condition of 'exhaustive' can be met by simply adding 'no drinks' in the choice set as an alternative or excluding the decision-makers who do not buy drinks. This also establishes the importance of available data while defining the choice set, since in the absence of a detailed dataset, excluding consumers who don't buy drinks makes more meaning.

The third condition of the choice set being finite is however restrictive because it distinguishes the discrete choice models from regression models. In regression models, the dependent variable is continuous and can therefore have an infinite number of possible outcomes. The discrete choice models cannot be applied in the presence of infinite alternatives. Therefore, discrete choice models are often said to examine 'which' of the dependent variable, and the regression models estimate 'how much' of the dependent variable.

Econometricians often interpret the discrete choice models in terms of the underlying behavioral model, known as the Random Utility Model (RUM). Under the random utility models, it is assumed that the decision-makers are utility maximizers, that is, the individual preferences among a set of discrete alternatives can be described with a utility function and the individual selects the alternative that provides the maximum utility [11]. The model can be mathematically outlined as a utility function, $\mathbf{U_{ij}}$ for individual i and alternative j, where $j = 1, 2, \cdots, J$ and $i = 1, 2, \cdots, J$. Since a researcher cannot observe all the factors that affect utility, the utility function can be represented as the sum of two components:

- The systematic component that consists of the observable factors. These components can be represented as $\mathbf{V_{ij}} = \mathbf{V(x_{ij}, s_i)}$, where $\mathbf{V_{ij}}$ is the systematic component function and it is the linear combination of the observable characteristics of the alternatives $\mathbf{x_{ij}}$ and the observable characteristics of the individuals $\mathbf{s_i}$.

- The random component that accounts for the unobserved factors. This random variable accounts for all the unobserved variables that have an impact on the utility of choosing a specific alternative. It can be represented as ϵ_{ij}

Hence, the utility function for an individual i for alternative j is,

$$\mathbf{U_{ij} = V_{ij} + \epsilon_{ij}}$$

Note that here onwards unless specified, the individual index i has been omitted to simplify the notation. Therefore, the utility function can be rewritten as,

$$\mathbf{U_j = V_j + \epsilon_j}$$

Given the utility function for an individual, the random utility model defines the individual's choice as the alternative that provides the highest utility. This implies that the individual chooses alternative k if and only if $\mathbf{U_k > U_j}, \forall \mathbf{j \neq k}$, that is,

$$\begin{cases} \mathbf{U_k > U_1 \implies V_k + \epsilon_k > V_1 + \epsilon_1} \\ \mathbf{U_k > U_2 \implies V_k + \epsilon_k > V_2 + \epsilon_2} \\ \vdots \\ \mathbf{U_k > U_J \implies V_k + \epsilon_k > V_J + \epsilon_J} \end{cases}$$

This can be rewritten as,

$$\begin{cases} \mathbf{U_k - U_1 = (V_k - V_1) + (\epsilon_k - \epsilon_1) > 0} \\ \mathbf{U_k - U_2 = (V_k - V_2) + (\epsilon_k - \epsilon_2) > 0} \\ \vdots \\ \mathbf{U_k - U_J = (V_k - V_J) + (\epsilon_k - \epsilon_J) > 0} \end{cases}$$

Since ϵ_j are not observable, a researcher can rewrite the utility function in terms of the random variable [7].

$$\begin{cases} \mathbf{\epsilon_1 < (V_k - V_1) + \epsilon_k} \\ \mathbf{\epsilon_2 < (V_k - V_2) + \epsilon_k} \\ \vdots \\ \mathbf{\epsilon_J < (V_k - V_J) + \epsilon_k} \end{cases}$$

Hence, the probability of an individual of selecting an alternative k conditional on ϵ_k is,

$$\mathbf{(P_k|\epsilon_k) = F_{-k}(\epsilon_1 < (V_k - V_1) + \epsilon_k, \epsilon_2 < (V_k - V_2) + \epsilon_k, \cdots, \epsilon_J < (V_k - V_J) + \epsilon_k)}$$

Where F_{-k} is the multivariate distribution of $J - 1$ error terms. Using the marginal density function of ϵ_k, $f_k(\epsilon_k)$, the unconditional probability of alternative k that depends on only the explained variables can be described as,

$$\mathbf{P_k = \int (P_k|\epsilon_k) f_k(\epsilon_k) d\epsilon_k}$$

That is,

$$\mathbf{P_K = \int F_{-k}(\epsilon_1 < (V_k - V_1) + \epsilon_k, \epsilon_2 < (V_k - V_2) + \epsilon_k, \cdots, \epsilon_J < (V_k - V_J) + \epsilon_k) f_k(\epsilon_k) d\epsilon_k}$$

3

The distribution of the unobserved component of the utility function is crucial in discrete choice modeling. Different types of discrete choice models are derived based on the specified density of the random variable, denoted as $\mathbf{f}(\epsilon)$. For certain specifications of $\mathbf{f}(\epsilon)$ and under the assumption that the unobserved component of the utility function is independently and identically distributed (iid) extreme value, the resulting integral has a closed-form solution, as exemplified by the multinomial logit model. Nested logit models, which are an extension of the multinomial logit, also yield a closed-form solution for certain specifications of $\mathbf{f}(\epsilon)$. However, unlike the multinomial logit model, the nested logit model assumes that the joint distribution of the unobserved components follows a generalized extreme value (GEV) distribution [21].

The multinomial logit (MNL) model is one of the most commonly used choice models due to its simple mathematical structure and ease of estimation [16]. The popularity of the multinomial logit can be inferred from its wide applications in econometrics and marketing. The model has been used to predict consumer choices, analyze product market shares based on aggregate data, and estimate price sensitivities in various sectors [14]. The MNL model relies on specific assumptions regarding the error component: (1) the errors follow a Gumbel distribution, (2) the errors are independently and identically distributed across alternatives, and (3) the errors are independently and identically distributed across individuals [12]. However, the popularity of multinomial logit models is diminished by the 'Independence of Irrelevant Alternatives (IIA)' property. The IIA property asserts that the relative odds of choosing between any two alternatives are unaffected by the presence or absence of other alternatives. The MNL model does not inherently satisfy the IIA property, which can lead to biased estimates. The limited applicability of the MNL model due to the IIA property underscores the need for other discrete choice models that can provide efficient estimates while addressing the limitations of the IIA property.

The Nested Logit (NL) model is a prominent alternative to the Multinomial Logit (MNL) model, designed to relax the restrictions imposed by the MNL model and to account for interdependence among pairs of alternatives within the same group. The NL model partitions alternatives into groups, or nests, based on similar characteristics. The error terms of alternatives within a nest are correlated, whereas the error terms of alternatives in different nests are uncorrelated. This structure allows the NL model to capture more realistic substitution patterns compared to the MNL model. The model is a specific case of the Generalized Extreme Value (GEV) model developed by McFadden, which ensures that the model is consistent with utility maximization and generates consistent estimates. This variant is known as the Utility Maximizing Nested Logit (UMNL). An alternative formulation, known as the Non-Normalized Nested Logit (NNNL), is similar to the UMNL except that it does not conform to the principles of utility maximization. Only by imposing restrictions on the nesting parameters or by introducing dummy nests can estimation results consistent with random utility theory be reached [20].

The application of discrete choice models often faces significant challenges due to the limited availability of individual-level data. Collecting detailed consumer data is not only difficult but also often prohibitively expensive, making it impractical for many studies. As a result, researchers frequently rely on market-level data, which is more accessible and can provide valuable insights into market behavior and consumer preferences.

The Berry, Levinsohn, and Pakes (BLP) model is commonly used in the analysis of differentiated product markets, allowing for the estimation of demand parameters by considering the complexities of consumer choice behavior. At its core, the BLP model operates under the framework of Random Utility Maximization (RUM). A key feature of the BLP model is the incorporation of random coefficients, which account for heterogeneity in consumer preferences that cannot be fully captured by observed characteristics alone. This aspect of the model enables it to address issues of price endogeneity and varying consumer responses to product attributes. When applied to aggregate data, the BLP model utilizes market-level information such as prices, market shares, and aggregate product attributes to infer individual-level demand patterns and elasticities. Despite the lack of individual-level data, the use of random coefficients allows the BLP model to capture a broad range of consumer behaviors and market dynamics. However, this robustness comes at the cost of computational intensity, which has led researchers to explore models like the Nested Logit within the BLP framework to streamline estimation while maintaining the

ability to account for the nested nature of consumer choices.

Hence, the objective of this study is to review discrete choice models and to develop a comprehensive understanding of the nested logit model within the BLP (Berry, Levinsohn, and Pakes) framework. To achieve this, Section 2 discusses the Multinomial Logit (MNL) model, followed by Section 3, which examines the significance of the Independence of Irrelevant Alternatives (IIA) property. Sections 4 and 5 provide detailed reviews of the Nested Logit (NL) model and the BLP model, respectively, with Section 5 also introducing the integration of MNL and NL models within the BLP framework. In Section 6, a simulation study is conducted to evaluate the performance of the model structure developed in the previous sections. Finally, Section 7 presents the empirical application of the methodology, demonstrating its potential to address questions regarding market efficiency.

2 Multinomial Logit Model

One of the most popular discrete choice models based on maximum liklihood estimation is the multinomial logit model. An extension of the binary logit, MNL is generally used to predict the dependent variable which is nominal and has more than two categories. Meanwhile, the independent variable in the model can be categorical as well as continuous. The model must satisfy the key assumption that the error terms are extreme-value independently and identically distributed. These three hypotheses have been discussed hereby:

- **Gumbel Distribution**

  Generally, the most common assumption about the distribution of the error terms in statistics is that the errors are normally distributed. However, in the multinomial logit model, a normality assumption of the error terms can lead to flawed choice analysis [12]. Koppelman and Bhat (2006) also highlight the preference for Gumbel distribution in the choice model due to its computational advantages when maximization is the objective, it closely approximates the normal distribution and generates closed-form solutions. A closed-form probabilistic choice model implies that the probability can be calculated without numerical integration or simulation methods.

  Hence, the Gumbel distribution has the following cumulative distribution:

$$\mathbf{P}(\mathbf{z} < \mathbf{t}) = \mathbf{F}(\mathbf{t}) = \int_{-\infty}^{\mathbf{t}} \frac{\mathbf{1}}{\theta} \exp\left(-\frac{\mathbf{z} - \mu}{\theta}\right) \exp\left(-\exp\left(-\frac{\mathbf{z} - \mu}{\theta}\right)\right) \mathbf{dz} = \exp\left(-\exp\left(-\frac{\mathbf{t} - \mu}{\theta}\right)\right)$$

  and probability density function:

$$\mathbf{f}(\mathbf{z}) = \frac{\mathbf{1}}{\theta} \exp\left(-\frac{\mathbf{z} - \mu}{\theta}\right) \exp\left(-\exp\left(-\frac{\mathbf{z} - \mu}{\theta}\right)\right)$$

  respectively. Here, μ is the location parameter and θ denotes the scale parameter which determines the variance of the distribution. Further, the first two moments of the distribution, that is, the mean and variance of the distribution are given by:

$$\mathbf{E}(\mathbf{z}) = \mu + (\mathbf{0.577})\theta$$

$$\mathbf{V}(\mathbf{z}) = \frac{\pi^2}{6}\theta^2$$

  respectively. Here, 0.577 is known as the Euler-Mascheroni constant.

  With respect to the MNL model, each ϵ follows a Gumbel distribution wherein the mean of ϵ_j's is not identified in the presence of an intercept in the $\mathbf{V_j}$. Then without any loss of generality, it can be assumed that $\mu_{\mathbf{j}} = 0 \ \forall j = 1, 2, \cdots, J$ alternatives. Moreover, the overall scale of utility is not identified, only $J - 1$ scale parameters may be identified. The natural choice of normalization is then to impose that of one of the $\theta_{\mathbf{j}}$ as equal to 1 [7]
  .

- **Independence of Errors**

  If the independence of errors holds, it implies a univariate distribution of the error term. Using the mathematical notation for the random utility models as described above, the conditional and unconditional probability of selecting an alternative can be written as:

$$(\mathbf{P_k}|\epsilon_{\mathbf{k}}) = \prod_{\mathbf{k} \neq \mathbf{j}} \mathbf{F_j}(\mathbf{V_k} - \mathbf{V_j} + \epsilon_{\mathbf{k}})$$

and

$$\mathbf{P_k} = \int \prod_{k \neq j} \mathbf{F_j}(\mathbf{V_k} - \mathbf{V_j} + \epsilon_k)\mathbf{f_k}(\epsilon_k)\mathbf{d}\epsilon_k$$

respectively, where F_j is the cumulative distribution for ϵ_j and f is the probability density function. This means that the computation of probabilities requires a solution for one-dimensional integral only.

- **Identically Distributed Errors**

Through the assumption of Gumbel distribution, we know that the location parameter is not identified for any of the error terms. Therefore, the hypothesis of identically distributed errors is essentially the case of homoscedasticity, wherein the scale parameter of Gumbel distribution is the same for all alternatives. Further, as the scale parameter for one of the alternatives has been normalized to 1, it can be assumed that in the case of homoscedasticity without any loss of generality, $\theta_j = 1 \ \forall j = 1, 2, \cdots, J$. The condition and unconditional probabilities can therefore be further simplified to:

$$(\mathbf{P_k}|\epsilon_k) = \prod_{k \neq j} \mathbf{F}(\mathbf{V_k} - \mathbf{V_j} + \epsilon_k)$$

and

$$\mathbf{P_k} = \int \prod_{k \neq j} \mathbf{F}(\mathbf{V_k} - \mathbf{V_j} + \epsilon_k)\mathbf{f_k}(\epsilon_k)\mathbf{d}\epsilon_k$$

where F and f are the Gumbel cumulative distribution and probability density functions with the location parameter, $\mu = 0$, and the scale parameter, $\theta = 1$.

Based on the assumptions on the distribution of the error terms and the utility function,

$$\mathbf{U_j} = \mathbf{V_j} + \epsilon_j$$

Where, $j = 1, \cdots, J$ and $J \geq 2$, we derive the multinomial logit model which gives the choice probabilities for each alternative as a function of the systematic portion of the utility of all the alternatives [12]. These probabilities are commonly expressed as:

$$\mathbf{Pr(k)} = \frac{\exp(\mathbf{V_i})}{\sum_{j=1}^{J} \exp(\mathbf{V_j})}$$

Where $Pr(k)$ is the probability of an individual choosing an alternative k and V_j is the systematic component of the utility of alternative j.

