Article

An Optimal Design through a Compound Criterion for Integrating Extra Preference Information in a Choice Experiment: A Case Study on Moka Ground Coffee

Rossella Berni *, Nedka Dechkova Nikiforova and Patrizia Pinelli

Department of Statistics Computer Science Applications “G. Parenti”, University of Florence, 50134 Florence, Italy; n.nikiforova@unifi.it (N.D.N.); patrizia.pinelli@unifi.it (P.P.)
* Correspondence: rossella.berni@unifi.it; Tel.: +39-055-2751500

Abstract: In this manuscript, we propose an innovative approach to studying consumers’ preferences for coffee, which integrates a choice experiment with consumer sensory tests and chemical analyses (caffeine contents obtained through a High-Performance Liquid Chromatography (HPLC) method). The same choice experiment is administered on two consecutive occasions, i.e., before and after the guided tasting session, to analyze the role of tasting and awareness about coffee composition in the consumers’ preferences. To this end, a Bayesian optimal design, based on a compound design criterion, is applied in order to build the choice experiment; the compound criterion allows for addressing two main issues related to the efficient estimation of the attributes and the evaluation of the sensorial part, e.g., the HPLC effects and the scores obtained through the consumer sensory test. All these elements, e.g., the attributes involved in the choice experiment, the scores obtained for each coffee through the sensory tests, and the HPLC quantitative evaluation of caffeine, are analyzed through suitable Random Utility Models. The initial results are promising, confirming the validity of the proposed approach.

Keywords: Mixed Logit; Coordinate-Exchange algorithm; sensory test; respondents’ heterogeneity; scoring card; caffeine

1. Introduction

Discrete Choice Experiments, or more briefly, Choice Experiments (CEs), are widely used by extant researchers to study consumers’ preferences for a given product or a service. They are successfully applied in various research fields like marketing, transportation, environmental, health, and political sciences. A fundamental issue when dealing with the method of CEs is the underlying experimental design, through which choice sets are composed and then administered to each respondent. In a CE, respondents are asked to express a preference considering different choice sets, each composed of several alternatives related to a product or a service; each alternative, in turn, is defined as a combination of different attributes’ levels. Respondents are supposed to be utility maximizers; therefore, they are asked to choose from each choice set the alternative that maximizes his/her utility. Preferences are subsequently analyzed through Random Utility Models (RUMs) [1,2].

The planning of the experimental design is fundamental for defining choice sets. To this end, the most widely accepted method in the literature is to build choice sets based on optimal design theory [3,4]. In the literature, also considering the Bayesian design framework, several seminal papers addressed the issue of how to build an optimal choice design according to the simplest RUM model, e.g., the Multinomial Logit (MNL), also called Conditional Logit [5,6].

In this paper, and starting from an optimal Bayesian design framework, we deal with Multinomial Discrete Choice models belonging to the class of RUM. More precisely, Mixed Logit (MIXL) models are applied to a real case study that concerns the analysis of consumers’ preferences for coffee.
consumers’ preferences for sustainable coffees, integrating CEs with a guided tasting for the sensory assessment. Through this experimental study, we also address the issue of how to incorporate efficiently, in the design matrix for the CE, extra preference information, which derives from a guided tasting and a High-Performance Liquid Chromatography (HPLC) analysis. For this purpose, the specific optimal design is based on a compound design criterion \[4,7,8\], achieving the twofold aim of efficiently estimating the CE attributes and detecting the extra preference information. In this direction, considering the case study on coffee, Pinelli et al. \[9\] applied Heteroscedastic Extreme Value (HEV) models to analyze consumers’ preferences for coffee. Following that, considering a real case study on coffee, Nikiforova et al. \[10\] proposed an innovative approach to build optimal heterogeneous choice designs under the Panel MIXL model based on an approximate design framework by using the General Equivalence Theorem (GET) \[11\].

Moreover, Berni et al. \[12\] applied the MIXL model to study consumers’ preferences for coffee consumption without involving socio-economic and respondents’ characteristics. In this manuscript, the modeling step includes some respondents’ characteristics jointly with issues related to the tasting, allowing a better understanding of the consumers’ preferences for coffee consumption. Moreover, this case study on coffee involves novel aspects of optimal designs for CE, e.g., the optimal design is built by specifically defining the program code in SAS IML by using the Coordinate-Exchange algorithm to obtain the final optimal design, in which a balance between the extra sensorial information and CEs is achieved.

In the literature, other studies deal with CEs integrated with sensorial tasting, such as Torquati et al. \[13\] and Boehm et al. \[14\], in which a study on tea is carried out.

The remainder of the manuscript is organized as follows. Section 2 includes the literature review. Section 3 reports the motivating example and the underlying theory on optimal designs and modeling for CEs. Section 4 includes a detailed case study description, while Section 5 presents the model results. Discussion and final remarks follow.

2. Literature Review

Many developments and improvements in consumer/user preferences by considering experimental designs and the statistical modeling were achieved starting from the 2000s, and considering the fundamental theory of the 1990s.

Undoubtedly, a further and precise distinction must be made when we refer to preference measurements or, more in general, to the preference theory and Stated Preferences (SPs), where we define as SP the preference of a respondent related to a scenario, hypothetical and/or based on a priori information, shown as an alternative in CEs.

In the standard situation, a CE is based on choice sets. Each choice set is formed by a set of alternatives, which are selected from an experimental design; the respondent is asked to give his/her preference within each choice set.

A further distinction is when considering conjoint analysis (CA), in which ranking (ordinal scale) is the classical response variable, owing to the different framework of profiles. In a CE and the corresponding Choice Modelling (CM) situation, choice (binary or not) is usually the response variable. Undoubtedly, CE is the preferred method in the literature and, consequently, the related theory was widely developed in the last few years by considering the experimental design, e.g., optimality criteria, and statistical models (RUMs), including the related developments, and considering the seminal papers by Train \[15\], McFadden and Train \[2\], Boxall and Adamowicz \[16\], Hynes et al. \[17\], Wen and Koppelman \[18\].