2.1 Choice Probability

The choice probability can be easily derived using the assumption of the error term that it has an i.i.d Gumbel distribution [7]. Let an alternative k be the alternative selected by the decision-maker over alternative j, then from the assumption of Gumbel distribution and identically distributed errors,

$$\mathbf{P}(\epsilon_j < \mathbf{V_k} - \mathbf{V_j} + \epsilon_k) = \mathbf{e}^{-\mathbf{e}^{-(\mathbf{V_k} - \mathbf{V_j} + \epsilon_k)}}$$

From the assumption of independence of errors, it can therefore be stated that the probability of choosing the alternative k conditional on ϵ_k is the product of probabilities of all alternatives except k, that is,

$$(\mathbf{P_k}|\epsilon_k) = \prod_{k \neq j} \mathbf{e}^{-\mathbf{e}^{-(\mathbf{V_k} - \mathbf{V_j} + \epsilon_k)}}$$

7

This implies the unconditional probability is,

$$P_k = \int_{-\infty}^{+\infty} \left(\prod_{k \neq j} e^{-e^{-(V_k - V_j + \epsilon_k)}} \right) e^{-\epsilon_k} e^{-e^{-\epsilon_k}} d\epsilon_k$$

Expanding the above equation for all alternatives including k, we get,

$$P_k = \int_{-\infty}^{+\infty} \left(\prod_{j} e^{-e^{-(V_k - V_j + \epsilon_k)}} \right) e^{-\epsilon_k} d\epsilon_k$$

$$P_k = \int_{-\infty}^{+\infty} \left(e^{-\sum_{j=1}^{J} e^{-(V_k - V_j + \epsilon_k)}} \right) e^{-\epsilon_k} d\epsilon_k = \int_{-\infty}^{+\infty} \left(e^{-e^{-\epsilon_k} \sum_{j=1}^{J} e^{-(V_k - V_j)}} \right) e^{-\epsilon_k} d\epsilon_k$$

Now let, $y = e^{-\epsilon_K} \implies dy = -e^{-\epsilon_K} d\epsilon_k$. Substituting the variable y in the above equation,

$$P_k = \int_0^{+\infty} \left(e^{-y \sum_{j=1}^{J} e^{-(V_k - V_j)}} \right) dy$$

This integral has a closed-form solution,

$$P_k = \left[-\frac{e^{-t \sum_{j=1}^{J} e^{-(V_k - V_j)}}}{\sum_{j=1}^{J} e^{-(V_k - V_j)}} \right]_0^{+\infty} = \frac{1}{\sum_{j=1}^{J} e^{-(V_k - V_j)}}$$

and can be rewritten using the logit probability as,

$$P_k = \frac{e^{V_k}}{\sum_{j=1}^{J} e^{V_j}}$$

2.2 Interpretation: Marginal Effects and Elasticities

2.2.1 Marginal Effects

Unlike the linear models where coefficients can be directly interpreted as the marginal effects of the independent variable on the dependent variable, the interpretation of the coefficient in the multinomial logit models is not easy. Marginal effects are one of the techniques to quantify the change in probabilities with respect to the change in the values of the characteristics that define the utility.

The explanatory variables can be classified as individual-specific (s_i) and alternative-specific (x_{ij}). Therefore, a researcher might be interested in knowing the change in the probability of choosing an alternative in response to a change in individual-specific characteristics or an alternative-specific characteristic. Both of these marginal effects have been covered hereby.

Alternative Specific: Consider P_{ik} be the probability for an individual i selecting alternative k and let x_{ik} be the alternative-specific explanatory variable. Then the marginal effect is calculated by differentiating P_{ik} with respect to x_{ik}. This measure is also known as the 'Direct Derivative' and is given by [21],

$$\frac{\partial P_{ik}}{\partial x_{ik}} = \frac{\partial \left[(e^{V_{ik}}) / \left(\sum_{j=1}^{J} e^{V_{ij}} \right) \right]}{\partial x_{ik}}$$

$$= \left[e^{V_{ik}} / \sum_{j=1}^{J} e^{V_{ij}} \right] \left(\frac{\partial V_{ik}}{\partial x_{ik}} \right) - \left[e^{V_{ik}} / \left(\sum_{j=1}^{J} e^{V_{ij}} \right)^2 \right] (e^{V_{ik}}) \left(\frac{\partial V_{ik}}{\partial x_{ik}} \right)$$

$$= \left(\frac{\partial V_{ik}}{\partial x_{ik}} \right) \left(P_{ik} - P_{ik}^2 \right)$$

Let $V_{ik} = \beta_k x_{ik}$, the observed component of the linear utility function. Then, the direct derivative can be written as

$$\frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{x_{ik}}} = \beta_{\mathbf{k}}(\mathbf{P_{ik}})(\mathbf{1} - \mathbf{P_{ik}})$$

The value of the derivative is the largest when $P_{ik} = 1 - P_{ik}$, that is, $P_{ik} = 0.5$ and becomes smaller as P_i approaches zero or one. This implies that the impact of a change in the observed variable is the highest when the choice probability has a high degree of uncertainty and as the probability of choosing an alternative improves (that is, moves closer to zero or one), the effect of the observed variable diminishes. It can also be explicitly stated that the sign of the marginal effect is the same as the sign of the coefficient β_k. Hence, for an increase in x_{ik}, the P_{ik} increase if β_k is positive and decreases if β_k is negative.

Often it is of more interest to calculate the effect of change in the observed variable of another alternative on the choice probability of the alternative of interest. This is known as the 'Cross Derivative' and is calculated as,

$$\frac{\partial \mathbf{P_{im}}}{\partial \mathbf{x_{ik}}} = \frac{\partial \left[\left(\mathbf{e^{V_{im}}} \right) \Big/ \left(\sum_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{e^{V_{ij}}} \right) \right]}{\partial \mathbf{x_{ik}}}$$

Here, P_{im} denotes the choice probability of choosing alternative m by an individual i and x_{ik} is the observed characteristic of alternative k. Solving the differentiation,

$$= - \left[\frac{\left(\mathbf{e^{V_{im}}} \right)}{\left(\sum_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{e^{V_{ij}}} \right)^{\mathbf{2}}} \right] \mathbf{e^{V_{ik}}} \frac{\partial \mathbf{V_{ik}}}{\partial \mathbf{x_{ik}}}$$

$$= - \frac{\partial \mathbf{V_{ik}}}{\partial \mathbf{x_{ik}}} \mathbf{P_{im}} \mathbf{P_{ik}}$$

Assuming the same linear utility representation as for the direct-derivative, $\dfrac{\partial V_{ik}}{\partial x_{ik}} = \beta_k$ and substituting the value in the marginal effect we get,

$$\frac{\partial \mathbf{P_{im}}}{\partial \mathbf{x_{ik}}} = -\beta_{\mathbf{k}} \mathbf{P_{im}} \mathbf{P_{ik}}$$

Where, P_{ik} is the choice probability of alternative k.

The above derivation tells informs that there exists an inverse relation between the choice probability of an alternative and the attribute of another alternative. If the observed variable x_{ik} is desirable and hence has a positive coefficient β_k, then an increase in x_{ik} reduces the probability of choosing alternative m.

An important outcome from the derivation of the direct and cross derivatives is that the sum of the marginal effects over all the alternatives must be zero. That is, if there is a change in the attribute of an alternative such that its choice probability increases, then the additional probability is drawn from other alternatives and hence, their choice probability necessarily decreases. This also stems from the fact that the sum of the probabilities is always one. This can be shown as,

$$\sum_{\mathbf{k=1}}^{\mathbf{J}} \frac{\partial P_{ik}}{\partial x_{ij}} = \frac{\partial P_{ij}}{\partial x_{ij}} + \sum_{k \neq j} \frac{\partial P_{ik}}{\partial x_{ij}}$$

$$= \beta_{\mathbf{j}}(\mathbf{P_{ij}})(\mathbf{1} - \mathbf{P_{ij}}) - \sum_{\mathbf{k \neq j}} \beta_{\mathbf{j}} \mathbf{P_{ik}} \mathbf{P_{ij}}$$

$$= \beta_{\mathbf{j}} \mathbf{P_{ij}}(\mathbf{1} - \mathbf{P_{ij}}) - \beta_{\mathbf{j}} \mathbf{P_{ij}} \sum_{\mathbf{k \neq j}} \mathbf{P_{ik}}$$

$$= \beta_j \mathbf{P_{ij}} (1 - \mathbf{P_{ij}}) - \beta_j \mathbf{P_{ij}} (1 - \mathbf{P_{ij}}) = 0$$

Individual Specific: Similar to alternative-specific attributes, individual-specific attributes, such as income, can also influence the choice probability of an alternative [12]. The impact of a change in such a variable on the choice probability of an alternative can be understood in two ways:

1. Direct Effect: The choice probability of an alternative k, denoted by P_{ik}, will change in response to a change in the individual-specific variable s_i, where s_i specifically affects alternative k for individual i. This effect can be quantified as the direct derivative of P_{ik} with respect to s_i.

2. Indirect Effect: The change in s_i also influences the choice probabilities of all other alternatives. Consequently, the change in the choice probability P_{ik} due to s_i needs to be computed considering the impact of s_i on alternatives other than k for individual i. This effect is captured as the cross derivative of P_{ik} with respect to s_i, reflecting how changes in s_i alter the relative choice probabilities across all alternatives.

The marginal effect of a change in s_i on P_{ik}, specific to alternative k is given by,

$$\frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{s_i}} = \beta_k \mathbf{P_{ik}} (1 - \mathbf{P_{ik}})$$

Meanwhile, the marginal effect of a change in s_i on P_{ik}, specific to another alternative j is given by,

$$\frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{s_i}} = -\beta_j \mathbf{P_{ik}} \mathbf{P_{ij}}$$

Aggregating the above marginal effect over all alternatives other than k,

$$\sum_{j \neq k} \frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{s_i}} = -\mathbf{P_{ik}} \sum_{j \neq k} \beta_j \mathbf{P_{ij}}$$

Therefore, the change in choice probability of an individual i for alternative k due to a change in an individual-specific attribute can be measured by summing up the marginal effects over all alternatives,

$$\frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{s_i}} = \beta_k \mathbf{P_{ik}} (1 - \mathbf{P_{ik}}) - \mathbf{P_{ik}} \sum_{j \neq k} \beta_j \mathbf{P_{ij}}$$

$$= \mathbf{P_{ik}} \left(\beta_k - \beta_k \mathbf{P_{ik}} - \sum_{j \neq k} \beta_j \mathbf{P_{ij}} \right)$$

$$= \mathbf{P_{ik}} \left(\beta_k - \left(\beta_k \mathbf{P_{ik}} + \sum_{j \neq k} \beta_j \mathbf{P_{ij}} \right) \right)$$

$$= \mathbf{P_{ik}} \left(\beta_k - \sum_{j=1}^{J} \beta_j \mathbf{P_{ij}} \right)$$

Hence, for an individual-specific variable, the marginal effect is given by

$$\frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{s_i}} = \mathbf{P_{ik}} \left(\beta_k - \sum_{j=1}^{J} \beta_j \mathbf{P_{ij}} \right)$$

Where $\sum_{j=1}^{J} \beta_j \mathbf{P_{ij}}$ is the weighted average of the coefficients for all the alternatives and the weights are given by the probability of choosing the alternatives.

Further, the sign of the marginal effect is ambiguous. It will be positive only if the coefficient of alternative k is larger than the weighted average of the coefficients of all the alternatives [7].

2.2.2 Elasticities

Elasticities are also a popular measure to interpret the choice probabilities since they are normalized for the variable units [21]. Elasticities are commonly defined as the percentage change in the outcome variable for a percentage change in the explanatory variable. In the context of the MNL model, the outcome variable is the choice probability of choosing an alternative, say P_{ik} and the explanatory variable is the alternative-specific, x_{ik} and individual-specific, s_i attribute. Hereby, the elasticity for each of the marginal effects has been calculated.

'Direct Elasticity' corresponds to the direct derivative derived above and it measures the proportional change in the choice probability of an P_{ik} for a proportional change in the explanatory variable of that alternative, x_{ik}.

$$\eta_{\mathbf{x_{ik}}}^{\mathbf{P_{ik}}} = \frac{\frac{\partial \mathbf{P_{ik}}}{\mathbf{P_{ik}}}}{\frac{\partial \mathbf{x_{ik}}}{\mathbf{x_{ik}}}} = \frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{x_{ik}}} \times \frac{\mathbf{x_{ik}}}{\mathbf{P_{ik}}}$$

$\frac{\partial P_{ik}}{\partial x_{ik}}$ is the direct derivative and hence its value can be substituted to get the 'Direct Elasticity'.

$$\eta_{\mathbf{x_{ik}}}^{\mathbf{P_{ik}}} = \beta_{\mathbf{k}}(\mathbf{P_{ik}})(\mathbf{1} - \mathbf{P_{ik}}) \times \frac{\mathbf{x_{ik}}}{\mathbf{P_{ik}}}$$

$$\eta_{\mathbf{x_{ik}}}^{\mathbf{P_{ik}}} = \beta_{\mathbf{k}}\mathbf{x_{ik}}(\mathbf{1} - \mathbf{P_{ik}})$$

Similarly, the 'Cross-Elasticity' corresponds to the cross-derivative and can be derived as the percent change in the choice probability of an P_{im} for a percent change in the attribute of another alternative x_{ik}.

$$\eta_{\mathbf{x_{ik}}}^{\mathbf{P_{im}}} = \frac{\frac{\partial \mathbf{P_{im}}}{\mathbf{P_{im}}}}{\frac{\partial \mathbf{x_{ik}}}{\mathbf{x_{ik}}}} = \frac{\partial \mathbf{P_{im}}}{\partial \mathbf{x_{ik}}} \times \frac{\mathbf{x_{ik}}}{\mathbf{P_{im}}} \qquad \forall \mathbf{k} \neq \mathbf{m}$$

Substituting the value for cross-derivative to obtain the 'Cross-Elasticity'.

$$\eta_{\mathbf{x_{ik}}}^{\mathbf{P_{im}}} = -\beta_{\mathbf{k}}(\mathbf{P_{im}})(\mathbf{P_{ik}}) \times \frac{\mathbf{x_{ik}}}{\mathbf{P_{im}}}$$

$$\eta_{\mathbf{x_{ik}}}^{\mathbf{P_{im}}} = -\beta_{\mathbf{k}}\mathbf{x_{ik}}\mathbf{P_{ik}}$$

An important property of the Multinomial Logit (MNL) model is that the cross-elasticity of choice probabilities is uniform across all alternatives in the choice set. Specifically, the cross-elasticity is independent of the particular alternative being analyzed, meaning that alternative *m* does not appear in the formula for cross-elasticity. This property implies that if an improvement in the attribute of one alternative leads to a decrease in the choice probability of another alternative by a certain percentage, then the choice probabilities of all other alternatives, excluding the improved alternative, will also decrease by the same percentage. This uniform response to changes in attributes reflects the Independence of Irrelevant Alternatives (IIA) property inherent to the MNL model. Thus, the cross-elasticity in the MNL model serves as a restatement of the IIA property, highlighting the model's limitation in capturing nuanced changes in choice probabilities when alternatives are added or removed from the choice set.

Finally, the elasticity for individual-specific attribute s_i can also be calculated similarly.

$$\eta_{\mathbf{s_i}}^{\mathbf{P_{ik}}} = \frac{\frac{\partial \mathbf{P_{ik}}}{\mathbf{P_{ik}}}}{\frac{\partial \mathbf{s_i}}{\mathbf{s_i}}} = \frac{\partial \mathbf{P_{ik}}}{\partial \mathbf{s_i}} \times \frac{\mathbf{s_i}}{\mathbf{P_{ik}}}$$

$$\eta_{\mathbf{s_i}}^{\mathbf{P_{ik}}} = \mathbf{P_{ik}}\left(\beta_{\mathbf{k}} - \sum_{\mathbf{j=1}}^{\mathbf{J}} \beta_{\mathbf{j}}\mathbf{P_{ij}}\right) \times \frac{\mathbf{s_i}}{\mathbf{P_{ik}}}$$

$$\eta_{s_i}^{P_{ik}} = \left(\beta_k - \sum_{j=1}^{J} \beta_j P_{ij} \right) \times s_i$$

It must be highlighted that the elasticities are calculated under the assumption that the explanatory variables are continuous. In the presence of categorical or ordinal variables, the elasticities cannot be derived, since discrete variables are not differentiable. An alternative is to measure the incremental change in the choice probability of an alternative for a unit change in a discrete attribute and then use the differences to compute the arc elasticities [12].