When a methodology includes several theoretical steps, these elements (mainly experimental designs and statistical models) are closely connected \[19,20\], and the properties of the design affect the corresponding model. If these properties are not contemplated in the design, this issue must be evaluated in the model formulation. In the past, the improvement in the design optimality was explicitly defined for a Mixed Multinomial Logit (MMNL) \[21\], or, when considering the respondents’ heterogeneity, a specific design matrix for each respondent was planned \[22\] by including the heterogeneity evaluation directly in the design step instead of the model step.
However, as said hereinabove, a different evolution characterized the Multi-Attribute Valuation methods (MAVs), e.g., Conjoint Analysis Choice Experiments and Contingent Ranking (in this review, we strictly pay attention to CA and CEs), even though there are some features in common, such as specific methods, algorithms, and models, in order to select alternatives, by considering the planning step ([23,24]) or the analysis of collected data, such as in [25], in a constantly developing situation, where practical and theoretical issues are combined involving both methods, CA and CE, sometimes jointly. Undoubtedly, the classical CA is an easier task; the theory and applications in the literature present above all developments and studies about the complexity and selection of profiles in the model step ([23,25]), i.e., some problems of complexity, such as preference uncertainty and conflicts solved through the evaluation of the judgment time and the response error in a rating task [26]. The design of experiments is involved in order to create an orthogonal design (sometimes optimal) where all the created profiles, according to the set of attributes considered, are eventually reduced by applying a fractional factorial design. The complexity of statistical models goes back to finite-mixture models and hierarchical Bayes models, such as in [27,28], to evaluate the respondents’ heterogeneity or the complexity of alternatives. Recently, CA is usually applied jointly with CE in a choice-based conjoint study, such as in Hong et al. [29], where the authors analyze meat reformulation and consumers’ Willingness-to-Pay (WTP) in a combined sensory evaluation plus a choice-based conjoint design that involves a simple full factorial design and a generalized mixed logit model.

Instead, a different situation is presented when evaluating the CE/CM field, namely the CE method, where optimality criteria, ad hoc algorithms, and specified information matrices for design and modeling were entirely defined starting in the 1990s [30]. The first developments were related to the construction of optimal or near-optimal designs with two-level attributes for binary choices in the presence of first-order interactions [31], or when optimal designs are defined for mixed-level attributes [32], or defining new criteria, such as the M-optimality [24], where the manager’s issues were involved.

Note that a common line of reasoning is related to the connection between designs and models jointly with the need for a guiding thread between manufacturers and consumers. In addition, the seminal paper of Sándor and Wedel [21] reflects the strict connection between experimental designs and statistical models through a Mixed Multinomial Logit. This model, belonging to the class of RUMs, is certainly the most widely applied and developed model in recent years for the CE situation [2,33–35]. Its success is easily explained when considering the theoretical results of McFadden and Train [2], Train [15], and the possibility, by adding additional random parameters, to study respondents’ heterogeneity and the correlation structures due to repeated choices. The last developments of this model include its relationship with the latent class model in order to create a finite number of respondent groups [16,17,36,37], also related to market segmentation, such as more recently, in Yeh et al. [38], where a Hierarchical Bayes mixed logit model jointly with latent class segmentation is applied to study red sweet peppers in Taiwan, and in [39], where the authors combine sensory test and CEs to identify specific quality signals in the sensorial field, focusing on fruits and vegetables.

When considering CEs jointly with CM, peculiar issues should be evaluated to discuss this method completely, such as the estimation step and the simulation algorithms to solve the model’s expression [40]. We mainly focus on the model step, by evaluating the solutions suggested in the literature to solve the effective problems when statistical models are applied. The role of the experimental design is not irrelevant when considering its properties; for example, the search for a D-optimal design implies the maximization of the determinant of the information matrix and, therefore, this maximization directly affects the variances of the coefficient estimates and the volume of the ellipsoid, confidence region of the estimated coefficients, strictly related to the design precision. This result, by using a D-optimal design, implies a larger efficiency in the estimates. Moreover, in the specific literature about CEs, D-optimal designs, starting from the simple fractional factorial designs, and by considering more complex designs [19,30], were defined through specific
algorithms for the trial-point selections, and then developed by using Bayesian optimal designs [41]. Furthermore, the use of alternative optimality criteria had a continuous improvement over time, considering, among others, Burgess and Street [32], up to the most recent developments [42], where optimal designs, and specifically the criteria D and I, are expounded and discussed for CEs involving mixture variables, e.g., ingredient proportions and process variables, as cooking features.

From an empirical point of view, a Bayesian efficient design is used in a combined study involving tasting and CEs, for extra-virgin olive oil [43]; in this case, a random parameter logit model is applied, aiming to estimate the consumers’ attitude towards purchasing and re-purchasing the product.

In the same line of reasoning, when considering the recent developments in the literature, Gatti et al. [44] implemented a D-efficient design for analyzing coffee consumers’ preferences towards eco-labels and biodiversity conservation. In this study, as in similar recent research, the CE (or, in general, the survey) is conducted online, and surely this aspect allows for a notable increase in the size of respondents, such as in Wuepper et al. [45], where consumers’ preferences towards sustainable coffees are analyzed. In this case, the authors deeply verify the possibility of a hypothetical bias due to the differences between the stated (expressed preferences by filling the CE) and the real behavior of the respondents versus the specific purchase. In general, this issue is particularly relevant when the WTP [29] and the purchase/re-purchase attitude are the main aim of the study.

In Torquati et al. [13], a joint study, e.g., sensory evaluation of meat and a CE, aims to analyze consumers’ behavior and the WTP. The authors used a different strategy, dividing the respondents into two groups: tasting and no-tasting. Both groups are asked to participate in the CE.

Sustainable coffees, labels, and certification are widely studied in the recent literature, as mentioned before [44]; other contributions are related to specific label certification, such as FairTrade and UTZ certifications [46], where the WTP is analyzed, and in Fuller and Grebitus [47], where an interesting distinction and analysis between social sustainability and environmental sustainability for coffee labeling is performed. In this case, only coffee buyers are interviewed, and to avoid hypothetical bias in estimating WTP, a cheap talk is applied.

3. Motivating Example and Underlying Theory

This section describes the CE theory applied to the real case study from which our innovative proposal is motivated. In particular, the study is articulated in three main parts, each related to a specific evaluation of the two sustainable coffees under research. The first relates to the results obtained from the HPLC evaluation method for the two coffees; the second regards the sensory assessment scores obtained through the guided tasting. The last part is the CE, which must be administered twice to the respondents before and after the guided tasting. Moreover, one main aim is to build an optimal CE that includes prior HPLC results and the scores obtained through the guided tasting.

3.1. Theory: The Optimal Compound Criterion for the Evaluation of Sustainable Coffees

The seminal contributions of Kiefer and Wolfowitz [11] and Kiefer [48] laid the basis for the optimal design theory. Differently from the classical design of experiments, optimal designs are generally model-dependent; that is, one statistical model should be defined a priori to build the design, considering one design optimality criterion. In this regard, design optimality criteria are specifically defined according to a design precision measure, based on the Fisher Information matrix (FIM) [3,4], that complies with one of the three GET theorem conditions [11].

When dealing with CEs, the most widely applied design criterion is D-optimality. Nevertheless, in the literature, other optimality criteria are also proposed; for instance, the A-criterion aims to minimize the average variance of the parameter estimates [5], and the M-criterion addresses the purpose of efficient estimation of linear functions of the model
parameters that are managerially relevant [20]. Furthermore, to accurately predict future choices, Kessels et al. [5] introduced the G- and V-optimality criteria, the latter also named I-optimality, comparing them with D- and A-optimality in a Bayesian design framework. Henderson and Liu [49] recently used a compound design criterion to construct CEs. Assuming a selective choice process composed of active and inactive attributes, the authors applied this criterion to incorporate prior information for the joint purpose of efficient estimation of the parameters for the active attributes and the detection of the effects for the inactive attributes.

Regardless of the optimality criterion chosen to build the optimal choice design, a further issue to consider is that the RUMs are not linear in the parameters, implying that the FIM depends on the unknown parameter values. In order to deal with this issue, it is possible to (i) assume nominal values, for example, from previous and pilot studies, or based on the opinion of experts or focus groups, or (ii) use Bayesian design techniques.

The Bayesian design framework for building optimal choice designs consists of integrating a design criterion over a prior distribution of likely parameter values [5,6,19,22,50]. This approach is preferred, especially when there is no a priori information about possible nominal values for the unknown parameters.

When considering the MIXL model, one of the main criteria focused on D-optimality, also in the Bayesian design framework. Let us define \( \theta = (\beta, \sigma) \), the 2K-dimensional vector of model parameters \( \theta = (\beta, \sigma) \) in a MIXL model, \( X \) the design matrix, and \( I(X, \theta) \) the FIM of the MIXL model. A Bayesian D-optimal choice design for the MIXL model minimizes the following design criterion \( D_B \) [19]:

\[
D_B = \int_{\mathbb{R}^K} \det \left[ I(X, \theta) \right]^{-\frac{1}{2}} \pi(\beta) d \beta
\]

(1)

where \( \text{det} \) is for determinant function, and \( \mathbb{R}^k \) denotes the real space of dimension \( K \); a prior distribution \( \pi(\beta) \) is defined for the parameter vector \( \beta \). Usually, a Normal prior is assumed for the parameters in the vector \( \beta \), while nominal values are used for the heterogeneity parameters in the vector \( \sigma \). For the detailed derivation of the FIM \( I(X, \theta) \), refer to Sándor and Wedel [21].

By following the motivating example, to take the result of the HPLC analysis and the scores obtained through the guided tasting into account, we apply a compound design criterion focused on two main issues related to the efficient estimation of the attributes and to the evaluation of sensorial part, e.g., the HPLC effects and the scores obtained through the guided tasting. It must be noted that, according to the plan of the CE administration, we present the same group of choice sets twice (first and second CE sessions).

In what follows, first, we illustrate the following general compound criterion \( \Phi_C \), Formula (2), and then we define, through Formula (3), the specific design criterion \( \Phi_{CB} \) [4,7,8].

The general compound criterion is defined as follows:

\[
\Phi_C(\xi) = \frac{\alpha}{k_1} \left[ \log |I_{11}(\xi)| \right] + \left( 1 - \frac{\alpha}{k_2} \right) \left[ \log |I(\xi)| - \log |I_{11}(\xi)| \right]
\]

(2)

where \( \xi \) is a design measure defined as a probability measure on the experimental region \( \chi \) [4,51]. In Formula (2), we can observe a decomposition of the FIM; more precisely, the first part of the FIM, \( I_{11} \), corresponds to the \( k_1 \) coefficients for the CE attributes. Therefore, \( I \) is the global FIM \( (k_1 + k_2 = k) \) that includes both the \( k_1 \) coefficients for the CE attributes and the \( k_2 \) coefficients related to the tasting. Furthermore, the coefficient \( \alpha (0 \leq \alpha \leq 1) \) establishes the relative weight of both objectives in the design criterion \( \Phi_C \). Moreover, if the value is \( \alpha = 1 \), a D-optimal choice design for the \( k_1 \) model coefficients related explicitly to the CE part is built; instead, \( \alpha = 0 \) allows obtaining a \( D_s \)-optimal design for the \( k_2 \) coefficient related to the tasting.