The Multinomial Logit (MNL) model is a widely used discrete choice model, valued for its straightforward mathematical formulation and ease of implementation. Its key advantages include a simple closed-form expression for choice probabilities, the ability to estimate coefficients efficiently using maximum likelihood estimation, and a clear interpretation of marginal effects. The MNL model's structure allows for the modeling of choice probabilities across multiple alternatives with a consistent and intuitive approach. However, a significant limitation of the MNL model is its inherent Independence of Irrelevant Alternatives (IIA) property. The following section will delve deeper into the implications of the IIA property

3 Independence of Irrelevant Alternatives

The Independence of Irrelevant Alternatives (IIA) is a fundamental property underlying the Multinomial Logit (MNL) model. The IIA property asserts that for any individual, the ratio of the probabilities of choosing between any two alternatives is unaffected by the presence or attributes of any other alternatives in the choice set [7]. To understand the concept better, consider P_l, P_m and P_n as the choice probabilities of alternatives l, m and n, where

$$P_l = \frac{e^{V_l}}{\sum_j e^{V_j}} = \frac{e^{V_l}}{e^{V_l} + e^{V_m} + e^{V_n}}$$

$$P_m = \frac{e^{V_m}}{\sum_j e^{V_j}} = \frac{e^{V_m}}{e^{V_l} + e^{V_m} + e^{V_n}}$$

$$P_n = \frac{e^{V_n}}{\sum_j e^{V_j}} = \frac{e^{V_n}}{e^{V_l} + e^{V_m} + e^{V_n}}$$

The ratio of the three choice probabilities can then be calculated as follows,

$$\frac{P_l}{P_m} = \frac{\dfrac{e^{V_l}}{\sum_j e^{V_j}}}{\dfrac{e^{V_m}}{\sum_j e^{V_j}}} = \frac{e^{V_l}}{\sum_j e^{V_j}} \times \frac{\sum_j e^{V_j}}{e^{V_m}} = \frac{e^{V_l}}{e^{V_m}} = e^{V_l - V_m}$$

Hence, the ratio of the choice probabilities are,

$$\frac{P_l}{P_m} = \frac{e^{V_l}}{e^{V_m}} = e^{V_l - V_m}$$

$$\frac{P_m}{P_n} = \frac{e^{V_m}}{e^{V_n}} = e^{V_m - V_n}$$

$$\frac{P_l}{P_n} = \frac{e^{V_l}}{e^{V_n}} = e^{V_l - V_n}$$

The ratios of probabilities for each pair of alternatives depend only on the explanatory variables of those alternatives and not on the characteristics of the third alternative. In other words, the relative odds of choosing between two alternatives remain constant regardless of the introduction or removal of other alternatives.

There are significant implications of the IIA property on the model framework, estimation, and the use of multinomial logit models [12]. The property allows the addition and removal of alternatives without affecting the parameters of the model. This flexibility allows the researcher to apply the model in cases where different samples face a different set of alternatives. For instance, in a study for the intercity mode of travel choice, individuals in a few sample sets would not have the choice for air travel while another sample might. Moreover, the IIA property facilitates the examination of choices among a subset of alternatives without considering the entire choice set. For instance, in the same intercity travel study, if the researcher believes that the IIA property holds, they can estimate a model using only a subset of alternatives, such as car and train, while excluding other options like air travel and bus services. This approach not only aligns with specific research requirements but also contributes to time and cost efficiencies by reducing the need for data collection on alternatives that are not relevant to the study [21].

While the Independence of Irrelevant Alternatives (IIA) property simplifies the application of the Multinomial Logit (MNL) model, it may not always hold in practice. This limitation arises because other alternatives can indeed be relevant to the ratio of probabilities between a specific pair of alternatives. When the IIA property does not hold, the MNL model can produce misleading or erroneous results. This issue is vividly illustrated by the well-known 'red bus/blue bus paradox'.

Consider a case where a consumer can choose between taking a car and a blue bus. The probability of taking a car is $P_c = 3/4$ and the probability of the consumer selecting the blue bus is $P_{bb} = 1/4$. That is,

$$P_c = \frac{e^{V_c}}{e^{V_c} + e^{V_{bb}}} = \frac{3}{4} \quad \text{and} \quad P_{bb} = \frac{e^{V_{bb}}}{e^{V_c} + e^{V_{bb}}} = \frac{1}{4}$$

Therefore, the ratio of probabilities of the two alternatives is given by,

$$\frac{P_c}{P_{bb}} = \frac{e^{V_c}}{e^{V_{bb}}} = \frac{3}{1}$$

Now consider that a new bus service is introduced, the 'red bus'. For simplicity, it is assumed that the red bus shares the same attributes as those of the 'blue bus' like it has the same schedule and route. The only difference between the two bus services is the color. The introduction of a new alternative provides the consumer with three choices, to drive in a car, take the blue bus, or take the red bus which can lead to a change in the choice probabilities of the alternatives. The IIA property states that the ratio of probabilities is unaffected by the presence of additional alternatives. This implies the ratio of probabilities between the car and the blue bus should remain the same $3 : 1$. Further, as the blue bus and red bus share the same attributes, the ratio of probabilities between the blue bus and red bus should be one. That is, since $e^{V_{bb}} = e^{V_{rb}}$,

$$\frac{P_{bb}}{P_{rb}} = \frac{e^{V_{bb}}}{e^{V_{rb}}} = 1$$

Where, P_{rb} is the probability of choosing the red bus and V_{rb} comprise of the attributes of the red bus.

The only probabilities for which $P_c/P_{bb} = 3 : 1$ and $P_{bb}/P_{rb} = 1 : 1$ is when $P_c = 3/5$, $P_{bb} = 1/5$, and $P_{rb} = 1/5$. These are the probabilities predicted by the multinomial logit model. However, in reality, the probability of taking a car would be expected to remain the same since consumers choose between a car and a bus, and the attributes of the buses are the same. This means that the introduction of the red bus would lead to a change in the probability of taking the blue bus which would now be split between the two buses. Hence, the choice probabilities expected post-initiation of the red bus are $P_c = 3/4$, $P_{bb} = 1/8$ and $P_{rb} = 1/8$.

In this example, the multinomial logit model, because of its IIA property, overestimates the probability of taking the buses and underestimates the probability of taking the car. It predicts that the introduction of the red bus shifts consumers away from the car and the blue bus towards the red bus. Whereas in reality, the introduction of the red bus leads to a change in the ratio of probabilities of the car and blue bus, P_c/P_{bb}, rather than remaining constant as required by the MNL model. Although an extreme case, the example illustrates the strong restrictions the IIA property places on the substitution patterns in the MNL model and also emphasizes the need for alternative models that can accommodate such interdependencies among choices.

The IIA property further relies on the hypothesis of the independence of the error terms. If this assumption is violated, resulting in correlated error terms, the Multinomial Logit (MNL) model can produce biased estimates. This issue can arise when two alternatives are similar, but the model fails to include variables that account for this similarity in the systematic portion of the utility function [7]. To see that suppose that the utilities for two alternatives are

$$U_{i1} = \alpha_1 + \beta_1 s_i + \gamma x_{i1} + \epsilon_{i1}$$

$$U_{i2} = \alpha_2 + \beta_2 s_i + \gamma x_{i2} + \epsilon_{i2}$$

with ϵ_{i1} and ϵ_{i2} uncorrelated. In this situation, the MNL model can be used as the independence of error terms assumption is satisfied. But if there exists a variable z_i that is unobserved in the utility function then the estimated model is,

$$\mathbf{U_{i1}} = \alpha_1 + \beta_1 \mathbf{s_i} + \gamma \mathbf{x_{i1}} + \zeta_{i1}$$

$$\mathbf{U_{i2}} = \alpha_2 + \beta_2 \mathbf{s_i} + \gamma \mathbf{x_{i2}} + \zeta_{i2}$$

$$\zeta_{i1} = \epsilon_{i1} + \mathbf{z_i}$$

$$\zeta_{i2} = \epsilon_{i2} + \mathbf{z_i}$$

Due to the omitted variable bias, the error terms are now correlated which is a violation of the assumption under which the MNL model is derived.

The undesirable characteristics of the Independence of Irrelevant Alternatives (IIA) property—such as biased estimates and flawed inferences—undermine the effectiveness of the standard Multinomial Logit (MNL) model. The assumption of independent error terms can lead to inaccuracies when alternatives are similar but not adequately accounted for in the utility function. This limitation highlights the need for alternative models that address these deficiencies. The Nested Logit (NL) model, a special case of the Generalized Extreme Value (GEV) model developed by McFadden (1972), provides a notable solution. By relaxing the IIA property, the NL model allows for correlated error terms within predefined nests of alternatives, accommodating the possibility of similar unobserved attributes among alternatives. This relaxation enables the model to better capture complex substitution patterns and interdependencies.

4   Nested Logit Model

The Generalized Extreme Value (GEV) models encompass a range of models that exhibit diverse substitution patterns among alternatives. In GEV models, the error terms for all alternatives are jointly distributed according to a generalized extreme value distribution, which generalizes the univariate extreme value distribution used in the Multinomial Logit (MNL) model. This distribution's advantage is its ability to capture correlations among alternatives, thereby accommodating more complex substitution behaviors.

When the correlation among alternatives is zero, the GEV distribution simplifies to the product of independent extreme value distributions, reducing to the standard MNL model [21]. Among the GEV models, the Nested Logit (NL) model is one of the most widely employed. The NL model is favored for its relatively simpler functional form and its ability to provide richer substitution patterns compared to the MNL model.

In the Nested Logit framework, the choice process is hierarchical: consumers first select a group of alternatives, termed "nests," and then choose an alternative within the selected nest. Nests are created based on shared characteristics or similarities in unobserved attributes. For instance, in a transportation mode choice scenario, nests could be categorized into "public transportation" (e.g., bus, train) and "private transportation" (e.g., car, bike). The NL model allows for correlated error terms within each nest, reflecting unobserved preferences or factors such as convenience, cost, or comfort, which are not explicitly included in the model. Conversely, the error terms across different nests remain uncorrelated. This structure enables the Nested Logit model to capture more nuanced substitution patterns and improve modeling flexibility compared to the standard MNL model.

The product alternatives are categorized into nests such that the following properties hold [21]:

- For any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes or existence of other alternatives, that is, the IIA property holds within each nest. In other words, this implies that within a nest the cross-elasticity between pairs of alternatives is the same for all pairs of alternatives in the nest.

- For any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests, that is, the IIA property does not hold for alternatives in different nests.

A two-level nested logit model is the most common choice among researchers and very few analysts estimate the nested logit models with more than three levels. The researcher decides the nesting pattern based on existing theory and casual observations [19]. A convenient method to visualize the nesting patterns is through the tree diagram. An upside-down tree diagram denotes the nests among which the alternatives have been divided and the alternatives within each nest. However, the 'decision tree' is purely for interpretation purposes and they do not necessarily imply that an individual makes choices in a certain order, that is, the derivation of the NL model makes no assumptions about the structure of the choice process [6]. For instance, for a two-level decision tree, the top part of the tree shows the 'upper model', the choice between the nests $Group\_0$ and $Group\_1$, and the bottom part of the tree shows the 'lower model', the choice among alternatives $Subgroup\_2$, $Subgroup\_3$ and $Subgroup\_4$ within a nest.

4.1   Choice Probability

Consistent with the utility maximization, the following nested logit model is based on McFadden (1972). Consider a set of alternatives $j = 1, 2, \cdots J$ partitioned into M subsets $B_m, m = 1, 2, \cdots M$ so that each alternative belongs to exactly one subset. These subsets are known as 'Nests'. As before, the utility for an individual i received from an alternative j can be represented as

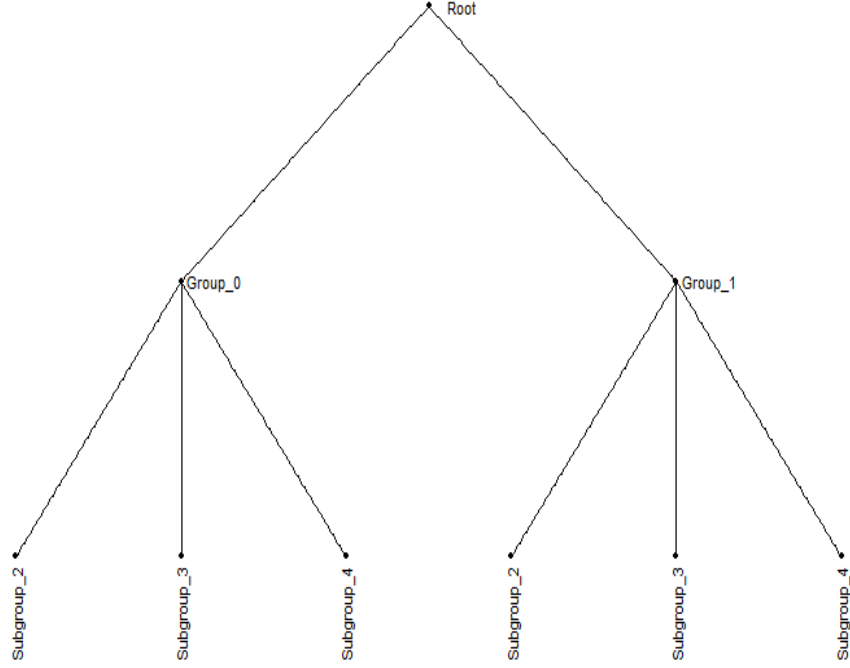$$\mathbf{U_{ij}} = \mathbf{V_{ij}} + \mathbf{\epsilon_{ij}}$$

16

Figure 1: Two Level Nested Logit Model Tree Structure

V_{ij} represents the systematic portion of the utility observed by the researcher, and ϵ_{ij} is the random variable whose value is not observed by the researcher. The choice probability of alternative $j \in B_m$ for individual i then can be decomposed into two parts: the probability that an alternative within the nest B_m is chosen and the probability that the alternative j is chosen given that an alternative in B_m is chosen. Hence, the choice probability is the product of two standard logit probabilities and can be represented as,

$$\mathbf{P_{ij} = P_{ij|B_m} \times P_{iB_m}}$$

Where, P_{ij} is the probability of an individual i selecting an alternative j and it is the product of $P_{ij|B_m}$ which is the conditional probability of choosing alternative j given that the nest B_m is selected and P_{iB_m} which is the probability of choosing some alternative in nest B_m.

The error terms in the nested logit model are a generalized version of the error terms distribution in the MNL model. This special form of the GEV distribution extends the Extreme Value Type 1 distribution by allowing the alternatives within a nest to have mutually correlated error terms. The mutual correlation between error terms of alternatives within a nest is represented by an additional parameter τ_m in the joint distribution of error terms. Some common parameterizations of τ_m are $\sigma_m = 1 - \tau_m$ (McFadden (1981) [17]) and $\mu_m = 1/\tau_m$ (Louviere et. al. (2000) [15]). Hereby, τ_m is specified to be equal to $\sqrt{1 - \rho_m}$ where ρ_m is the correlation coefficient [10]. Since τ_m is an inverse measure of the correlation or an index of the dissimilarity of alternatives included in the nest m, it is known as the 'dissimilarity parameter', the 'nesting parameter', and the 'logsum parameter'.