To account for the FIMs’ dependence on the unknown parameter values, we employ Bayesian design techniques by integrating the design criterion \( \Phi_C \) (Formula (2)) over a prior
distribution of likely parameter values \([4,6,19]\). Therefore, for obtaining the compound D-optimal design for the MIXL model, we assume a Normal prior distribution for the vector \(\beta\) and nominal values for the heterogeneity coefficients. Thus, the specific compound design criterion is expressed as follows:

\[
\Phi_{CB}(\xi) = \int_{\mathbb{R}^k} \frac{\alpha}{k_1} \left| \log |I_{11}(\xi)| \right| + \frac{(1-\alpha)}{k_2} \left[ \log |I(\xi)| - \log |I_{11}(\xi)| \right] \pi(\beta) d\beta \tag{3}
\]

The compound D-optimal design for the CE is obtained through the SAS software, Interactive Matrix Language (IML) (SAS Version 9.4; Windows Platform). We specifically developed the program code in SAS IML by following the underlying theory, in order to define the final optimal design, through the Coordinate-Exchange algorithm \([6,52,53]\). The \(k\)-dimensional integral in Formula (2) is approximated through Quasi-Monte Carlo methods of Halton sequences \([19,54]\). This algorithm is computationally more efficient for constructing optimal choice design, especially when the design involves many attributes and levels \([6]\). Unfortunately, as pointed out by Liu and Tang \([55]\), a common problem with these computer search algorithms is that there is no guarantee that we can find a globally optimal design. Consequently, the search may rise to a far less efficient design than the globally optimal one unless all possible choice designs are covered in the computer search. For this reason, we perform multiple searches (also called tries) by exploring different initial designs, and the best design is then selected.

To illustrate the value of \(\alpha\) for the compound design criterion, Formula (3), in Figure 1, we plot the efficiencies against different values of \(\alpha\). According to the results shown in this plot, we choose a value of \(\alpha = 0.75\), which allows us to achieve a balance equal to 68% efficiency for D and Ds optimalities.

![Figure 1](image-url). Efficiencies according to different values of \(\alpha\) for achieving the final optimal design (authors’ source).

The final optimal design for the planned CE includes 24 choice sets with binary alternatives. These choice sets are randomly subdivided into groups of eight choice-sets, and subsequently, each group, including eight choice sets, is administered to each
respondent. Moreover, the planned CE is unlabeled without including the no-choice option. For example, we provide one of the 24 supplied choice sets as a Supplementary Materials.

The whole sequence of the study follows this order: background questionnaire, Choice1, tasting of the two coffees, and Choice 2. The same CE is administered before (Choice1) and after (Choice 2) the guided tasting, as detailed in Section 4.3.

3.2. Theory: The Choice Modeling

In the RUM class, the simplest model is the Conditional Logit model, in which the main limitation is related to the Independent and Irrelevant Alternatives (IIA) assumption [2]. In particular, this assumption considers the preference of an alternative in a choice set independent of the presence of other alternatives within the same choice set, and from the other choice sets, about which the consumer is asked to express a preference. This assumption has implications for the cross-elasticities of two alternatives, in the sense that these are constant with respect to the presence and characteristics of the other alternatives in the choice set. To address these issues, we apply the Mixed Logit (MIXL) model [2], which allows us to account for the heterogeneity across consumers.

In general, we suppose that the $i$-th statistical unit (assumed to be a respondent) expresses his/her preference by choosing the $j$-th alternative; $C_i = (1, \ldots, j, \ldots, J)$ is the general choice set supplied to a consumer $i$.

A general formulation for a single decision, when the choice is binary, is based on the assumption of a general continuous distribution for the $\psi_{ij}$, called the mixing term, so that the utility index $U_{ij}$ becomes:

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

where $V_{ij}$ is the deterministic part of utility, while $\epsilon_{ij}$ is the random component, assumed independent and Gumbel-distributed.

A density for $\psi_{ij}$ is defined as:

$$g(\psi|\zeta)$$

where the parameter $\zeta$ contains the parameters of the error distribution, such as Normal, Uniform, and Log-Normal. If $\psi$ is not evaluated, then the MIXL reduces to the simple conditional logit; moreover, the unconditional probability is equal to:

$$P(y_i = j) = P_i(j) = \int_{\psi} L_i(j|\psi_{ij}) g(\psi_{ij}|\zeta) d\psi_{ij}$$

where $C_i$ is the general choice-set supplied to the $i$-th consumer; $x_{ij}$ is the vector of attribute values for the $j$-th alternative corresponding to the decision for the $i$-th consumer; $\psi$ is the random parameter term expressing the respondents’ heterogeneity. The term $L_i(\cdot)$, Formula (7), is the general expression for a MIXL, where the $g(\psi_{ij}|\zeta)$ is the mixing component.

Concerning Formulas (6) and (7), we note that the general expression of a RUM model changes, since we include the component $\psi_{ij}$ to highlight the aspect of heterogeneity and to represent the unobserved characteristics of the respondent.

4. The Case Study

4.1. HPLC Analysis and Guided Tasting

Two Italian brands of coffee grounds for moka are chosen with different organoleptic characteristics: an intense, soft, and aromatic blend (100% Arabica) and a round coffee with high aftertaste intensity (a blend of Arabica and Robusta varieties). The two selected coffees are previously analyzed concerning their caffeine and polyphenolic antioxidant content through an HPLC method. Therefore, for both types of coffee, we obtained the HPLC measurement results (three replicates) related to the following quantities:
– chlorogenic acid: the main antioxidant compound, derivative of caffeic acid;
– caffeine;
– sum of other antioxidant compounds and caffeic acid derivatives.

Subsequently, a guided tasting is planned in order to analyze the role of the taste in consumers’ preferences. In particular, two scorecards, one for each type of coffee, are defined for the organoleptic evaluation. They are composed of ten different descriptors: the color related to the sight, the intensity of the aroma, and the quality of the aroma related to the olfaction; four different descriptors related to the taste (bitter, acidic, sweet, and aroma); the touch related to the body, the aftertaste, and the general equilibrium of the coffee. For each type of coffee, a respondent gives his/her evaluation by scores in an interval from 1 to 7, with the possibility of also giving “intermediate” scores (e.g., 3.5, 2.5, etc.). For each scorecard, the scores related to the ten descriptors are summarized in a unique value. In the Supplementary Materials, we provide the general blank scorecard used for the guided tasting.

4.2. Attributes and Levels for the Planned CE

Before and during the research planning, CEs, and guided tasting, we carefully evaluated all the available coffee products with similar characteristics (e.g., packaging, price, and taste qualities). To this end, we chose the CE attributes (and levels) based on a careful assessment of the types of coffees available from mass-market retailers in Italy and by considering Italian coffee brands [56]. In particular, attributes were chosen following a prior investigation of the main Italian coffee brands present in large-scale retail trade, including local roasters, premium store brands, and high-quality Italian brands with different coffee blends. Furthermore, the two types of coffee, tasted by the respondents, were blinded and accurately chosen in order to be completely representative of the scenarios faced in the choice experiment.

Regarding the CE, we identified six attributes, reported in Table 1: the “Type of coffee”, the “Type of packaging”, two attributes related to the taste (“Soft-Velvety taste” and “Intense-Aromatic taste”), the sustainability and the geographical origin, included in the attribute “Geographical origin and sustainability”, where the certification of sustainability is defined as the presence (in the coffee label) of some certification related to economic, social, and/or environmental sustainability, and the indication of the geographical origin of the coffee. Finally, the last attribute is the price (for a quantity of 250 g of coffee), established at EUR 4.50, 6.00, and 7.50.

Table 1. Attributes and attribute levels for the CE.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Type of coffee</td>
<td>1: Blend of Arabica and Robusta 2: Blend of 100% Arabica</td>
</tr>
<tr>
<td>II. Type of packaging</td>
<td>1: Package in a protective atmosphere 2: Tin in a protective atmosphere</td>
</tr>
<tr>
<td>III. Intense-Aromatic Taste</td>
<td>1: Medium 2: High</td>
</tr>
<tr>
<td>IV. Geographical origin and</td>
<td>1: Geographical origin 2: Certification of sustainability</td>
</tr>
<tr>
<td>sustainability</td>
<td></td>
</tr>
<tr>
<td>V. Soft-Velvety Taste</td>
<td>1: Medium 2: High</td>
</tr>
<tr>
<td>VI. Price (euro)</td>
<td>1: 4.50 2: 6.00 3: 7.50</td>
</tr>
</tbody>
</table>

4.3. Plan for the CE Administration

Data are collected by administering the same CE in two consecutive sessions, before and after the guided tasting, to compare if the tasting could affect consumers’ preferences. The survey steps are defined as follows:

1. A background questionnaire containing structural items (age, marital status, respondents’ habits on coffee consumption and buying) is completed by each respondent.
2. Subsequently, the choice sets are administered to the respondents (first CE session). Each choice set is composed of binary alternatives; therefore, for each supplied choice set, the respondents are asked to choose the alternative which maximizes his/her utility. Once the first CE session is completed, we perform the guided tasting.

3. First, an expert in this field provides detailed information on coffee. More precisely, the expert illustrates the two coffee blends (e.g., Arabica and Robusta and 100% Arabica), the type of taste that characterizes each, and the corresponding quantities of caffeine and antioxidants contained. Moreover, the expert also describes the ten descriptors reported in the scoring card and explains how to fill in them. In the meantime, the first coffee type, blinded, is prepared in moka. The respondents tasted it and expressed their sensory assessment scores to each coffee descriptor reported on the scoring card. The same procedure, e.g., preparation, tasting, and scoring card-filling, is repeated for the second type of coffee. Both coffee blends are supplied for tasting without sugar. Moreover, a cup of water is also given before tasting the second type of coffee. The two types of coffee were prepared in a moka coffee pot for 18 cups (thus, a maximum of 18 people at a time) and administered in a randomized sequence. Considering each survey occasion, we conducted the study for a maximum of 18 persons, primarily in Florence, Tuscany. The duration of each survey occasion was about an hour and a half.

4. Once the guided tasting is completed, the same group of choice sets (second CE session) is supplied to each respondent to compare if some differences can occur in the consumers’ preferences between the two CE sessions given the effect of the guided tasting.

Therefore, the supplied choice sets are related to the CE per se and are composed of binary alternatives. Differently, the scoring cards are related in strictu sensu to the guided tasting; then, two scoring cards are supplied, one for each selected coffee blend.

5. Model Results

In this section, we illustrate the estimation results obtained by applying the MIXL models considering Choice 1 (Section 5.1) and Choice 2 (Section 5.2) sessions. The data are related to 107 respondents [12]. More precisely, 68% were younger than 35 years, and 32% were older than 35 years; 48% were men, and 52% were women; 78% were unmarried, and the rest (22%) included all the other civil statuses [12]. As regards the perceived economic well-being, the questionnaire item was stated with regard to the respondent’s judgment of his/her economic status; 47% perceived themselves as having a well-off or affluent economic status, while 53% perceived themselves as having a medium or modest economic status. In relation to purchasing habits, e.g., how often they purchase sustainable products, the results indicate that 51% of respondents always or regularly purchase sustainable products, 30% do so often, and 19% rarely or never. Furthermore, respondents were also asked to express their preferences for sustainable products; 84% consider sustainable products very/fairly important, while 16% declared that they are unimportant.

In our experience with the investigated coffee samples, Arabica and Robusta blends’ coffee composition is characterized by higher caffeine than the coffee obtained from 100% Arabica blends. Therefore, the caffeine content allows us to better evaluate the consumers’ preferences towards the type of coffee, e.g., more or less strong coffee. To this end, we include the caffeine in both MIXL models, opportunely standardized so that more caffeine means a blend of Arabica and Robusta varieties, and less caffeine means a 100% Arabica blend. For both models, we report the estimated heterogeneity effects in terms of $\sigma^2$ rather than $\sigma^2$; for this reason, the estimated heterogeneity coefficients can take values on $\mathbb{R}$ space. Moreover, we also report several goodness-of-fit measures for both estimated models. Both models are estimated in SAS by applying the procedure Multinomial Discrete Choice (MDC) (SAS System, SAS Campus Drive, Cary, NC 27513, USA; Windows Platform v. 9.4).
5.1. Mixed Logit Model Results for Choice 1 Coffee Session

When considering the estimation results related to the Choice 1 coffee session (Table 2), the positive sign of the estimated coefficient related to the “Geographical origin and sustainability” indicates that consumers’ preferences go towards the certification of product sustainability with respect to the indication of geographical origin. The negative sign of the “Type of packaging” means that consumers prefer soft bags with a modified atmosphere with respect to the jars in a modified atmosphere.