Based on the logsum parameter, the utilities can be rescaled by its inverse since a lower τ_m implies a higher correlation and vice versa. Hence, the choice probability of choosing alternative j given some alternatives in its nest B_m,

$$\mathbf{P_{j|B_m} = \frac{e^{V_j/\tau_m}}{\sum_{k \in B_m} e^{V_k/\tau_m}}}$$

The denominator of the above equation represents the measure of the attractiveness of the nest B_m. The log of the expression is known as the 'logsum variable' or the 'Inclusive Value (IV)'. It states the expected value an individual i obtains from the alternatives in the nest B_m. The variable can be written as follows:

$$\mathbf{IV_m} = \ln \sum_{k \in B_m} e^{V_k / \tau_m}$$

Finally, the probability of choosing some alternative from nest m is the MNL probability for the choice between the sets. The rescaled inclusive value takes the role of the deterministic part of the utility.

$$\mathbf{P_{B_m}} = \frac{e^{\tau_m \mathbf{IV_m}}}{\sum_{k=1}^{M} e^{\tau_k \mathbf{IV_k}}}$$

It is known that the RUM choice probabilities are based on the utility differences. This implies that adding a constant or scaling the utilities by a constant factor does not change the probabilities. Hence, the RUM models have to normalize the utilities. In the MNL model, the utilities are scaled such that the error terms have a variance of $\sigma^2 = \pi^2/6$. Since the error terms are independent by assumption in the MNL model, their differences have a variance of $2\sigma^2$. However, in the NL model, the error terms are positively correlated within a nest and a higher correlation implies a lower variance of their difference. Based on the relationship between the dissimilarity parameter τ_m and the correlation coefficient ρ_m, the variance of the difference of the error terms is $2\sigma^2\tau_m^2$. The difference is further normalized by normalizing the utilities by the factor $1/\tau_m$ to get the variance of normalized differences as $2\sigma^2$. This normalization is necessary since without it the utilities in each nest would be scaled by a different factor and hence would not be comparable across nests [10].

The 'logsum parameter' (or the dissimilarity parameter or the nesting coefficient'), τ_m, measures the degree of correlation between the unobserved components of alternatives in the nest m and hence it also characterizes the degree of substitutability between those alternatives. To ensure consistency with the random utility maximization principles, the nesting parameter is bounded between zero and one. The interpretation of the different values of the parameter is as follows [12]:

- $\tau_m > 1$, this is not consistent with the theoretical derivation, Hence, the NL model is rejected.

- $\tau_m = 1$, this implies that there is zero correlation between the alternatives in the nest and so the NL model collapses to the MNL model. That is, in this case $\sum_{m=1}^{M} e^{IV_m} = \sum_{k=1}^{J} e^{V_k}$ must hold.

- $0 < \tau_m < 1$, implies a non-zero correlation among pairs of alternatives in a nest. This range of values is considered appropriate for the nested logit model. A lower value of τ_m indicates increased substitution among alternatives in a nest.

- $\tau_m = 0$, this case implies a perfect correlation among the pairs of alternatives in a nest. That is, the choice between the nested alternatives, conditional on the nest is deterministic.

- $\tau_m < 0$, is as well not consistent with the theoretical derivation and hence, the NL model can be rejected.

4.2 Estimation

The parameters of the nested logit model can be estimated in two ways: simultaneously and sequentially [21]. Simultaneous estimation implies that the parameters are estimated using the standard maximum likelihood techniques. The choice probabilities are substituted in the log-likelihood function to obtain an explicit function of the parameters of the model. The values of the parameters that maximize this function under general conditions are consistent and efficient. The maximization of the function is sometimes difficult since the log-likelihood function is not globally concave and even in concave areas is not close to a quadratic.

An alternative means to estimate the parameters of the NL model is to exploit the fact that the choice probabilities can be decomposed into marginal and conditional probabilities that are logit. Such a sequential estimation

is conducted bottom-up, that is, the lower models (for the choice of alternatives within a nest) are estimated first. Then the inclusive values are calculated for each lower model using the estimated coefficients. Finally, the upper model (for choice of the nest) is estimated using inclusive values as the explanatory variable. However, there are two drawbacks of sequential estimation. First, while consistent, the standard errors of parameters of the upper model are biased downwards because the variance of the inclusive value terms is not included in the calculation of standard errors. This leads to smaller confidence intervals and larger t-statistics for the parameter estimates than true and the upper model will appear better than it is. The second drawback is that because several parameters appear in several sub-models, estimating the various upper and lower models separately provides separate estimations of the common parameters. In simultaneous estimation using maximum likelihood, the common parameters are constrained to be the same wherever they appear in the model.

Therefore, these two complications imply that the sequential estimates of the nested logit mode, while consistent, are not as efficient as simultaneous estimation. A researcher should use sequential estimation if problems arise in simultaneous estimation. The estimates of the sequential estimation can be used as starting values in the simultaneous estimation. The advantage of the decomposition of the choice probability is hence, not in the estimation but in understanding the structure of the NL model.

4.3 Elasticities

Similar to the Multinomial Logit (MNL) model, elasticities are crucial for interpreting choice probabilities in the Nested Logit (NL) model. As in the MNL model, direct elasticity measures the effect of a change in the attribute of an alternative on the choice probability of that alternative, while cross elasticity measures the change in the choice probability of one alternative due to a change in the attribute of another alternative. In the NL model, these elasticities are further distinguished based on whether the alternatives are in the same nest or different nests. The elasticity expressions for the NL model, along with those for the MNL model, are presented in Table 1 [13].

Firstly, the direct elasticities for alternatives not within any nest are structurally identical to the MNL direct elasticities. For alternatives within a nest, the direct elasticities are greater than those for alternatives not in a nest. This reflects the intuitive notion that alternatives within a nest face more competitive pressures from similar alternatives compared to those outside any nest.

Cross elasticities between pairs of alternatives depend on whether the pairs are within the same nest or in different nests. When alternatives are in different nests, the cross elasticity is structurally identical to that in the MNL model. However, when alternatives are within the same nest, the cross elasticity is larger than that for alternatives in different nests. This consistency aligns with the higher substitutability between alternatives within the same nest.

The dissimilarity parameter, τ_m, plays a crucial role in differentiating elasticity equations for nested alternatives. When, $\tau_m = 1$, the maximum value, the direct and cross elasticities for nested alternatives converge to the corresponding elasticities for non-nested alternatives. As τ_m decreases from one to zero, the direct and cross elasticities within the nest become larger than those between nests. This indicates that the sensitivity of a nested alternative to changes in its own attributes or the attributes of other alternatives within the same nest increases significantly compared to the sensitivity of non-nested alternatives.

The Nested Logit (NL) model not only aligns with utility maximization principles but also accounts for substitutability between alternatives. By nesting similar alternatives, the NL model relaxes the Independence of Irrelevant Alternatives (IIA) property inherent in the MNL model, leading to more realistic and interpretable estimates. This enhanced flexibility and accuracy make the NL model particularly valuable in market research and competition analysis, where understanding complex substitution patterns is crucial. Consequently, the NL model is widely adopted for its ability to provide robust, easily estimable insights into consumer choice behavior.

Table 1: Direct and Cross-Elasticities the Nested Logit and MNL models

Model Structure	Direct-Elasticity (change in $\mathbf{P_j}$ due to change in $\mathbf{x_j}$)	Cross-Elasticity (change in $\mathbf{P_{j'}}$ due to change in $\mathbf{x_j}$)
MNL Model	$(\mathbf{1} - \mathbf{P_j})\beta_{\mathbf{j}}\mathbf{x_j}$	$-\mathbf{P_j}\beta_{\mathbf{j}}\mathbf{x_j}$
NL Model	j not in nest $(\mathbf{1} - \mathbf{P_j})\beta_{\mathbf{j}}\mathbf{x_j}$	j and j' not in the same nest $-\mathbf{P_j}\beta_{\mathbf{j}}\mathbf{x_j}$
	j in nest m $\left[(\mathbf{1} - \mathbf{P_j}) + \left(\dfrac{\mathbf{1}}{\tau_{\mathbf{m}}} - \mathbf{1}\right)(\mathbf{1} - \mathbf{P_{j|m}})\right]\beta_{\mathbf{j}}\mathbf{x_j}$	j and j' in nest m $-\left[\mathbf{P_j} + \left(\dfrac{\mathbf{1}}{\tau_{\mathbf{m}}} - \mathbf{1}\right)\mathbf{P_{j|m}}\right]\beta_{\mathbf{j}}\mathbf{x_j}$

5   BLP: Berry, Levinsohn and Pakes Model

Estimation of demand has been of focus in recent studies. One of the reasons being the essential role demand plays in examining questions regarding competition, market power, and mergers.

One common method for estimating demand for a set of closely related but not identical products is to specify a system of demand equations, one for each product, using market-level data. Each equation estimates the demand for a product as a function of its own price, the prices of other products, and additional variables [18]. However, there are several fundamental challenges in demand estimation that must be addressed [3].

The first challenge is the endogeneity of prices, which refers to the statistical dependence between prices and unobserved factors that also affect demand. Price endogeneity can result in biased and inconsistent parameter estimates, complicating the inference of causal relationships. This issue can arise due to reverse causality or omitted variable bias. A common technique to address price endogeneity is the use of instrumental variables (IV). Instrumental variables are correlated with the endogenous variable (price) but uncorrelated with the error term. By identifying and employing appropriate instruments, it is possible to isolate the exogenous variation in prices and obtain consistent estimates. However, the challenge to demand identification extends beyond price endogeneity.

Another challenge in demand estimation is that the demand for a good depends on multiple latent demand shocks. While it is straightforward to infer the impact of prices and characteristics of all related goods on the demand for a given good, the demand for a specific good also shifts in response to variations in unobserved demand shifters associated with other goods. Since prices, characteristics, and demand shocks of the focal good and all related goods cannot be excluded, they must be held constant as 'ceteris paribus' effects to measure the impact of a price change on demand. Consequently, conventional econometric techniques cannot be used to estimate demand without relying on strong functional form assumptions for identification.

Aggregate data is often used for demand estimation due to the lack of availability of consumer-level data. A few commonly observed variables in market-level data are the number of goods available to consumers in each market, the prices and other characteristics of the goods, their observed market shares, and consumer characteristics. This approach to estimating discrete choice demand using market-level data was developed by Berry, Levinsohn, and Pakes (1995) [1]. The model can be defined as follows:

- There are $j = 1, \cdots, J$ goods available in each market/periods $t = 1, \cdots, T$. Further, $j = 0$ represents the outside good. Hence, consumer i makes choice $j \in J$ in market t.

- M is the market size and given by the number of consumers in that market. Then the observed market share is the quantity of good j sold in market t divided by the market size, that is, $s_{jt} = q_{jt}/M$ and $s_{0t} = 1 - \sum_{j=1}^{J} s_{jt}/M$.

- x_{jt} is a vector of observable product characteristics and p_{jt} is the price of product j in market t.

- Finally, let β be the vector of coefficients on the observed characteristics, α be the price coefficient, ξ_{jt} be the product characteristics unobserved to the researcher and ϵ_{ijt} an i.i.d unobserved random shock.

The indirect utility function can then be written as,

$$\mathbf{u_{ijt}} = \beta_{\mathbf{i}}\mathbf{x_{jt}} + \alpha_{\mathbf{i}}\mathbf{p_{jt}} + \xi_{\mathbf{jt}} + \epsilon_{\mathbf{ijt}}$$

The above equation informs three things. First, the indirect utility can be derived from a quasilinear utility function, which is free from wealth effects. Second, the unobserved product characteristics are identical for all the consumers. Since the price coefficient is allowed to vary among consumers, this assumption is consistent with the theoretical literature on vertical product differentiation. Finally, the equation assumes that all consumers face the same product characteristics. In particular, they are offered the same price. If different consumers face different prices, using an average transaction price will lead to measurement error bias. This might also lead the prices to

be correlated to the error term which can be accounted for by introducing instrumental variables [18].

Nevo (2000) further defines individual characteristics as being composed of two components, the observed component or demographics, represented by D_i, and the unobserved component, represented by v_i. Since no individual data is observed, neither of the two components is directly observed. The distinction then between the components of the individual characteristics is that the researcher has a bit of knowledge on the demographics, D_i, like income, age, family size, and education, while no such information exists on additional characteristics, v_i, in the surveys. Formally, this can be modeled as

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \mathbf{\Pi D_i} + \mathbf{\Sigma v_i}, \qquad v_i \sim \mathbf{P}_v^*(v), \mathbf{D_i} \sim \hat{\mathbf{P}}_\mathbf{D}^*(\mathbf{D})$$

Where $P_v^*(v)$ is a parametric distribution and $\hat{P}_D^*(D)$ is a nonparametric distribution known from other data sources or a parametric distribution with parameters estimated elsewhere. The relation between the coefficients and the demographic variable, D_i, can be described as linear and affects the model's heterogeneity. By letting the parameters vary with D_i, the researcher includes more information about the distribution of the demographics in the analysis. Further, it reduces the alliance on parametric distribution.

To incorporate individual characteristics in the model, let $\theta = (\theta_1, \theta_2)$ be a vector containing all the parameters of the model. The vector $\theta_1 = (\alpha, \beta)$ contains the linear parameters and $\theta_2 = (\Pi, \Sigma)$ contains the nonlinear parameters. Including these vectors in the indirect utility model,

$$\mathbf{u_{ijt}} = \delta_\mathbf{jt}(\mathbf{x_{jt}}, \mathbf{p_{jt}}, \xi_\mathbf{jt}; \theta_1) + \mu_\mathbf{ijt}(\mathbf{x_{jt}}, \mathbf{p_{jt}}, v_\mathbf{i}, \mathbf{D_i}; \theta_2) + \epsilon_\mathbf{ijt},$$

$$\delta_\mathbf{jt} = \mathbf{x_{jt}}\beta - \alpha \mathbf{p_{jt}} + \xi_\mathbf{jt}, \qquad \mu_\mathbf{ijt} = [-\mathbf{p_{jt}}, \mathbf{x_{jt}}](\mathbf{\Pi D_i} + \mathbf{\Sigma v_i}),$$

Where $[-p_{jt}, x_{jt}]$ is a row vector, δ_{jt} represents the mean utility and is common to all the consumers, $\mu_{ijt} + \epsilon_{ijt}$ refer to the mean-zero heteroskedastic deviation from the mean utility and it captures the effects of the random coefficients.

Consumers are assumed to choose the option which offers the highest utility. Here, the consumers' utility depends on individual and product characteristics and their choice can be represented through the following choice set,

$$\mathbf{A_{jt}}(\mathbf{x_{.t}}, \mathbf{p_{.t}}, \delta_\mathbf{.t}; \theta_2) = \{\mathbf{D_i}, v_\mathbf{i}, \epsilon_\mathbf{i0t} \cdots \epsilon_\mathbf{iJt} | \mathbf{u_{ijt}} \geq \mathbf{u_{ilt}} \qquad \forall \mathbf{l} = \mathbf{0}, \mathbf{1}, \cdots, \mathbf{J}\}$$

Where $x_{.t} = (x_{lt}, \cdots, x_{Jt})'$, $p_{.t} = (p_{lt}, \cdots, p_{Jt})'$ and $\delta_{.t} = (\delta_{lt}, \cdots, \delta_{Jt})'$ are the observed characteristics, prices and mean utilities of all products respectively. Therefore, the choice set A_{jt} defines the consumer choice for product j in market t over all the other products since it provides the highest utility. The error term ϵ_{ijt} is assumed to be i.i.d and distributed according to a Type-1 extreme value (Gumbel) distribution and v_i is an i.i.d draw from the standard normal distribution. Further, it is also assumed that the consumer distinctly chooses a single product, that is, ties in the consumer choosing an alternative occurs with zero probability. Then the market share of the jth product is given by an integral over the mass of consumers in the choice set [18], [3].

$$\mathbf{s_{jt}}(\mathbf{x_{.t}}, \mathbf{p_{.t}}, \delta_\mathbf{.t}; \theta_2) = \int_{\mathbf{A_{jt}}} \mathbf{dP}^*(\mathbf{D}, v, \epsilon)$$

$$= \int_{\mathbf{A_{jt}}} \mathbf{dP}^*(\epsilon | \mathbf{D}, v), \mathbf{dP}^*(v | \mathbf{D}), \mathbf{dP}_\mathbf{D}^*(\mathbf{D})$$

$$= \int_{\mathbf{A_{jt}}} \mathbf{dP}_\epsilon^*(\epsilon) \mathbf{dP}_v^*(v) \mathbf{d\hat{P}}_\mathbf{D}^*(\mathbf{D})$$

Where $P^*()$ denotes the population distribution functions.