Table 2. Model results for Choice 1: estimated coefficients, standard errors, p-values.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>St. Error</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of coffee</td>
<td>1.0237</td>
<td>0.2306</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Type of packaging</td>
<td>−0.2230</td>
<td>0.1162</td>
<td>0.0550</td>
</tr>
<tr>
<td>Geographical origin and sustainability</td>
<td>0.7894</td>
<td>0.1796</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intense-Aromatic Taste</td>
<td>−0.8542</td>
<td>0.1811</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Soft-Velvety Taste</td>
<td>1.2012</td>
<td>0.2658</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Caffeine</td>
<td>1.5389</td>
<td>0.4035</td>
<td>0.0001</td>
</tr>
<tr>
<td>Price—1st level</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Price—2nd level</td>
<td>−0.6261</td>
<td>0.3165</td>
<td>0.0479</td>
</tr>
<tr>
<td>Price—3rd level</td>
<td>−1.0208</td>
<td>0.4751</td>
<td>0.0317</td>
</tr>
<tr>
<td>Type of coffee × gender</td>
<td>0.2176</td>
<td>0.3999</td>
<td>0.5864</td>
</tr>
<tr>
<td><strong>Heterogeneity effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intense Aromatic Taste</td>
<td>−0.0199</td>
<td>2.0355</td>
<td>0.9922</td>
</tr>
<tr>
<td>Soft-Velvety Taste</td>
<td>0.0758</td>
<td>2.1544</td>
<td>0.9719</td>
</tr>
<tr>
<td>Caffeine</td>
<td>2.0270</td>
<td>0.5818</td>
<td>0.0005</td>
</tr>
<tr>
<td>Price—1st level</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Price—2nd level</td>
<td>1.7119</td>
<td>0.6560</td>
<td>0.0091</td>
</tr>
<tr>
<td>Price—3rd level</td>
<td>0.8766</td>
<td>1.1853</td>
<td>0.4596</td>
</tr>
</tbody>
</table>

*St. error: standard error.

When considering the “Type of coffee”, its positive estimated coefficient indicates a preference for the 100% Arabica blend. For the soft and velvety taste, the positive sign indicates a preference for a high presence of this taste. In contrast, for the intense and aromatic taste, the negative sign indicates a preference for a fair presence of this one. The “Caffeine” estimated coefficient is positive, indicating a preference for coffee with more caffeine, which means a preference for the Arabica and Robusta blend. Moreover, the two price attribute coefficients are negative and decreasing, meaning consumers’ WTP decreases when the price rises. Lastly, the “Coffee Type × Gender” coefficient evaluates the first-order interaction between the type of coffee and the gender. Its positive sign indicates that females prefer the Arabica and Robusta blend with respect to males.

The heterogeneity coefficients are included at the bottom of Table 2. In this first modeling step, only two attributes show a highly significant heterogeneity, e.g., the caffeine and the second price level. Notably, the heterogeneity effects of intense, aromatic, soft, and velvety tastes are not statistically significant.

5.2. Mixed Logit Model Results for Choice 2 Coffee Session

When observing the model results related to the Choice 2 session (Table 3), carried out after the guided tasting, in general, the results show estimated coefficients with significant or almost significant p-values.
Table 3. Model results for Choice 2: estimated coefficients, standard errors, p-values.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>St. Error *</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of coffee</td>
<td>−2.1689</td>
<td>0.6971</td>
<td>0.0019</td>
</tr>
<tr>
<td>Type of packaging</td>
<td>−0.2467</td>
<td>0.1482</td>
<td>0.0959</td>
</tr>
<tr>
<td>Geographical origin and sustainability</td>
<td>1.1989</td>
<td>0.2914</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intense-Aromatic Taste</td>
<td>0.8767</td>
<td>0.3023</td>
<td>0.0037</td>
</tr>
<tr>
<td>Soft-Velvety Taste</td>
<td>−1.0596</td>
<td>0.2843</td>
<td>0.0002</td>
</tr>
<tr>
<td>Caffeine</td>
<td>1.5222</td>
<td>0.5109</td>
<td>0.0029</td>
</tr>
<tr>
<td>Tasting</td>
<td>−2.8631</td>
<td>1.4170</td>
<td>0.0433</td>
</tr>
<tr>
<td>Price—1st level</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Price—2nd level</td>
<td>−0.9160</td>
<td>0.3958</td>
<td>0.0206</td>
</tr>
<tr>
<td>Price—3rd level</td>
<td>−2.8381</td>
<td>0.8539</td>
<td>0.0009</td>
</tr>
<tr>
<td>Type of coffee × Tasting</td>
<td>3.7221</td>
<td>1.5970</td>
<td>0.0198</td>
</tr>
<tr>
<td>Type of coffee × Gender</td>
<td>−1.1434</td>
<td>0.7066</td>
<td>0.1056</td>
</tr>
<tr>
<td><strong>Heterogeneity effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intense Aromatic Taste</td>
<td>1.5249</td>
<td>0.5572</td>
<td>0.0062</td>
</tr>
<tr>
<td>Soft-Velvety Taste</td>
<td>0.7634</td>
<td>0.4982</td>
<td>0.1254</td>
</tr>
<tr>
<td>Caffeine</td>
<td>0.7209</td>
<td>2.0959</td>
<td>0.7309</td>
</tr>
<tr>
<td>Tasting</td>
<td>1.6804</td>
<td>6.3712</td>
<td>0.7920</td>
</tr>
<tr>
<td>Price—1st level</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Price—2nd level</td>
<td>−2.0503</td>
<td>1.0098</td>
<td>0.0423</td>
</tr>
<tr>
<td>Price—3rd level</td>
<td>−2.5678</td>
<td>1.1341</td>
<td>0.0236</td>
</tr>
</tbody>
</table>

*St. error: standard error.

Consumers confirm their preferences for product sustainability and the soft bag in a modified atmosphere for the coffee packaging. For the rest of the estimated coefficients, we may observe a relevant change in the consumers’ preferences with respect to Choice 1 modeling. More specifically, the “Type of coffee” is negative in this modeling step, indicating a preference for the Arabica and Robusta blend. The “Soft-Velvety Taste” is negative, indicating a preference for its fair presence level, while the “Intense-Aromatic Taste” is positive, meaning a preference towards its high presence level. Both of these results confirm the tendency to choose the Arabica and Robusta blend. The caffeine coefficient still remains positive, indicating a preference for coffee containing more caffeine, which is the Arabica and Robusta blend. Moreover, in this second modeling step (Choice 2 modeling), the guided tasting information is included through the variable “Tasting” [10], which is computed considering all the judgments collected through the scorecard. The estimated coefficient is negative and further confirms the respondents’ preferences for the Arabica and Robusta blend. The two coefficients related to the price attributes show the same pattern obtained in Choice 1 modeling, confirming that consumers’ willingness to pay decreases when the price rises.

Two first-order interaction terms are included in the Choice 2 model estimation. More precisely, the “Type of coffee × Gender” interaction is now negative, indicating that in this second modeling step, females’ preferences go towards the Arabica and Robusta blend with respect to the males. The “Type of coffee × Tasting” coefficient evaluates the interaction between the type of coffee and the tasting variable. Its positive sign indicates that respondents choosing the Arabica and Robusta blend during the tasting coherently prefer the same blend with the regard to the type of coffee in the Choice 2 session.

When considering the heterogeneity coefficients (see Table 3, bottom), the results show a very low heterogeneity effect, given that only three heterogeneity coefficients are significant: the intense and aromatic taste, the second and the third price levels, while the heterogeneity effect related to the soft and velvety taste, even though not significant, shows a relevant p-value (i.e., p-value = 0.1254). However, the heterogeneity coefficient related to the “Tasting” variable shows a very large standard error.
In Table 4, we report several relevant diagnostic measures to show the goodness-of-fit for the estimated MIXL models. For each MIXL model, we report the Log-Likelihood (LogLik), the Akaike’s information criterion (AIC) [57], and the Schwarz’s or Bayesian information criterion (BIC) [58]; both the AIC and BIC indexes are used for model selection (e.g., the lower their value, the better the model’s goodness-of-fit). We also report the McFadden’s Likelihood Ratio Index (LRI, also called Pseudo-R^2) [1] and the Veall–Zimmermann index [59]. The approximate interval values 0.30–0.35 and 0.50–0.55 for McFadden’s LRI and Veall–Zimmermann indexes, respectively, are usually considered good. In the case study, the goodness-of-fit is very satisfactory for both estimated MIXL models (Table 4).

Table 4. MIXL models: diagnostic measures.

<table>
<thead>
<tr>
<th>Model</th>
<th>LogLik a</th>
<th>AIC b</th>
<th>BIC c</th>
<th>McFadden’s LRI d</th>
<th>Veall–Zimmermann</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIXL model for Choice 1</td>
<td>−375.2284</td>
<td>778.4568</td>
<td>844.9886</td>
<td>0.3676</td>
<td>0.5811</td>
</tr>
<tr>
<td>MIXL model for Choice 2</td>
<td>−422.2938</td>
<td>878.5876</td>
<td>959.3762</td>
<td>0.2883</td>
<td>0.4915</td>
</tr>
</tbody>
</table>

a LogLik: Log-Likelihood; b AIC: Akaike’s Information Criterion; c BIC: Bayesian Information Criterion; d LRI: Likelihood Ratio Index.

6. Discussion and Final Remarks

The choice modeling results illustrated in the previous section clearly indicate that the tasting session and the information provided on chemical composition (caffeine as a selected parameter in this investigation) play a relevant role in unequivocally guiding the consumers’ preferences. In fact, in Choice 1 modeling, the consumers’ preferences are collected in a situation where the respondents express the preferences according to their previous knowledge about coffee and their own lifestyle, deriving quality expectations from credence attributes rather than experience attributes. In addition, the controversial results obtained in Choice 1, related to the choice of 100% Arabica blend jointly with a preference for more caffeine, confirm the respondents’ “misinformation”, especially when considering the relevant heterogeneity effect of the caffeine in Choice 1 modeling with a highly significant p-value.

Instead, in Choice 2 modeling, the respondents are also exposed to the information provided during the guided tasting about the taste properties and chemical composition of the two types of coffee (100% Arabica and an Arabica/Robusta blend). In fact, the guided tasting of the two types of coffee (administered in a randomized sequence in each tasting step) is carried out by specifically giving information on how to assign higher or lower scores for perceptions. Since our study is addressed to non-trained people, the guided tasting consists of explaining how to fill out the scoring card, explaining that the lowest score (equal to one) corresponds to the worst perception.

The sensory attributes of bitterness, acidity, sweetness, and aroma are chosen, along with aftertaste and general equilibrium.

To confirm the role of both the guided tasting sessions and the information provided [60], we should note that the change in the consumers’ preferences between Choice 1 and Choice 2 modeling only concerns the type of coffee and the two attributes related to its taste properties, e.g., the soft and velvety taste and the intense and aromatic taste. Moreover, unlike Choice 1, in Choice 2, the respondents’ preferences about the quantity of caffeine are correctly defined; the consumers choose the Arabica and Robusta blend, and in line with this, they show a preference for more caffeine.