22

As the heterogeneity enters the model via D_i, v_i, and ϵ_{ijt}, the integral defining the market share has to be computed using Monte Carlo simulation. One of the closest approximate is

$$\mathbf{s_{jt}}(\mathbf{x_{.t}}, \mathbf{p_{.t}}, \delta_{.t}, \mathbf{P_{ns}}; \theta_2) = \frac{1}{\mathbf{ns}} \sum_{\mathbf{i=1}}^{\mathbf{ns}} \mathbf{s_{ijt}}$$

$$= \frac{1}{\mathbf{ns}} \sum_{\mathbf{i=1}}^{\mathbf{ns}} \frac{\exp\left[\delta_{\mathbf{jt}} + \sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{x_{jt}^k}\left(\sigma_{\mathbf{k}} v_{\mathbf{i}}^{\mathbf{k}} + \mathbf{\Pi_{k1}D_{i1}} + \cdots + \mathbf{\Pi_{kd}D_{id}}\right)\right]}{1 + \sum_{\mathbf{m=1}}^{} \mathbf{J} \exp\left[\delta_{\mathbf{mt}} + \sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{x_{mt}^k}\left(\sigma_{\mathbf{k}} v_{\mathbf{i}}^{\mathbf{k}} + \mathbf{\Pi_{k1}D_{i1}} + \cdots + \mathbf{\Pi_{kd}D_{id}}\right)\right]}$$

Where $i = 1, \cdots, ns$ are random draws and x_{jt}^k, $\quad k = 1, \cdots, K$ are the variables that have random slope coefficients.

5.1 BLP Estimator

The random coefficients model accounts for the heterogeneity of consumer choice and endogeneity of prices due to the use of market data through nonlinear parameters $\theta_2 = (\beta_i, \alpha_i)$ and instrumental variables respectively. As the identification of the model relies on instrumental variables the estimator is computed using moment conditions [3]. Berry, Levinsohn, and Pakes (1995) proposed a generalized method of moments (GMM) estimation approach to compute the BLP estimator. The steps for BLP estimation can be sketched as follows

1. Calculate the market shares conditional on δ_t and θ_2. This can be attained using the Monte Carlo simulation approach.

2. For each market t compute δ_t that equates the observed and predicted market shares, that is, $S_t = s_t(\delta_t; \theta_2)$. BLP proposes contraction mapping to compute the mean valuations δ_t. That is, given a guess of θ_2 and initial δ_t^0, iterate on

$$\delta_{\mathbf{jt}}^{\mathbf{h+1}}(\theta_2) = \delta_{\mathbf{jt}}^{\mathbf{h}}(\theta_2) + \log \mathbf{S_{jt}} - \log \mathbf{s_{jt}}(\delta_{\mathbf{jt}}^{\mathbf{h}}; \theta_2), \quad \mathbf{j = 1, \cdots, J}$$

The iteration stops when $\left\|\delta_{\mathbf{j}}^{\mathbf{h}} - \delta_{\mathbf{j}}^{\mathbf{h-1}}\right\| \le \epsilon_{\mathbf{in}}$.

3. Next step involves the computation of the GMM objective function conditional on δ_t and θ_2. The GMM criterion function based on the moment condition is defined as

$$\mathbb{E}\left[\xi(\theta)'\mathbf{Z}\right] = \mathbf{0}$$

The BLP estimator is obtained by minimizing the GMM objective function

$$\min_{\theta}\mathbf{Q}(\theta) = \xi(\theta)'\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}'\xi(\theta),$$

Where, Z is a vector of appropriate instrumental variables for price, $\mathbf{A} = \mathbb{E}\left[\mathbf{Z}'\xi(\theta)\xi(\theta)'\mathbf{Z}\right]$, and $\xi(\theta) = \delta(\theta_2) - (\bar{\beta}'\mathbf{x_j} - \bar{\alpha}\mathbf{p_j})$ is calculated using the estimates of δ and guessed θ.

4. As the final goal is to minimize the objective function $\mathbf{Q}$, repeat from step 1 until a minimum is found.

The BLP (Berry, Levinsohn, and Pakes) random coefficients model addresses consumer heterogeneity and price endogeneity, allowing for flexible substitution patterns without the constraints imposed by prior segmentation, as seen in logit and nested logit models. However, these advantages come with certain costs. The model lacks an analytic closed-form solution, making the computation of the integral challenging. The solution for the integral in the BLP estimator is typically obtained through simulation methods and Generalized Method of Moments (GMM) estimation.

Moreover, the inclusion of consumer heterogeneity necessitates additional information on the distribution of consumer heterogeneity to accurately compute market shares. This can be achieved by utilizing supplementary data sources or making parametric assumptions regarding the functional form of the distribution. These additional requirements highlight the complexity and data demands associated with the BLP model, despite its improved flexibility and accuracy in capturing consumer preferences and substitution patterns.

5.2 Logit Model under the BLP Framework

The calculation of the market shares can be simplified by easing the assumptions of the BLP model such that the utility function has a closed-form analytic solution. This would imply that the consumer heterogeneity enters the model only through separable additive random shocks, ϵ_{ijt} while maintaining the distribution assumptions specified above. Therefore, we obtain an aggregate homogenous logit model wherein, $\theta_2 = 0$, or $\beta_i = \beta$ and $\alpha_i = \alpha$. The utility function can be therefore written as

$$\mathbf{u_{ijt}} = \alpha \mathbf{p_{jt}} + \mathbf{x_{jt}}\beta + \xi_{\mathbf{jt}} + \epsilon_{\mathbf{ijt}}, \qquad \text{where,} \quad \mathbf{i} = 1, \cdots, \mathbf{I}, \quad \mathbf{j} = 1, \cdots, \mathbf{J}, \text{and} \quad \mathbf{t} = 1, \cdots, \mathbf{T}$$

And the market share of brand j in market t is given by

$$\mathbf{s_{jt}} = \frac{\exp(\mathbf{x_{jt}}\beta - \alpha\mathbf{p_{jt}} + \xi_{\mathbf{jt}})}{1 + \sum_{\mathbf{k}=1}^{\mathbf{J}} \exp(\mathbf{x_{kt}}\beta - \alpha\mathbf{p_{kt}} + \xi_{\mathbf{kt}})} = \frac{\exp(\delta_{\mathbf{jt}})}{1 + \sum_{\mathbf{k}=1}^{\mathbf{J}} \exp(\delta_{\mathbf{kt}})}, \qquad \text{where,} \quad \delta_{\mathbf{jt}} = \mathbf{x_{jt}}\beta - \alpha\mathbf{p_{jt}} + \xi_{\mathbf{jt}}$$

By normalizing the mean utility of the outside good, δ_{0t} to zero, its market share can be expressed as

$$\mathbf{s_{0t}} = \frac{1}{1 + \sum_{\mathbf{k}=1}^{\mathbf{J}} \exp(\delta_{\mathbf{kt}})}$$

Unlike the random coefficients model, the mean utility, δ_{jt} can be calculated by inverting the market shares. Note that, $\dfrac{\mathbf{s_{jt}}}{\mathbf{s_{0t}}} = \exp(\delta_{\mathbf{jt}})$. This implies that $\delta_{\mathbf{jt}} = \log(\mathbf{s_{jt}}) - \log(\mathbf{s_{0t}})$. Hence,

$$\log(\mathbf{s_{jt}}) - \log(\mathbf{s_{0t}}) = \delta_{\mathbf{jt}} = \mathbf{x_{jt}}\beta - \alpha\mathbf{p_{jt}} + \xi_{\mathbf{jt}}$$

The above equation represents a regression equation for demand and can be estimated straightforwardly [3]. To account for price endogeneity, a single market-level instrument can be included in the equation. In the absence of endogeneity, the demand equation could be estimated using Ordinary Least Squares (OLS). However, due to the potential for persistent price endogeneity, the inclusion of an instrument necessitates the use of the Two-Stage Least Squares (2SLS) method for estimation. Consequently, compared to the random coefficients model, demand estimation using the aggregate logit model is simpler and more straightforward.

However, there are a few problems with the model. These issues can be explained using the elasticities of the logit model.

$$\text{Own Elasticity :} \quad \eta_{\mathbf{j}} = \left(\frac{\partial \mathbf{s_j}}{\partial \mathbf{p_j}}\right) \frac{\mathbf{p_j}}{\mathbf{s_j}} = \alpha \mathbf{p_j}\left(1 - \mathbf{s_j}\right)$$

$$\text{Cross Elasticity :} \quad \eta_{\mathbf{jk}} = \left(\frac{\partial \mathbf{s_j}}{\partial \mathbf{p_k}}\right) \frac{\mathbf{p_k}}{\mathbf{s_j}} = -\alpha \mathbf{p_k}\mathbf{s_k}$$

In the own-price elasticity of demand, given that market shares are typically small, elasticity is directly proportional to the product's own price. Consequently, lower-priced brands exhibit lower elasticity, leading to higher markups, whereas high-priced goods tend to be more price elastic. This relationship, however, may not always hold true. The functional form of the price in the model underpins this relation. If the indirect utility were a function of the logarithm of price, the implied elasticity would align more consistently with theoretical expectations. Thus, the functional form significantly influences price elasticity patterns.

The cross-price elasticity in the logit model indicates that the cross-elasticity of product j with respect to the price of product k depends solely on the price and market share of product k. This implies that cross-elasticity is identical across products, irrespective of differences in product characteristics. This limitation exemplifies the failure of the IIA property in the multinomial logit model. Nevo (2000) illustrates the cross-elasticity issue through the cereal data. Consider a children's cereal product 'A' and a nutrition cereal product 'B' with similar market shares, then the cross-elasticity of the logit model suggests that the substitution from another children's cereal product 'C' toward either of them will be the same. Intuitively, the substitution between product 'A' and 'C' would be expected to be higher than between 'B' and 'C'. Hence, the logit model restricts the substitution towards brands in proportion to the market shares regardless of the characteristics.

5.3 Nested Logit Model under the BLP Framework

To incorporate the possibility of market segmentation, an aggregate nested logit model can be adopted. The two-level nested logit model is frequently used in research studies and is also employed in the subsequent sections of 'simulation study' and 'empirical application'. Therefore, the model specification discussed here is an aggregate two-level nested logit model as well.

Incorporating a nested logit model within the BLP (Berry, Levinsohn, and Pakes) framework offers several advantages, notably the transformation into a linear estimating equation for a cross-section of products, which simplifies estimation as demonstrated by Berry (1994) [2] and further extended by Björnerstedt and Verboven (2016) [4]. Björnerstedt and Verboven (2016) also address the issue of own-price elasticity within this framework.

The BLP model typically employs a utility function where price enters linearly or additively with income, implying that consumers purchase only one unit of their preferred good. While this assumption is suitable for certain goods like automobiles, it may not be as appropriate for frequently purchased commodities. To address this, Björnerstedt and Verboven propose an alternative utility specification known as the constant expenditures specification. This specification differs from the unit demand approach in three key ways: (1) prices enter logarithmically rather than linearly, (2) market shares are measured in terms of value rather than volume, and (3) the potential market is defined by the potential budget instead of the potential number of consumers. Therefore, the aggregate two-level nested logit model proposed by Björnerstedt and Verboven (2016) is illustrated here.

Consider the indirect utility function,

$$\mathbf{u_{ij}} = \mathbf{x_j}\beta + \alpha\mathbf{f}(\mathbf{y_i}, \mathbf{p_j}) + \xi_\mathbf{j} + \epsilon_\mathbf{ij}$$

Where, $i = 1, \cdots, I$ are the consumers, and each consumer chooses one out of $J + 1$ differentiated products, $j = 0, 1, \cdots, J$. Here good 0 represents the outside good or the no-purchase alternative. As before x_j is the vector of observed product characteristics of product j, ξ_j is a vector of product characteristics of product j unobserved by the researcher and ϵ_{ij} is the random utility term. $f(y_i, p_j)$ represents the function through which income and price enter the model non-additively and enable consumers to purchase multiple units of the product of their choice. For simplification purposes, the time subscript t has been dropped.

The next step in the estimation of the aggregate demand system would involve the individual consumers' decision regarding the number of units of the preferred product they buy. The demand for the preferred good j by consumer i can then be calculated using Roy's identity.

$$\mathbf{d_{ij}} = -\left(\frac{\partial \mathbf{f}(\mathbf{y_i}, \mathbf{p_j})}{\partial \mathbf{p_j}} \middle/ \frac{\partial \mathbf{f}(\mathbf{y_i}, \mathbf{p_j})}{\partial \mathbf{y_i}}\right)$$

The price function $f(y_i, p_j)$ for two utility specifications is expressed as follows

$$\text{Unit Demand}: \quad \mathbf{f(y_i, p_j) = y_i - p_j} \implies \mathbf{d_{ij} = 1}$$

$$\text{Constant Expenditures}: \quad \mathbf{f(y_i, p_j) = \gamma^{-1} \ln y_i - \ln p_j} \implies \mathbf{d_{ij} = \gamma \frac{y_i}{p_j}}$$

The individual buys a fixed unit of good j in case of unit demand and a constant fraction of her budget, γ, in case of constant expenditures specification.

Given the demand decision, the next step in the process is to identify which product the consumer chooses. This can be achieved through random utility maximization. Assuming that the consumer i chooses product j to maximize the utility level u_{ij}, the utility function can be rewritten as,

$$\mathbf{u_{ij} = K_i + \delta_j + \epsilon_{ij}}$$

where '$K_i = \alpha y_i$' in unit demand specification and '$K_i = \alpha \gamma^{-1} \ln y_i$' in constant expenditures specification. δ_j as before denotes the mean utility of product j.

While in the case of outside good, the mean utility is normalized to zero, $\delta_0 = 0$, the mean utility, δ_j, for the other utility specifications is represented as follows

$$\text{Unit Demand}: \quad \mathbf{\delta_j = x_j \beta + \alpha p_j + \xi_j}$$

$$\text{Constant Expenditures}: \quad \mathbf{\delta_j = x_j \beta + \alpha \ln p_j + \xi_j}$$

The model structure of the two-level nested logit model is the same as defined before. ϵ_{ij} follows the extreme type-1 value distributional assumption. The products are partitioned into G groups, $g = 0, 1, \cdots, G$, where $g = 0$ only consists of the outside good. Each group is thereby further partitioned into H_g subgroups, $h = 1, \cdots, H_g$. Each subgroup h of group g contains J_{hg} products so that $\sum_{g=1}^{G} \sum_{h=1}^{H} J_{hg} = J$.