The two estimated coefficients for the attributes of soft and velvety and intense and aromatic tastes are also coherent with the choice of the Arabica and Robusta blend; both attributes show a low heterogeneity effect by also considering the corresponding lower standard errors achieved in Choice 2 compared to Choice 1 modeling. Moreover, in Choice 2 modeling, the tasting coefficient confirms the consumers’ preferences for the Arabica and Robusta blend. Furthermore, the interaction that associates the type of coffee with the tasting variable further confirms the role of the guided tasting and the choice of the Arabica and Robusta blend. In addition, the interaction between type of coffee and gender shows a
preference for the Arabica and Robusta blend by females, which becomes relevant in Choice 2 modeling. It is also relevant to note the non-significant heterogeneity coefficient for the “Tasting”. Nevertheless, a comment should be made when observing the correspondent standard error, which is very high; this is probably due to the overall synthesis of all collected judgments related to taste descriptors.

When considering the price attributes, the estimated coefficients show the same pattern in both Choice 1 and Choice 2 model results, confirming that the consumers’ WTP decreases when the price rises. However, it must be noted that the two price levels have different relevance for the respondents. More precisely, unlike the second price level, which is more or less equally important in both Choice 1 and Choice 2 results, the third price level changes its importance in Choice 2 with respect to Choice 1 modeling. Moreover, the significant heterogeneity effect related to the third price level attribute obtained in Choice 2 modeling could indicate an increasing uncertainty of the respondents regarding the WTP after the guided tasting.

As far as the ethical aspects are concerned, “fair trade” and “organic” are credence attributes [61] that constitute an important dimension adding value to the product and contributing to the definition of their price. Consumers associate organic food with naturally produced and environmentally sustainable foodstuffs, while the fair-trade label mainly focuses on the well-being of workers and farmers in developing countries. However, many European consumers value environmental performance, and consequently, a substantial part of fair-trade-labeled coffee is also produced and certified in line with organic standards [62].

Globally, according to Fairtrade International (https://www.fairtrade.net/ accessed on 17 April 2024), 30% of all fair-trade coffee is also organically certified.

In a previous study [63], the most sensitive market segment for fair-trade coffee is represented by young women with a higher education level. Also, younger customers are sensitive to price and fair-trade issues. Another recent study that uses a CE [64] reports that Italian consumers usually prefer organic coffee to fair-trade, providing evidence for the need for more information about fair-trade products and their ethical traceability.

A very recent Spanish work carried out in [46] shows evidence that sustainability labels on coffee can significantly enhance the development of a more sustainable market, stimulating demand for sustainable coffee and expanding the consumer demographic. Williams et al. [65] indicated that coffee taste and flavor influenced purchasing decisions more than any sustainable attribute. Taste is likely to be one of the most important factors in purchasing certified coffee as it is for organic food [66].

Through this research, we try to emphasize the role of the consumer as an active element in understanding the sensory characteristics of a highly consumed and variable beverage, in terms of its sensory characteristics, such as coffee. On the consumer side, greater awareness of the sensory profile of coffee increases the importance of this aspect in product choice. On the producer and retailer side, it is possible to implement a variety of products, and it is possible to stimulate a segmentation of the offer, for example, from the basic level of the discount/supermarket to the specialty store, up to sales on online channels. Finally, in packaging, it is possible to implement supply chain storytelling, and this then can also be substantiated by the choice of obtaining product certifications.

In our study, we combine sustainability-related certifications and indications of origin because these aspects are relatively close, especially for the coffee producers considered in this research, namely major supermarket brands necessarily characterized by communication transparency and product traceability. A recent study suggests that to attract consumers, ecolabels should emphasize agrochemical standards, such as pesticide-free production, along with biodiversity preservation (Gatti et al. [44]).

Geographical origins also represent an approach to food production that places social, cultural, and environmental considerations at the center of the value chain. From this perspective, both indications of origin and sustainability-related certifications refer to communicating the company’s sustainability.
When specifically considering the label indication, there is no change in the consumers’ preference between Choice 1 and Choice 2 modeling. On both occasions, the respondents’ preferences tend towards product sustainability rather than the indication of the geographical origin.

In conclusion, this study enhances the role of the consumer who, with a greater awareness of the sensorial aspects of the product, is more willing to choose between a broader range of products, and is more interested in the story behind the product. Moreover, to increase the variety of the offer and to obtain greater segmentation of the coffee market, producers and retailers are encouraged to “tell” the story of the product more through packaging and, in the future, by the acquisition of certifications that could link the sensorial/hedonistic aspects of the drink with its environmental and social sustainability. One limitation of the study could be the consideration of non-trained subjects; however, administering a professional scoring card to consumers jointly with onsite training allows for obtaining satisfactory and, mainly, coherent results considering all the survey steps.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/stats7020032/s1, File S1: choice-set; File S2: scoring-card.

Author Contributions: The manuscript is a joint work of all the authors. However, R.B., P.P. and N.D.N. jointly carried out Section 1. R.B. carried out Section 2. N.D.N. carried out Section 3.1, while R.B. carried out Section 3.2. P.P. carried out Section 4.1, while N.D.N. carried out Section 4.2 and R.B. carried out Section 4.3. N.D.N. carried out Section 5.1, while R.B. carried out Section 5.2. P.P. carried out Section 6. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

15. Train, K.E. Recreation demand models with taste differences over people. Land Econ. 1998, 74, 230–239. [CrossRef]


38. Yeh, C.-H.; Hartmann, M.; Langen, N. The Role of Trust in Explaining Food Choice: Combining Choice Experiment and Attribute Best-Worst Scaling. *Foods* 2020, 9, 45. [CrossRef]


42. Becerra, M.; Goos, P. Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables. *Food Qual. Prefer.* 2023, 110, 104928. [CrossRef]

43. Balco, P.; Gracia, A. An extended approach combining sensory and real choice experiments to examine new product attributes. *Food Qual. Prefer.* 2020, 80, 103830.


63. Ostteveer, P.; Sonnenfeld, D.A. *Food, Globalization and Sustainability*; Earthscan: London, UK, 2012; Section III.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.