Based on the two-level nested logit model structure and the assumption of random utility maximization, the probability that the consumer i chooses product j is given by

$$\mathbf{s_j = s(\delta, \sigma)} \equiv \frac{\exp((\delta_j)/(1 - \sigma_1))}{\exp(I_{hg}/(1 - \sigma_1))} \frac{\exp(I_{hg}/(1 - \sigma_2))}{\exp(I_g/(1 - \sigma_2))} \frac{\exp(I_g)}{\exp(I)},$$

where I_{hg}, I_g, and I are the 'inclusive values'. The inclusive values are expressed as,

$$I_{hg} \equiv \mathbf{(1 - \sigma_1)} \ln \sum_{k=1}^{J_{hg}} \exp((\delta_k)/(1 - \sigma_1))$$

$$I_g \equiv \mathbf{(1 - \sigma_2)} \ln \sum_{h=1}^{H_g} \exp(I_{hg}/(1 - \sigma_2))$$

$$I \equiv \ln \left(\mathbf{1} + \sum_{g=1}^{G} \exp(I_g) \right),$$

In the choice probability, $s(\delta, \sigma)$, δ is a $J \times 1$ vector containing the mean utilities δ_j, and $\sigma = (\sigma_1, \sigma_2)$ are the nesting parameters of the nested logit model. They measure the correlation across the products of the same subgroup (σ_1) and group (σ_2).

The aggregate demand for product j can now be derived as the probability that a consumer buys product j, 's_j', times the quantity purchased, 'd_{ij}', summed over all the consumers. Under the two utility specifications the aggregate demand for $j = 1, \cdots, J$ is as follows,

$$\text{Unit Demand}: \quad \mathbf{q_j} = \sum_{i=1}^{I} \mathbf{s_j d_{ij}} = \mathbf{s_j I}$$

$$\text{Constant Expenditures}: \quad \mathbf{q_j} = \sum_{i=1}^{I} \mathbf{s_j d_{ij}} = \mathbf{s_j \frac{B}{p_j}}$$

where, $B = \gamma Y$ and $Y = \sum_{i=1}^{I} y_i$. Hence, B is the total budget allocated to the products and is calculated as a fraction γ of the total income of all the consumers.

The final goal is to estimate the parameters (α, β, σ) that enter the aggregate demand. A linear estimable equation with ξ_j as the error term is required to calculate the parameters, which can be derived in two steps. Firstly, following Berry (1994), the system choice probabilities are inverted to solve for the mean utilities δ_j.

The inverted choice probability system for a two-level aggregate nested logit model is given by,

$$\delta_\mathbf{j} = \ln(\mathbf{s_j/s_0}) - \sigma_\mathbf{1} \ln(\mathbf{s_{j|hg}}) - \sigma_\mathbf{2} \ln(\mathbf{s_{h|g}}),$$

where s_j is the market share of product j in the market, $s_{j|hg}$ is the market share product j in its subgroup h of group g given that an alternative is chosen and $s_{h|g}$ is the market share of subgroup h in group g given that group g is chosen.

Substituting the inverted choice probability system into the expression for the mean utilities, δ_j to get the estimating equations.

$$\text{Unit Demand}: \quad \ln(\mathbf{s_j/s_0}) - \sigma_\mathbf{1} \ln(\mathbf{s_{j|hg}}) - \sigma_\mathbf{2} \ln(\mathbf{s_{h|g}}) = \mathbf{x_j}\beta + \alpha \mathbf{p_j} + \xi_\mathbf{j}$$

$$\ln(\mathbf{s_j/s_0}) = \mathbf{x_j}\beta + \alpha \mathbf{p_j} + \sigma_\mathbf{1} \ln(\mathbf{s_{j|hg}}) + \sigma_\mathbf{2} \ln(\mathbf{s_{h|g}}) + \xi_\mathbf{j}$$

$$\text{Constant Expenditures}: \quad \ln(\mathbf{s_j/s_0}) - \sigma_\mathbf{1} \ln(\mathbf{s_{j|hg}}) - \sigma_\mathbf{2} \ln(\mathbf{s_{h|g}}) = \mathbf{x_j}\beta + \alpha \ln \mathbf{p_j} + \xi_\mathbf{j}$$

$$\ln(\mathbf{s_j/s_0}) = \mathbf{x_j}\beta + \alpha \ln \mathbf{p_j} + \sigma_\mathbf{1} \ln(\mathbf{s_{j|hg}}) + \sigma_\mathbf{2} \ln(\mathbf{s_{h|g}}) + \xi_\mathbf{j}$$

The next step is to use the aggregate demand expressions to write the unobserved choice probabilities as observables. In the case of the unit demand specification, the choice probabilities are equal to the market shares in volume terms.

$$\mathbf{s_j} = \frac{\mathbf{q_j}}{\mathbf{I}}, \quad \mathbf{s_{j|hg}} = \frac{\mathbf{q_j}}{\sum_{\mathbf{j} \in \mathbf{H_{hg}}} \mathbf{q_j}}, \quad \mathbf{s_{h|g}} = \frac{\sum_{\mathbf{j} \in \mathbf{H_{hg}}} \mathbf{q_j}}{\sum_{\mathbf{h=1}}^{\mathbf{H_{hg}}} \sum_{\mathbf{j} \in \mathbf{H_{hg}}} \mathbf{q_j}}, \quad \mathbf{s_0} = \mathbf{1} - \frac{\sum_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{q_j}}{\mathbf{I}}$$

and in the constant expenditures specification, they are equal to the market shares in terms of value.

$$\mathbf{s_j} = \frac{\mathbf{p_j q_j}}{\mathbf{B}}, \quad \mathbf{s_{j|hg}} = \frac{\mathbf{p_j q_j}}{\sum_{\mathbf{j} \in \mathbf{H_{hg}}} \mathbf{p_j q_j}}, \quad \mathbf{s_{h|g}} = \frac{\sum_{\mathbf{j} \in \mathbf{H_{hg}}} \mathbf{p_j q_j}}{\sum_{\mathbf{h=1}}^{\mathbf{H_{hg}}} \sum_{\mathbf{j} \in \mathbf{H_{hg}}} \mathbf{p_j q_j}}, \quad \mathbf{s_0} = \mathbf{1} - \frac{\sum_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{p_j q_j}}{\mathbf{B}}$$

The estimating equation for both of the utility specifications can be estimated using instrumental variables regression of market shares on the product characteristics, prices, subgroup market shares, and group market shares. Note that based on McFadden (1972), the model is consistent with random utility maximization if $\alpha < 0$ and $1 \geq \sigma_1 \geq \sigma_2 \geq 0$. A negative price coefficient implies the expectation that higher prices reduce the utility and thus the probability of choosing an alternative. When σ_1 is high, the consumer preferences are strongly correlated among the products of the same subgroup, and when σ_2 is high, consumer preferences are also highly correlated across the products of the same group. If $\sigma_1 = \sigma_2$, the model reduces to a one-level nested logit model wherein groups are the nests. If $\sigma_2 = 0$, the model reduces to a one-level nested logit model with the subgroups as the nests. Finally, if $\sigma_1 = \sigma_2 = 0$, then the model reduces to a simple logit model without nests.

Björnerstedt and Verboven (2016) [4] also report the formula to calculate the own-price elasticity and cross-price elasticity for a two-level nested logit model. The cross-price elasticity can be further segmented into three categories: (1) when both the products j and k are in the same subgroup, (2) when products j and k are in different subgroups of the same group, and (3) when products j and k are in different groups. The expression in the case of unit demand specification is,

$$\frac{\partial \mathbf{q_j}}{\partial \mathbf{p_k}} \frac{\mathbf{p_k}}{\mathbf{q_j}} = \alpha \left(\frac{1}{1-\sigma_1} \mathbf{D^1_{jk}} - \left(\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2} \right) \mathbf{s_{j|hg}} \mathbf{D^2_{jk}} - \frac{\sigma_2}{1-\sigma_2} \mathbf{s_{j|g}} \mathbf{D^3_{h|g}} - \mathbf{s_j} \right) \mathbf{p_j}$$

Here, $s_{j|g}$ defines the market share of product j in group g. $D^1_{jk} = 1$ if $j = k$ else 0, $D^2_{jk} = 1$ if j and k are in the same subgroup otherwise 0 and $D^3_{jk} = 1$ if j and k are in the same group and otherwise 0.

This implies that the own-price elasticity for the two-level nested logit model for the unit demand specification is given by,

$$\frac{\partial \mathbf{q_j}}{\partial \mathbf{p_j}} \frac{\mathbf{p_j}}{\mathbf{q_j}} = \alpha \left(\frac{1}{1-\sigma_1} - \left(\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2} \right) \mathbf{s_{j|hg}} - \frac{\sigma_2}{1-\sigma_2} \mathbf{s_{j|g}} - \mathbf{s_j} \right) \mathbf{p_j}$$

The cross-elasticity for the case where products j and k are in the same subgroup can also be calculated in a similar manner,

$$\frac{\partial \mathbf{q_j}}{\partial \mathbf{p_k}} \frac{\mathbf{p_k}}{\mathbf{q_j}} = \alpha \left(- \left(\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2} \right) \mathbf{s_{j|hg}} \mathbf{D^2_{jk}} - \frac{\sigma_2}{1-\sigma_2} \mathbf{s_{j|g}} \mathbf{D^3_{h|g}} - \mathbf{s_j} \right) \mathbf{p_j}$$

Then the cross-elasticity for the case, when j and k are in the same groups, is,

$$\frac{\partial \mathbf{q_j}}{\partial \mathbf{p_k}} \frac{\mathbf{p_k}}{\mathbf{q_j}} = \alpha \left(- \frac{\sigma_2}{1-\sigma_2} \mathbf{s_{j|g}} \mathbf{D^3_{h|g}} - \mathbf{s_j} \right) \mathbf{p_j}$$

Finally, the cross-elasticity for the case, when j and k are in different groups, is

$$\frac{\partial \mathbf{q_j}}{\partial \mathbf{p_k}} \frac{\mathbf{p_k}}{\mathbf{q_j}} = \alpha \left(-\mathbf{s_j} \right) \mathbf{p_j}$$

The elasticities for constant expenditures specification can also be calculated in a similar manner. The own-price elasticity in the nested logit model is still driven by the functional-form assumption, necessitating the use of theoretically sound functional forms to ensure accurate elasticity estimates. Due to the error term, ϵ_{ijt}, being decomposed into an i.i.d. random shock and a group-specific component, it would be expected that the products within the same subgroup are more highly correlated than those in different subgroups and groups. This could possibly lead to a comparatively higher cross-price elasticity for pairs of alternatives within the same groups than those in different groups.

5.4 Instrumental Variables

Due to the endogeneity issues in the BLP model, the parameter estimates can be biased and inconsistent. A method to address the issue is by adding exogenous instrumental variables in the model. However, there is no distinct set of variables that can be used as an instrument for all data sets belonging to different industries. Rather, the literature offers a few possible solutions that can provide exogenous variation in prices and other endogenous variables [3].

- Cost Shifters and Their Proxies - A standard approach to search for demand-side instrumental variables is to look for variables that shift cost and are uncorrelated with the demand shock, ξ_t. These can for instance be exogenous marginal costs, a tax, or a tariff. Proxies of a producer's actual cost shifters can also be used as an instrument. For instance, the average wage in a labor market can be a proxy for the producers' labor cost and be included as an instrument in the demand model as long as it is uncorrelated with the demand shocks conditional on the exogenous variables and consumer-specific measures (like income, education) which are also included in the model. However, the rarity of detailed data on cost shifters which vary by brand makes it challenging to use them as instruments.

- Hausmann Instrument - Under this, the price for the same good in another geographic market is used as an instrument for the price in the model. It assumes that although the producer-specif cost shifters are not observable, variation in costs exists and is likely the reason for variation in prices of the product in the markets a producer serves in. Hence, a price increase in market t' can signal a change in the producer's cost that leads to a shift in the price in market t. Since price change also depends on demand shocks, the instrument can fail if the demand shocks of the two markets are correlated. Therefore, to use a proxy for an exogenous change in a firm's cost as an instrument requires scrutiny of the exogeneity of the instrument.

- BLP Instruments - One of the most commonly used instruments is the exogenous characteristics of competing products. Although the demand for good j depends on the characteristics of all related goods, as long as the assumption of mean independence between x_t and ξ_t is satisfied, the entire set of competitors product characteristics, x_{-jt}, can be used as instruments creating exogenous variation in good j's market share. Since these characteristics also affect the product j's price through the demand system, they can also provide exogenous variation in prices.

6 Simulation Study

The objective of the simulation study is to analyze the performance of the two-level nested logit model and the logit model under the BLP framework. Based on the models developed in Section 5, a Monte Carlo experiment is conducted for the two discrete choice models. Under each experiment, 1000 datasets are generated and each dataset consists of $T = 25$ independent markets and $J = 100$ products per market. Further, in each market for product j there exists a vector of observed product characteristics, x_j, and the corresponding price, p_j. The price is drawn from the lognormal distribution as it ensures positive realizations and roughly mimics the distribution of prices in empirical datasets [9]. The product characteristic is drawn from a random normal distribution with a mean of 10 and a standard deviation of 3. Finally, ξ_{jt} represents unobserved product characteristic which is drawn from the standard normal distribution.

To include the issue of endogeneity in the model, the price variable is assumed to be correlated with x_{jt} and ξ_{jt}. Instrumental variables are incorporated in the model to address the endogeneity. Based on the theory, the product characteristic x_{jt} is used as one of the instruments. Two additional instruments are developed by adding a random disturbance to the observed and unobserved product characteristics.

The mean utility δ can then be represented as

$$\delta_{\mathbf{j}} = \beta_1 + \beta_2 \mathbf{x_{jt}} + \alpha \mathbf{p_{jt}} + \xi_{\mathbf{jt}}$$

Where, β_1 is the constant, β_2 is the coefficient of product characteristic and α is the price parameter. For the simulation, the mean valuations are assumed to be $(-1, -0.5, -2)$ for β_1, β_2 and α respectively.

The market share for product j is represented by s_{jt} and calculated as follows

$$\mathbf{s_{jt}} = \frac{\exp(\delta_{\mathbf{jt}})}{\mathbf{1} + \sum_{\mathbf{k=1}}^{\mathbf{J}} \exp(\delta_{\mathbf{kt}})}$$

The market share of the products can be used to calculate the mean utility, δ_{jt} and thereby estimate the demand through the following linear regression equation

$$\text{Multinomial Logit Model}: \quad \ln(\mathbf{s_{jt}}/\mathbf{s_{0t}}) = \beta_1 + \beta_2 \mathbf{x_{jt}} + \alpha \mathbf{p_{jt}} + \xi_{\mathbf{jt}}$$

$$\text{Nested Logit Model}: \quad \ln(\mathbf{s_{jt}}/\mathbf{s_{0t}}) = \beta_1 + \beta_2 \mathbf{x_{jt}} + \alpha \mathbf{p_{jt}} + \sigma_1 \ln(\mathbf{s_{j|hg}}) + \sigma_2 \ln(\mathbf{s_{h|g}}) + \xi_{\mathbf{jt}}$$

Where, s_{0t} is the market share of outside good, $s_{j|hg}$ is the market share of good j in subgroup h of group g and $s_{h|g}$ is the market share of subgroup h in group g. Further, σ_1 and σ_2 represent the nesting parameters.

This estimation process is based on Berry (1994) and Björnerstedt and Verboven (2016). Under the nested logit, the model estimation is based on the unit demand utility specification with the parameter estimate $\alpha, \sigma_1, \sigma_2$ expected to satisfy the following restrictions from economic theory: $\alpha < 0$ and $1 \geq \sigma_1 \geq \sigma_2 \geq 0$. In the model, for each market, there are two groups 0 and 1 and each group has three further subgroups 2,3, and 4 among which the products are randomly divided. The decision tree for the nested logit model would be a two-level tree with the upper level consisting of nests '$Group\_0$' and '$Group\_1$' and the lower level consisting of the alternatives '$Subgroup\_2$', '$Subgroup\_3$' and '$Subgroup\_4$' for each of the nests. This structure is similar to the tree displayed in Figure 1.

The first-stage regression assesses the correlation between instruments and endogenous regressors, ensuring the instruments are correlated with the regressors but not with the error term. A highly significant average F-statistic of 345.576 indicates strong evidence in favor of relevant instruments. Table 2 reports the estimation results for the two models based on the data-generating process. Considering the parameter estimates first, the price parameter α is -1.002 in the case of the multinomial logit model which is close to the true value of -2. Meanwhile, the

estimate in the case of the two-level nested logit model is −0.295. The estimates of the nesting parameters are $\sigma_1 = 0.724$ and $\sigma_2 = 0.551$. A high value for the nesting parameters imply a strong degree of correlation among the product alternatives within each nest. Along with $\alpha = -0.295 < 0$, the nesting parameters also satisfy the restrictions laid down for the nested logit model to be consistent with the random utility maximization principles. Values within this range ensure that the model correctly captures the correlation structure without violating the theoretical foundations of utility maximization.

The reason for the price parameter α to differ significantly from the true value is because of the significance of the nesting parameters as they rescale all utility parameters [9]. These nesting parameters capture different degrees of correlation among choices at each nesting level. This can lead to the need to adjust the estimated utility parameters for the combined effect of both nesting parameters to align them with their true values. Indeed, the rescaled price parameter $\alpha/(1 - \sigma_1)(1 - \sigma_2) = -2.38$ is very close to the true value of −2.

The own and cross-price elasticity are one of the ways to interpret the demand estimates. The own-price elasticity on average is approximately the same for the multinomial logit model, −6.117, and the nested logit model, −6.362. Since the functional form of price determines the price elasticity patterns, the similarity in the own-price elasticity of the two models can be linked to the fact that the price enters linearly in the models. The cross-price elasticity on the other hand helps to identify the substitution patterns. The average cross-price elasticity for the MNL is approximately 0.00003, which is the same across all the product alternatives. This uniformity arises from the restrictive assumption of Independence of Irrelevant Alternatives (IIA) and implies that a 1% increase in the price of one product would result in only a negligible increase in the demand for another product. Hence, the model does not adequately capture the substitution patterns.

The average cross-price elasticity of the NL model validates the same. As the IIA assumption is relaxed and the products are categorized into groups and subgroups based on their characteristics, this leads to varying cross-price elasticity depending on the relative similarity between the products. When both the products are in the same subgroup, the cross-price elasticity on average is around 0.092, a strong substitution effect. There exists a moderate substitution effect of 0.0217 when the products are in the same group but not as closely related, that is, in different subgroups. At last, if the products are in different groups, the cross-price elasticity indicates a very weak substitution effect. Therefore, the NL model provides a more nuanced understanding of substitution patterns compared to the MNL model.

Model-selection criteria, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the Likelihood Ratio test, have been employed to compare the performance of the MNL and NL models. The AIC and BIC are statistical measures used to assess the relative quality of models, with lower values indicating a better balance between goodness of fit and model complexity. For the MNL model, the AIC and BIC values are 5320.905 and 5338.378, respectively. In contrast, the NL model exhibits significantly lower AIC and BIC values, at 4557.034 and 4586.154, respectively, suggesting that it captures the data more efficiently while maintaining model simplicity. Additionally, the Likelihood Ratio test, which evaluates whether the more complex model provides a significantly better fit than the simpler model, strongly rejects the null hypothesis in favor of the NL model. This test provides robust statistical evidence that the NL model better fits the data compared to the MNL model, despite its additional complexity. Thus, the NL model demonstrates an improved performance in capturing the underlying structure of the data.

In summary, the analysis reveals that the Nested Logit (NL) model under the BLP framework provides a more nuanced and accurate depiction of consumer choice compared to the Multinomial Logit (MNL) model. The NL model's capacity to capture complex substitution patterns by relaxing the Independence of Irrelevant Alternatives (IIA) assumption enables it to better reflect the correlation between alternatives within the same nest, leading to more precise parameter estimates and elasticities. This is evident from the significantly lower AIC and BIC values for the NL model, which suggest a more favorable trade-off between goodness of fit and model complexity. Addi-

tionally, the Likelihood Ratio test strongly supports the NL model's superiority, indicating that it fits the data more effectively than the MNL model. Thus, the NL model offers a more comprehensive and accurate framework for analyzing market dynamics and consumer behavior within the BLP framework.

Table 2: Simulation Study Results

Parameters	Multinomial Logit (MNL)	Nested Logit (NL)
β_2	−0.7994	−0.2379
	(0.0005)	(0.0013)
α	−1.0023	−0.2948
	(0.0015)	(0.0023)
σ_1	*n/a*	0.724
		(0.0017)
σ_2	*n/a*	0.5509
		(0.0024)
R^2	0.8817	0.9447
Own and Cross-Price Elasticity		
Own-Price Elasticity	−6.1167	−6.3622
Cross: same subgroup	0.00003	0.0927
Cross: different subgroup	0.00003	0.0217
Cross: different group	0.00003	0.000009
Model Selection Criteria		
AIC	5320.906	4557.034
BIC	5338.378	4586.154
Likelihood Ratio Test		
Test Statistic	767.872	
P-Value	0.000	

Note: The table reports empirical means and standard deviations (in parenthesis) for selected parameters, the implied price elasticities, and two model selection criteria, AIC and BIC. The simulation study is based on 1000 random samples of 25 independent markets and 100 products in each market.

7 Empirical Application

This section focuses on applying the Nested Logit model to a real aggregate dataset, emphasizing its utility in post-estimation analysis. The Nested Logit model is instrumental in analyzing market dynamics such as price changes and competitive interactions, particularly in response to market shocks. For instance, the model can be employed to evaluate the potential effects of hypothetical firm mergers on market structure. Additionally, it can incorporate regulatory interventions or remedies implemented by government or competition authorities to mitigate anti-competitive practices. Thus, the Nested Logit model offers valuable insights into real-world market dynamics and policy decisions.

The dataset on the European car market collected by Björnerstedt and Verboven (2014) [5] is used for the study. The dataset contains firm and brand-wise information on car sales between 1970 and 1999 in five countries (markets): Belgium, France, Germany, Italy, and the United Kingdom. There are 11,483 observations in total and for each observation, the horsepower, width, height, and fuel efficiency are a few of the product characteristics. The price is measured in '1,000 Euro' based on the 1999 purchasing power. The panel dataset is also unbalanced since not all car brands are available throughout the period or across all the countries. A summary of the dataset is given hereby.

Table 3: Summary Statistics

	Sum	Mean	SD	Min	Max	N
country	33,513	3	1	1	5	11,483
segment	29,386	3	1	1	5	11,483
domestic	2,166	0	0	0	1	11,483
firm	166477	14	9	1	34	11,483
qu	2.3e+08	19,911	37,804	51	433694	11,483
price	212399	18	9	5	150	11,483
horsepower	657562	57	24	13	170	11,483
fuel	77,268	7	2	4	19	11,483
width	1888465	164	10	122	188	11,483
height	1612711	140	5	118	174	11,483
pop	5.5e+11	48057602	21815495	9660000	82060000	11,483
ngdp	2.0e+18	1.8e+14	4.7e+14	5.2e+10	2.1e+15	11,483

The empirical application is conducted through the help of the Stata package 'Mergesim' developed by Björnerstedt and Verboven (2014). The package is advantageous in the post-estimation of the demand parameters, that is, the price and nesting parameters. The package provides quick and direct commands to compute the elasticity and market conditions like price and marginal costs. Given the aggregate data, it is imperative to assess the impact of market competitiveness on consumer choice. Market shocks like firm mergers are commonly visible and affect product prices and market shares. The 'Mergesim' package provides the scope to simulate mergers between firms and quantify the effect of the merger on factors like marginal costs and consumer surplus.

The nested logit model is defined by a two-level decision tree with the upper level (segment) and the lower level (domestic). The five car segments: subcompact, compact, intermediate, standard, and luxury are used as the nests. Each nest holds two subgroups according to the origin of the car: Foreign and Domestic. For instance, the cars manufactured by BMW in Germany will be domestic in the German market and foreign in other countries markets. Figure 2 depicts the decision tree for the dataset.
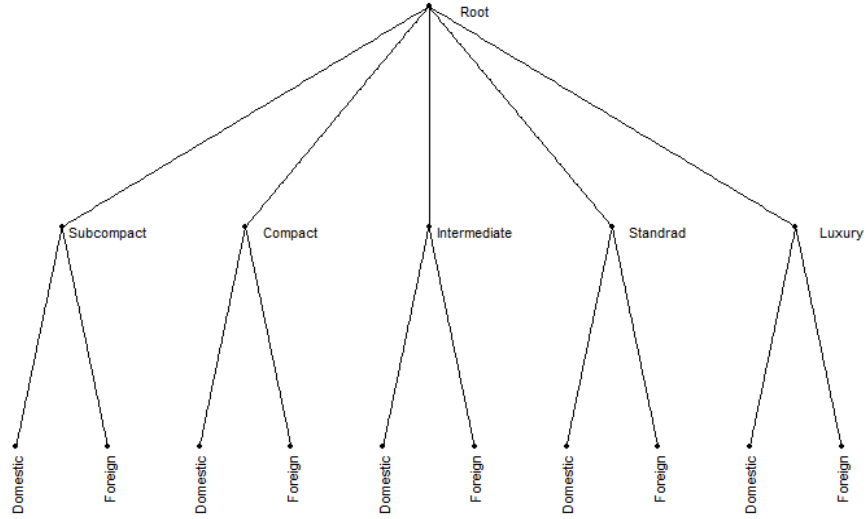
Figure 2: Nested Logit Structure for European Car Sales: Two-level Decision Tree

The utility model is assumed to follow the unit demand specification which implies the prices enter the model linearly. The market share of a product j, the market share of the product in its subgroup h of group g, and the market share of the subgroup h in group g are calculated based on the method as described by Björnerstedt and Verboven (2016) in section 5.3.

$$s_j = \frac{q_j}{I}, \quad s_{j|hg} = \frac{q_j}{\sum_{j \in H_{hg}} q_j}, \quad s_{h|g} = \frac{\sum_{j \in H_{hg}} q_j}{\sum_{h=1}^{H_{hg}} \sum_{j \in H_{hg}} q_j}, \quad s_0 = 1 - \frac{\sum_{j=1}^{J} q_j}{I}$$

While q_j the quantity of each type of car sold is available, the potential market size I is defined through a proxy for the number of households, namely, the total population divided by 4. These values are thereby used to calculate the market shares of the products.

The nested logit model can now be estimated using linear regression with instrumental variables to account for the endogeneity of the price variable. Firstly, the vector product characteristics, x_{jt} are used as instruments. The other instrument used in the estimation process can be derived from the exogenous characteristics of competing products which act as a proxy for the intensity of competition firms face. This is the sum of the product characteristics of products sharing the same cluster (that is market segment and foreign/domestic status) [8]. Therefore, the estimation model is as follows

$$M\_ls = \beta_1 \alpha price + \sigma_1 M\_lsjh + \sigma_2 M\_lshg + \beta_2 horsepower + \beta_3 fuel + \beta_4 width + \beta_5 height + \beta_7 log\_pop$$

$$x_{jt} = (horsepower, fuel, width, height)$$

$$instr = horsepower + fuel + width + height$$

M_ls is the dependent variable $ln(s_j/s_0$, M_lsjh is the log of the product's market share in the selected subgroup $ln(s_{j|hg})$, and M_lshg is the log of the market share of the selected subgroup $ln(s_{h|g})$. $horsepower$, $fuel$, $width$, $height$ and $weight$ are the product characteristics x_{jt} and log_pop is the logarithmic expression of the population variable. z_{jt} represents the instruments included in the estimation to account for the price endogeneity and $instr$ is the sum of the product characteristics.

The parameters of interest are the price parameter, α and the nesting parameters, σ_1 and σ_2. The estimated values of the parameters are $\alpha = -0.038$, $\sigma_1 = 0.907$ and $\sigma_2 = 0.739$. These estimates satisfy the conditions of random utility maximization and indicate a high degree of correlation among the product alternatives within each nest. The first-stage test to check the validity of the instruments also gives significant results, an F-statistic of 130.023.

```
Instrumental variables 2SLS regression          Number of obs   =      11,483
                                                 Wald chi2(8)    =    76307.09
                                                 Prob > chi2     =      0.0000
                                                 R-squared       =      0.8836
                                                 Root MSE        =      .51207


                        Robust
       M_ls │ Coefficient  std. err.      z     P>|z|    [95% conf. interval]

      price │  -.0381801   .0082896    -4.61   0.000    -.0544275   -.0219327
     M_lsjh │   .9070285   .0056866   159.50   0.000     .895883     .918174
     M_lshg │   .7396119   .0142599    51.87   0.000     .711663    .7675607
 horsepower │  -.0094749   .0021789    -4.35   0.000    -.0137455   -.0052043
       fuel │  -.0690276   .0104977    -6.58   0.000    -.0896026   -.0484526
      width │   .0016932   .0008586     1.97   0.049     .0000104    .0033761
     height │  -.0063827   .0010622    -6.01   0.000    -.0084645   -.0043009
    log_pop │  -.1626125   .0138978   -11.70   0.000    -.1898517   -.1353733
      _cons │   1.125736   .3434154     3.28   0.001     .4526539    1.798818

Instrumented: price
 Instruments: M_lsjh M_lshg horsepower fuel width height log_pop instr
```

Figure 3: Parameter Estimates for Nested Logit Model

Before simulating a merger between the firms, it is necessary to evaluate the premerger market conditions. For this discussion, it is assumed that the firms' equilibrium behavior is defined by multiproduct Bertrand competition. The elasticity and firms' prices and marginal costs in the year 1990 for France have been discussed hereby. The cross-price elasticity values reveal distinct patterns in consumer preferences: products within the same subgroup exhibit a high cross-price elasticity of 0.643, indicating that consumers view them as close substitutes and are highly likely to switch between them in response to price changes. In contrast, the elasticity between products from different subgroups is moderate at 0.129, showing some degree of substitutability but less pronounced. Products from entirely different groups have an extremely low cross-price elasticity of 0.001, reflecting minimal substitution and suggesting that consumers perceive these products as distinct and unrelated. Overall, these elasticity patterns highlight that consumer preferences are strongly aligned with the location where the cars are manufactured and the segment they belong to. The cross-price elasticities highlight The Lerner index denotes the percentage markup over marginal cost. It varies between 12% and 35% with the higher markup being associated with firms with lower-priced car models. The premerger market details for France for the year 1990 can be found below in Table 4 and Fig 8 in the Appendix.

Table 4: France: Own- and Cross-Price Elasticities: Unweighted Market Averages

Variable	Mean	SD	Min	Max
Own-Price Elasticity	−6.266	2.989	−18.461	−2.660
Cross: same subgroup	0.643	0.845	0.004	3.094
Cross: different subgroup	0.129	0.215	0.001	0.902
Cross: different group	0.001	0.002	0.000	0.006
Observations : 78				

Finally, a horizontal merger between two domestic firms in France is simulated to comprehend the effect of the anti-competitive practice on market conditions for the year 1990. In France, it is assumed that a merger will take place between PSA and Renault, where PSA will sell the brands Peugeot and Citröen. Further, the merger participants have been assumed to not have any marginal cost savings. Horizontal mergers are often scrutinized due to their tendency to accumulate market power and influence product prices. The results show the average pre and post-merger prices (Figure 8 in the Appendix) and the percentage price change for the firms operating in France. The merger predicts that on average the prices of PSA will increase by 43.1% and of Renault by 52.3%. Figure 4 below visualizes the percentage change in prices.
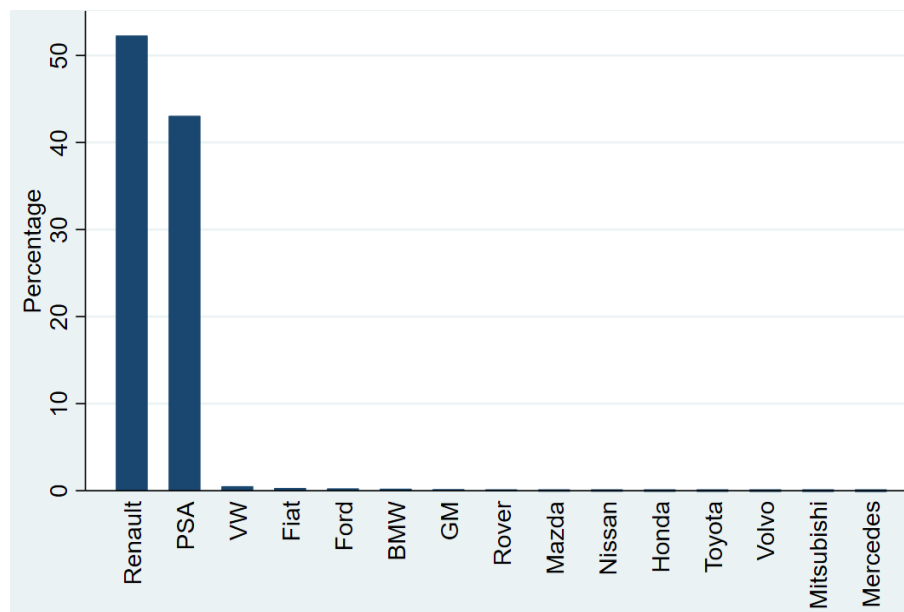

Figure 4: Average Percentage Price Increase Per Firm in France

The post-merger results also convey information on the market share of the 4 largest firms (C4) which increases to 80.65%. Herfindahl-Hirschman Index (HHI) is used to assess the level of market concentration before and after the merger. The index increases from 2209 to 2346. Since the simulation is based on a market with differentiated products, the index reports a concentrated market even without the merger. Post-merger, the market is seen to become more uncompetitive. This result is also corroborated by the change in market welfare. The consumer surplus falls by about 6.57 billion Euro which is partly compensated by the rise in producer surplus of around 3.69 billion Euro (Figure 9 in the Appendix).

The competition authorities often undertake remedy measures like 'Divestiture' to mitigate the price effects of a merger and maintain competitiveness in the market. Under divestiture, the governing body allows the merger under the condition that the firm sells some of its products or brands. In the above example for France, the merger between PSA and Renault led to a significant increase in prices which consequently caused a huge consumer loss. Divestiture in this scenario would imply that the merger between the two firms would occur if PSA considers selling one of its brands to another firm. It is assumed that PSA sells its brand of Peugeot to the firm Ford. The merger results after incorporating divestiture differ from the previous results without any anticompetitive remedies as follows. In the new merger, the price effects are more dampened with no significant increase. Divestiture raises the average price by only 5% for Renault and 2% for Ford, which now includes Peugeot as well. On the other hand, the average price of cars sold by PSA falls by 4%. The impact of these price effects is also visible on market welfare, where the consumer surplus increases by $115,774$ Euro and the producer surplus falls by $24,089$ Euro.

The nested logit model within the BLP framework plays a pivotal role in analyzing consumer choice and market dynamics. By leveraging the model's parameter estimates, researchers can calculate cross-elasticities, which

provide insights into the correlation between products and the degree of substitution among them. These estimates are instrumental in assessing how market competition influences consumer welfare. For instance, the example provided demonstrates how the model can simulate firm mergers—an important market shock—and evaluate its effects on market competition through market share calculations and on overall welfare by assessing changes in consumer and producer surplus. This comprehensive approach highlights the model's relevance for both academic research and practical market analysis.
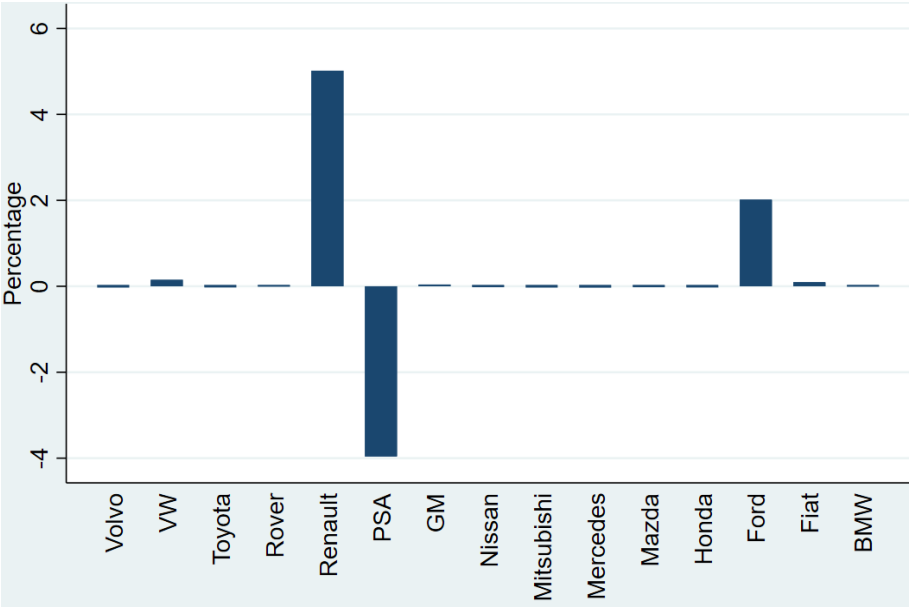

Figure 5: Average Percentage Price Increase Per Firm in France post Divestiture

firm code	Pre-merger	Post-merger	Difference
BMW	0.012	0.012	-0.000
Fiat	0.074	0.074	0.001
Ford	0.072	0.065	-0.008
Honda	0.006	0.006	0.000
Mazda	0.008	0.008	0.000
Mercedes	0.012	0.012	-0.000
Mitsubishi	0.001	0.001	0.000
Nissan	0.007	0.007	0.000
GM	0.052	0.053	0.000
PSA	0.347	0.415	0.068
Renault	0.273	0.211	-0.062
Rover	0.018	0.018	0.000
Toyota	0.005	0.005	0.000
VW	0.107	0.107	0.001
Volvo	0.005	0.005	0.000

	Pre-merger	Post-merger
HHS:	2209	2593
C4:	80.10	87.25
C8:	95.56	96.77

	Change
Consumer surplus:	115,774
Producer surplus:	-24,089

Figure 6: Average Change in Market Shares and Market Welfare in France post Divestiture

8 Conclusion

Discrete choice models are commonly used to answer research questions where the dependent variable is categorical. These models are based on the assumption of 'Random Utility Maximization' (RUM) which states the choice maker's primary objective is to maximize his utility. The multinomial logit (MNL) and nested logit (NL) models are the two most commonly used discrete choice models due to their simplistic mathematical structures and ease of estimation. The assumption that the error terms in the models follow iid Type 1 Gumbel distribution, leads to a closed-form analytic solution for the choice probability. However, The difference between the two models arises through the property of 'Independence and Irrelevant Alternatives'. The IIA property implies that the relative odds of choosing between any two options should not change if a third option is added or removed, which can lead to unrealistic predictions in situations where adding or removing options significantly affects the relative attractiveness of the other choices. The nested logit model overcomes the challenge of the IIA property in the MNL by accounting for the substitution patterns of consumers between alternative products. These substitution patterns help the researcher to incorporate the complexity of decision-making processes to a greater extent than the MNL model. The BLP model is commonly used to estimate the demand parameters in a differentiated products market. The model estimates these parameters while considering the consumer heterogeneity and price endogeneity that is not fully explained by observed characteristics. However, this model tends to be computationally intensive and is also one of the motivations for applying the nested logit model under the BLP framework.

Based on the simulation study, the Nested Logit (NL) model demonstrates clear advantages over the Multinomial Logit (MNL) model within the BLP framework. The empirical application also underscores the NL model's potential for analyzing market competition.

In the simulation study, the nesting parameters under the NL model are correctly predicted in line with the principle of RUM as seen in Figure 7 in the Appendix. The results from the study corroborate the economic theory, primarily highlighting the flawed inference of the cross-elasticity estimates in the MNL model. It underscores the NL model's superior ability to handle endogeneity and capture complex substitution patterns among products. This implies that the segments capture the important role unobserved characteristics play in consumers' choice of alternatives. Even in terms of statistical performance, the MNL model is rejected in favor of the NL model. The NL model yields lower AIC and BIC values, indicating a better fit and more efficient use of data. Furthermore, the likelihood ratio test strongly supports the NL model's fit over the MNL model, reinforcing its robustness.

The empirical application analyzes the use of the target model structure in interpreting the market outcomes. Using the market-level dataset on car sales and specifying the two-level nesting structure of the NL model, demand estimates were drawn. These estimates were thereby used to simulate firm mergers to understand the impact of change in market competition on the overall market welfare. A particularly relevant aspect of consumer heterogeneity was the cars domestic/foreign origin, which the NL model captured reasonably well. The difference in cross-price elasticity of alternatives belonging to the same and different subgroups within the same nest highlights that the origin of the car defines consumer preference as well. The merger also informs that the merger will entail high price effects if the merging firms sell close substitutes with respect to each other. Therefore, the price effects are more pronounced on domestic firms than the foreign firms in France. Divestiture as a remedy for monopoly behavior does curtail the huge price effects and ensures the limited impact of the merger on overall market efficiency. Thus, the two-level nested logit model can be a valuable tool for understanding the dynamics of market shocks such as firm mergers.

The topic presents substantial opportunities for further application and research. Potential avenues include extending the model with alternative functional forms of demand, such as the random coefficients nested logit model and the cross-nested logit model. Additionally, employing constant expenditures utility specification could be investigated for enhanced accuracy of predictions. The logarithmic form of price in this indirect utility function may generate different results for elasticity and merger effects compared to unit demand.

References

[1]   Arthur Berry, James Levinsohn, and Ariel Pakes. "Automobile Prices in Market Equilibrium". In: *Econometrica* 63.4 (1995), pp. 841–890.

[2]   Steven T Berry. "Estimating discrete-choice models of product differentiation". In: *The RAND Journal of Economics* (1994), pp. 242–262.

[3]   Steven T Berry and Philip A Haile. "Foundations of demand estimation". In: *Handbook of industrial organization*. Vol. 4. 1. Elsevier, 2021, pp. 1–62.

[4]   Jonas Björnerstedt and Frank Verboven. "Does merger simulation work? Evidence from the Swedish analgesics market". In: *American Economic Journal: Applied Economics* 8.3 (2016), pp. 125–164.

[5]   Jonas Björnerstedt and Frank Verboven. "Merger simulation with nested logit demand". In: *The Stata Journal* 14.3 (2014), pp. 511–540.

[6]   Axel Börsch-Supan. *Econometric analysis of discrete choice: with applications on the demand for housing in the US and West-Germany*. Vol. 296. Springer Science & Business Media, 2012.

[7]   Yves Croissant et al. "Estimation of multinomial logit models in R: The mlogit Packages". In: *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf* (2012).

[8]   Pinelopi Koujianou Goldberg and Frank Verboven. "The evolution of price dispersion in the European car market". In: *The Review of Economic Studies* 68.4 (2001), pp. 811–848.

[9]   Laura Grigolon and Frank Verboven. "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation". In: *Review of Economics and Statistics* 96.5 (2014), pp. 916–935.

[10] Florian Heiss. "Structural choice analysis with nested logit models". In: *The Stata Journal* 2.3 (2002), pp. 227–252.

[11] Joel L Horowitz. "Modeling the choice of choice set in discrete-choice random-utility models". In: *Environment and Planning a* 23.9 (1991), pp. 1237–1246.

[12] Frank S Koppelman and Chandra Bhat. "A self instructing course in mode choice modeling: multinomial and nested logit models". In: (2006).

[13] Frank S Koppelman and Chieh-Hua Wen. "Alternative nested logit models: structure, properties and estimation". In: *Transportation Research Part B: Methodological* 32.5 (1998), pp. 289–298.

[14] Hongmin Li and Woonghee Tim Huh. "Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications". In: *Manufacturing & Service Operations Management* 13.4 (2011), pp. 549–563.

[15] Jordan J Louviere, David A Hensher, and Joffre D Swait. *Stated choice methods: analysis and applications*. Cambridge university press, 2000.

[16] Daniel McFadden. "Conditional logit analysis of qualitative choice behavior". In: (1972).

[17] Daniel McFadden. "Econometric models of probabilistic choice". In: *Structural analysis of discrete data with econometric applications* 198272 (1981).

[18] Aviv Nevo. "A practitioner's guide to estimation of random-coefficients logit models of demand". In: *Journal of economics & management strategy* 9.4 (2000), pp. 513–548.

[19] Nalin Kumar Ramaul and Pinki Ramaul. "Regional incentives and location choice of new firms in India: a nested logit model". In: *Journal of Quantitative Economics* 16.2 (2018), pp. 501–525.

[20] Nadja Silberhorn, Yasemin Boztuğ, and Lutz Hildebrandt. "Estimation with the nested logit model: specifications and software particularities". In: *Or Spectrum* 30.4 (2008), pp. 635–653.

[21] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

9 Appendix

Figure 7: Nesting Parameters for random 100 datasets in the simulation study to corroborate the validity of the 'Random Utility Maximization' principles

firm code	price	Marginal costs	Pre-merger Lerner
BMW	27.036	23.806	0.130
Fiat	15.447	12.626	0.230
Ford	16.057	13.240	0.194
Honda	25.549	23.058	0.115
Mazda	12.712	10.218	0.198
Mercedes	31.482	26.791	0.150
Mitsubishi	15.376	12.928	0.165
Nissan	11.544	9.060	0.222
GM	18.585	15.925	0.174
PSA	16.127	11.261	0.347
Renault	14.052	10.213	0.309
Rover	13.762	11.228	0.213
Toyota	14.023	11.555	0.184
VW	13.848	10.699	0.264
Volvo	22.456	19.910	0.124

Figure 8: Pre-Merger Market Conditions in France

firm code	Pre-merger	Post-merger	Difference
BMW	0.012	0.019	0.006
Fiat	0.074	0.114	0.040
Ford	0.072	0.113	0.040
Honda	0.006	0.009	0.003
Mazda	0.008	0.013	0.005
Mercedes	0.012	0.014	0.002
Mitsubishi	0.001	0.002	0.001
Nissan	0.007	0.010	0.004
GM	0.052	0.083	0.031
PSA	0.347	0.262	-0.086
Renault	0.273	0.155	-0.118
Rover	0.018	0.028	0.011
Toyota	0.005	0.008	0.003
VW	0.107	0.163	0.057
Volvo	0.005	0.008	0.002

	Pre-merger	Post-merger
HHS:	2209	2346
C4:	80.10	80.65
C8:	95.56	95.05

	Change
Consumer surplus:	-6,574,984
Producer surplus:	3,689,366

Figure 9: Average Change in Market Shares and Market Welfare in France

10 Statement of Authorship

"Ich versichere hiermit, dass ich die vorstehende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass die vorgelegte Arbeit noch an keiner anderen Hochschule zur Prüfung vorgelegt wurde und dass sie weder ganz noch in Teilen bereits veröffentlicht wurde. Wörtliche Zitate und Stellen, die an- deren Werken dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall kenntlich gemacht."

"I hereby confirm that the work presented has been performed and interpreted solely by myself, except for where I explicitly identified the contrary. I assure that this work has not been presented in any other form for the fulfillment of any other degree or qualification. Ideas taken from other works in letter and in spirit are identified in every single case."

Date: _____   Signature: _____